LTY OF GRADUATE STUDIES

[redacted]

E 17 May 94

Rubisco LSU amino acid composition as a phylogenetic indicator

by

Bruce Thomas Jeffs
B.Sc., University of Victoria, 1988

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Biology

We accept this thesis as conforming
to the required standard

[redacted]

Dr. L.A. Hobson, Supervisor (Department of Biology)

[redacted]

Dr. D.J. Ballantyne, Departmental Member (Department of Biology)

[redacted]

Dr. R.W. Olafson, Outside Member (Department of Biochemistry)

[redacted]

Dr. J. N. C. Whyte, External Examiner (Fisheries and Oceans Canada,
Pacific Biological Station)

© BRUCE THOMAS JEFFS, 1994

University of Victoria

All rights reserved. Thesis may not be reproduced in whole or in part, by photocopy or other means, without the permission of the author.

Supervisor: Dr. Louis A. Hobson

ABSTRACT

Difficulty in explaining the phylogenetic relationships of the groups of algae has been based in failure to allow adequately for the endosymbiotic theory, which considers the history of the organelles as different from the cell body. This study examines the phylogeny of algal chloroplasts, using the carbon fixing enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco). Rubisco amino acid composition data were obtained from a range of marine algae by denaturing polyacrylamide gel electrophoretic (SDS-PAGE) separation of cell lysate, electroblotting onto polyvinylidine difluoride (PVDF) membrane, and reversed-phase high-performance liquid chromatographic separation and quantification of the amino acids. The SΔn distance statistic ($S\Delta n = \frac{1}{2}\Sigma(n_{iA} - n_{iB})$, where n_{iA}=the number of the ith type of amino acid residue in Rubisco from species A) was calculated for pair-wise comparisons of the taxa. These distances were clustered using the unweighted pair-group using arithmetic means method (UPGMA) and a dendrogram produced. The data were also examined by principal components analysis (PCA). Rubisco amino acid compositions and sequences from a range of bacteria, algae, and land plants in the literature were included in other comparisons. Four groups of organisms were separated by the dendrograms: 1) The α-purple bacterium, *Rhodospirillum rubrum* was near the base of the tree. 2) *Amphidinium carterae,* a dinoflagellate, possessed the most different Rubisco. 3) Land plants, a euglenophyte, a prochlorophyte, *Cyanophora paradoxa,* and green algae grouped with the cyanobacteria. 4) A group of chromophyte algae, including cryptophytes, chrysophytes, and diatoms, were intermixed with rhodophytes and purple bacteria of the α, β, and, γ subdivisions. These general relationships were observable in the reduced-dimension representation of the original amino acid composition data in the PCA. The PCA also served to point out some possible sources of error in the analyses. These included contamination of the membrane with glycine. Also, the combining of amino acid compositions derived from gene sequences with those derived from amino acid compositional analysis may have introduced an error in the methionine levels.

In spite of these difficulties, these results agree well with those obtained from analysis of Rubisco gene and derived-protein sequences, at least at the higher taxonomic levels. However, some of these results disagree with phylogenies suggested by other molecular data, including protein sequences deduced from *atpB,* and *psbA* genes, and16S ribosomal RNA sequences. These other data support the hypothesis that cyanobacteria are the most recent common ancestors of all plastids. In contrast, the Rubisco data indicate that purple bacteria are the precursors of rhodophyte and chromophyte plastids. Several hypotheses have been advanced to account for this discrepancy. There may have been a lateral transfer of a Rubisco operon from a purple bacterium to the ancestor of the rhodophytes that possessed a cyanobacterial plastid, either directly or via a precursor to mitochondria. Another hypothesis is that the endosymbiont possessed two forms of Rubisco, a cyanobacteria-like and a purple bacteria-like, losing either in the two lineages.

Several aspects of algal and eubacterial photosynthetic machinery are examined in an effort to reveal the origin of cyanobacteria. Most of the evidence points to a link with purple bacteria, and hence origin of cyanobacterial Rubisco in this group. On this basis, another hypothesis to explain the discrepancy of the Rubisco-based phylogeny is explored. Possibly, the endosymbiosis that resulted in rhodophyte plastids was of a proto-cyanobacterium. In this organism, a characteristic cyanobacterial photosystem had become established, but the Rubisco was still purple bacterial in nature. This endosymbiosis may have occurred at an earlier time when CO_2 levels were high and O_2 low. These conditions resulted in little evolutionary pressure on Rubisco from the point of view of the inefficiency of its oxygenase reaction, and hence its retention of a purple bacteria-like sequence. Subsequently, upon achievement of more modern CO_2 and O_2 levels, during which time there was considerable evolutionary pressure on Rubisco, an endosymbiosis resulting in chlorophyte plastids occurred. The Rubisco at this stage was more like that found in modern cyanobacteria, and subject to the oxygenase reaction that modern purple bacteria still avoid by inhabiting anoxic environments.

This is the first report of a dinoflagellate in Rubisco phylogenetic analysis. The highly divergent Rubisco of *Amphidinium carterae* suggests a unique origin for its plastids, with no indication of its predecessor.

[redacted]

Dr. L.A. Hobson, Supervisor (Department of Biology)

[redacted]

Dr. D.J. Ballantyne, Departmental Member (Department of Biology)

[redacted]

Dr. R.W. Olafson, Outside Member (Department of Biochemistry)

[redacted]

Dr. J. N. C. Whyte, External Examiner (Fisheries and Oceans Canada,
Pacific Biological Station)

Table of Contents

List of Tables

List of Figures

Acknowledgements

I am grateful for personal funding to the University of Victoria, NSERC, the Melburn-Brenton Scholarship, the Lewis J. Clark Memorial Fellowship, and Dr. Louis A. Hobson. Dr. Hobson also supplied funds, space, and equipment to do the work.

I am indebted to S. Kielland of Dr. Olafson's lab for performing the hydrolyses and amino acid analyses. I thank Dr. R. D. Burke for use of the ultracentrifuge.

Thanks to Mavis Pareis for encouragement to return to university and for personal support for this and my previous degree. Patrick Lucey was helpful in seeing this project to its completion. Melissa McQuoid aided in many ways.

Abbreviations

3-PGA	3-phosphoglyceric acid
$\delta^{13}C$	ratio of ^{13}C to ^{12}C relative to a standard (PDB)
CA1P	carboxy arabinitol 1-phosphate
cpn60	Rubisco subunit binding protein
D1, D2	proteins in PSII
Disc-SDS-PAGE	discontinuous SDS-PAGE
DTT	dithiothreitol
EGTA	ethylene glycol bis(β-aminoethyl ether) *N,N,N`,N`* -tetraacetic acid
ETC	electron transport chain
$K_{act}(CO_2)$	CO_2 concentration for half-maximal activation
K_c	K_m of carboxylase reaction of Rubisco
K_m	substrate concentration for half-maximal activity
K_o	K_m of oxygenase reaction of Rubisco
L, M, H	proteins in RCII
LHC	light harvesting complex
LSU	large subunit of Rubisco
N	number of amino acids in protein
OEC	oxygen evolving complex
OTU	operational taxonomic unit
PAGE	polyacrylamide gel electrophoresis
PCA	principal components analysis
PSI	photosystem I
PSII	photosystem II
PTC	phenyl-thio-carbamyl
PVDF	polyvinylidine difluoride

rbcL	gene for Rubisco LSU
rbcS	gene for Rubisco SSU
RCI	reaction center I
RCII	reaction center II
RP-HPLC	reversed phase high performance liquid chromatography
rRNA	ribosomal RNA
Rubisco	ribulose-1,5-bisphosphate carboxylase/oxygenase
RuBP	ribulose-1,5-bisphosphate
SAHN	sequential agglomerative hierarchical nested methods
SΔn	index of Cornish-Bowden (1983)
SΔQ	index of Marchalonis and Weltman (1971)
SDS	sodium dodecyl sulfate
SSU	small subunit of Rubisco
TEMED	*N,N,N`,N`* -tetramethyl-ethylenediamine
Tris	tris(hydroxymethyl) aminomethane
UPGMA	unweighted pair group method with arithmetic means
V_c	maximum velocity of carboxylase reaction of Rubisco
V_o	maximum velocity of oxygenase reaction of Rubisco
WPGMA	weighted pair group method with arithmetic means

Introduction

This study uses primary structure of the carbon fixing enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase, or Rubisco, to investigate the evolution of the algae. This molecule has the potential to clarify the origin of chloroplasts and the origins of different algal groups from prokaryotic ancestors. Molecular data can be especially useful where morphological data have not given a clear picture, in addition to supporting morphological phylogenies or overturning false ones. One such instance concerns the chromophytes, or chlorophyll *c* -containing algae, another is the dinoflagellates. The relationships within and among these groups and to other algal and bacterial groups have been a matter of speculation. This work helps to reveal ties among the algae and prokaryotes that have previously remained hidden, and allows postulates for early stages of evolution of life to be advanced.

Why study algal evolution and taxonomy?

Taxonomy is essential for understanding ecosystems, and taxonomy amongst aquatic primary producers is as challenging as for any other group. In order to adequately understand the dynamics of primary production one must first be able to identify the primary producers. A classification system facilitates identification. It also aims to represent the relatedness, in a broad sense, of the organisms. A classification built on relatedness allows generalizations to be made about its groups; taxa embody a summary of information (morphological, functional, ecological) about its constituent members. The classification system also allows predictions to be made about a novel or uninvestigated taxon based on what is known about presumed related taxa. The relationships, embodied in the classification system, between these higher order taxa are also usually intended to reflect theories of the phylogeny of the taxa. Possibly the most important function of a taxonomy is to generate hypotheses about evolutionary theory, as well to provide a context for testing these hypotheses. The necessity for work on the taxonomy of primary producers extends to the highest taxonomic level.

Another connection between ecology and taxonomy is through comparative physiology and biochemistry. To understand how an organism is interacting with its environment one can look at its physiology. Different physiological responses to environment determine different levels of success and hence shape ecological relationships. At the molecular level important enzymes can be compared functionally and structurally. These functional and structural characters can be compared in a taxonomic structure. For example kinetic parameters can constitute characters for comparison. In turn, a taxonomy can represent and allow prediction of ecological relationships through incorporation of physiological and biochemical data. Molecular data are very important in current taxonomy and phylogeny, especially in issues that resist resolution using morphological data. Molecular data become particularly relevant when coupled with biochemistry and physiology.

In addition to studying algal taxonomy for the purposes of identification in ecological studies and for building a framework for summarizing and refining knowledge about the algae, there is also a more academic interest in that not much is known about larger issues: the origin and shaping of modern life forms, and the relationship algae have to others; it is suspected that the algae play a major role. Aquatic primary producers represent an enormous diversity: these are usually collected under the poorly, or difficult to define term, algae, but also include a range of prokaryotes of different nutritional strategies. In spite of the scientific/academic opportunities presented by this diversity, or perhaps because of the enormity of the task, the phylogeny of primary producers is not well worked out.

As an illustration, a central question of algal evolution is that of origin of chloroplasts. The debate as to the mechanism for their arising is almost entirely resolved on the side of the endosymbiotic theory (Gray 1989). But the problem of multiple *versus* a single endosymbiosis is still appearing in the literature. Until this question is resolved, relationships among higher order taxa cannot be definitively determined. Most textbook taxonomies (eg. Bold and Wynne 1985) lump together all the characteristics of the algae, both chloroplastic and "host". This is a potential source of confusion, especially if there have been multiple endosymbiotic events, either from closely or distantly related endosymbionts. It would be prudent to consider both types of characters separately until monophyly of chloroplasts has been demonstrated more conclusively. Trying to determine the history of chloroplast characters will help to clarify the lineages of the algae as a whole.

The remainder of this introduction outlines the taxonomy of the algae as it is currently accepted and highlight some of the phylogenetic issues of concern in each of the groups. Following this is some background material on numerical taxonomic methods, and an introduction of the molecule of primary interest in this study.

The present state of Algal taxonomy

The Algae

Problems in algal taxonomy begin at the outset; phycologists have difficulty deciding what the algae are. Part of this problem arises when trying to apply the criterion of photosynthetic ability. Some groups of "algae" clearly contain members that are not photosynthetic. Maintenance of the algal designation is easily accepted if it can be demonstrated that photosynthesis has been secondarily lost; but is more difficult if the group is apparently primarily non-photosynthetic. The boundaries of the definition of algae are blurred also when somewhat unique characteristics are shared by members of another more or less well defined group such as the heterokont flagellae that occur in the chromophyte algae and oomycete fungi.

Chlorophyta

The green algae form a monophyletic assemblage, with little controversy about what the chlorophytes are, except that some have wanted to include the charophytes (eg. Fritsch 1935), or the euglenophytes (eg. Christensen 1962); but the current consensus (eg. Bold and Wynne 1985) is to exclude them based on complex vegetative and reproductive morphology of the former, and cellular organization and biochemistry of the latter. It could also be argued that the vascular plants should also be grouped with the chlorophytes because of their presumed origin from this group, although they are not algae.

Within the Chlorophyta, a dramatic reorganization has been proposed based on ultrastructural analysis of cell division (eg. Pickett-Heaps 1975). From this it has been deduced that the most advanced form ancestral to all green algae was likely a scaly green flagellate, similar to the modern *Pedinomonas* (Pickett-Heaps 1975). The ultimate origin of the chlorophytes is postulated (van den Hoek 1981) to be a heterotrophic, colorless, isokontan flagellate that acquired endosymbionts possessing chlorophylls *a* and *b*. A prochlorophyte, because they have these pigments, seems an attractive candidate for such an endosymbiont.

It is proposed that the green line be split by the separation of a phycoplast mode of cell division from the more universal phragmoplast mode. The phycoplast is a bundle of microtubules at telophase crossing at right angles to the axis of cell division. This structure, involved in cytokinesis, is associated with either formation of a cell plate (vesicles of material that coalesce into a new septum), or by furrowing (extension of the plasmalemma from the equatorial periphery toward the center). The phycoplast may have arisen from a need to maintain accurate partitioning of daughter nuclei and other cell components (Pickett-Heaps 1975). Phycoplast formation occurs in algae where daughter nuclei approach one another after telophase. This latent problem is correlated with (likely the result of) an early collapse of the interzonal spindle.

The phragmoplast consists of microtubules oriented parallel to the spindle, serving as a site for cell plate formation by coalescing of plasmalemma/cell wall vesicles. This feature is dependent on the presence of a persistent spindle and widely separated daughter nuclei.

The distinction between modes of cell division and also ultrastructure of the flagellar apparatus has resulted in a proposal for three classes of green algae (Stewart and Mattox 1978). The Chlorophyceae have early dispersing spindle microtubules, close telophasic nuclei, transverse phycoplast microtubules with furrowing or cell plate formation, and a cruciate flagellar root system in motile cells. The Charophyceae possess a persistent spindle, distant telophasic nuclei, furrowing or phragmoplast cell plate formation, and a multilayered structure (flat band of microtubules) at the base of laterally inserted flagellae in zoids. Thirdly, the Ulvophyceae have a distinctive spindle that remains into late anaphase, absence of a phycoplast, early furrowing that serves to hold the daughter nuclei apart, and a different type of cruciate flagellar root system. This classification causes considerable difficulty because it would separate a formerly homogeneous group, the Ulotrichales, even splitting up a genus, *Ulothrix*, amongst all three orders. While the arguments of this thesis have been generally accepted, the conclusions have not yet been incorporated, likely due to nomenclatural rules; and recent classifications retain a traditional structure (eg. Bold and Wynne 1985).

Another phylogenetic issue concerning the green line algae is that of the origin of higher plants. Using characteristics such as multicellularity, heterotrichous and parenchymatic growth, and enclosure of gametangia in sterile cells, the Coleochaetales and Charales are likely related to the ancestors of land plants. Enclosed gametangia is taken to be significant because, it can be hypothesized, a retained zygote might be able to draw on gametophyte resources for protection during adverse conditions, and more importantly, this retained zygote could grow and develop as an alternate sporophyte generation (Graham 1985). This interpolated generation may have been different enough, functionally and probably morphologically, to allow persistence on exposed land, with subsequent evolution occurring as a decreasing dependence on a gametophyte of aquatic habit. Heterotrichy or parenchymatous growth, in either or both generations, could allow an erect habit without the buoyancy of water, in turn allowing more efficient gas exchange. To provide rigidity for support, cell walls must remain associated following cell division. For greater strength, new cross walls would be better as single rather than double, and new wall material at the other pole, if laid down, should be continuous with the parental wall. This mode of cell division is termed desmoschisis and is found in the Chlorosarcinales, Ulotrichales, Chaetophorales, and Ulvales as well as higher plants. While multicellularity can arise by the contrasting means, eleutheroschisis (new wall material is not contiguous with the parental wall), as in the tetrasporalean line, they are not thought to have been ancestral to the higher plants. Single cross walls as found in desmoschisis is more likely to be the outcome of cell plate formation than the pinching off of furrowing. Vascular tissue was likely made possible by plasmodesmata as they develop from cell plates and persistent interzonal spindles. This phragmoplast mode of cell division, as stated above, must have split very early in evolution of the green line. The phycoplast mode can be traced from multicellular parenchymatous and heterotrichous forms such as *Fritschiella,* through *Chlamydomonas,* back to scaly green monads that some (eg. Pickett-Heaps 1975) would designate as prasinophytes. *Pedinomonas* may be the simplest extant member of these scaly green monads, having a phycoplast. The phragmoplast mode of cell division extends along the chlorophyte line to the land plants also from simple prasinophytes. Thus, the prasinophytes have both modes of cell division which may have arisen in this group or previously. The phycoplast mode seems to be a side branch experiment of the green algae that has met with limited evolutionary success.

One could look for methods to help distinguish between two alternate, and equally difficult to accept, hypotheses to explain the contradiction of splitting a genus such as *Ulothrix* using a presumably conservative character, mode of cell division, that would seem to divide the chlorophyte line back to a very early stage. These hypotheses are: that the *Ulothrix* "genus", in spite of all the characters that serve to unite them, is an unnatural assemblage whose similarity arose by convergent evolution from different branches; or alternately, that the suite of characters that comprise coordination of all the events of cell division are not conservative and the phycoplast arose several times. The latter hypothesis suggests a regulatory control of the many processes of cell division that is switched back and forth between the two modes of cell division along the various branches of green line evolution. The former hypothesis implies a phylogeny that cannot be revealed by conventional morphological means, a problem not expected in more complex multicellular forms but common to smaller, simpler unicells, especially prokaryotes. This would likely be an issue whose clarification could be found in analysis of appropriate macromolecules.

Phaeophyta

The brown algae, or phaeophytes, are a group unto themselves, with characteristics that serve to identify and distinguish them from other groups. Their homogeneity is reflected in recognition of a single class (eg. Bold and Wynne 1985). While members of this group, in the order Laminariales, have achieved the largest size and greatest structural complexity of all algae, there are no major divisions within, nor significant offshoots from this line. The phaeophytes have been included with other groups under more encompassing terms: The phaeophytes have been considered a class of the Heterokontophyta (van den Hoek and Jahns 1978) which also includes the chrysophytes. They have also been designated Class Phaeophycea, beside Chrysomonadea and Xanthophycea under the phylum Heterokonta subkingdom Chromophyta (Cavalier-Smith 1989). Several have included them as a class of the Chromophyta, which can include the chrysophytes (Bourrelly 1968), or the chrysophytes and cryptophytes (Lee 1980), or the chrysophytes, cryptophytes and dinoflagellates (Christensen 1962), among others. In spite of the distinctiveness of the phaeophytes they also have many characters in common with other groups, particularly the Chrysophytes, and so a question of ranking of the group arises. Currently (Clayton 1989) they are being considered as close to the chrysophytes, diatoms, and Oomycetes (fungi) based on rRNA sequences. Because of similarity of motile cells, it is thought that the phaeophytes arose from the Sarcinochrysidales of the Chrysophytes (Gayral and Billard 1986).

Rhodophyta

The red algae, or rhodophytes constitute a distinct group. Their complete lack of flagellated forms correlates with 5S rRNA analysis (Hori and Osawa 1987) suggesting that the rhodophytes have a very early origin, possibly the earliest of all eukaryotes. By comparative immunology of phycobiliproteins, the reds are linked most closely to the cyanophytes rather than to the cryptophytes (Glazer *et al.* 1971) Within the group, homogeneity is illustrated by recognition of a single class by Bold and Wynne (1985), and two subclasses, the Bangiophycidae and the Floridiophycidae, but the former contains only 1% of the red algal genera. Among the Bangiophycidae there are no shared derived characters to define them (Gabrielson *et al.* 1985); they are judged (Gabrielson *et al.* 1990) to be a paraphyletic group, hence, no subclass distinction is valid. A possible distant relative of the rhodophytes may be *Glaucocystis,* an organism of otherwise uncertain affinity (Bold and Wynne 1985). Cavalier-Smith (1982) suggests that *Glaucocystis, Glaucosphaera, Gloeochaete,* and *Cyanophora* constitute a separate class with the Rhodophyceae in his proposed phylum and kingdom Biliphyta, based on plastid characters (discounting the presence of the bacteria-like peptidoglycan membrane layer), although others (Kies 1979, Lee 1980) would place these four genera in a group separate from the other phycobilin-containing eukaryotes.

Pyrrhophyta

Pyrrhophytes, or dinoflagellates, are presumed to be an ancient group (Loeblich 1976) because of their distinctive nucleus. This nuclear condition has been termed mesokaryotic, to indicate a state intermediate between pro- and eukaryotic, but, since they are often considered a unique form of eukaryote, the dinoflagellate nucleus is sometimes referred to as dinokaryotic. Prokaryotic features of this nucleus include chromosomes that are attached to a membrane, remain condensed in interphase with a bacterial-like fibrillar arrangement, and DNA that is generally associated with, at most, low levels of histone proteins. These basic proteins, present in only a few species, bear no amino acid composition resemblance to histones found in all other eukaryotes (Taylor 1990). In contrast, eukaryotic features include: large amounts of repeated DNA, a discrete DNA synthesis phase, and a unique form of spindle where cytoplasmic tunnels containing extranuclear microtubules perforate the nucleus with its persistent envelope. Because of this combination of nuclear and other characteristics it is postulated (Loeblich 1976) that the dinoflagellates diverged from the early eukaryote line after the evolution of repeated DNA but before the presence of true eukaryotic chromatin. Loeblich (1984) has since revised his interpretation of these features and concluded that dinoflagellates are truly eukaryotic, with some degenerate and some advanced cytological characteristics. Still, a long evolutionary history is indicated for the dinoflagellates by their range of complexity of ultrastructure and mode of nutrition.

The ancestral dinoflagellate is a matter of controversy, depending on the presumed direction of evolution within the dinoflagellates. At one end of the continuum, the desmokont, or *Prorocentrum* -type is favoured as the ancestor (Loeblich 1976, Taylor 1980, Bujak and Williams 1981, Goodman 1987). At the other end, candidates for the ancestral type are a gymnodinoid, or a free living form similar to the parasitic syndinians, or *Oxyrrhis.* The last is favoured by Loeblich (1984) based on observations of histone proteins, intranuclear spindle, lack of bacterial-like chromosome fibrils, and a scaly cell surface, all of which connect this pyrrhophyte to other flagellate eukaryotes. In contrast, Dodge (1983) evaluates the prorocentroids as being the most advanced dinoflagellate group.

Dinoflagellates have been grouped with chrysophytes and phaeophytes in the division Chromophyta (eg. Christensen 1962). Whereas Klein and Cronquist (1967) and Bourrelly (1970) have included the cryptophytes with the dinoflagellates. They are now recognized as a separate and distinct division (Loeblich 1984, Bold and Wynne 1985, Taylor 1987), but matters of their phylogeny remain unresolved. Cavalier-Smith (1982) and Whatley and Whatley (1981) have proposed that the dinoflagellates arose as a branch from the same line as the chromophytes (Kingdom Chromista includes chrysophytes, phaeophytes, and cryptophytes, among others). Cavalier-Smith (1982) believes that a now non-extant ancestor common to these groups arose by endosymbiosis of a non-extant eukaryote containing chlorophyll *c* and phycobilins on the outer surface of the thylakoid membrane. The dinoflagellates are proposed to have arisen by loss of phycobilins, and cryptophytes by movement of the phycobilins into the thylakoid lumen allowing appression of the thylakoids. The remaining chromophyte line was thought to have been derived from cryptophytes by loss of phycobilins. Whatley (1989) revised her conception to agree with Dodge (1987, 1989), based on pigment, ultrastructural, and macromolecular sequence data. They view the dinoflagellate chloroplasts as polyphyletic in origin and the result of relatively recent (or less integrated) endosymbioses. Whatley (1989) allows the possibility that peridinin-containing species may represent a line with early well established chloroplasts that could be viewed as uniquely dinoflagellate.

Roughly half of the dinoflagellates have photosynthetic capacity while the remainder are primarily parasitic, saprobic, or phagotrophic (Dodge 1983). There is a variety of chloroplast types with respect to pigmentation (red, green, blue-green, chrysophyte-like, and cryptophyte-like), pyrenoid structure (stalked, embedded), chloroplast envelope (two or three membranes, or two and three membranes possibly anastamosing in a complex manner (Dodge 1983)), and degree of reduction of an endosymbiont into the host (from true chloroplasts, through a chloroplast with a putative persistent endosymbiont (chrysophyte) eukaryotic nucleus (Tomas and Cox 1973), to phagocytic vacuoles containing cellular contents where the ingested chloroplast may remain photosynthetically active for some time (Larsen 1988)). There seems to be a preponderance of less advanced endosymbioses and unusual pigmentation in the unarmored groups, possibly correlated with easier phagotrophy afforded by the lack of a rigid covering (Dodge 1983). Many species of dinoflagellates have no vestiges of chloroplasts. This evidence points to dinoflagellates as being primarily non-photosynthetic, with plastids acquired endosymbiotically over a broad range of time from various sources (Dodge 1983). Furthermore, dinoflagellates become symbionts in other organisms such as corals. The ability to form intimate associations thus extends in both directions, serving both as host and symbiont, with the capacity for accommodation seeming to reside with the dinoflagellate.

The dinoflagellates are grouped as five classes (eg. Bold and Wynne 1985), with two classes, the Desmophyceae and the Dinophyceae being the dominant two (numerically) and are combined by some (eg. Loeblich 1976). The issue of major distinction within the dinoflagellates is the direction of the series with respect to the number of thecal plates. Three models have been proposed.

The plate increase model

Loeblich (1976), using neontological arguments, places the prorocentroids, armoured with two large plates and a swimming direction placing the flagellae anteriorly (a situation common in other algal groups), as ancestral to dinoflagellates with several series of plates and a swimming direction placing the flagellae in a distinctively dinoflagellate lateral arrangement, to thinning and eventual loss of the plates, but with vestigial vesicles in many unarmored species.

The plate reduction model

It is proposed that the ancestral form is an unarmored gymnodinoid (Dörhöfer and Davies 1980) or *Oxyrrhis* -like (Loeblich 1984), and the most derived is the prorocentroid. This model draws on evidence showing a reduction in plate number in the geological record. While some fossil dinoflagellate series illustrate this progression, there are still wide gaps in the geological record and proposed connections between groups have little supportive evidence (Goodman 1987). Taylor (1990) suggests that *Oxyrrhis* should not be included in the pyrrhophyte division because of anomalous features such as internal mitotic spindle, transverse fission, and the typical differentiation of one of the flagellae is absent.

The plate fragmentation model

Bujak and Williams (1981) propose the ancestral form to be a prorocentroid with two plates or valves. Evidence of their antiquity may reside in the organic-walled microfossils, the acritarchs. Coincident with a change in swimming direction, the valves may have become fragmented into many undifferentiated plates. These plates then may have either thinned and disappeared or become reduced in number and modified, thus forming two major branches in the dinoflagellates.

This controversy in evolution of the morphology of dinoflagellate lines seems an ideal issue for resolution by biochemical means. Chloroplast characters must be excluded because of the promiscuity that dinoflagellates exhibit in taking up endosymbionts. Chloroplast characters are proposed to be unreliable above the genus level in dinoflagellates (Taylor 1987), thus "host" characters must be used to interconnect dinoflagellate lineages.

A valid question that arises is whether there is a characteristically dinoflagellate type of chloroplast. As indicated above, there are several instances of dinoflagellate chloroplasts in which recent endosymbioses are apparent. These display a range of incorporation, from ingested organisms whose chloroplasts are slower to be assimilated, to chloroplasts whose division is coordinated with that of the cell. These more or less recently acquired chloroplasts can be easily traced to members of well defined groups, occasionally to a group not usually phylogenetically associated with dinoflagellates (eg. the chlorophytes, Watanabe *et al.* 1990). These recent endosymbioses represent a horizontal linkage to other groups that can be contrasted with a more vertical lineage of other dinoflagellate chloroplasts. However, most dinoflagellate chloroplasts bear no marks of recent endosymbioses. Peridinin is the dominant photosynthetic carotenoid in some species of dinoflagellates and in no other groups, and might be a marker indicating a chloroplast that is primarily dinoflagellate.

Cryptophyta

Cryptophytes are a small well-defined distinctive group, but often termed "enigmatic" because of their combination of characters that seems to draw alliances in opposite directions. These include possession of chlorophyll *c,* indicating an affinity with chrysophytes and most other chromophytes, contrasted with possession of phycobiliproteins and pigments, indicating an affinity with rhodophytes and cyanophytes. Unique features include tubular mastigonemes on both flagellae, a nucleomorph, and an extensive periplastidal space between the two inner chloroplast membranes with an endoplasmic reticulum. As a result, this group has been placed with others. Lee (1980) put them with the chromophyte, and Bourrelly (1970), with the Pyrrhophyta. Others (eg. Chapman and Chapman 1973) have given them their own division. Discovery of the nucleomorph, with its DNA, has fuelled discussion of endosymbiotic origin of chloroplasts, and is thought to be a vestigial nucleus of an endosymbiont (Gillott and Gibbs 1980).

Chrysophyta

The Chrysophyte division constitutes a large and diverse assemblage, whose classes have often been raised to the division status. However, food reserve (chrysolaminarin) and pigmentation (high levels of carotenoids and lack of chlorophyll *b*) serve to unite the group. There are several features of chrysophytes which link them to phaeophytes. These include presence of chrysolaminarin which resembles laminarin, and the pigment, fucoxanthin, which occurs in several chrysophyte groups, and the distinctive morphology of the flagellated stage of certain chrysophytes. Also, the chloroplast endoplasmic reticulum is present in the chrysophytes as well as the cryptophytes and phaeophytes. Several classes of chrysophytes are less typical than others. The class Eustigmatophyceae lack chlorophyll *c* and have extraplastidal eyespots, in common only with the euglenophytes. The class Raphidophyceae have only small amounts of chlorophyll *c,* and the chloroplast endoplasmic reticulum is not continuous with the outer nuclear membrane. They also possess trichocysts, as found in the pyrrhophytes. The class Prymnesiophyceae have a haptonema, an appendage of unique ultrastructure that distinguishes them from other chrysophytes. The Class Bacillariophyceae, or diatoms, are linked to the Mallomonads of the Class Chrysophyceae by their siliceous cell covering, which is formed in a similar fashion. Chrysophytes of the Class Chrysophyceae have a diversity of cell covering, including none and those composed of calcium carbonate and cellulose, but a distinctive feature of this class, the statospore, is always siliceous. The silicoflagellates differ from other Chrysophyceae in having an internal siliceous skeleton. The diatoms are easily distinguished from other chrysophytes by the presence of a siliceous frustule. Also the dominant life stage is diploid, rather than haploid as for most other chrysophytes. Some centric diatoms where sexuality has been observed produce a sperm with a flagellum that lacks the central microtubule doublet seen in all other eukaryotic flagellae.

The Chrysophytes have been included in a division, or higher taxon, Chromophyta that includes dinoflagellates, phaeophytes, euglenophytes, and cryptophytes (eg. Lee, 1980) or in the Heterokontophyta that includes the phaeophytes (van den Hoek and Jahns 1978). They have also been separated at the division level into the Xanthophyta, Bacillariophyta, and Chrysophyta (Chapman and Chapman 1973), and also by separating the Haptophyta (Round 1977).

The above understanding of algal phylogeny is based primarily on morphological, ultrastructural and biochemical characters. However analysis of macromolecules are providing a wealth of characters that are giving new insights into phylogenetics, most notably ribosomal RNA (rRNA). While the methods used for the analysis of sequences, such as rRNA, are different from those used for amino acid analysis, much of the theory of numerical analysis is common to both. The numerical methods for interpreting amino acid compositions and background for them are presented below. The main purpose of numerical methods is to allow an unbiased assessment of a large number of characters, such as are found in sequences and amino acid compositions.

Evolutionary study, theory and practice

Taxonomic approaches

The dominant three operational approaches to the realm of taxonomy are evolutionary taxonomy, cladistics, and numerical taxonomy. The first does not have a consistently used name. Evolutionary taxonomy (eg. Mayr 1965) undertakes to reveal relative genetic relationships via comparison of extant organisms and the fossil record. The resulting phylogenetic relationships are interpreted as a branching and diverging evolution and can include sequential speciation along a branch. The recognition and assignment of a specific group to a taxonomic level is based on the intuition of the expert, using phenetic and cladistic elements and various character weighting according to their presumed information content (Mayr 1965). The evolutionary approach allows paraphyly, where the descendants of the most recent common ancestor are not all contained in the taxon (that is, some descendants from a stem taxon can give rise to another taxon whose rank is the same as or higher than the stem). For example, while land plants are considered derived from green algae, and hence by cladistics a sub group, they are afforded kingdom status apart from and above the Division Chlorophyta because of their innovations and radiational success in the terrestrial habitat.

Cladistics also traces genealogical histories but strictly prohibits paraphyly - all descendants of an ancestor must be contained in a monophyletic group (Rothschild and Heywood 1987). Grouping is done on the basis of shared derived (synapomorphic) traits, with the fossil record having overriding importance in ascribing derivation. Shared primitive traits (symplesiomorphies) are disregarded because they do not distinguish the descendants from the ancestral group and hence are uninformative, even misleading. Phenetics would regard this as an introduction of bias. Cladistic relationships are expressed as branching patterns. Evolution without branching, anagenesis, is not easily explainable by cladistics (Sneath and Sokal 1973).

The terms phenetics and numerical taxonomy are sometimes confused. Numerical taxonomy refers merely to the analysis of data by numerical means for the purposes of constructing a taxonomy. Both cladistics and phenetics can use numerical means, evolutionary taxonomy does not. Phenetics differs from cladistics in that it only uses numerical analysis, whereas cladistics employs the fossil record when available in addition to numerical techniques. The term phenetics, unfortunately, is also associated with the term phenotype, the morphology and physiology of an organism, which is distinguished from the genotype. Phenetics is the study of relationships based on a number of phenetic characters (all available, or a subset) without any weighting of one character as more important than another (Sneath and Sokal 1973). Phenetic characters are any that have a physical manifestation in the organism as distinguished from any proposed phylogeny. Phenetics is in opposition to the other two approaches, by regarding the fossil record as unable to reliably reveal phylogenies because too many gaps exist in most, if not all, records. More fundamentally, phenetics asserts that relatedness cannot be established from a supposed determination of phylogeny. Rather, phenetics must come first to establish degree of relatedness and phylogeny can be then inferred. In contrast to cladistics there is no rejection of characters based on their perceived primitiveness; primitive or derived is a question not asked in phenetics. The down side of this approach is that limited data sets may overstate relatedness where parallelism or convergence has occurred, and there is little opportunity to employ reliable genealogical data.

In a phenetic analysis, as defined here, the data are obtained in a numerical format and are analyzed by numerical techniques, of which there is a diversity. Some (eg. McNeill 1979) would argue that phenetics can include analysis of many characters both by numerical and "neural" means. Neural analysis uses visual macro-characters assessed intuitively (McNeill 1979). The neural analysis employs phenetic criteria supposedly objectively, but bypasses the coding and calculations of numerical phenetics. Rather than following a predefined procedure, the neural pheneticist adapts the methods to the current situation, and would be able to describe them after the fact, as implied by McNeill (1979). Sneath and Sokal (1973) argue that the purpose of phenetics is to employ objective techniques and such neural assessments of character arrays is not valid, and some numerical coding and analysis must be applied to remove as much subjectivity as possible. Selection of organisms and characters, and coding of these occasionally introduce biases, but Sneath and Sokal (1973) propose that all known characters and all species of a proposed genus (or all the genera of a family etc.) should be used. This has proved difficult in practice and is seldom attempted. Past experience, previous classifications, and/or subjective assessment determines which organisms or groups are included in a taxonomic study. In this, the prudent pheneticist must side with the evolutionary school; a study constitutes the testing of hypotheses concerning a prior classification. The selection of characters is also a matter of prejudging what is likely to appropriately resolve a group, but the pheneticist would weight characters equally and may use a broader range. Previous criticism of phenetics (eg. Mayr 1965) seems mainly because of a presumed emphasis of morphological characters, and lack of recognition of phylogenetic evidence. Recent use of biochemical data for phenetic analysis (eg. gene and protein sequences) counters this criticism: These data types are at or close to the genotype, the molecular basis of evolution, and hence largely unclouded by developmental and environmental effects on the phenotype. These data can indicate phenetic and genetic relationships simultaneously.

Protein primary structure

When one thinks of the primary structure of proteins or peptides one is typically understood to mean the type and order of the amino acids along the length of the protein (amino acid sequence) and the disulfide bridges between cysteine residues. But primary structure can also apply to characteristics such as pI (the pH at which the protein/peptide has a net neutral charge) and amino acid composition (the number or proportion of each type of amino acid).

Representation of phenetic proximity

Amino acid compositions are phenetic characters and must be analyzed phenetically. Primitive character states cannot be distinguished from derived in amino acid compositions, therefore a cladistic approach cannot be applied to these data. Generally, phenetic proximity is determined by comparison of characters that are expressed as numbers. These data can be obtained as numbers, eg. lengths or counts, or need conversion to a number as occurring in a category, eg. shape, or presence/absence. These numerical data can be conceived of as representing a position in N-dimensional space. Structure in this space can be evaluated in two different ways. Distances between different operational taxonomic units (OTU) can be calculated, effectively reducing the N dimensions to one. These OTUs can then be clustered. The other method involves a more direct analysis of the original character states, usually employing a method that highlights more informative characters (Li and Graur 1991).

Indices of amino acid composition

The indices, or distance measures, employed here are those of Marchalonis and Weltman (1971), SΔQ, and Cornish-Bowden (1977), SΔn. The calculation of these are outlined in the Methods and Materials section. SΔQ can theoretically be used to compare proteins of unequal length, but conclusions based on such comparisons can be questioned. In spite of this, in cases of unequal length, because of the error bias of the statistic, the likelihood of incorrectly assigning a relation when none exists is small, but so also is the likelihood of correctly assigning one that does exist (Cornish-Bowden 1983). If the proteins are of equal length SΔQ can be converted to SΔn. An essential feature of SΔn is that it is an unbiased estimator of the number of sequence differences between two proteins. Both of these statistics fall under the phenetic taxonomic approach.

Use of these indices is based on the following assumptions: The proteins being compared are random amino acid sequences independent of each other. The probability of occurrence of a particular type of amino acid residue at a position in the protein is the same for all positions and for both proteins.

An expression of the variance of SΔn can be derived (Cornish-Bowden 1977) from theoretical considerations. Utilizing the average frequencies of amino acids observed in proteins (Dayhoff and Hunt 1972), the standard deviation is evaluated as 0.35 times the number of amino acids in the protein, which is equivalent to 0.38 times the number of sequence differences between the proteins. The variance has only a weak dependence on the assumed amino acid frequencies, which may vary from protein to protein, so this serves as a reasonable approximation (Cornish-Bowden 1977).

Significance can be tested, assuming that SΔn is normally distributed, with the null hypothesis being that identities found between two proteins arise only by chance. A strong test of relatedness, at the 95% confidence level, would have a cutoff at 0.35 times the number of amino acids in the protein (Cornish-Bowden 1983). A weak test of relatedness would be if SΔn were below its expected value of 0.93 times the number of amino acids in the protein (N) (calculated from the average frequencies of amino acids in proteins), beyond which correspondence is due only to chance (Cornish-Bowden 1979). To summarize, for the dissimilarity measure, SΔn, if $S\Delta n < 0.35N$ there is strong evidence that the proteins are related; if $0.35N < S\Delta n < 0.93N$ there is weak evidence of a relationship; if $S\Delta n > 0.93N$ there is little or no evidence of a relationship.

The SΔn index is applied here to related proteins (Rubisco large subunit) from a variety of eubacteria, algae, and land plants, and as such the assumptions can be relaxed. Randomness need only be assumed for those positions where sequences differ. Further, sequence length can vary by a maximum of 18 residues (Cornish-Bowden 1983). From what is now known about structure-function relationships, which include secondary structure formation and operation of active and allosteric sites, obviously distribution of residues in a sequence is not random. However, when considering dissimilarity indices such as these, one could reason that differences occur mainly at those sites with a low functional constraint and that there they vary randomly. Random substitution of one amino acid for another is seen to be untrue when considering the protein as a whole, and the non-randomness is tabulated in the mutation data matrix of Dayhoff *et al.* (1979). To counter this theoretical difficulty there are several instances where this analysis has succeeded, insofar as it agrees with other, possibly more rigorous, tests (*eg.* Black and Harkins 1977). Cornish-Bowden (1979) has shown that SΔn agrees with empirically observed sequence-difference values when considering related proteins, but underestimates similarity between unrelated proteins. Hence, the chances of assigning a relationship erroneously are reduced. The correction to SΔn needed because of difference in size is ±2 if they differ by less than 18 residues. If they differ by more than 18 residues, SΔQ and SΔn can be used, but the length is taken to be the shorter of the two for the purposes of calculation (Cornish-Bowden 1977), but taken as the larger for the purposes of determining the limits of the significance tests (Cornish-Bowden 1980). Experimental error should also be considered but this source of error would most likely increase SΔn and therefore decrease the chance of finding a relationship. The effect of experimental error can be reduced with careful work and restriction to data from one lab by one operator. Analyses by reversed-phase high performance liquid chromatography (RP-HPLC) of phenyl-thio-carbamyl (PTC)-derivatized amino acids in the sub-nanomole range can be accurate to within ±10% (Stone and Williams 1986), relatively small compared to the ±38% of SΔn.

Cluster analysis and graph theory

Once distances between OTUs has been determined, the next step towards a phylogeny/classification is to search for patterns using cluster analysis.

From the graphical viewpoint, the OTUs are seen to be represented by vertices that are connected by edges if a relationship exists. The presence of an edge signifies a similarity greater than a specified cutoff value. The degree of connectivity of OTUs can be specified. A minimally connected graph has only one direct or indirect path between each vertex; these are often termed trees, and single linked clusters (below) are equivalent to this kind of graph (Sneath and Sokal 1973). A network is a type of graph where the edges have directionality, the relationship of A to B is not the same as that of B to A. A dendrogram or cladogram is a directed tree with the branch tips, the furcation points, and the root of the tree as vertices, termed nodes in cladistics (Sneath and Sokal 1973). A length dimension can be applied to the edges to indicate degree of dissimilarity. This allows total tree length to be minimized by manipulating some of the variables within certain assumptions; this can be applied to both directed and undirected trees. For higher orders of connectivity, relationships are often more easily represented on a page using a graphical technique (Sneath and Sokal 1973).

Cluster analysis can be a combination of alternate states of eight parameters: agglomerative vs. divisive, hierarchic vs. nonhierarchic, overlapping vs. nonoverlapping, and sequential vs. simultaneous methods; local vs. global criteria; direct vs. iterative solutions; weighted vs. unweighted, and adaptive vs. nonadaptive clustering (Sneath and Sokal 1973). Agglomerative techniques may not yield a classification where a group is defined by a unique set of characters (necessary and sufficient), but divisive techniques can, and hence the latter are useful for drawing up identification keys. Nonhierarchical techniques include ordination and those that produce nonrooted trees; these may be better for a more faithful representation of the relationships, but are less effective at summarizing them. Imposition of a nested hierarchy can result in distortion of the relationships; on the other hand, allowing overlapping of clustering confuses drawing and interpretation of the relationships. Direct methods are computationally more feasible than iterative but are likely to result in a suboptimal conclusion; attempts to ease the computational burden of iterative techniques can also remove the certainty of obtaining the one best solution.

Of the possible permutations, a group of clustering methods widely used in biology are the sequential, agglomerative, hierarchical, nested methods (SAHN). Of these the UPGMA (unweighted pair-group method with arithmetic averages) is the most frequently used. To complete the definition with respect to the eight parameters, UPGMA is a nonoverlapping method, as indicated by the nested criterion; it is usually applied globally but a shortcoming of this is that it may be a better estimator of similarity within a cluster than at larger levels of clustering; it is, overall, a direct technique because no self-correcting or optimizing procedure is an integral part of this method; it is a nonadaptive method when used alone but could be part of an adaptive routine.

UPGMA operates on a distance matrix and proceeds by first selecting the most similar element (or least dissimilar). These OTUs are then clustered and a new matrix calculated by averaging the elements for the now clustered OTUs. For example, in an analysis of four OTUs, A B C D, if A and B are to be clustered, the elements for A-C and B-C are averaged to yield a new element, AB-C, and A-D and B-D are averaged to yield AB-D, while element C-D remains unchanged at this stage. The unweighted aspect arises when combining clusters of different size. The simplest description of the UPGMA procedure is that values from the original matrix are used for the calculation of new averages, rather than, as for the W(weighted)PGMA method, which treats equally each cluster to be combined, and this has the effect of weighting late joining members more heavily. Proceeding with the illustration, if cluster AB is to be combined with C by UPGMA, element ABC-D is calculated by averaging three elements A-D, B-D, and C-D, rather than averaging two elements, AB-D and C-D, as by WPGMA. UPGMA yields cluster distances intermediate between single linkage (where the distance between two clusters is equal to the distance between the two nearest neighbours, the two members of each cluster that are closest) and complete linkage (the distance between two clusters is equal to the distance between the furthest neighbours). UPGMA has no easy geometric interpretation, while other averaging techniques such as the centroid methods do; for them the distance is taken as that between the geometric centers of the clusters.

Ordination

Often, with an analysis with many characters, an ordination procedure is employed to reduce the number of dimensions to something manageable, usually in order to be able to present a summary approximation of the original data in a graph, almost certainly 3 dimensions maximum. Other rationales for employing ordination procedures are as an alternative for clustering, either for looking for different types of trends using different assumptions or biases, or for reducing a large number of calculations. Also, unrevealing characters can be factored out to allow concentration on informative ones. For example, an interspecific morphometric study may wish to eliminate size in order to allow for different ages of individuals; the remainder would be shape or proportions.

One such manipulation, principal components analysis (PCA), involves ascertaining those characters that are most phenetically significant by determining a reduced number of new axes by least-squares linear regression analysis, such that the first dimension accumulates the largest amount of the variance in the data; a subsequent iteration will account for the next largest part of the remainder and so on until a sufficient amount of the variance is accounted for or the desired number of dimensions are reached. The OTUs are then collapsed onto this reduced-dimensional space. This often results in combining characters that covary; for example, length, width, etc. can be transformed into a dimension that could be interpreted as general size, possibly with shape constituting the bulk of the remainder. In the case of amino acid compositions of Rubisco the phenetic space is 16-dimensional. In performing PCA one is looking for amino acids or families of amino acids that may be particularly important for distinguishing taxa. Due to the general tendency of some amino acids to be rare and others common it is necessary to standardize the data. Thus, a change of 3 rare amino acids is weighted more heavily than one of 3 common amino acids. Each of the new dimensions resulting from PCA is a linear combination of the original 16 amino acids with the more significant amino acids having a larger coefficient. In collapsing the taxa in 16 dimensions onto the 3-dimensional graph an approximation is made of the original hyperspace.

Other methods of ordination are available with different sets of assumptions, but with the same goal of summarizing data to allow extraction and visualization of important trends. If one can assume independence of each of the original N dimensions, as is the case for amino acid composition data, then this lack of correlation allows visualization of the N axes as being orthogonal, or all at right angles to each other; and as such, distances between species are Euclidean. Under these conditions principal coordinate analysis is equivalent to principal component analysis. In other situations principal coordinate analysis has utility in allowing representation and further analysis in a Euclidean space of a non-Euclidean metric. A relevant example would be the expression of amino acid substitutions taking into account nucleotide substitutions, where one requiring three nucleotide changes is three times greater than one requiring one. These data would be best expressed as the non-Euclidean city-block metric (eg. if the route between two points is 1 block east, 1 block north and 1 block west; the distance would be 3 blocks not 1).

From organisms to a classification

To tie the above theory together and to indicate the direction of this study the following series is outlined. Organisms are selected and characters observed - in this case the characters are Rubisco amino acid composition. For phenetic analysis these are tabulated as a matrix of t taxa by a amino acids. These compositions are reduced to a dissimilarity distance measure SΔQ and converted to the more meaningful SΔn; the result is a symmetric matrix of dimensions t by t of which the elements are distances. These are then cluster analyzed to yield a representation of the distances as a t by 1 matrix, that is displayed as a dendrogram. Also, the t by a matrix are analyzed by principal components analysis yielding an extraction from the axes (a) of a reduced number of axes ($a' < a$), usually 2 or 3 for graphical representation of the t taxa.

Rubisco

Rubisco has several properties that suit it for a phylogenetic study but first aspects of its structure and function will be introduced.

Function

Of the two series of reactions of photosynthesis, the light reactions and the dark reactions, a central enzyme of the latter is ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco). Under conditions of adequate light, Rubisco is considered to be the limiting enzyme of the dark reactions of photosynthesis (Miziorko and Lorimer 1983). The importance of this enzyme is highlighted by the fact that it is probably the most abundant protein in nature. Rubisco is a bifunctional enzyme: As an oxygenase it catalyses the committed step of photosynthetic carbon oxidation, also called photorespiration. And, as a carboxylase it catalyses the committed step of photosynthetic carbon reduction, known as the Calvin cycle. Subsequent steps of these dark reactions of photosynthesis utilize the products of the light reactions, ATP and NADPH, to reduce the products of the Rubisco carboxylase reaction, 3-phosphoglyceric acid (3-PGA). These now reduced products (triose-phosphates) eventually go on to simple sugars and then to carbohydrates, lipids, amino and nucleic acids, and so on, by the usual metabolic pathways. ATP from the light reactions is also used to regenerate the substrate of Rubisco, RuBP, that is common to both photorespiration and photosynthesis.

Because of its important and unique role, the activity of Rubisco could be a significant factor in an explanation of dominance of any algal group. The objective of this study is to examine the structure of Rubisco LSU, its amino acid composition, from representatives of diverse algal groups, but emphasizing marine chromophytes, and to compare these structures, phenetically to reveal possible phylogenetic relationships and chloroplast origins.

Structure

Rubisco is a multimeric protein. In higher plants, algae, and most photosynthetic microorganisms it consists of two types of subunits, known as the large subunit (LSU) (mol. wt. near 56 kD) and the small subunit (SSU) (mol. wt. near 14 kD), with eight copies of each (LSU_8SSU_8) (Miziorko and Lorimer 1983). The only other well established alternate configurations are those for *Rhodospirillum rubrum,* which is a dimer of large subunits (LSU_2) (Tabita and McFadden 1974), and for *Rhodobacter sphaeroides,* which has two forms of Rubisco, the LSU_8SSU_8 enzyme and an LSU_6 enzyme (Gibson and Tabita 1977, Chory *et al.* 1985). Those configurations that have both types of subunits are designated form I Rubisco, and those that have only the LSU are form II. The large subunit contains the catalytic and activation sites; the small subunit is required for catalysis in species possessing it, otherwise its function is not known but is thought to be regulatory (Miziorko and Lorimer 1983).

The large subunit is coded for by chloroplast DNA, translated by 70S chloroplast ribosomes to a larger precursor, which is cleaved to its mature length. The primary structure of the LSU protein has been compared for evolutionarily distant species. An 85% homology was observed between the eukaryotes *Chlamydomonas* and *Zea mays* (Miziorko and Lorimer 1983). Homology is reduced to 80% when cyanobacteria are included (McFadden *et al.* 1986). However, comparing LSU amino acid sequences of the anoxygenic bacterium, *Rhodospirillum rubrum,* with those of higher plants gives only a 31% homology but active site residues are highly conserved (McFadden *et al.* 1986). It is suggested that the gene for the LSU in *R. rubrum* is a distinct gene, different from that of higher plants, and that this gene arose before the appearance of O_2 (McFadden *et al.* 1986). By physical characteristics (molecular size and subunit composition), Rubisco from a range of chromophytic algae is similar to cyanobacterial, green algal, and vascular plant Rubisco (Plumley *et al.* 1986), but by immunological characteristics, they are distinguishable.

In plants and green algae the small subunit is coded, as a small family of genes, by nuclear DNA (McFadden and Small 1988, Berry-Lowe *et al.* 1982). Translation by cytoplasmic 80S ribosomes yields a higher molecular weight precursor that is transported into the chloroplast by an ATP-dependent mechanism. In the stroma the amino-terminal extension is cleaved by a specific protease, forming the mature subunit now able to assemble with the LSU to form the holoenzyme. Light has a significant role in stimulating the transcription of SSU DNA; this effect seems to be mediated via phytochrome and possibly other light detector molecules (Nagy *et al.* 1986). In contrast with the LSU, the SSU protein sequence generally is not as well conserved: the homology amongst different higher plants is 70 to 75% (Akazawa *et al.* 1984) whereas a 40% homology was seen when comparing a cyanobacterium with higher plants (McFadden *et al.* 1986).

In photosynthetic prokaryotes such as the cyanobacteria, obviously, the SSU is coded by the single genome. For two species, at least, the SSU and LSU genes are adjacent and cotranscribed (Nierzwicki-Bauer *et al.* 1984, Shinozaki and Sugiura 1985). A recent discovery has shown that the SSU in a chromophytic alga (*Olisthodiscus luteus*) is coded by chloroplast DNA, again adjacent to the LSU gene (Reith and Cattolico 1985, Hardison and Cattolico 1989). In addition, workers in the same lab (Chesnick and Cattolico 1989) have shown that for two species of dinoflagellates (*Peridinium balticum* and *P. foliaceum*) the SSU is coded by chloroplast DNA of the photosynthetic symbiont, and that the gene for the *O. luteus* SSU is structurally similar to those for the *Peridinium* species and also to that of a purple photosynthetic bacterium (*Alcaligenes eutrophus*), but is distinct from that of other plant groups as diverse as cyanobacteria and the higher plants. The phylogenetic and evolutionary implications are significant and many.

Assembly

One of the tenets of biochemistry holds that all of the information necessary for assembly of a functional protein is resident in its primary sequence. Reasonably, this was thought to apply to Rubisco. However, recent work (eg. Ellis and van der Vies 1988; Roy *et al.* 1988) has illuminated the role of a "chaperone molecule", dubbed the Rubisco subunit binding protein, whose function appears to be to non-covalently bind newly synthesized large and small subunits of Rubisco to form the competent holoenzyme. The Rubisco subunit binding protein is now called cpn60 and is one member of a class of proteins called chaperonins, which in turn are members of the larger class of molecular chaperones. These chaperonins are likely present in all living organisms and are highly conserved. The mechanism appears not to be an active direction of folding but rather a binding of non-native forms and inhibition of aggregation (Gatenby 1991). Mg-ATP has been seen to promote the reversible dissociation of the binding protein, but not via a phosphorylation or adenylation mechanism. These observations could help to explain the reported requirement for ATP of the assembly of Rubisco LSU into the holoenzyme. An obligatory nature of this mechanism *in vivo* has not yet been established, but this protein may help to explain several *in vitro* observations. Rubisco large subunits isolated from higher plants self-aggregate and are insoluble in aqueous media, whereas LSU from prokaryotes remain soluble and are capable of assembly with SSU to form active enzyme (Ellis and van der Vies 1988). This Rubisco subunit binding protein is a multimer of two kinds of subunits (α and β). There are six copies of each, both encoded by nuclear DNA.

Like the subunits of Rubisco, the binding protein subunits are synthesized in a precursor form larger than the mature form. Light also plays a part in accumulation of the binding protein but not to the same degree as for Rubisco. Chaperonin production is also stimulated by stresses, particularly heat, hence their previous designation, as one group of the heat shock proteins (Gatenby 1991).

Activation

Another factor in the net activity of Rubisco is the activation state of the enzyme. Until aspects of the requirements for activation of Rubisco were elucidated, an anomaly of its kinetics troubled students of Rubisco. The K_m (CO_2) for chloroplasts and leaves was in the range of atmospheric [CO_2], but for isolated Rubisco was ten times higher (Ogren *et al.* 1986). Then it was realized that CO_2 and Mg^{2+} participated in activation of the functional enzyme. As it is now understood, CO_2 carbamylates an uncharged epsilon amino group of a lysine residue ($^{\#}$201 in the LSU of spinach). The divalent metal ion (*in vivo,* Mg^{2+}) then coordinates with the resultant negatively charged carbamate. The functional significance of these activation steps is that the metal ion stabilizes a negatively charged transition state intermediate upon carboxylation of RuBP. Various phosphorylated esters (eg. 6-phosphogluconate, fructose-1,6-bisphosphate, or NADPH) are conceived of as effectors of Rubisco by their interaction with the active site (Ogren *et al.* 1986). The substrate, RuBP, can bind to the active site before it is activated. Under these conditions it acts as an inhibitor because dissociation is slow, and hence activation is slowed. *In vitro* the addition of activating CO_2 precedes the addition of Mg^{2+} and is the slow step of activation. Light plays a role in the activation of Rubisco by increasing the concentration of Mg^{2+} (to balance cations, countering the effect of photosynthetic efflux of H^+ from the stroma) and the pH (necessary for maintenance of the uncharged epsilon-amino group). However, a discrepancy still exists: *In vitro* the K_{act} (CO_2) under these conditions is 25-30 μM CO_2, while *in vivo* full activation is observed at 10 μM CO_2.

Another level of Rubisco activity regulation has been identified in several species of plants. Carboxyarabinitol 1-phosphate (CA1P) can operate as an inhibitor by tightly binding the active site. Structurally it resembles an intermediate of the Rubisco carboxylase reaction, 2-carboxy-3-keto-arabinitol 1,5-bisphosphate. CA1P seems to be confined to plants but not all, and its distribution follows no taxonomic pattern (Seemann *et al.* 1990). The amount of CA1P accumulated in the dark varies greatly amongst these species. In those species where it occurs, light is again the controlling factor for its effect, with light being required for both degradation of CA1P and recovery of Rubisco activity (below).

There has recently come to light the existence of a protein which promotes the activation of Rubisco (Salvucci *et al.* 1985). This Rubisco activase decreases the level of CO_2 required for activation to within physiological range (ca. 8 μM) (Werneke *et al.* 1988). The activity of Rubisco activase is dependent on Mg^{2+} and ATP in an ATP hydrolysis reaction (Streusand and Portis 1987). One of its functions appears to be to aid carbamylation of the enzyme-RuBP complex, thus contributing to activation (Portis 1990). Also important to activation, it removes both RuBP from the inactivated (non-carbamylated) form of Rubisco (Wang and Portis 1992), and CA1P from the carbamylated form (Robinson and Portis 1988). As mentioned above, both of these are endogenous inhibitors of Rubisco (Jordan and Chollet 1983, Seemann *et al.* 1990). It also prevents and reverses the decline in Rubisco activity with time (fallover) due to isomerization of RuBP in the carbamylated active site (Edmondson *et al.* 1990). Another possible function may be to change concentrations of phosphate ester effectors of Rubisco (Parry *et al.* 1988). Thus Rubisco activase and molecules like RuBP and CA1P have opposing regulatory roles, activation and deactivation of Rubisco.

Rubisco as a phylogenetic indicator

Taxonomy and phylogenetics must deal with whole organisms and find characters with which to define and distinguish groupings of these organisms. For eukaryotic algae, many of the defining characters are chloroplast-based and others are those of the cell body or nucleus. The endosymbiotic theory postulates that the organelles have had a different history from that of the "host". This theory has been generally accepted, at least for mitochondria and chloroplasts. Some of the implications of the endosymbiotic theory have been incorporated into taxonomic and phylogenetic schemes. However, the genealogical relationships among the higher order taxa, division and above, for the algae are still a matter of dispute, as evidenced in the uncertainty of grouping of the algal phyla as outlined above. The root of this problem may be in failure to adequately allow for the endosymbiotic theory. The phylogeny of the chloroplast should be separated from that of the cell body to resolve these issues. Because of the antiquity of the endosymbiotic event(s), and because of the morphological simplicity of chloroplasts, partly as a result of their high degree of reduction since endosymbiosis, molecular techniques are required to ascertain the histories of chloroplasts. The Rubisco large subunit is an ideal molecule for this task. Rubisco has a long history: it is present in many bacterial forms as well as in all photosynthetic eukaryotes. In the latter, the Rubisco gene (*rbcL*) is chloroplast encoded. It is sufficiently large to provide enough characters in analysis of sequences. It is sufficiently conserved to follow both similarities and differences within the taxa concerned, representatives of the ancestors of the endosymbionts as well as the descendants. This may be related to its central role in carbon metabolism.

The chromophyte algae have been particularly difficult with respect to yielding Rubisco protein for study (Rothschild and Heywood 1987). The alternative, gene sequencing, involves considerable investment of resources, and is also subject to many of the same problems. Recent advances have accelerated the process somewhat, and the

Rubisco sequence data base is expanding (Martin *et al.* 1992). The method presented here, of phenetic analysis of Rubisco LSU amino acid compositions for the determination of phylogenetic relations, represents a substantial shortcut over gene and protein sequencing and allows examination of many groups that have resisted analysis by other methods. The relations indicated by this method will be compared with those from analysis of Rubisco sequences and of other molecules. These results will be examined for information they can provide about the origin of chloroplasts.

Methods and Materials

Algal species

(and clonal designation) Classification according to Bold and Wynne 1985.

- D. Pyrrhophyta
  - C. Dinophyceae
    - *Amphidinium carterae* Hulburt (NEPCC clone "Amphi")

- D. Chrysophyta
  - C. Prymnesiophyceae
    - O. Isochrysidales
      - *Isochrysis* sp. (Parke) (clone T-ISO)

- D. Cryptophyta
  - *Chroomonas salina* (Wislouch) Butch. (Woods Hole clone "3C")

Diatom classification according to Round, Crawford and Mann 1990.

- D. Bacillariophyta
  - C. Coscinodiscophyceae
  - s.C. Thalassiosirophycidae
    - O. Thalassiosiarales
      - F. Thalassiosiraceae
        - *Thalassiosira weissflogii* Grun. (Woods Hole clone "Actin")
      - F. Skeltonemataceae
        - *Skeletonema costatum* (Grev.) Cl. (from Saanich Inlet, B. C.)
  - s.C. Chaetocerotophycidae
    - O. Chaetoceratales
      - F. Chaetoceraceae
        - *Chaetoceros similis* Cl. (from NEPCC)

Cell culture

Media

All cultures were maintained in nutrient-enriched filtered seawater collected from Saanich Inlet. *Amphidinium carterae* Hulburt and *Chaetoceros similis* Cl. were grown with HESNW (modified) medium (Harrison *et al.* 1980) with nitrate level at 550 μM. All other species were grown in f/20 (Guillard and Ryther 1962) enriched medium, with 50 μM nitrate. Air was bubbled through cultures larger than 2 litres. Typically, 42 L cultures were used for amino acid analyses. This volume was achieved by doubling the culture daily to ensure actively growing cells (Hobson *et al.* 1985).

Light

The light source was a bank of 300W cool-white fluorescent tubes, giving 10^{16} quanta cm^{-2} s^{-1} at the vessel surfaces. The photoperiod was 12:12 (h light:h dark).

Temperature

Temperature was maintained in the growth chamber at 20°C. This temperature was chosen as a compromise between 20°C for best growth for *Skeletonema costatum* (Grev.) Cl. (Smayda 1973) and 25°C, the best temperature for maximum effect of photoperiod on growth (Hobson 1974).

Isolation and purification techniques.

All procedures under this heading were performed at 0 to 8°C.

Cell harvest and disruption

Cells were harvested in later stages of exponential growth to balance the needs for maximum yield and actively growing cells. They were collected by centrifugation through a Sorvall KSB "Szent-Gyorgyi and Blum" Continuous Flow System in a Sorvall RC2-B centrifuge, at 3000 x g. The majority of the supernatant was decanted and the pellets were combined and centrifuged at 3000 x g for 15 minutes. The cells were resuspended in minimal cold Buffer A (Newman *et al.* 1989) (25mM 2-amino-2-

(hydroxymethyl)-propane-1,3-diol (Tris) pH 8.1, 10mM $MgCl_2$, 10 mM KCl, 1mM dithiothreitol (DTT), 0.5mM ethylene glycol bis(β-aminoethyl ether) *N,N,N`,N`* -tetraacetic acid (EGTA), 0.7 μg/mL leupeptin, 0.7 μg/mL pepstatin). Collected cells were lysed by sonication, using pulses of 1 s on,1 s off (Heat Systems-Ultrasonics Inc. model W-375). Some cells, *Isochrysis* sp. (Parke) and *Chroomonas salina* (Wislouch) Butch. were easily broken open by 3 minutes of this treatment. Other cells, the diatoms and *A. carterae,* were resistant and required longer periods of sonication for adequate yield. The lysate was then centrifuged at 18000 x g for 20 minutes to clear it of cells and subcellular debris. The lysate was further centrifuged at 100000 X g for 40 minutes.

Ammonium-sulfate differential precipitation

Saturated ammonium sulfate in Buffer A was added (to 10% for diatoms and 20% for others) to the lysate, allowed to precipitate with gentle shaking for 40 minutes, then centrifuged at 10000 x g for 20 minutes. The supernatant was increased to 55% saturated ammonium sulfate, precipitated and centrifuged.

Sucrose density gradient ultracentrifugation

Lysate proteins were partially purified from other soluble cellular components, such as phenolics, by sucrose density ultracentrifugation. The pellet was redissolved in a minimal amount of Buffer A ($\leq$ 2 mL/24mL tube of sucrose density gradient) and sucrose added to 2%. This was layered on top of a linear 10 to 40% sucrose density gradient. The sucrose density gradient was formed by layering 6 mL each of 10%, 20%, 30%, and 40% sucrose in Buffer A, each applied at the bottom of the non-wettable 33 mL bell-top quick-seal ultracentrifuge tube. The tubes were laid on their sides for 2 h to convert the step gradient to linear (Griffith 1986). The gradient and sample were then overlayed with Buffer A to fill the tube. The gradient was balanced and centrifuged at 184,000 x g for 2 h after slow acceleration.

Analysis of fractions

Fractionation

The top of the tube was cut, the bottom of the tube was pierced and the contents allowed to flow from the bottom through a fraction collector where 1-2 mL fractions were collected. Each fraction was monitored for UV absorbance (280 and 260 nm). Rubisco was found in fractions where A_{280} exceeded A_{260}.

Disc-SDS-PAGE

Aliquots from selected fractions were mixed 1 to 1 with 2X sample buffer (125mM Tris pH 6.8, 4% sodium dodecyl sulfate (SDS), 0.02% bromphenol blue, 10% β–mercaptoethanol, 17.5% glycerol) and heated to 37°C for 10-15 minutes. The discontinuous SDS-polyacrylamide gel electrophoresis (Disc-SDS-PAGE) system consisted of a 12% separating gel (12% acrylamide, 0.11% bis-acrylamide, 0.1% SDS, 375 mM Tris pH 8.8, 0.03% *N, N, N*`, *N*` -tetramethyl-ethylenediamine (TEMED), 0.03% ammonium persulfate) and a 3% stacking gel (3% acrylamide, 0.1% bis-acrylamide, 125mM Tris pH 6.8, 0.1% SDS, 0.17% TEMED, 0.17% ammonium persulfate). After 24-72 h polymerization, the separating gels were pre-electrophoresed 2 h at 3 mA with 5μM glutathione in electrophoresis buffer (Towbin 1979) (25 mM Tris, 192 mM glycine, pH=8.2, 0.1% SDS w/v). The stacking gels were poured and used immediately upon polymerization. Pre-stained molecular weight standards and a partially purified spinach Rubisco standard (Sigma Chemical Company) were run alongside the 20 μL prepared sucrose density fractions. Gels were developed, using electrophoresis buffer with 0.1 mM thioglycolate, at 8-10 mA through the stacking gel then 12-15 mA until the dye front approached the bottom of the gel. When not electroblotting, gels were Coomasie stained for 2h (45% methanol; 9% acetic acid; 0.25% Coomasie blue, v:v:w) and destained (45% methanol; 9% acetic acid, v:v).

The fraction with the highest Rubisco concentration, judged by the strength of the Coomasie stained band that dominated near the size of the spinach standard Rubisco LSU, was loaded in consecutive wells on a similarly prepared gel and run as above.

The analysis of spinach Rubisco used semi-purified Rubisco supplied by Sigma Chemical Company. It was dissolved in 1X sample buffer and analyzed as for all others.

Electroblotting

The second gel was equilibrated 5 minutes in transfer buffer (12.5 mM Tris, 86 mM glycine, 10% methanol, pH=8.2). A pair of polyvinylidine difluoride (PVDF) sheets were wetted briefly in 100% methanol and then equilibrated 5 minutes in transfer buffer. The transfer sandwich was assembled using a nylon scrubbie pad and double layers of Whatman #3 filter paper on either side of the gel and membranes, with the membranes to the anode side. The proteins were transferred at 90 V for 18 minutes at 20°C. They were rinsed 3 times in H_2O for 5 minutes each, stained 5 minutes in 0.025% Coomasie R-250, 40% methanol, and destained 10-15 minutes in 50% methanol. The bands were cut out while still wet and stored at -20°C until analyzed.

Amino acid analysis

The excised PVDF-bound protein band was overlaid with argon and hydrolyzed in approximately 425μL 6N HCl with 5% phenol at 165°C for 60-70 minutes. The digested protein on PVDF was then analyzed on an automated Applied Biosystems amino acid analyzer (Model 420) via phenylisothiocyanate derivatization of the free amino acids followed by reversed-phase HPLC separation of the phenylthiocarbonyl (PTC) residues. The hydrolyzed and derivatized sample was loaded onto a C-18 microbore column using a sodium acetate buffer (29 mM, pH 5.0). The derivatized amino acids were eluted with a mixed slope gradient ranging from 5 to 70% acetonitrile (Buffer A: 50 mM sodium acetate, pH 5.4; Buffer B 70% acetonitrile, 30% sodium acetate (32 mM, pH 6.1)) over 20 minutes. Detection was at 254 nm and peak heights were used to convert to pmol.

Each run was standardized using 100 pmol of a synthetic peptide. If there was less than 90% recovery of each amino acid from the standard, which was rare, the experimental data were corrected to 100%. Therefore the amounts of amino acid reported are within 10%. Because the concentration of Rubisco on the band was not known, the length of membrane strip required to result in a reasonable amount of protein in the analyzer was estimated by eye. To give some allowance for error in this estimate, several analyses from the same electroblot were performed with varying amounts of membrane. This method was not successful in avoiding low amounts of protein. These results should be considered pseudo-replicates for a single preparation.

Where data were obtained as mole percent values or proportions relative to a standard amino acid (eg. Ala or Phe) these data were used directly and not converted to number of residues. Thus, the SΔn values were calculated via the SΔQ statistic where proportions could be applied.

Amino acid compositions were also gathered from the literature (Table 1) (Appendices 2, 3). In some cases these were obtained as amino acid composition data via similar methods. Others were determined from protein sequenced by enzymatic and chemical cleavages and subsequent Edman degradation. Many amino acid compositions were extracted from deduced protein sequences from sequencing of the Rubisco LSU gene. The number of amino acid residues ranged from 454 to 508 when considering those residues in common with the amino acid analyses. The quantities for the amino acid pairs aspartate, asparagine and glutamate, glutamine were combined in all cases because amino acid analyses cannot distinguish them (due to acidification of the amide group). Tryptophan cannot be determined in amino acid composition analysis because the acid hydrolysis destroys it. Cysteine also is not reliably determined. Thus, tryptophan and cysteine were not included in the analysis, even if the information was available.

Table 1. Code and taxonomic classification of species whose Rubisco LSU amino acid compositions were extracted from the literature. Type of source data: .A) amino acid composition, .G) DNA sequence, .P) protein sequence. References below.

Genus species (strain)	Name code & data type	Classification	ref.
Acidovorax facilis = *Hydrogenomonas facilis*	Acido.A	β purple bacterium	17
Alcaligenes eutrophus	Alcal.G	β purple bacterium	4
A. eutrophus	Alcal1.A	β purple bacterium	5
A. eutrophus = *Hydrogenomonas eutropha*	Alcal2.A	β purple bacterium	17
Anabaena sp. 7120	Anab.G	cyanobacterium	6
Antithamnion spec.	Antith.G	rhodophyte	16
Beta vulgaris	Beta1.A	chlorophyte plant	11
B. vulgaris	Beta2.A	chlorophyte plant	21
Chlamydomonas reinhardii	Chlamy.G	chlorophyte alga	8
Chlorella ellipsoidea	Chlore.G	chlorophyte alga	39
C. ellipsoidea	Chlore.A	chlorophyte alga	31
Chlorella sorokiniana	Chlors.A	chlorophyte alga	33
Chromatium sp. D	ChromD.A	γ purple bacterium	32
Chromatium vinosum	Chromv.G	γ purple bacterium	37
Cryptomonas Φ	Crypto.G	cryptophyte chromophyte	7
Cyanidium caldarium	Cyanid.G	rhodophyte	35
Cyanophora paradoxa	Cyanoph.G	uncertain affinity	36
Cylindrotheca sp. N1	Cylindro.G	chrysophyte chromophyte	14
Euglena gracilis	Euglena.G	euglenoid	10
Marchantia polymorpha	March.G	chlorophyte plant	25
Medicago sativa	Medic.G	chlorophyte plant	2

Table 1. cont.

Genus species	Name code	Classification	ref.
Nicotiana tabacum	Nico.P	chlorophyte plant	3
N. tabacum	Nico.G	chlorophyte plant	28
N. tabacum	Nico.A	chlorophyte plant	15
Olisthodiscus luteus	Oli.A	chrysophyte chromophyte	24
O. luteus	Oli.G	chrysophyte chromophyte	12
Oryza sativa	Oryza.G	chlorophyte plant	20
Petunia hybrida	Petunia.G	chlorophyte plant	1
Phaseolus vulgarus	Phaseo.A	chlorophyte plant	11
Pisum sativum	Pisum.G	chlorophyte plant	41
Porphyridium aerugineum	Porphy.G	rhodophyte	34
Prochlorothrix hollandica	Prochl.G	prochlorophyte	22
Rhodobacter sphaeroides form I	Rbactfl.G	α purple bacterium	9
R. sphaeroides form II	Rbactf2.G	α purple bacterium	38
Rhodospirillum rubrum	Rrub.P	α purple bacterium	13
R. rubrum	Rrub.G	α purple bacterium	23
R. rubrum grown on CO_2	Rrub2.A	α purple bacterium	27
R. rubrum grown on butyrate	Rrub1.A	α purple bacterium	27
Spinacea oleracea	Spin.G	chlorophyte plant	40
S. oleracea	Spin4.A	chlorophyte plant	30
S. oleracea	Spin5.A	chlorophyte plant	26
S. oleracea	Spin6.A	chlorophyte plant	15
Synechococcus sp. 6301	Synech.G	cyanobacterium	29
Xanthobacter flavus H4-14	Xbactfl.G	α purple bacterium	19
Zea mays	Zea.G	chlorophyte plant	18
Cow glutamate dehydrogenase	CowGDH.A	outgroup	42

References to Table 1.

1. Aldrich. J. *et al.* 1986a. Nucl. Acids Res. 14:9534.
2. Aldrich. J. *et al.* 1986b. Nucl. Acids Res. 14:9535.
3. Amiri, I. *et al.* 1984. Biochim. Biophys. Acta 784:116-123.
4. Andersen, K. and Caton, J. 1987. J. Bacteriol. 169:4547-58.
5. Bowien, B. and Mayer, F. 1978. Eur. J. Biochem. 88:97-107.
6. Curtis, S.E. *et al.* 1983. Proc. Nat. Acad. Sci. U.S.A. 80:1835-9.
7. Douglas, S. E. *et al.* 1990. J. Phycol. 26:500-8.
8. Dron, M. *et al.* 1982. J. Mol. Biol. 162:775-793.
9. Gibson, J. L. *et al.* 1991. J. Biol. Chem. 266:14646-14653.
10. Gingrich, J. C. and Hallick, R. B. 1985. J. Biol. Chem. 260:16162-8.
11. Gray, J. C. and Kekwick, R. G. O. 1974. Eur. J. Biochem. 44:481-489.
12. Hardison, L. K. *et al.* 1992. Plant Mol. Biol. 18:595-599.
13. Hartman, F. C. *et al.* 1982. Arch. Biochem. Biophys. 219:422-437.
14. Hwang, S.-R. and Tabita, F. R. 1991. J. Biol. Chem. 266:6271-79
15. Kawashima, N. 1969. Plant and Cell Physiol. 10:31-40.
16. Kostrzewa, M. *et al.* 1990. Curr. Genet. 18:465-469.
17. Kuehn, G. D. and McFadden, B. A. 1969. Biochem. 8:2403-2408.
18. McIntosh, L. *et al.* 1980. Nature (Lond.) 288:556-560.
19. Meijer, W. G. *et al.* 1991. Mol. Gen. Genet. 225:320-330.
20. Moon, E. *et al.* 1987. Nucl. Acids Res. 15:611-630.
21. Moon, K. E. and Thompson, E. 1969. Aust. J. Biol. Sci. 22:463-70.
22. Morden, C. W. and Golden, S. S. 1991. J. Mol. Evol. 32:379-395.
23. Nargang, F. *et al.* 1984. Mol. Gen. Genet. 193:220-4.
24. Newman, S. M. and Cattolico, R. A. 1987. Plant Phys. 84:483-490.
25. Ohyama, K. *et al.* 1986. Nature (Lond.) 322:572-4.

26. Rutner, A. and Lane, M. 1967. Biochem. Biophys. Res. Com. 28:531-7.

27. Schloss, J. V. *et al.* 1979. J. Bacteriol. 137:490-501.

28. Shinozaki, K. and Sugiura, M. 1982. Gene 20:91-102.

29. Shinozaki, K. *et al.* 1983. Proc. Nat. Acad. Sci. U.S.A. 80:4050-4.

30 Sugiyama, T. and Akazawa, T. 1970. Biochem. 9:4499-4504.

31. Sugiyama, T. *et al.* 1971. Biochem. 10:3406-11.

32. Takabe, T. and Akazawa, T. 1975. Arch. Biochem. Biophys. 169:686-94.

33. Toman, P. D. *et al.* 1985. Plant Phys. 79:806-814.

34. Valentin, K. and Zetsche, K. 1989. Curr. Genet. 16:203-9.

35. Valentin, K. and Zetsche, K. 1990a. Mol. Gen. Genet. 222:425-30.

36. Valentin, K. and Zetsche, K. 1990b. Curr. Genet. 18:199-202.

37. Viale, A. M. *et al.* 1989. J. Bacteriol. 171:2391-2400.

38. Wagner, S. J., *et al.* 1988. FEMS Microbiol. Lett. 55:217-222.

39. Yoshinaga, K. *et al.* 1988. Plant Mol. Biol. 10:245-250.

40. Zurawski, G. *et al.* 1981. Nucl. Acids Res. 9:3251-70.

41. Zurawski, G. *et al.* 1986. Nucl. Acids Res. 14:3975.

42. Smith. E. L. *et al.* 1970. Proc. Natl. Acad. Sci. U.S.A. 67:724-730.

Data Analyses

Dendrogram

Amino acid compositions were compared using the SΔQ index of Marchalonis and Weltman (1971) and the SΔn index of Cornish-Bowden (1977).

$$S\Delta Q = 10^4\ \Sigma(^{niA}/_{N_A} - {}^{niB}/_{N_B})^2$$

where, for proteins A and B, whose amino acid residues number N_A and N_B, respectively, n_i is the number of the ith type of residue in the protein, and these (N and n_i) are summed over only those types of amino acid that are in common to analyses for both proteins. The factor of 10^4 shifts the decimal place to yield more readable numbers and can be rationalized as a conversion from decimal fraction to percentage (x 10^2) and then squared ($(10^2)^2 = 10^4$).

The SΔn statistic can be calculated from SΔQ values.

$$S\Delta n = 0.5 \times 10^{-4}\ N^2\ S\Delta Q$$

where, N is the number of amino acid residues in the sequences being compared. SΔn is also restricted to the amino acid types common to all amino acid compositions. Assuming a uniform amino acid residue total of 462 (N), as calculated from the average of the 31 gene sequences used (mean=461.94; $s.d._{(n-1)}$=8.53; max=482; min=450), then SΔn = 10.67 X SΔQ. In original terms,

$$S\Delta n = {}^1/_2\ \Sigma(n_{iA} - n_{iB})^2$$

This distance was calculated comparing every taxon with every other, and a matrix of SΔn values was thus obtained. This statistic is used here for presentation and discussion because its property of estimating the number of sequence differences allows comparisons with literature using sequences only. Phylogenetic trees were constructed from matrices of SΔn values using the unweighted pair-group method with arithmetic averages (UPGMA) (eg. Sneath and Sokal 1973). Calculation was aided by entering the matrix in a computer program, NTSYS-pc Version 1.50 (Applied Biostatistics Inc. 1989), that performed the UPGMA cluster analysis under the heading, sequential agglomerative hierarchical nested methods, SAHN. This program also produced a graphical output of the tree (Treeg).

The scale of the tree includes confidence limits that were calculated from theoretical considerations (Cornish-Bowden 1979, 1983). These limits are to provide evidence that Rubisco from different organisms are related. The cut off for a strong test of significance (95%) is 0.35 X N, which in this case is $S\Delta n$=161.7. A weak test of significance, or sequence identity due only to chance, results in a cut off at 0.93 X N, or $S\Delta n$= 429.7.

Data sets analyzed by this method were: 1) separate pseudo-replicate values for the seven species of this study; 2) the averages of pseudo-replicates for the seven species; 3) these seven species plus all amino acid compositions found in the literature; 4) as for 3, but including amino acid compositions deduced from protein sequences.

PCA

Principal components analysis (PCA) was performed on the first 3 data sets used for cluster analysis. Also, a subset of the fourth data set was analyzed. The software used (Rohlf 1988) had a limit of 400 variables; thus, with 16 amino acids, a maximum of 25 species could be included. In addition, to represent these graphically a small number of species is more practical. For the fourth data set, species were chosen to represent as broad a taxonomic range of organisms as possible, while including a balance of data types.

Input data were the percent of each amino acid in Rubisco LSU. These were standardized by computing the mean and standard deviation for the range of percentage of each amino acid and the difference of each percentage from the percentage mean; this was expressed in standard deviation units by dividing the difference by the standard deviation and computed using NTSYS-pc ("Stand") (Rohlf 1988). A correlation matrix was constructed by calculating the correlation between each amino acid from the standardized data (NTSYS-pc "Simint" c=corr). Axes, or eigenvectors, were extracted from this correlation matrix (NTSYS-pc "Eigen"); eigenvalues were also computed. The species were projected onto the first three axes (NTSYS-pc "Proj") and their coordinates presented in table form and graphically (NTSYS-pc "Mod3dg").

The importance of each amino acid in the PCA was calculated from the eigenvalues and the eigenvector coefficients for each amino acid. The absolute value of the coefficient for an amino acid in the eigenvector equation for a given dimension represents the relative contribution of that amino acid to the dimension. To reflect the weight of the dimension relative to others, that is the amount of variance in the original data that the dimension accumulates, each eigenvector coefficient was multiplied by the eigenvalue (%) for a given dimension. The importance of the amino acid to the PCA representation of the original data was calculated by summing over the three dimensions:

$$\text{importance}_A = \Sigma_{D=1,2,3}(\text{eigenvalue\%}_D \text{ X } |\text{eigenvector}_A|)$$

where A=amino acid and D=dimension. These values were ordered and the amino acids ranked.

Results

The dendrogram of the mean compositions (Appendix 1) separates three groups (Figure 1): 1) *Amphidinium carterae,* 2) *Spinacea oleracea,* and 3) the chromophytes. The PCA also allows recognition of these three (Figure 2).

Variance in these means is represented in the separate pseudo-replicate amino acid compositions (Appendix 1). The dendrogram of these data delineates five groups (Figure 3): 1) *A. carterae,* 2) *S. oleracea,* 3) four species of chromophytes, 4) one analysis of *Skeletonema costatum* (Skel4.A*), and 5) three other analyses of *S. costatum* with one of *Thalassiosira weissflogii.* In contrast, the PCA of these data could be interpreted as separating out only three groups: 1) *A. carterae,* 2) *S. oleracea,* and 3) the chromophytes (Figure 4).

Protein preparations

While the analytical SDS-PAGE gel for determination of presence of Rubisco (Figure 5) shows the LSU for *Isochrysis* sp. to be faint in this gel, there was a strong band in the preparative gel used for electroblotting. This could not be shown because the gel was consumed in the electroblot and analysis. For all other species what is shown in Figure 5 is representative of what was analyzed. These results show that the preparation was dominated by one band near the size of the *S. oleracea* Rubisco LSU.

Cluster analysis and PCA

Means

In more detail, the dendrogram of the averages of the pseudo-replicates of the seven analyses shows three groups (Figure 1): *A. carterae* at SΔn=311, *S. oleracea* at SΔn=254, and the chromophytes at SΔn=198. As for the chromophytes, three chrysophytes, *C. similis, Isochrysis* sp., and *T. weissflogii,* cluster at SΔn=113. *S. costatum* separates from the other diatoms at SΔn=198, and *Chroomonas salina* appears between *S. costatum* and the other chromophytes at SΔn=185.

Figure 1. UPGMA cluster analysis of SΔn distances between Rubisco amino acid compositions of *Spinacea oleracea* and 6 species of marine algae. Means of pseudo-replicates. Codes: (Skel) *Skeletonema costatum;* (Thal) *Thalassiosira weissflogii;* (Chaet) *Chaetoceros similis;* (Spin) *Spinacea oleracea;* (Iso) *Isochrysis* sp.; (Chroo) *Chroomonas salina;* (Amphi) *Amphidinium carterae;* (M) mean of pseudo-replicates; (.A) amino acid analyses; (*) analyses added by this study.

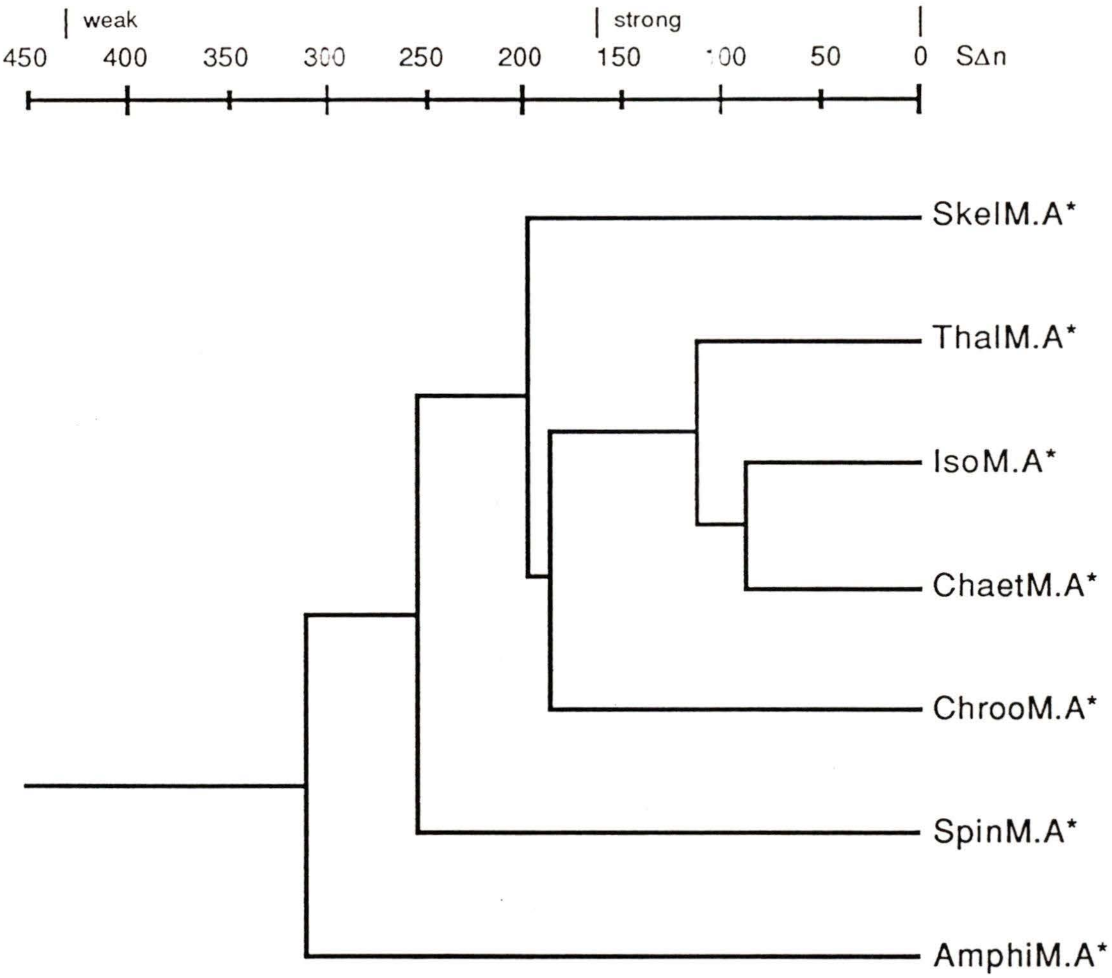
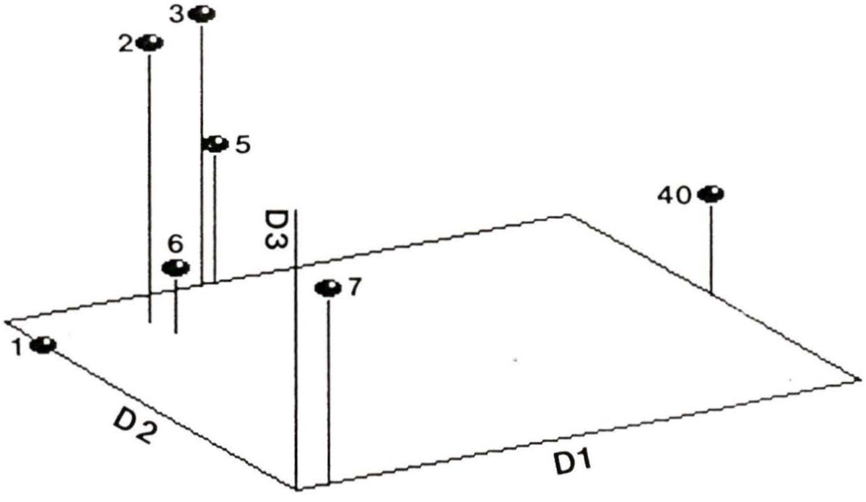

weak
strong
450
400
350
300
250
200
150
100
50
0
SΔn
SkelM.A*
ThalM.A*
IsoM.A*
ChaetM.A*
ChrooM.A*
SpinM.A*
AmphiM.A*

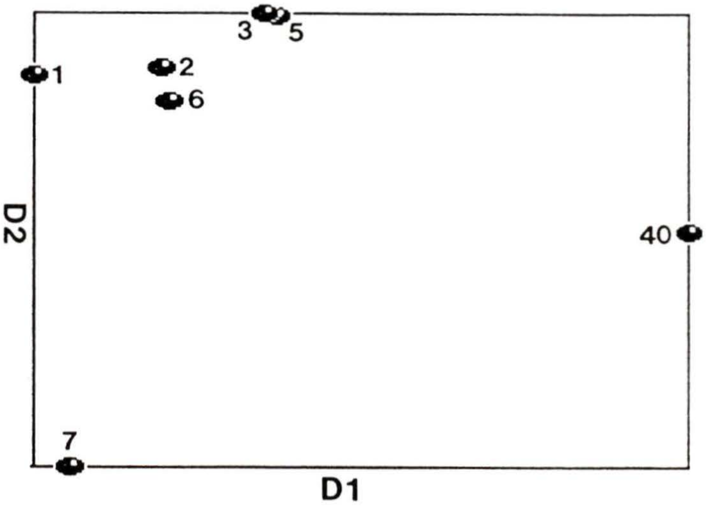
Figure 2. PCA of Rubisco amino acid compositions of *Spinacea oleracea* and 6 species of marine algae. Means of pseudo-replicates. Codes are as for Figure 1: (1) SkelM.A*; (2) ThalM.A*; (3) ChaetM.A*; (40) SpinM.A*; (5) IsoM.A*, (6) ChrooM.A*; (7) AmphiM.A*.

a. Orthogonal projection. b. Dimensions 1 and 2. c. Dimensions 1 and 3.

a.



b.



c.

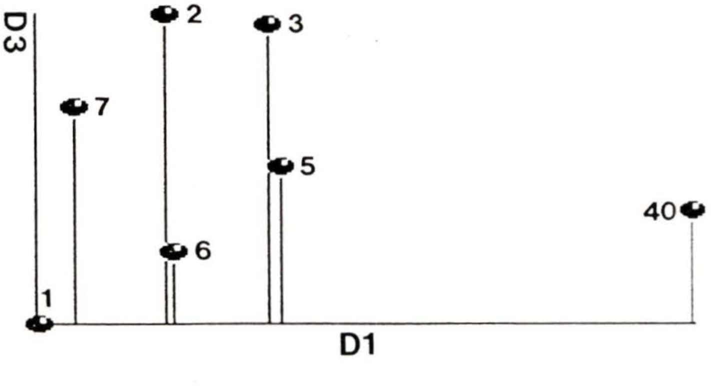


Figure 3. UPGMA cluster analysis of SΔn distances between Rubisco amino acid compositions of *Spinacea oleracea* and 6 species of marine algae. Pseudo-replicates. Codes are as for Figure 1; (1), (2), (3), etc. are separate pseudo-replicate analyses.

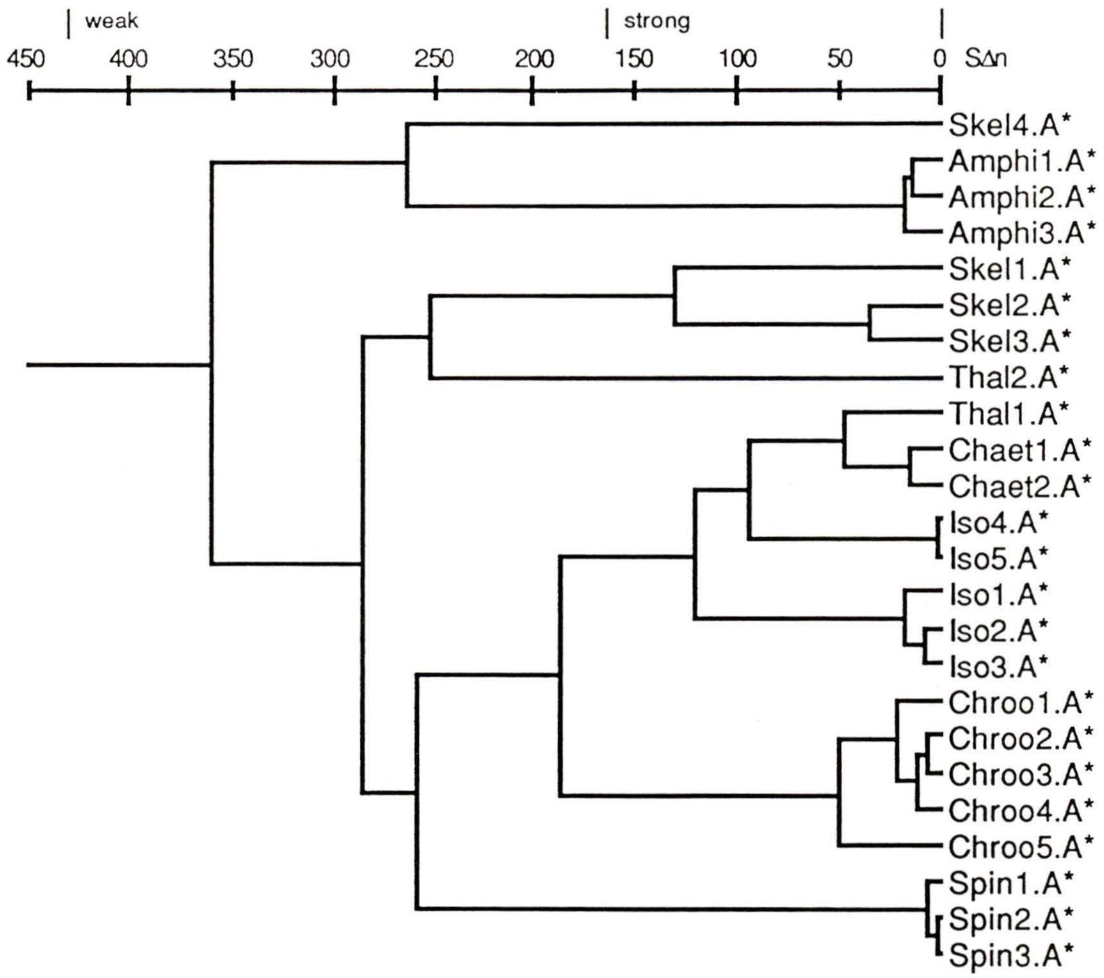

weak
strong
450
400
350
300
250
200
150
100
50
0
SΔn
Skel4.A*
Amphi1.A*
Amphi2.A*
Amphi3.A*
Skel1.A*
Skel2.A*
Skel3.A*
Thal2.A*
Thal1.A*
Chaet1.A*
Chaet2.A*
Iso4.A*
Iso5.A*
Iso1.A*
Iso2.A*
Iso3.A*
Chroo1.A*
Chroo2.A*
Chroo3.A*
Chroo4.A*
Chroo5.A*
Spin1.A*
Spin2.A*
Spin3.A*

Figure 4. PCA of Rubisco amino acid compositions of *Spinacea oleracea* and 6 species of marine algae. Pseudo-replicates. Codes as for Figure 3. Species: (1) Skel1-4.A*; (2) Thal1-2.A*; (3) Chaet1-2.A*; (40) Spin1-3.A*; (5) Iso1-5.A*; (6) Chroo1-5.A*; (7) Amphi1-3.A*.

a. Orthogonal projection. b. Dimensions 1 and 2. c. Dimensions 1 and 3.

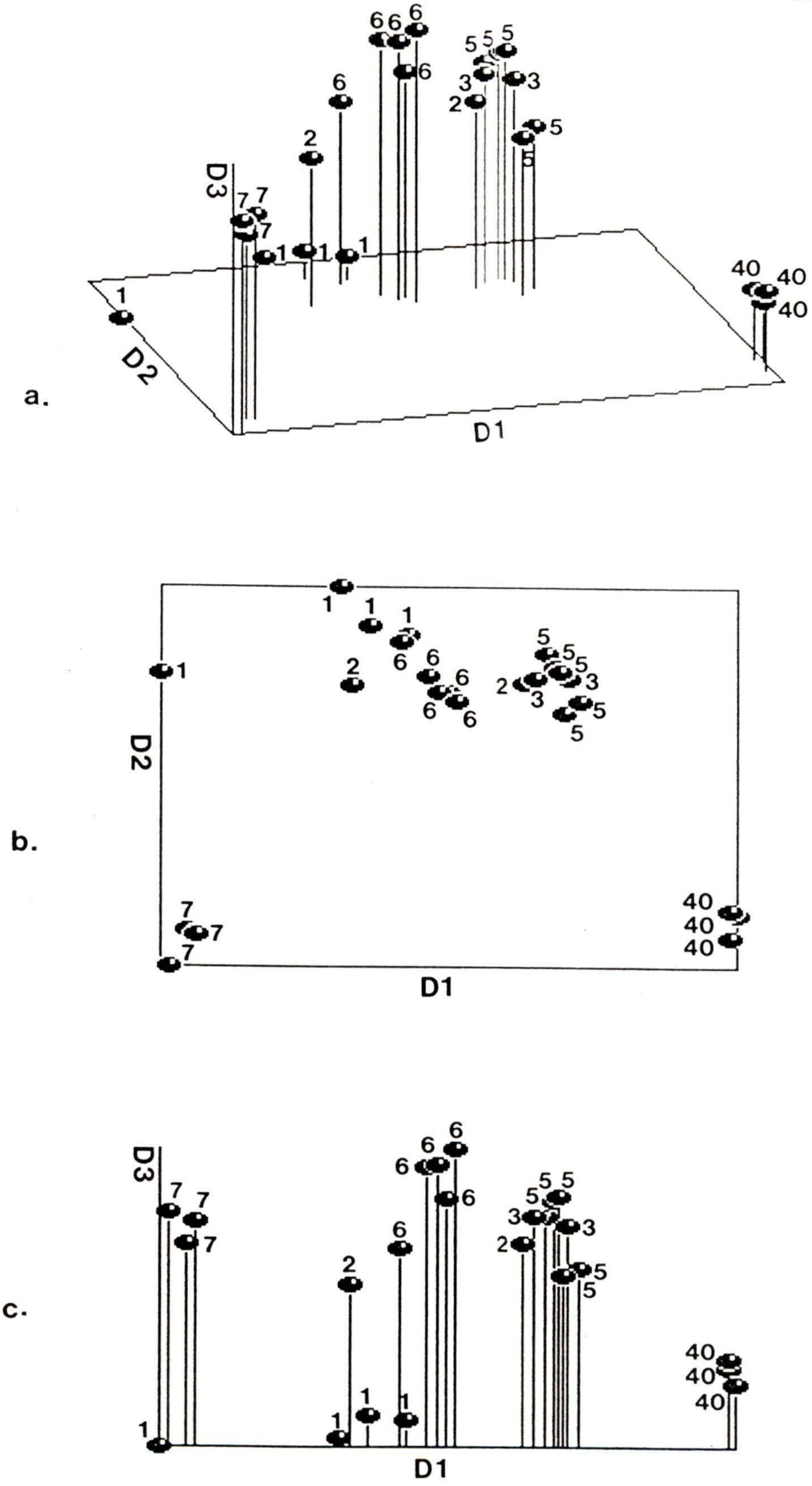
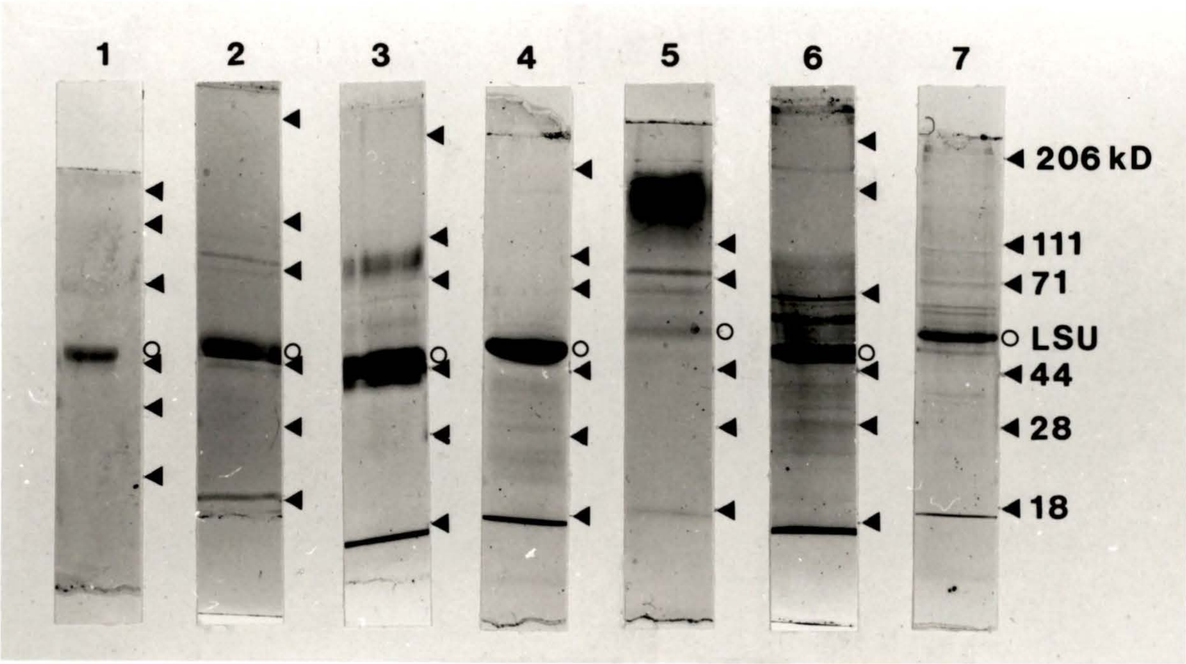

a.
D3
D2
D1
6 6 6
6
6
2
7 7
7
1 1 1
1
5 5 5
3 3
2
5
5
40 40
40
b.
D2
D1
1
1 1 1
2
6 6
6 6
6 6
5 5 5
2 3 3
5
5
7 7
7
40
40
40
c.
D3
D1
7 7
7
6 6 6
6
6
2
1 1 1 1
5 5 5
3 3
2
5
5
40
40
40

Figure 5. SDS-PAGE separation of Rubisco-containing fractions from sucrose density preparations. Lanes: (1) *Skeletonema costatum.* (2) *Thalassiosira weissflogii.* (3) *Chaetoceros similis.* (4) *Spinacea oleracea.* (5) *Isochrysis* sp.. (6) *Chroomonas salina.* (7) *Amphidinium carterae.* Molecular weight markers indicated by arrowheads. Rubisco LSU indicated by (O). The *Spinacea oleracea* preparation was from Sigma Chemicals.


1
2
3
4
5
6
7
206 kD
111
71
LSU
44
28
18

Like the dendrogram, the PCA analysis of the average of pseudo-replicates (Tables 2, 3, 4; Figure 2) allows recognition of three groups, *A. carterae,* the chromophytes, and *S. oleracea.* Fully 75% of the variance in the 16-dimensional space is accounted for by the first three dimensions (Table 2), and both dimension 1 and dimension 2 accumulate near 30%. There is considerable variability in the chromophytes in dimension 3 (Figure 2c).

Pseudo-replicates

The dendrogram of the pseudo-replicates (Figure 3) shows a somewhat different arrangement than the averages. Three of the five groups correspond to those found in Figure 1: *A. carterae* (separating from the others at SΔn=360), four species of chromophytes (the greatest distance between them SΔn=186), and *S. oleracea* (separate from the chromophytes at SΔn=259). There are two other groups consisting of analyses of several diatoms that are widely separated (SΔn=286) from the other chromophytes. One *S. costatum* result (Skel4.A*) more closely aligns with *A. carterae* (SΔn=264) than with any other. Several of the pseudo-replicates are in good agreement with one another: *S. oleracea, C. salina, C. similis,* and *A. carterae.* However, *Isochrysis* sp. agrees less well, separating at SΔn=122. Two diatoms are very inconsistent, with *T. weissflogii* differing at SΔn=286, and *S. costatum* separating to the base of the tree (SΔn=360).

Unlike the cluster analysis for the separate pseudo-replicates, the PCA analysis could be interpreted as separating out only three groups (Figure 4; Tables 5, 6, 7). In the plot of the first two dimensions (Figure 4b) *A. carterae* and *S. oleracea* are off in separate corners while the chromophytes are more scattered along the other side. One *S. costatum* analysis (Skel4.A*) could be separated from other chromophytes as a fourth group along dimension 1. All of the *S. costatum* analyses are more negative in dimension 3 than other chromophytes (Figure 4c; Table 7), which would suggest that they form a fifth group. The two *T. weissflogii* analyses differ in dimension 1 (Figures 4b, 4c) but

Table 2. Dimension eigenvalues, their percentages of total variance, and cumulative percentages for the PCA of Rubisco LSU amino acid composition, pseudo-replicates averaged for each of 7 species. The first 3 axes are plotted in Figure 2.

Dimension	Eigenvalue	Percent	Cumulative %
1	5.06375	31.65	31.65
2	4.67574	29.22	60.87
3	2.34293	14.64	75.52
4	1.97603	12.35	87.87
5	1.20978	7.56	95.43
6	0.73177	4.57	100

Table 3. Coefficients of the amino acid components, eigenvectors, of the first 3 dimensions of the PCA of Rubisco LSU amino acid compositions, averages of pseudo-replicates of 7 species. These axes are plotted in Figure 2.

Dimension 1		Dimension 2		Dimension 3	
Amino acid	coefficient	Amino acid	coefficient	Amino acid	coefficient
His	0.871	Arg	0.741	Tyr	0.784
Pro	0.848	Leu	0.713	Phe	0.611
Val	0.703	Asx	0.557	Ile	0.423
Thr	0.629	Tyr	0.521	Lys	0.404
Lys	0.407	Met	0.442	Leu	0.389
Arg	0.374	Ala	0.366	Ala	0.309
Leu	0.371	Thr	0.164	Ser	0.056
Phe	0.216	Val	0.108	His	0.026
Tyr	0.077	Ile	-0.050	Pro	0.008
Glx	-0.199	His	-0.211	Val	0.000
Met	-0.268	Pro	-0.446	Asx	-0.123
Asx	-0.527	Ser	-0.514	Met	-0.258
Gly	-0.551	Gly	-0.616	Glx	-0.295
Ala	-0.643	Lys	-0.693	Thr	-0.389
Ser	-0.648	Phe	-0.727	Arg	-0.449
Ile	-0.820	Glx	-0.879	Gly	-0.491

Table 4. Projection of 7 species, averages of pseudo-replicates, onto the first 3 eigenvectors of the Rubisco LSU amino acid composition PCA. The analyses have been sorted for each dimension (D). These data are plotted in Figure 2.

Species	D1	Species	D2	Species	D3
SpinM.A*	1.163	ChaetM.A*	0.421	ThalM.A*	0.472
IsoM.A*	0.100	IsoM.A*	0.409	ChaetM.A*	0.439
ChaetM.A*	0.070	ThalM.A*	0.235	AmphiM.A*	0.167
ChrooM.A*	-0.178	SkelM.A*	0.210	IsoM.A*	-0.032
ThalM.A*	-0.197	ChrooM.A*	0.127	SpinM.A*	-0.180
AmphiM.A*	-0.429	SpinM.A*	-0.310	ChrooM.A*	-0.315
SkelM.A*	-0.527	AmphiM.A*	-1.094	SkelM.A*	-0.550

Table 5. Dimension eigenvalues, their percentages of total variance, and cumulative percentages for the PCA of Rubisco LSU amino acid composition, pseudo-replicate analyses of 7 species. The first 3 axes are plotted in Figure 4.

Dimension	Eigenvalue	Percent	Cumulative %
1	4.47906	27.99	27.99
2	3.59771	22.49	50.48
3	2.25330	14.08	64.56
4	2.15360	13.46	78.02
5	1.67314	10.46	88.48
6	0.81103	5.07	93.55
7	0.41928	2.62	96.17
8	0.23876	1.49	97.66
9	0.13596	0.85	98.51
10	0.09285	0.58	99.09
11	0.06672	0.42	99.51
12	0.03897	0.24	99.75
13	0.02532	0.16	99.91
14	0.00802	0.05	99.96
15	0.00627	0.04	100

Table 6. Coefficients of the amino acid components, eigenvectors, of the first 3 dimensions of the PCA of Rubisco LSU amino acid compositions, pseudo-replicate analyses of 7 species. These axes are plotted in Figure 4.

Dimension 1		Dimension 2		Dimension 3	
Amino acid	coefficient	Amino acid	coefficient	Amino acid	coefficient
Leu	0.663	Asx	0.563	Tyr	0.671
Thr	0.656	Arg	0.562	Met	0.554
His	0.600	Met	0.536	Lys	0.451
Pro	0.582	Ala	0.479	Phe	0.366
Arg	0.542	Leu	0.319	Ile	0.313
Val	0.493	Tyr	0.253	Thr	0.174
Tyr	0.443	Ile	0.203	Asx	0.105
Lys	0.055	Thr	-0.043	Ser	0.071
Met	0.037	Gly	-0.151	His	0.022
Phe	-0.024	Ser	-0.170	Pro	-0.131
Asx	-0.202	Val	-0.275	Leu	-0.187
Ala	-0.367	His	-0.351	Glx	-0.242
Glx	-0.639	Glx	-0.535	Ala	-0.305
Ile	-0.669	Pro	-0.707	Arg	-0.369
Ser	-0.747	Lys	-0.735	Gly	-0.403
Gly	-0.765	Phe	-0.814	Val	-0.715

Table 7. Projection of 7 species, pseudo-replicates analyses, onto the first 3 eigenvectors of the Rubisco LSU amino acid composition PCA. The analyses have been sorted for each dimension (D). These data are plotted in Figure 4.

Species	D1	Species	D2	Species	D3
Spin3.A*	0.864	Skel3.A*	0.591	Chroo2.A*	0.497
Spin2.A*	0.840	Skel1.A*	0.437	Chroo3.A*	0.439
Spin1.A*	0.838	Skel2.A*	0.396	Chroo4.A*	0.427
Iso5.A*	0.393	Chroo5.A*	0.369	Iso3.A*	0.310
Chaet2.A*	0.353	Iso1.A*	0.328	Chroo1.A*	0.305
Iso4.A*	0.341	Iso2.A*	0.267	Iso2.A*	0.288
Iso3.A*	0.329	Iso3.A*	0.255	Amphi3.A*	0.245
Iso2.A*	0.315	Skel4.A*	0.247	Chaet1.A*	0.229
Iso1.A*	0.284	Chroo4.A*	0.234	Iso1.A*	0.228
Chaet1.A*	0.251	Chaet1.A*	0.230	Amphi1.A*	0.213
Thal1.A*	0.215	Chaet2.A*	0.222	Chaet2.A*	0.195
Chroo2.A*	-0.003	Thal1.A*	0.209	Thal1.A*	0.121
Chroo1.A*	-0.032	Thal2.A*	0.196	Amphi2.A*	0.116
Chroo3.A*	-0.058	Chroo1.A*	0.175	Chroo5.A*	0.104
Chroo4.A*	-0.093	Chroo3.A*	0.167	Iso5.A*	0.019
Skel2.A*	-0.156	Chroo2.A*	0.139	Iso4.A*	-0.007
Chroo5.A*	-0.184	Iso5.A*	0.134	Thal2.A*	-0.043
Skel1.A*	-0.282	Iso4.A*	0.089	Spin2.A*	-0.344
Thal2.A*	-0.335	Spin1.A*	-0.689	Spin1.A*	-0.380
Skel3.A*	-0.370	Spin3.A*	-0.711	Spin3.A*	-0.441
Amphi1.A*	-0.822	Amphi2.A*	-0.774	Skel1.A*	-0.571
Amphi2.A*	-0.848	Amphi1.A*	-0.793	Skel2.A*	-0.591
Amphi3.A*	-0.906	Spin2.A*	-0.799	Skel3.A*	-0.662
Skel4.A*	-0.936	Amphi3.A*	-0.920	Skel4.A*	-0.696

this is contrasted with their wider separation in the dendrogram (Figure 3) and the clustering of one of them, Thal2.A*, with *S. costatum.* In spite of these considerations, this PCA could be interpreted to indicate three groups, in agreement with the dendrogram and PCA of the averages of pseudo-replicates (Figures 1, 2).

Discussion

These results expand the data set for Rubisco LSU primary structure, adding several chrysophytes, a cryptophyte, and a dinoflagellate. To the best of this author's knowledge no other results with dinoflagellates, in particular *Amphidinium carterae,* are extant.

In order to examine what they might add to concepts of algal evolution they were combined with data present in the literature. These literature data are of several types: amino acid compositions, protein sequences, and protein sequences deduced from gene sequences. The following are new analyses of these data, in that protein sequences have been reduced to amino acid compositions, and the methods are those suitable for compositions.

Phylogeny based on Rubisco LSU amino acid composition

Comparisons of these results with those of others

Amino acid composition data

With inclusion of amino acid compositions from the literature (Appendix 2) cluster analysis separated out 1) *A. carterae,* 2) *Rhodospirillum rubrum,* 3) four of the chromophytes added by this study, 4) a cluster of chlorophyte algae and land plants, and 5) a cluster of chromophyte algae and purple bacteria (Figure 6). The PCA of these data also recognizes *A. carterae* and the chlorophytes and land plants group (Figure 7). However, this analysis groups all the chromophytes together and places the purple bacteria, including *R, rubrum,* between the chromophyte and chlorophyte groups (Figure 7b).

In more detail, cluster analysis of these data (Figure 6) identifies *A. carterae* as a unique Rubisco (SΔn=361). *R. rubrum* is also unique (SΔn=272). Several of the chromophyte analyses differ greatly (SΔn=265) from other chromophytes, specifically, *S. costatum, T. weissflogii, Isochrysis* sp., and *C. similis* differ from *O. luteus* and

Figure 6. UPGMA cluster analysis of SΔn distances between Rubisco amino acid compositions of *Spinacea oleracea*, 6 species of marine algae, and amino acid compositions reported in the literature.

Code as for Figure 1 and Table 1.

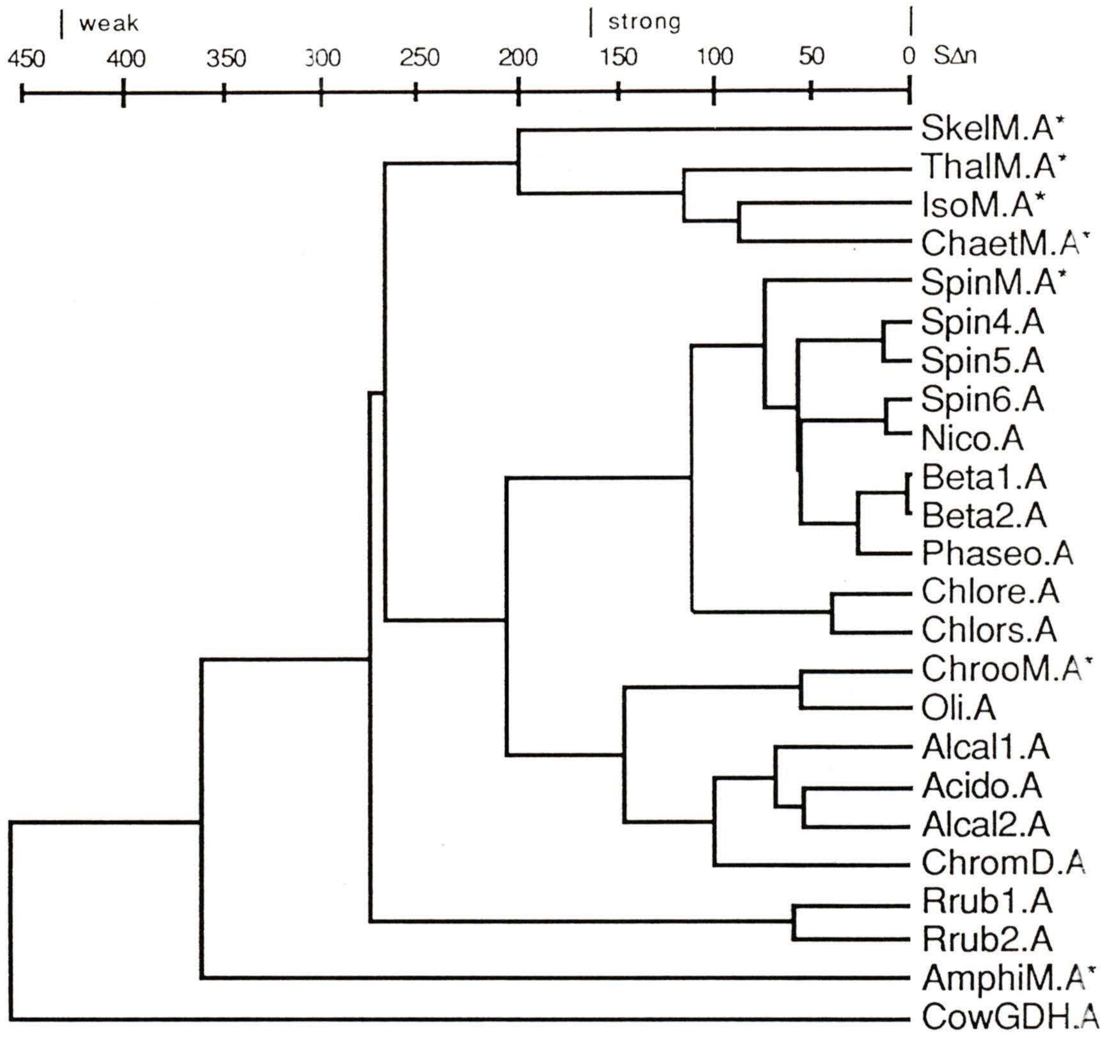
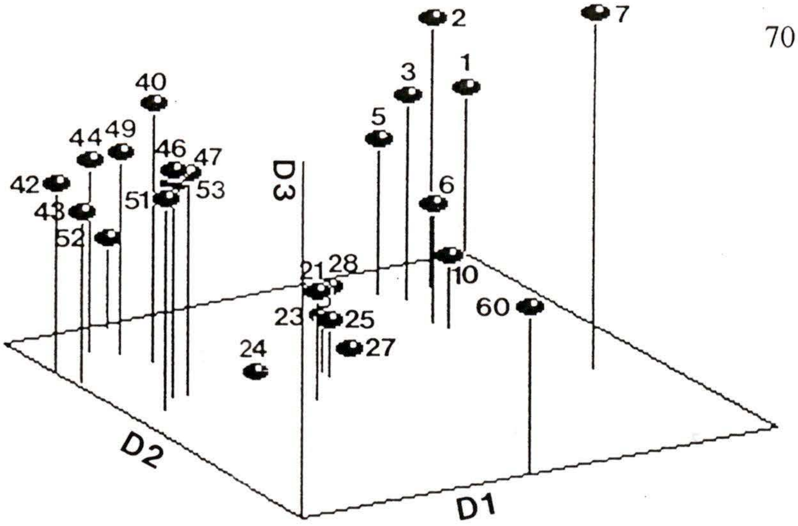

weak
strong
450
400
350
300
250
200
150
100
50
0
SΔn
SkelM.A*
ThalM.A*
IsoM.A*
ChaetM.A*
SpinM.A*
Spin4.A
Spin5.A
Spin6.A
Nico.A
Beta1.A
Beta2.A
Phaseo.A
Chlore.A
Chlors.A
ChrooM.A*
Oli.A
Alcal1.A
Acido.A
Alcal2.A
ChromD.A
Rrub1.A
Rrub2.A
AmphiM.A*
CowGDH.A

Figure 7. PCA of Rubisco amino acid compositions of *Spinacea oleracea* and 6 species of marine algae, and amino acid compositions reported in the literature. Code as for Figure 1 and Table 1.

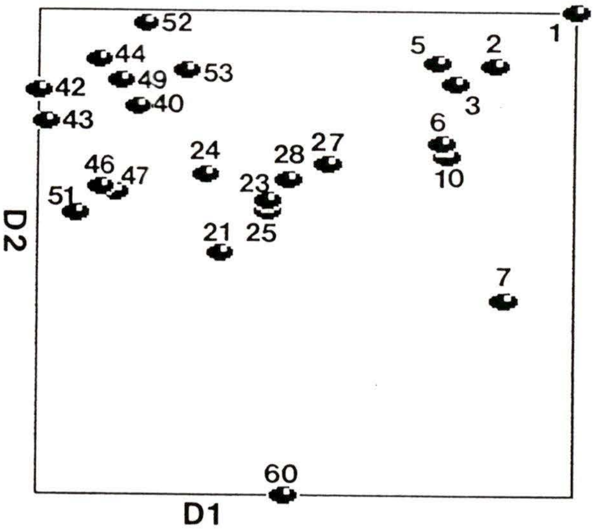
Species: (1) SkelM.A* (2) ThalM.A* (3) ChaetM.A* (5) IsoM.A*
(6) ChrooM.A* (7) AmphiM.A* (10) Oli.A
(21) ChromD.A (23) Alcal1.A (24) Alcal2.A (25) Acido.A (27) Rrub1.A
(28) Rrub2.A
(40) SpinM.A* (42) Spin4.A (43) Spin5.A (44) Spin6.A (46) Beta1.A
(47) Beta2.A (49) Nico.A (51) Phaseo.A (52) Chlore.A (53) Chlors.A
(60) CowGDH.A.

a. Orthogonal projection. b. Dimensions 1 and 2. c. Dimensions 1 and 3.

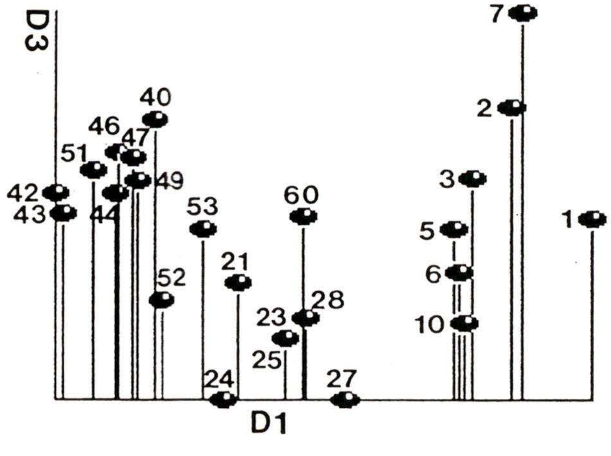
a.


2
7
1
3
40
5
44
49
46
47
42
53
51
43
6
52
21
28
10
60
23
25
27
24
D3
D2
D1

b.


52
1
5
2
44
53
49
42
3
40
43
6
24
28
27
46
47
10
23
51
D2
21
25
7
60
D1

c.


D3
7
40
2
46
47
51
3
42
49
43
44
53
60
1
5
6
21
52
23
28
10
25
24
27
D1

C. salina. The cryptophyte, *C. salina,* is most closely related to the chrysophyte, *O. luteus,* and both of these cluster with the purple bacteria (SΔn=146). The *S. oleracea* result added by this work, SpinM.A*, appears on the branch with the land plants, and the green algae join them at SΔn=111.

In contrast, the PCA of the amino acid compositions of Rubisco LSU separates all of the chromophytes from all of the chlorophytes along dimension 1 , which accounts for the dominant proportion of the variance (28%) in the original data (Tables 8, 9, 10; Figure 7). The purple bacteria are placed between these two groups. *A. carterae* occurs at the extremities of all three dimensions (Table 10; Figure 7). The chromophytes are more diffuse than other groups in dimension 3 (Figure 7c). The chlorophytes are clustered in dimensions 1 and 3 (Figure 7c), except for *Chlorella ellipsoidea* (Chlore.A) and *C. sorokiniana* (Chlors.A). However close they are to one another, the chlorophytes and purple bacteria can be enclosed in non-overlapping space. Thus, the PCA of these data distinguishes four groups *A. carterae,* the chromophytes, the purple bacteria, and the chlorophytes.

Amino acid and gene sequence data

Amino acid compositions deduced from protein and gene sequences (Appendix 3) were compared with the above data sets. The dendrogram of the SΔn distances separated five major groups (Figure 8): 1) *A. carterae;* 2) four chromophytes of this study; 3) the remaining chromophyte amino acid analyses with *R. rubrum;* 4) chromophyte gene-based data with other purple bacteria; and 5) chlorophytes and land plants with cyanobacteria, a prochlorophyte, a euglenophyte, and *Cyanophora paradoxa.* The PCA uses a data set of reduced size (Figure 9). It allows recognition of four groups: 1) *A. carterae;* 2) all chromophytes; 3) chlorophytes, land plants, and cyanobacteria; and 4) the purple bacteria, with *R. rubrum* somewhat distant from the others.

Table 8. Dimension eigenvalues, their percentages of the total variance, and cumulative percentages for the PCA of Rubisco LSU amino acid composition of all species where data originate as amino acid compositions. The first 3 axes are plotted in Figure 7.

Dimension	Eigenvalue	Percent	Cumulative %
1	4.53392	28.34	28.34
2	2.89754	18.11	46.45
3	2.50468	15.65	62.10
4	1.97578	12.35	74.45
5	1.07115	6.69	81.14
6	0.83684	5.23	86.37
7	0.67811	4.24	90.61
8	0.45631	2.85	93.46
9	0.37870	2.37	95.83
10	0.31948	2.00	97.83
11	0.11508	0.72	98.55
12	0.11067	0.69	99.24
13	0.06509	0.41	99.65
14	0.04111	0.26	99.90
15	0.01553	0.10	100.00

Table 9. Coefficients of the amino acid components, eigenvectors, of the first 3 dimensions of the PCA of Rubisco LSU amino acid compositions of all species where data originate as amino acid compositions. These axes are plotted in Figure 7.

Dimension 1		Dimension 2		Dimension 3	
Amino Acid	Coefficient	Amino Acid	Coefficient	Amino Acid	Coefficient
Asx	0.684	Leu	0.795	Glx	0.459
Ser	0.650	Ala	0.399	Tyr	0.440
Ile	0.586	Arg	0.386	Gly	0.421
Gly	0.519	Tyr	0.374	Lys	0.418
Ala	0.470	Gly	0.203	Leu	0.408
Tyr	0.259	Thr	0.155	Phe	0.378
Met	0.142	Asx	0.134	Val	0.276
Leu	-0.153	Glx	0.083	Ile	0.105
Val	-0.156	Pro	0.022	His	-0.054
Phe	-0.329	Val	-0.042	Thr	-0.118
Lys	-0.439	His	-0.312	Asx	-0.161
Arg	-0.464	Met	-0.323	Ser	-0.179
Glx	-0.468	Ser	-0.555	Ala	-0.311
Pro	-0.687	Phe	-0.581	Pro	-0.438
Thr	-0.814	Ile	-0.637	Arg	-0.591
His	-0.897	Lys	-0.688	Met	-0.800

Table 10. Projection of all species where data originate as amino acid compositions onto the first 3 eigenvectors of the Rubisco LSU amino acid composition PCA. The species have been sorted for each dimension (D). These data are plotted in Figure 7.

Species	D1	Species	D2	Species	D3
SkelM.A*	1.001	SkelM.A*	0.579	AmphiM.A*	0.921
AmphiM.A*	0.785	Chlore.A	0.531	ThalM.A*	0.519
ThalM.A*	0.752	Spin6.A	0.377	SpinM.A*	0.465
ChaetM.A*	0.625	IsoM.A*	0.364	Beta1.A	0.328
Oli.A	0.599	ThalM.A*	0.354	Beta2.A	0.305
ChrooM.A*	0.582	Chlors.A	0.335	Phaseo.A	0.256
IsoM.A*	0.563	Nico.A	0.298	ChaetM.A*	0.220
Rrub1.A	0.224	ChaetM.A*	0.277	Nico.A	0.208
Rrub2.A	0.103	Spin4.A	0.245	Spin4.A	0.159
CowGDH.A	0.095	SpinM.A*	0.190	Spin6.A	0.159
Alcal1.A	0.033	Spin5.A	0.122	Spin5.A	0.070
Acido.A	0.033	ChrooM.A*	0.041	CowGDH.A	0.064
ChromD.A	-0.121	Oli.A	-0.019	SkelM.A*	0.052
Alcal2.A	-0.171	Rrub1.A	-0.046	IsoM.A*	0.009
Chlors.A	-0.242	Alcal2.A	-0.096	Chlors.A	0.005
Chlore.A	-0.379	Rrub2.A	-0.118	ChrooM.A*	-0.170
SpinM.A*	-0.404	Beta1.A	-0.149	ChromD.A	-0.215
Nico.A	-0.464	Beta2.A	-0.169	Chlore.A	-0.289
Beta2.A	-0.482	Alcal1.A	-0.197	Rrub2.A	-0.368
Beta1.A	-0.531	Acido.A	-0.237	Oli.A	-0.382
Spin6.A	-0.538	Phaseo.A	-0.249	Alcal1.A	-0.449
Phaseo.A	-0.612	ChromD.A	-0.419	Acido.A	-0.451
Spin5.A	-0.710	AmphiM.A*	-0.609	Alcal2.A	-0.706
Spin4.A	-0.740	CowGDH.A	-1.406	Rrub1.A	-0.710

Figure 8. UPGMA cluster analysis of SΔn distances between Rubisco amino acid compositions of *Spinacea oleracea,* 6 species of marine algae, amino acid compositions reported in the literature, and sequences from the literature reduced to amino acid compositions. Codes as for Figure 1 and Table 1.

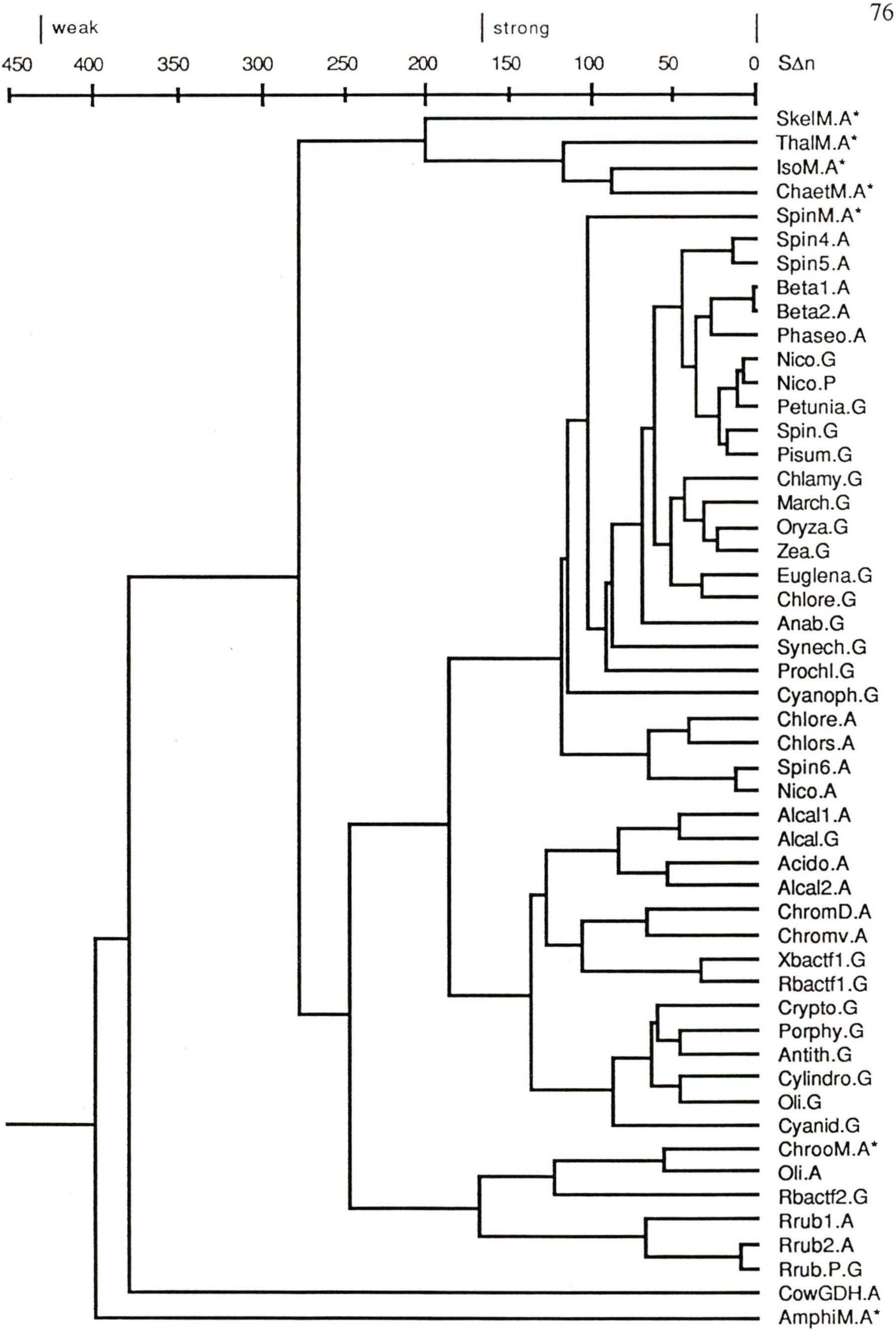
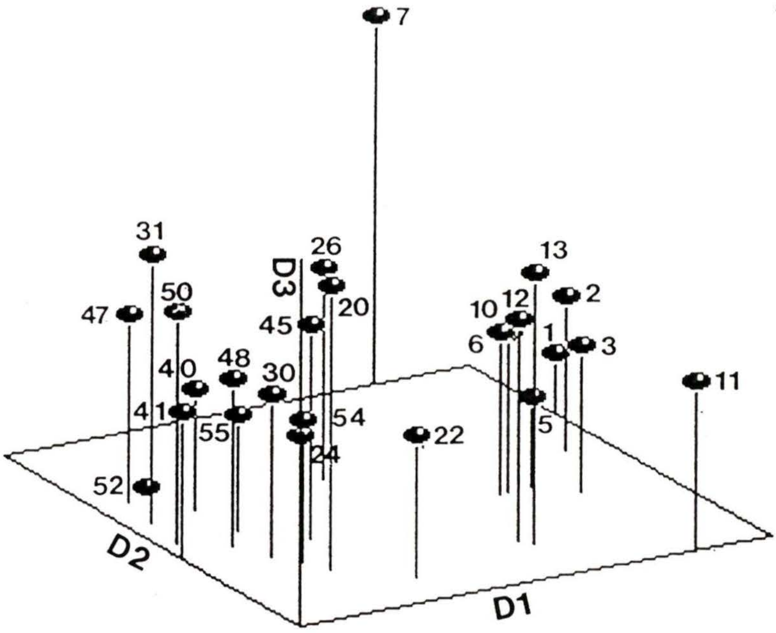
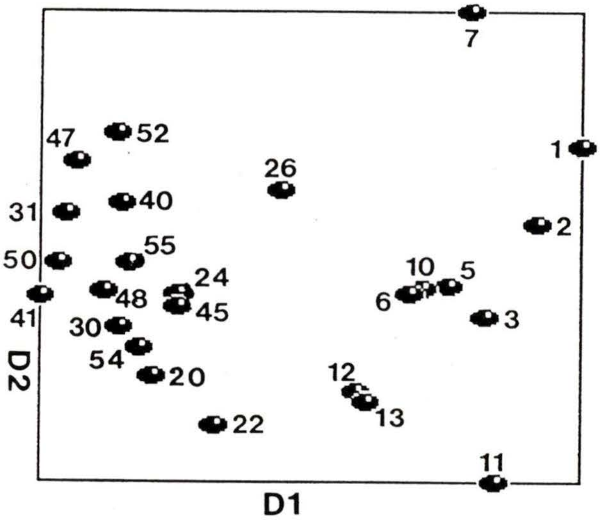

weak
strong
450
400
350
300
250
200
150
100
50
0
SΔn
SkelM.A*
ThalM.A*
IsoM.A*
ChaetM.A*
SpinM.A*
Spin4.A
Spin5.A
Beta1.A
Beta2.A
Phaseo.A
Nico.G
Nico.P
Petunia.G
Spin.G
Pisum.G
Chlamy.G
March.G
Oryza.G
Zea.G
Euglena.G
Chlore.G
Anab.G
Synech.G
Prochl.G
Cyanoph.G
Chlore.A
Chlors.A
Spin6.A
Nico.A
Alcal1.A
Alcal.G
Acido.A
Alcal2.A
ChromD.A
Chromv.A
Xbactf1.G
Rbactf1.G
Crypto.G
Porphy.G
Antith.G
Cylindro.G
Oli.G
Cyanid.G
ChrooM.A*
Oli.A
Rbactf2.G
Rrub1.A
Rrub2.A
Rrub.P.G
CowGDH.A
AmphiM.A*

Figure 9. PCA of Rubisco amino acid compositions of *Spinacea oleracea* and 6 species of marine algae, amino acid compositions reported in the literature, and sequences from the literature reduced to amino acid compositions. Code as for Figure 1 and Table 1. Species: (1) SkelM.A*; (2) ThalM.A*;
(3) ChaetM.A*; (5) IsoM.A*; (6) ChrooM.A*; (7) AmphiM.A*;
(10) Oli.A; (11) Oli.G; (12) Porphy.G; (13) Crypto.G;
(20) Chromv.G; (22) Alcal.G; (24) Alcal2.A; (26) Rrub.P.G;
(30) Anab.G; (31) Synech.G;
(40) SpinM.A*; (41) Spin.G; (45) Zea.G; (47) Beta2.A; (48) Nico.P;
(50) Oryza.G; (52) Chlore.A; (54) Chlamy.G; (55) Euglena.G.
a. Orthogonal projection. b. Dimensions 1 and 2. c. Dimensions 1 and 3.

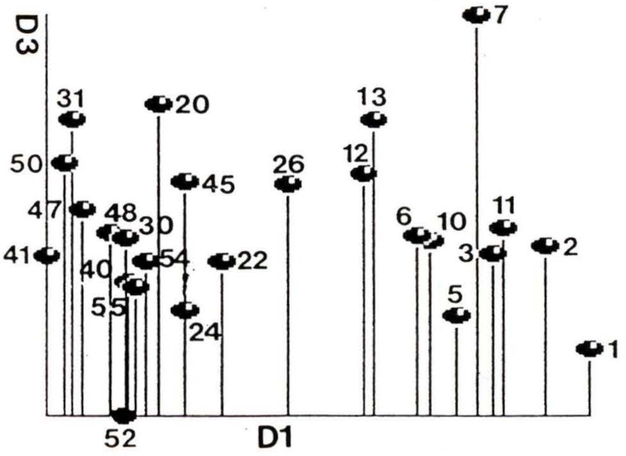
a.


7
31
26
13
D3
50
20
2
47
12
10
45
1
6
3
48
40
30
11
41
55
54
5
22
24
52
D2
D1

b.


7
52
47
1
26
31
40
2
55
50
10
5
24
6
48
3
41
45
30
54
D2
20
12
13
22
11
D1

c.


7
D3
31
20
13
12
50
45
26
11
47
48
30
6
10
41
54
22
3
2
40
5
55
24
1
52
D1

For the analyses that include sequence data, the first organism that the dendrogram separates out (Figure 8), like all of the other dendrograms, is *A. carterae* (SΔn=398). Again, several of these chromophyte results are separated next (SΔn=278). At SΔn=247 the tree splits, with two chromophyte amino acid-based compositions, *O. luteus* (Oli.A) and *C. salina* (ChrooM.A*), present on one branch with form II Rubiscos from *R. rubrum* and *Rhodobacter sphaeroides.* These eubacteria are represented by both gene and amino acid composition data. Another subdivision separates two large groups (SΔn=184), chlorophytes from chromophytes and purple bacteria. The chlorophyte group consists of land plants, green algae, *Euglena gracilis, C. paradoxa,* a prochlorophyte, and cyanobacteria. Some of the clustering within this group follows accepted phylogenetic lines: Many of the land plant analyses group together, and the prokaryotes and *C. paradoxa* are different from these. In this tree, with both amino acid- and gene-based data, within the chlorophyte branch, several of the chlorophyte amino acid compositions are separated (Chlore.A, Chlors.A, Spin6.A, and Nico.A) (SΔn=116). But, other amino acid compositions are mixed in with the gene-based data (eg. Spin4.A and Pisum.G) (the greatest distance here being SΔn=46). The *S. oleracea* analysis of this study, SpinM.A*, branches off between these at SΔn=101. Note that Nico.P and Nico.G, when reduced to compositions, actually differ by 7 amino acids (Appendices 2, 3), but SΔn overestimates this difference as SΔn=7.5.

The other major group is subdivided into two groups, the chromophytes and the purple bacteria. The purple bacteria separate into α, β, and γ types, and include amino acid- and gene-based data. The chromophyte group on this branch consists only of gene based data, but from a variety of taxa, including a cryptophyte, rhodophytes, and chrysophytes.

Comparison of the tree based on amino acid compositions (Figure 6) with that based on amino acid and gene data (Figure 8), reveals differences between them. In the former tree, the chromophyte amino acid compositions, *C. salina* and *O. luteus* (ChrooM.A*, Oli.A), are grouped with the purple bacteria having form I Rubisco, while the *R. rubrum* form II Rubisco is quite different. In the latter tree, these chromophytes separate off with the bacteria having form II Rubisco, while the chromophytes that are gene-based data, are placed along side the purple bacteria with form I Rubisco.

In order to run PCA, it was necessary to reduce the size of the data set. To do this, all of the results of this work were retained and an attempt was made to keep a balance of prokaryote and eukaryote organisms, and a balance of gene-based and amino acid composition-based data (Figure 9; Tables 11, 12, 13). *A. carterae* is unique, especially in dimensions 2 and 3 (Figures 9b, 9c). Dimension 1, accounting for 31% of the total variance, separates chromophytes from all chlorophytes (Figure 9b). The cyanobacteria, *Synechococcus* and *Anabaena,* are embedded in the chlorophyte group in dimensions 1 and 2 (Figure 9b), but *Synechococcus* is at the positive end of dimension 3 (Figure 9c). The purple bacteria appear between the chromophyte and chlorophyte groups, separated from the chromophytes on one side, but overlapping with the chlorophytes on the other. The γ purple bacterium, *Chromatium vinosum,* is closer to the chlorophytes in dimension 1. In dimension 2, the purple bacteria are less tightly clustered, with *R. rubrum* more positive than *A. eutrophus* or *C. vinosum.* The chlorophytes and chromophytes are also scattered along this dimension (Table 13; Figure 9b). Dimension 3 also disperses these three groups (Figure 9c). Note that dimension 2 is weighted by glycine and methionine content, ranking 2 and 3 respectively (Table 12), but the overall ranking places them 11 and 13 respectively (Table 14).

Thus four groups can be distinguished (Figure 9). *A. carterae* is clearly set apart, and the chlorophytes are distinguished from the chromophytes. The purple bacteria form the fourth group, with *R. rubrum* somewhat separate. The cyanobacteria are not differentiated from the chlorophytes.

Table 11. Dimension eigenvalues, their percentages of the total variance, and cumulative percentages for the PCA of Rubisco LSU amino acid composition of 25 selected species. The first 3 axes are plotted in Figure 9.

Dimension	Eigenvalue	Percent	Cumulative %
1	5.01602	31.35	31.35
2	2.82890	17.68	49.03
3	1.91829	11.99	61.02
4	1.76948	11.06	72.08
5	0.98431	6.15	78.23
6	0.90420	5.65	83.88
7	0.75606	4.73	88.61
8	0.53807	3.36	91.97
9	0.41255	2.58	94.55
10	0.30859	1.93	96.48
11	0.28220	1.76	98.24
12	0.14198	0.89	99.13
13	0.08459	0.53	99.66
14	0.03527	0.22	99.88
15	0.01948	0.12	100.00

Table 12. Coefficients of the amino acid components, eigenvectors, of the first 3 dimensions of the PCA of Rubisco LSU amino acid compositions of 25 selected species. These axes are plotted in Figure 9.

Dimension 1		Dimension 2		Dimension 3	
Amino acid	coefficient	Amino acid	coefficient	Amino acid	coefficient
Asx	0.823	Glx	0.776	Lys	0.656
Ser	0.643	Gly	0.715	Phe	0.554
Ile	0.576	Phe	0.425	Ile	0.527
Ala	0.545	Ser	0.299	His	0.167
Tyr	0.508	Pro	0.220	Met	0.129
Gly	0.408	Ala	0.214	Ser	0.122
Leu	0.125	Leu	0.177	Val	0.029
Met	0.008	His	-0.041	Gly	0.019
Val	-0.005	Asx	-0.101	Asx	-0.093
Glx	-0.271	Lys	-0.126	Glx	-0.095
Phe	-0.411	Arg	-0.306	Ala	-0.100
Arg	-0.462	Thr	-0.327	Thr	-0.138
Lys	-0.604	Ile	-0.449	Tyr	-0.164
Pro	-0.758	Val	-0.456	Pro	-0.230
Thr	-0.770	Tyr	-0.460	Leu	-0.396
His	-0.931	Met	-0.686	Arg	-0.749

Table 13. Projection of 25 selected species onto the first 3 eigenvectors of the Rubisco LSU amino acid composition PCA. The species have been sorted for each dimension (D). These data are plotted in Figure 9.

Species	D1	Species	D2	Species	D3
SkelM.A*	1.015	AmphiM.A*	1.077	AmphiM.A*	0.875
ThalM.A*	0.877	Chlore.A	0.590	Chromv.G	0.498
Oli.G	0.742	SkelM.A*	0.538	Synech.G	0.440
ChaetM.A*	0.711	Beta2.A	0.473	Crypto.G	0.439
AmphiM.A*	0.662	Rrub.P.G	0.352	Oryza.G	0.266
IsoM.A*	0.601	SpinM.A*	0.301	Porphy.G	0.220
Oli.A	0.514	Synech.G	0.261	Zea.G	0.191
ChrooM.A*	0.477	ThalM.A*	0.218	Rrub.P.G	0.180
Crypto.G	0.336	Oryza.G	0.064	Beta2.A	0.076
Porphy.G	0.307	Euglena.G	0.061	Oli.G	-0.005
Rrub.P.G	0.056	IsoM.A*	-0.038	Nico.P	-0.021
Alcal.G	-0.150	Oli.A	-0.041	ChrooM.A*	-0.033
Alcal2.A	-0.266	Nico.P	-0.054	Anab.G	-0.045
Zea.G	-0.268	ChrooM.A*	-0.065	Oli.A	-0.054
Chromv.G	-0.352	Alcal2.A	-0.065	ThalM.A*	-0.081
Chlamy.G	-0.393	Spin.G	-0.078	ChaetM.A*	-0.107
Euglena.G	-0.426	Zea.G	-0.122	Spin.G	-0.114
SpinM.A*	-0.447	ChaetM.A*	-0.162	Chlamy.G	-0.139
Anab.G	-0.453	Anab.G	-0.204	Alcal.G	-0.140
Chlore.A	-0.462	Chlamy.G	-0.286	SpinM.A*	-0.223
Nico.P	-0.505	Chromv.G	-0.402	Euglena.G	-0.246
Beta2.A	-0.591	Porphy.G	-0.467	Alcal2.A	-0.335
Synech.G	-0.625	Crypto.G	-0.513	IsoM.A*	-0.364
Oryza.G	-0.653	Alcal.G	-0.606	SkelM.A*	-0.502
Spin.G	-0.709	Oli.G	-0.831	Chlore.A	-0.775

Table 14. Relative importance of amino acids in first 3 dimensions (D) of PCAs. Rank determined from eigenvalues and eigenvectors by ordering the calculation for each amino acid (A) of:

$$\Sigma_{D=1,2,3}(\text{eigenvalue}_D \times |\text{eigenvector}_A|).$$

	—	Rank of amino acid in PCA		—
	pseudo-replicates	mean of pseudo-replicates	amino acid compositions	amino acid & sequences
amino acid	Table 6	Table 3	Table 1	Table 12
Val	1	2	1	16
Tyr	2	4	4	15
Ala	3	1	2	9
Thr	4	5	5	6
Arg	5	3	3	5
Asx	6	6	6	12
Ser	7	8	7	3
Glx	8	7	8	14
Gly	9	10	10	11
His	10	11	12	1
Pro	11	9	9	7
Ile	12	13	14	4
Phe	13	12	11	8
Met	14	14	13	13
Lys	15	16	15	2
Leu	16	15	16	10

Comparison of the PCA results with those in the dendrogram shows that they disagree in some respects. In the PCA (Figure 9), separation of the chromophytes into two groups with chlorophytes between is not seen, as it is in the dendrogram (Figure 8). More specifically, *C. salina* and *O. luteus* (ChrooM.A* Oli.A) group with two other chromophytes, *C. similis* and *Isochrysis* sp. (ChaetM.A* IsoM.A*), by PCA (Figure 9), but in the dendrogram they are separated at SΔn=278 (Figure 8). The grouping of the form II Rubisco of *R. rubrum* with *O. luteus* and *C salina* (Oli.A ChrooM.A*), as occurs in the dendrogram (Figure 8), is not seen in the PCA (Figure 9). Instead, the form I Rubiscos of the other purple bacteria are closer to the chromophytes.

Data quality

The results of this study are based solely on amino acid compositions, and as such are subject to some limitations. These include analysis of contaminated protein due to co-migration of other proteins with Rubisco, amino acid contamination from buffers, and sensitivity of the amino acid analyzer to the quantity of injected protein. In addition, comparison of composition data to those predicted by gene sequences may result in errors.

Spurious results for the amino acid composition for *A. carterae* Rubisco LSU could have been obtained if the wrong protein had been excised from the electroblot. However, the protein band in the gel was at the correct position and was intensely stained (Figure 5) as was expected for the Rubisco LSU. Nevertheless, the identity of this band as authentic Rubisco needs to be independently confirmed. This may be done by isolation of non-denatured Rubisco holoenzyme and a test for Rubisco activity. However, Rubisco from dinoflagellate and chromophyte algae has proven difficult to isolate, purify, and activate due to insolubility and instability (Rothschild and Heywood 1987). Overcoming these difficulties could be the subject of future work, but at the moment require investments of time and resources beyond the scope of this study.

Co-migration of other proteins with Rubisco is a possible source of error. Because of the intensity of staining of the Rubisco band in the SDS-PAGE gel relative to others in the lane (Figure 5) it was presumed that any contaminant would have contributed only a small fraction to the protein analyzed.

Using these methods of protein isolation, it is difficult to estimate the optimal amount of protein to load on an amino acid analyzer. Trial and error allows some adjustment. The major problem throughout most of this research was to obtain sufficient Rubisco to work with. In several cases not enough Rubisco protein was present to give optimal readings from the HPLC. However, this problem did not seem to affect clustering of the amino acid compositions as judged by variances among pseudo-replicates, except for those for *S. costatum* and *T. weissflogii,* which were quite variable (Figure 3). The PCA results for these pseudo-replicates were affected less dramatically (Figure 4).

Glycine contamination may be suspected because it is a component of the transfer buffer for blotting the protein. This source of contamination was minimized by washing the blot numerous times before and during the staining. However, comparison of compositional data with sequence data from the literature (Table 15) showed that for some species glycine levels were higher in the compositions. The amino acid analysis for spinach (SpinM.A*) (this work) was higher in glycine than that for the gene sequence (Spin.G). With regard to literature data, the amino acid compositions for *N. tabacum* and *O. luteus* were also higher in glycine than for the sequences, but for *A. eutrophus and R. rubrum* the sequences were richer in glycine than the amino acid compositions. This may have biased the results. However, the ranges for glycine content were similar for both data types (Table 16). Also, glycine ranks eleventh (out of the sixteen amino acids) in the PCA that includes both data types (Table 14).

Table 15. Percent content of methionine and glycine in Rubisco. Comparison of data types. Code as for Figure 1 and Table 1. Data types: (.A) amino acid composition; (.P) protein sequence; (.G) gene sequence.

Amino acid	Genus species	code name	data type	%
Met	*S. oleracea*	SpinM.A*	A	0.35
		Spin.A (mean of 4,5,6)	A	1.70
		Spin.G	G	2.18
	N. tabacum	Nico.A	A	1.59
		Nico.P	P	1.74
		Nico.G	G	1.74
	O. luteus	Oli.A	A	3.60
		Oli.G	G	3.60
	A. eutrophus	Alcal.A (mean of 1,2)	A	3.06
		Alcal.G	G	3.59
	R. rubrum	Rrub.A (mean of 1,2)	A	2.77
		Rrub.P.G	P, G	3.08
Gly	*S. oleracea*	SpinM.A*	A	10.90
		Spin.A (mean of 4,5,6)	A	10.02
		Spin.G	G	9.83
	N. tabacum	Nico.A	A	10.77
		Nico.P	P	10.20
		Nico.G	G	10.00
	O. luteus	Oli.A	A	10.87
		Oli.G	G	9.32
	A. eutrophus	Alcal.A (mean of 1,2)	A	9.34
		Alcal.G	G	9.51
	R. rubrum	Rrub.A (mean of 1,2)	A	10.77
		Rrub.P.G	G	10.99

Table 16. Percent content of methionine and glycine in Rubisco. Range. Code as for Figure 1 and Table 1. Data types: (.A) amino acid composition; (.P) protein sequence; (.G) gene sequence.

Amino acid	Genus species	code name	data type	% (min	max)
Met	*S. oleracea*	SpinM.A*	A	0.35	
	C. salina	ChrooM.A*	A		3.85
	N. tabacum	Nico.P Nico.G	P, G	1.74	
	Anabaena sp. 7120	Anab.G	G	1.74	
	R. sphaeroides form II	Rbactf2.G	G		4.44
Gly	*C. similis*	ChaetM.A*	A	10.02	
	A. carterae	AmphiM.A*	A		13.14
	A. eutrophus	Alcal.A	A	8.73	
	C. sorokiniana	Chlors.A	A		11.44
	C. caldarium	Cyanid.G	G	8.30	
	E. gracilis	Euglena.G	G		11.01

There may have been some error in comparing amino acid compositions with deduced amino acid sequences. Methionine is a rare amino acid (Doolittle 1987) and is also the initiating amino acid in the sequence of a translated protein. Methionine was found to be relatively rare in Rubisco (Table 16). However, often amino termini of proteins are post-translationally processed, as has been documented for Rubisco of some species, so far all of them chlorophytes (Houtz *et al.* 1992). Thus, cleavage of an amino terminal methionine can result in a significant change in total methionine. There was some evidence of amino terminal processing in these analyses. Methionine is a major component of dimension 2 of the PCA that includes both compositions and deduced amino acid sequences (Table 12). In the PCA using sequence data, and for those species that are represented by both data types (*O. luteus, A. eutrophus,* and *S. oleracea*), the amino acid compositions were more positive than the sequences in dimension 2 (Figure 9, Table 13). Because methionine is a large negative component of this dimension, the placement of these amino acid compositions more positively than the sequences indicates that the mature protein may be more deficient in methionine than the sequences would predict. Reference to the original data (Appendices 1, 2, 3) shows this to be the case for *A. eutrophus* and *S. oleracea,* but not for *O. luteus.* The PCA dimensions for this analysis were very different from the PCAs that did not include sequence data (Table 14). However, methionine ranks 13 out of 16 in this analysis (and 13 or 14 in the other PCAs), indicating its lack of importance in distinguishing the taxa. Thus, while amino-terminal cleavage of methionine may have been a factor in the determination of the phylogenetic relationships presented here, its effect appears variable and minor.

In this work the analyses of two species, *S. costatum* and *A. carterae* were suspicious. The *S. costatum* analyses were variable (Appendix 1). For example, glutamate/glutamine (Glx) in Skel1.A* and Skel3.A* differs by 3.18%. This amino acid alone contributes 108 to the total SΔn between them. Alanine differs by 1.85%, contributing SΔn=36.5 between these. This variation may have been due to the experimental error of underloading, and thus, the *S. costatum* results should be considered uncertain.

The *A. carterae* analyses were more uniform. They differed from the other species mainly in the following respects: Glycine was high (Table 16). Methionine levels were low (0.64%) relative to both sequences(μ=2.81%, n=28, $\sigma_{(n-1)}$=0.75) and amino acid compositions (μ=2.11%, n=24, $\sigma_{(n-1)}$=0.94). Arginine was lower than any other species, 3.92% as opposed to 6.25% (=mean, n=52, $\sigma_{(n-1)}$=0.44). Serine was high, 5.84% as opposed to 4.16% (=mean, n=52, $\sigma_{(n-1)}$=0.64), but several analyses were in the same range (Cylindro.G=5.68%, Rrub1.A= 5.31%, Oli.A=5.23%) (Appendices 1, 2, 3). Thus, in spite of the potential experimental errors of glycine and methionine, some apparently real differences remain between *A. carterae* Rubisco and others.

Phylogenetic implications

Results from *Amphidinium carterae* suggest that its evolutionary history, and perhaps that of the dinoflagellates in general, is quite different from the other algae. However, a hypothesis concerning the evolution of dinoflagellates remains unforthcoming in spite of these new data. More specifically, they do not allow any of the other algal or prokaryotic organisms to be considered ancestral.

Further, these results find, by both analysis of principal components (PCA) and SΔn distances between Rubisco LSU, that the plastids of other photosynthetic eukaryotes have two different lineages. This pattern is observed in the results of this work (Figures 1, 2, 3, 4) and in comparison of these with literature values (Figures 6, 7, 8, 9). This finding differs from the other major hypothesis, which states that all plastids have a single lineage.

One of the lineages indicated by analysis of Rubisco LSU amino acid compositions includes plastids from the divisions Rhodophyta, Cryptophyta, and

Chrysophyta. Rubisco from these organisms are most closely related to the Rubisco from purple bacteria, which form a group alongside the chromophytes (Figures 6, 7, 8, 9). However, the integrity of this result is sensitive to the statistical technique used to compare the results of this work with those from the literature. For example, some of the chromophyte analyses of this work (SkelM.A*, ThalM.A*, ChaetM.A*, IsoM.A*) are separated from other chromophytes by the cluster analysis of SΔn distances (Figures 6, 8), but all chromophytes group together by PCA (Figures 7, 9). Also, where data from both amino acid compositions and from sequence sources are analyzed using the SΔn distance statistic, clustering is observed that is not supported by either the PCA of these data or the analyses that use only composition-based data. Specifically, the amino acid compositions for *C. salina* and *O. luteus* (ChrooM.A*, Oli.A) group with the form II Rubisco of the α purple bacteria in the dendrogram using both data types (Figure 8), but by all other analyses these two chromophytes group with other chromophytes or with purple bacteria with form I Rubisco (Figures 6, 7, 9). Also, the purple bacteria assemble as a group between the chromophytes and chlorophytes by PCA (Figures 7, 9), but by cluster analysis, the form I Rubiscos of purple bacteria are closer to the chromophytes than to the chlorophytes (Figures 6, 8). These PCA results are more in accord with what is expected than are the dendrograms. The chromophytes are expected to form a coherent group as are the purple bacteria. Thus, the dendrograms may be less reliable than the PCA representations of the data. However, analysis of Rubisco sequences finds the form II Rubisco of purple bacteria to be least like any other (McFadden *et al.* 1986). The amino acid composition dendrograms of this work are in agreement with this result, separating form II Rubisco from all others. Also, in the dendrogram that includes both Rubisco sequence- and composition-based data (Figure 8) the purple bacteria are separated into their subdivisions (α, β, γ) as predicted by 16S rRNA (Woese 1987).

The second lineage distinguished by analysis of Rubisco includes plastids from chlorophyte algae, land plants, and euglenophytes, and the cyanelle of *Cyanophora paradoxa.* Rubisco from these organisms most closely resembles that from cyanobacteria and prochlorophytes. The prochlorophytes bear no closer resemblance to land plants than do cyanobacteria. These prokaryotes are scattered among the eukaryotic members of this group. The clustering of cyanobacteria with these chlorophyll *b* - containing eukaryotes is consistently observed in both analyses (Figures 8, 9).

Thus, it can be concluded that two primary endosymbiotic events resulted in Rubisco of eukaryotic algal chloroplasts, given that the endosymbiotic theory is correct (Gray 1989). The nature of these endosymbioses is a matter of speculation, however (eg. Martin *et al.* 1992). But before discussion of these speculations, more established aspects of endosymbiosis will be addressed.

Primary endosymbioses are defined as the endosymbiosis of a prokaryote and indicated by two limiting chloroplast membranes, as seen in the rhodophytes and chlorophytes. Secondary endosymbioses as denoted by more than two membranes (3 or 4), are defined as the uptake of a eukaryote, which subsequently may be reduced to chloroplasts in the eukaryotic host, as observed in euglenophytes, cryptophytes, chrysophytes, and dinoflagellates (eg. Whatley and Whatley 1981). The primary endosymbiont of rhodophytes is indicated to be a cyanobacterium, mainly on the basis of common possession and arrangement of phycobilins. The primary endosymbiont of chlorophytes may be more closely related to cyanobacteria than prochlorophytes, despite the presence of chlorophyll *b* in this latter prokaryote (Whatley 1989). The criterion of accessory photosynthetic pigment also fails when considering eukaryotes with chlorophyll *c*. All of these have more than 2 chloroplast membranes, and neither a eukaryote with chlorophyll *c* and 2 membranes, nor a prokaryote with chlorophyll *c* are known (Whatley and Whatley 1981). Using cytochrome c_6 data, chromophytes may be linked to red algae (Hunt *et al.* 1985). Clearly, these Rubisco data indicate that rhodophytes are phylogenetically related to chromophytes and may be interpreted as their eukaryotic endosymbiont. As already stated, the endosymbiont for dinoflagellates is not indicated by these or any other data, except that common presence of chlorophyll *c* would link them to chromophytes and would suggest the two groups had a common origin for their plastids. However, if confirmed, the Rubisco result for *A. carterae* would negate this claim.

Chlorophyte lineage

Prokaryotic cyanobacteria and prochlorophytes, *Cyanophora paradoxa,* a eukaryote of uncertain lineage, and the eukaryotic chlorophytes, euglenophytes, and land plants form a consistent group in these Rubisco analyses. The presence of oxygenic photosynthesis requiring two photosystems, I and II, with the accessory light-absorbing pigment, chlorophyll *b,* and α-linked starches, also unite all of these organisms, with the exception of the cyanobacteria and *Cyanophora,* whose accessory pigments are phycobilins. The more logical progenitor for chloroplasts would be prochlorophytes because of the common presence of chlorophyll *b.* However, based on results obtained in studies of 16S rRNA (Turner *et al.* 1989, Douglas and Turner 1991, Urbach *et al.* 1992) and the gene for a subunit of RNA polymerase (rpoC1) (Palenik and Haselkorn 1992), some cyanobacteria appear more closely related to chlorophyte chloroplasts, and the prochlorophytes seem to have multiple origins within the cyanobacterial radiation (Urbach *et al.* 1992, Palenik and Haselkorn 1992). The cyanelle of *Cyanophora* has also been postulated to occupy an intermediate position between endosymbiosis of a cyanobacterium and chloroplasts. The cyanobacterial characters of the cyanelle include a peptidoglycan cell wall between the cyanelle bounding membrane and the host vacuolar membrane, unstacked thylakoids, and phycobilin pigments (Trench 1982). The green chloroplast-like characters of the cyanelle include a genome of similar size, gene content, and gene arrangement (Lockhart *et al.* 1992).

Chloroplasts of euglenophytes have been generally considered a secondary endosymbiosis of a chlorophyte because of the three membranes surrounding their chloroplasts and the common presence of chlorophyll *b* (Gibbs 1978).

Chromophyte lineage

Purple bacteria, rhodophytes, cryptophytes, and chrysophytes form a consistent group by these Rubisco analyses. Thus, the plastids of these eukaryotes appears to have originated from a purple bacterium. In some respects this seems illogical because purple bacteria do not possess oxygenic photosynthesis because they lack a second photosystem (a type I reaction center (Olson and Pierson 1987)). Further, purple bacteria possess neither chlorophyll *a* nor phycobilins as photosynthetic pigments. In contrast a cyanobacterium seems a better candidate for endosymbiont of rhodophytes and cryptophytes because of common possession of oxygenic photosynthesis and these pigments. Nevertheless the former conclusion is supported by Rubisco LSU composition (Figures 6, 7, 8, 9) and sequence analysis of Rubisco LSU and SSU (Morden *et al.* 1992). These latter analyses also include phaeophytes with this group. The eukaryotic cells of rhodophytes and cryptophytes are also connected by their common storage product α-linked starches; while cryptophytes, phaeophytes, and chrysophytes have chlorophyll *c;* and these latter two groups have β-linked starches and distinctive flagellar structures. Thus, this array of characters would both unite and separate the diverse members of this group, but the Rubisco analyses suggest a common heritage.

One could speculate that the first rhodophyte, the product of the endosymbiosis of a eubacterium (purple bacterium or cyanobacterium) into a host (amoeboid), may have resembled the modern-day *Cyanidium.* This is because this primitive rhodophyte is the first organism to separate off the chromophyte sub-branch (Figure 8). This pattern is also observed in analysis of Rubisco LSU sequence (Martin *et al.* 1992). Further, this eukaryote may have been similar to the organism that gave rise to the plastids of other chromophyte lines because the chromophyte line splits above the branching point of *Cyanidium* (Figure 8). The sequence tree does not make a clear distinction between the cryptophytes and more advanced rhodophytes (Martin *et al.* 1992). This suggests that a relatively recent endosymbiosis of a rhodophyte more advanced than *Cyanidium* resulted in the cryptophyte plastid. This suggestion is supported by the observation that nucleomorphs, redundant vestigial nuclei, are found in the periplasmic space of cryptophytes (Gillott and Gibbs 1980), but not in plastids of other chromophytes. Work by Douglas *et al.* (1991) showed that the nucleomorph is closely related to rhodophyte nuclei, while the cryptophyte nucleus is most closely related to that of an amoeboid protozoan and unrelated to other chromophyte nuclei.

A difference between composition and sequence analysis occurs when results with LSU from purple bacteria are examined. In composition analysis, form I Rubisco LSU of the γ subdivision clusters with the chromophytes (Figure 8), while in analysis of Rubisco sequences (both LSU and SSU) they cluster with chlorophytes and cyanobacteria (Martin *et al.* 1992, Morden *et al.* 1992). The Rubisco sequence results suggest that the γ subdivision is more closely related to cyanobacteria than are any other purple bacteria. However, in consideration of Rubisco SSU data, Assali *et al.* (1991) regard the linkage of γ purple bacteria to chlorophytes to be artifactual because of technical considerations and lack of characteristic signatures of green-plastid Rubisco SSU. These Rubisco composition analyses supports this view. Also in support are 16S rRNA data that indicate that all subdivisions of purple bacteria are more closely linked to each other than they are to cyanobacteria and chloroplasts (Woese 1987, Markowicz and Loiseaux-de Goër 1991).

Phylogeny based on other molecular data

Gene sequences for 16S rRNA *psbA, rbcL, rbcS, tufA,* and *atpB* (respectively, the genes for the small subunit ribosomal RNA, the D1 protein of PSII, the LSU of Rubisco, the SSU of Rubisco, elongation factor Tu, and the β subunit of ATP synthase) have been examined in an attempt to arrive at an integrated understanding of plastid origins (Morden *et al.* 1992). All of these except for the *rbcL* and *rbcS* sequences support the hypothesis that a single symbiotic event between a host cell and a cyanobacterium was the progenitor of all photosynthetic eukaryotic cells. As already indicated for *rbcL* and Rubisco LSU protein, and reinforced by *rbcS,* these data indicate separate chlorophyte and chromophyte lineages. Other aspects of these comparisons provide useful information.

Several gene sequences, including *tufA, rbcS,* and *rbcL,* have served to place the cyanelle of *Cyanophora* adjacent to cyanobacteria and chlorophyte chloroplasts (Morden *et al.* 1992), as does Rubisco amino acid composition (Figure 8). However, these two organelles are separated by analysis of other genes, including *psbA* and 16S rRNA. The 16S rRNA analysis suggests closer affinities between the cyanelle and rhodophyte plastids. Further, results from a study of GC *vs.* AT contents of genes of cyanelles and green chloroplasts indicate that similarities of cyanelles and chlorophyte chloroplasts may be artifactual (Lockhart *et al.* 1992) and are interpreted as consistent with the hypothesis that cyanelles arose from an endosymbiotic event separate from the one(s) that gave rise to green chloroplasts.

Some molecular data support the theory that euglenophyte plastids represent an endosymbiosis of a eukaryotic chlorophyte, including *psbA, tufA, rbcS,* and *rbcL* (Morden *et al.* 1992), as do Rubisco amino acid compositions (Figure 8). However, 16S rRNA analysis places the euglenoid chloroplast as a sister group to the rhodophytes and chromophytes (Douglas and Turner 1991). This and several other distinctive features of euglenoid chloroplasts put their genealogy in question (Morden *et al.* 1992).

Analysis of *rbcS* data support the interpretation that rhodophytes are ancestral to purple bacteria (Morden *et al.* 1992), a conclusion that is intuitively almost certainly wrong. If, however, it can be concluded that purple bacteria are ancestral to rhodophyte plastids, as Rubisco LSU data indicate, then the SSU result may be due to slower evolution of the rhodophyte gene relative to the bacterial. This result may be due to sequestering of genes, such as *rbcS,* in the plastids of rhodophytes (Valentin and Zetsche 1989) where selective pressures could be reduced. For example, Wolfe *et al.* (1987) demonstrated that plant chloroplast genomes evolve at half the rate of plant nuclei.

Summary and implications

Two different evolutionary histories are indicated for the algae and land plants by the results of different molecular analyses. In one, a cyanobacterium is considered to be the donor of plastids in all photosynthetic eukaryotic cells; while in the other, a cyanobacterium is associated with green-line organisms and a purple bacterium is related to the chromophyte line. Reconciling these two will be the main subject of the rest of this thesis. Several hypotheses have been put forward in the literature to rationalize these two different histories.

To achieve a unified hypothesis, Morden *et al.* (1992) proposed that a eubacterium, ancestral to the cyanobacteria, was the progenitor of the plastids of all photosynthetic eukaryotes. The Rubisco results could be explained as a lateral transfer of the Rubisco operon from a purple bacterium to an early rhodophyte thereby replacing the cyanobacterial Rubisco. Others had suggested this previously in slightly different forms and contexts (Boczar *et al.* 1989, Douglas *et al.* 1990, Morden and Golden 1991). Assali *et al.* (1991) cite support for the hypothesis in the finding that a strain of the purple bacterium, *Alcaligenes,* carries a plasmid that has a Rubisco gene that is highly homologous to that in a phaeophyte. Thus a source and a possible vector for horizontal transfer is indicated.

Cavalier-Smith (1989) proposes that a plastid received Rubisco by lateral transfer from a purple bacterium that was evolving towards a mitochondrion and had not yet lost its photosynthesis. However, 16S rRNA data for purple bacteria and mitochondria do not allow this possibility (Morden *et al.* 1992). Further, green line plastids exhibit lack of genetic transfer from mitochondria to chloroplasts, but the red line plastids are not well enough studied to indicate either way (Palmer *et al.* 1988). Another hypothesis suggests that the pre-endosymbiotic cyanobacterium had two distinct copies of Rubisco, losing different ones in the two plastid lines (Morden *et al.* 1992). However, no modern organism with suitable Rubiscos exists to support this conjecture.

Another hypothesis states that cyanobacteria may have been the progenitors of purple bacteria and that oxygenic photosynthesis preceded anoxygenic (Martin *et al.* 1992). The Rubisco sequence data support this in that both anoxygenic (purple bacteria) and oxygenic (cyanobacteria and eukaryotic algae) photosynthetic organisms are intermixed on widely separated branches (Martin *et al.* 1992). They also cited evidence that a PSI-like photosynthetic reaction center in *Heliobacterium chlorum* has a quinone receptor (Brok *et al.* 1986). This was taken to mean that the common ancestors of cyanobacteria and purple bacteria possessed both photosystems I and II. However, this evidence of a quinone is unconfirmed and regarded as inconclusive (Amesz 1989). Further, the wide separation of anoxygenic photosynthetic forms in the Rubisco sequence analysis is not supported by 16S rRNA data that instead finds these purple bacteria to be tightly clustered and separate from cyanobacteria and chloroplasts (Woese 1987, Markowicz and Loiseaux-de Goër 1991). The Rubisco composition analyses also counter this hypothesis, in that they unite all anoxygenic purple bacterial subdivisions (Figures 8, 9).

The order of accumulation of replication and metabolic components in the evolution of life is the subject of much speculation perhaps without sufficient evidence to support convincingly any one view. The following is discussed in the context of the hypothesis that oxygenic photosynthesis followed anoxygenic. A number of biochemical studies support the idea that the light reactions of photosynthesis evolved prior to the separation of the eubacterial line (Dickerson 1980, Schwartz and Dayhoff 1978, Woese 1987). It seems more parsimonious that the simpler photosystem with a single reaction center would have arisen first, and a double photosystem from that. There is evidence that the early environment contained sufficient reduced sulfur and organic compounds as electron donors to support photosynthesis using a single photosystem (Schidlowski *et al.* 1983).

The differences between these two disparate evolutionary histories are predicated on the assumption of a significant distinction between the cyanobacteria and purple bacteria. The cyanobacteria possess oxygenic photosynthesis and a concomitant double photosystem that the purple bacteria lack. These are certainly substantial differences, but the cyanobacteria and these innovations had to arise from somewhere and that question is addressed in the next section.

A step back, the origin of cyanobacteria

Through homologies of 16S rRNA sequences and photosynthetic reaction centers and electron transport it will be shown that cyanobacteria very likely arose from the purple bacteria. Evidence will be given that the second photosystem may have been contributed by a green sulfur bacterium or a gram positive bacterium. The origin of oxygen evolution remains unknown. Given this derivation of cyanobacteria, their Rubisco likely came from purple bacteria via direct ancestry.

16S rRNA

Cyanobacteria possess a distinctive rRNA signature, but its relationship to those from other eubacteria is ambiguous (Woese 1987). Several of the eubacterial lineages radiate from a central point in the rRNA tree in a fan-like manner. While, by rRNA analysis it cannot be stated with confidence that the cyanobacteria arose from the gram-positive or the purple bacteria, these data do indicate that they are at least as closely related as any of the eubacterial groups. Woese (1987) allows a closer relation between the cyanobacteria and the gram-positive bacteria, including *Heliobacterium,* based on a common and unique sequence identity. The green non-sulfur bacteria, the Chloroflexaceae, are not included in this fan-like radiation and by 16S rRNA have no close relatives (Woese 1987).

Photosystem II

The terms "PSII" and "PSI" are used in the literature to refer only to oxygenic photosynthetic systems and "reaction center" is the preferred term for anoxygenic photosynthesis. The homologies of PSII to reaction centers of photosynthetic bacteria are many, and are so well accepted that the bacterial reaction center, which is more fully studied in some respects, is used as a model to make predictions about PSII (Bryant 1986). However, the levels of relatedness within and between these two groups are variable. The PSII reaction centers of land plants, green algae, and cyanobacteria are nearly identical. This evaluation comes from the polypeptide compositions of the reaction centers, number and location of chlorophyll molecules relative to these polypeptides, and immunological relatedness.

In the purple bacteria, *Rhodobacter capsulatus, R. sphaeroides,* and *Rhodopseudomonas viridis,* the reaction center proteins include L, M, and H subunits. These proteins are homologous amongst both the different subunits and the different species. A photosystem protein in cyanobacteria and land plants, D1, is highly homologous within these oxygen-evolving groups (85-100%), and is also related to the purple bacterial L and M proteins (near 25%). A companion protein in PSII, D2, is homologous to D1, and also to L and M. Michel and Deisenhofer (1988) have examined the similarities between L, M, D1, and D2 using the structural relationships revealed by crystallization of the reaction center of a purple bacterium, *Rhodopseudomonas viridis.* These findings significantly strengthen the arguments of homology based on sequence data and thus the conclusion that these proteins are phylogenetically related.

The much less studied green non-sulfur bacteria, the Chloroflexaceae, also have a reaction center that resembles PSII and those of the purple bacteria in overall functional and structural arrangement (Glazer and Melis 1987, Olson and Pierson 1987).

Summing up, PSII of oxygenic photoautotrophs, at least in the lineages so far examined, appears conserved and bears sufficient resemblance to the reaction center of purple bacteria to indicate a common ancestry.

Photosystem I

The general electron transport pathways of PSI have properties in common with reaction centers of the green sulfur bacterium, *Chlorobium limicola* (Hurt and Hauska 1984). More specifically, the reaction center apoproteins (Hurt and Hauska 1984) and the Fe-S centers (Scheller and Møller 1990) have functional and structural homologies. The gram- positive bacterium, *Heliobacterium chlorum* also has some homologies to PSI (Olson and Pierson 1987). However, sequences of reaction center proteins are not yet available for these bacteria with which to make more definite comparisons.

Structural and functional parallels have also been made between PSI and PSII (Bryant 1986, Olson and Pierson 1987). A sequence homology (31%) has been observed between a 4 kD protein of PSI and helix E of the D2 protein of PSII in barley, however they are predicted to have opposite orientations in the membrane (Scheller and Møller 1990). Overall the two photosystems are taken as analogous but not homologous because there are no convincing protein sequence similarities (Bryant 1986, Scheller and Møller 1990).

There is another issue that concerns PSI and the question of origin of the cyanobacteria. Some cyanobacteria can conduct anoxygenic photosynthesis that uses only PSI and H_2S as the electron donor. It is induced by H_2S and involves inhibition of PSII. This could be explained as a holdover from an earlier evolutionary stage before the evolution of the double photosystem, or as a transition phase where either mode was facultatively available. Alternatively it could be a new adaptation to high H_2S.

Light harvesting complexes

There is a trend that could be called "mix and match" with regard to the association of light harvesting complexes (LHCs) with photosynthetic reaction centers (Glazer and Melis 1987). In some groups there may have been a grafting of different types of LHC on a single type of photosystem. The chlorophyll *b* -containing LHC of PSII of green line algae and plants, on one hand, and the phycobilisomes of cyanobacteria and rhodophytes, on the other, are both associated with their PSII reaction centers. Cryptophytes use a mixture of both phycobilins and chlorophyll. Instead of transferring energy from phycoerythrin to the PSII reaction center via allophycocyanin, they use chlorophyll *c* as an intermediary (Raven *et al.* 1989). It may be phylogenetically significant that the physical bonding of both types of antennae to the reaction center is relatively weak, in that there are also probably low functional constraints on the interaction of these components, perhaps facilitating interchange of antennae in the evolutionary sense.

A different type of grafting is apparent in the green sulfur versus the green non-sulfur bacteria. They have highly homologous light harvesting structures, called chlorosomes, grafted onto different types of reaction centers: a PSI-like one in the green sulfur, and a PSII-like one in the green non-sulfur bacteria.

With regard to the question of origin of cyanobacteria it would seem that there is some promiscuity in the associations of light harvesting structures with reaction centers. It has been assumed that characters of light harvesting complexes were primary in linking groups above the phylum level. In particular, phycobilisomes were used to ally cyanobacteria and rhodophytes to the exclusion of chlorophytes and other algae. On the molecular and ultrastructural level, as indicated above, these structures seem to be associated rather loosely. As such they may be less reliable indicators of phylogeny than has been assumed, and the affiliations inferred from them could then be considered less certain.

Cytochrome $b/c_1/f$ complexes

Cytochrome b/c_1 complexes of electron transport chains have been found to be homologous between a purple bacterium (*Rhodobacter capsulatus*), a mitochondrion (*Saccharomyces cerevisiae*) and a spinach chloroplast (*Spinacea oleracea*) (Gabellini 1988). The structure of the cytochrome *b* component has been highly conserved: that for *R. capsulatus* bears 41% homology with the yeast mitochondrial sequence and 26% identity to the corresponding cytochrome b_6 plus subunit IV from the cytochrome complex of spinach chloroplasts. The Reiske Fe-S protein of *R. capsulatus* is 46% homologous with the mitochondrial, and 14% with the chloroplastic, mainly at the C-terminus. A central region of this protein is highly homologous between the bacterium and the mitochondrion; the difference seen in the chloroplastic protein is related to the interaction of this region with the considerably different cytochrome *f*. Cytochrome c_1 from *R. capsulatus* is 31% identical to the mitochondrial cytochrome c_1, but only 12% to cytochrome *f* from the chloroplast. Despite this, the chloroplastic cytochrome *f* bears conserved heme binding sites and general topology (hydropathy profiles).

The conclusion of this investigation (Gabellini 1988) is that the *b* cytochromes and the C-terminus of the Fe-S proteins are strictly conserved in purple bacteria, mitochondria, and green chloroplasts, and arose only once. These regions interact with the quinones. In contrast, the sites involved in the cytochrome reductase activity, cytochrome c_1 and the central region of the Fe-S protein, are homologous between the purple bacterium and mitochondria but not with the plastocyanin reductase site of the chloroplast Fe-S protein and cytochrome *f*. Whether the reductase protein sequences diverged in a splitting of cyanobacteria from purple bacteria, or whether the similar catalytic function arose convergently is not known, at least from these chloroplast data.

Electron transport for photosynthesis and aerobic respiration are performed by the same cytochrome b/c_1 complex of purple bacteria. The homologies to corresponding mitochondrial proteins suggest that aerobic respiration arose once in the purple bacteria, and support the hypothesis that mitochondria derived from an endosymbiosis of a purple bacterium. The evolution of cyanobacteria from purple bacteria, directly or as a sister group, is less well established by these data (Bryant 1986).

Cytochrome *c* / Plastocyanin

This function of mobile carrier from the cytochrome *b/c /f* complex can be performed by one of two possible structures: a *c* cytochrome or plastocyanin. The *c* cytochromes are iron-containing heme proteins. Plastocyanin is a soluble protein carrying copper coordinated to several amino acid residues. It occurs only in photosynthetic electron transport chains (ETCs). Cytochrome *c* is the only mobile electron carrier fulfilling this role in the mitochondrion ETC (Jones and Poole 1985). There are several other pertinent observations concerning the distribution of these mobile carriers.

Plastocyanin appears to occur only in the green and cyanobacterial lines and not in the rhodophyte and chromophyte line. Cyanobacteria and many green algae can facultatively switch between plastocyanin and cytochrome *c*, depending on copper availability. The more advanced forms on the green line use only plastocyanin. Euglenophytes use only cytochrome *c*. All chromophytes and rhodophytes tested so far apparently use only cytochrome *c*. There are no data yet on the type of mobile carrier used by cryptophytes or dinoflagellates (Raven *et al.* 1989).

When compared with other *c* -type cytochromes, those associated with oxygenic photosynthesis are quite distinct from others, including mitochondrial and purple bacterial (Schwartz and Dayhoff 1978). As far as information on the origin of cyanobacteria there are no clear indications in this cytochrome, but they may have some significance regarding subsequent evolution, as described above.

Plastocyanin is related to several other copper-containing proteins, including azurin and plantacyanin (Aitken 1988). Azurin is a bacterial respiratory protein not occurring in photosynthetic bacteria. A sequence comparison including both plastocyanin and azurin clearly separate them into two clusters. There is a large phenetic distance between these two clusters, the plastocyanins of land plants, algae, and cyanobacteria on the one hand and the azurins of aerobic bacteria on the other. However, at copper-binding sites in both molecules there appears to be sequence conservation. This relationship between copper proteins is taken as evidence that cyanobacteria arose from purple bacteria (Hunt *et al.* 1985). But Aitken (1988) disagrees allowing only a functional similarity resulting in convergence on similar motifs and not common ancestry.

Oxygen evolving complex

The oxygen evolving complex (OEC) is perhaps the most significant feature of photosynthesis in cyanophytes, prochlorophytes, and eukaryotic photoautotrophs. In spite of this importance the OEC has remained the "inner sanctum" of photosynthesis that only recently is beginning to be opened up. With regard to revealing the phylogenetic derivation of water splitting and oxygen evolution even less information has been forthcoming. The 33 kD extrinsic protein of the OEC is central to the function of water splitting (Ghanotakis and Yocum 1990). The 33 kD protein of spinach was determined to have considerable local homology to the manganese-liganding site of Mn-superoxide dismutases of the purple bacterium, *E. coli,* and the gram positive bacterium, *Bacillus sterothermophilus* (Bryant 1986), but these relationships have not been confirmed by sequences from other species (Ghanotakis and Yocum 1990). Overall, the limited understanding of the structural and functional complexities of the OEC does not yet allow a clear indication of the origin of the complex. However, if the apparent trend from cyanobacteria to land plants of adding components to the OEC, and their likely role in modification of function, can be extrapolated backwards, then the modern OEC may have arisen from a step-wise accumulation and modification of components and functions. Perhaps a more complete understanding of how the OEC works will help reveal its origin.

Implications of origin of cyanobacteria from purple bacteria

Photosystems

The two photosystems of cyanobacteria probably originated from one photosystem. An early use of photosynthetic reaction centers was likely to employ a charge separation to drive the transfer of ions across membranes. Elaborations in the reaction centers likely came about involving coordinating metals, modifications of the porphyrin-derivative, and participation of membrane-bound proteins, as well as integration with other electron transport chain components. Under evolutionary pressures of competition for light and reducible substrates, two different reaction center types may have evolved by developing different metabolic strategies for energy conversion. The type of reaction center homologous to PSI (RCI) directly produces reducing power. The PSII-like reaction center (RCII) establishes a proton gradient that is used for ATP production. (Chapman and Schopf 1983, Gest and Schopf 1983, Olson and Pierson 1987, Alberts *et al.* 1989, Kondratieva *et al.* 1991).

These different photosynthetic reaction centers probably arose in different organisms, rather than differentiating within one organism. In addition to the evidence of reaction center homologies recounted above, several other factors influence this view. The diversity of RC types persists today in separate organisms occupying environments that were probably more widespread in the ancient world. The main advantage of the double photosystem is to allow extraction of electrons from low electron potential compounds, such as water, and still promote them to ferredoxin. Evolutionary pressure to turn to increasingly recalcitrant electron donors was likely not felt until supply of readily reducible substrate became limiting, a situation that likely did not arise until after reduced forms of sulfur had supported life and evolution for some time. The geological record indicates that oxygenic photosynthesis did not come about until later, likely after much diversification had already taken place (Chapman and Schopf 1983).

With regard to the issue of two organisms contributing to a double photosystem in one organism, mechanisms of genetic transfer between prokaryotes are relatively common (Amábile-Cuevas and Chicurel 1993). Also, symbiotic associations of prokaryotes with single photosystems can be observed today in the consortia of green sulfur bacteria with chemotrophic eubacteria of the family Chromatiaceae, purple bacteria (Trüper and Pfennig 1991). Thus the opportunity for genetic transfer or endosymbiosis likely existed.

Rubisco

In the matter of origin of cyanobacterial Rubisco recall that while there are homologies of the green non-sulfur bacterial reaction centers to PSII and of green sulfur bacterial reaction centers to PSI, neither possesses Rubisco. Therefore these two could not have amalgamated to form cyanobacteria. On the other hand purple bacteria do possess Rubisco and their reaction centers are highly homologous to PSII. Thus it seems reasonable that the origin of cyanobacterial Rubisco is from the purple bacteria, because the candidate for contributor of the other photosystem, the green sulfur bacteria, did not have Rubisco.

Another hypothesis

The above evidence indicates that the cyanobacteria likely arose out of the purple bacteria. Also described above are indications by some molecules that chromophyte plastids arose from cyanobacteria, and conflicting evidence from Rubisco that they arose from purple bacteria. The hypotheses to resolve this, mentioned in a previous section, have assumed the differences between cyanobacteria and purple bacteria were primary. The following is an alternate explanation to resolve this conflict based on timing of evolution of cyanobacteria from purple bacteria and on timing of evolution of these molecular markers.

Evidence of the antiquity of Rubisco exists in carbon isotope ratios of ancient sediments, where a Rubisco-like signal (high $\delta^{13}C$) goes back 3.5 $\times 10^9$ years before present (Schidlowski *et al.* 1983). The enzyme had almost certainly evolved before this date, as suggested by a report of occurrence of Rubisco in an archaebacterium (Altekar and Rajagopalan 1990), thus indicating it was present before the archaebacteria and eubacteria diverged.

The earliest estimate for onset of oxygenic photosynthesis is based on microfossils of probable cyanobacterial nature and set at 3.5 $\times 10^9$ years before present (Schopf 1993). Abundant ferrous iron sequestered oxygen as it was produced, creating a time lag before oxygen began to accumulate and a stable oxygen environment became established by about 1.7 $\times 10^9$ years ago (Walker *et al.* 1983).

Rubisco evolution may have tracked the same pattern as oxygen accumulation. That is, a slow evolutionary change prior to the appearance of free oxygen, a rapid change as oxygen accumulated in the atmosphere, and subsequent slow rate of change once modern oxygen tensions had been achieved. Further, during this evolution of

Rubisco, two endosymbioses may have taken place, widely separated in time, that resulted in different lineages of plastids. In the early stages of oxygen evolution by cyanobacteria an endosymbiosis may have resulted in rhodophyte plastids. It is possible that at this time CO_2 levels had not yet dropped nor O_2 levels gone up. As a result Rubisco could have been still much like the ancestral purple bacteria form. Subsequently, following full oxidation of reduced minerals, the CO_2 to O_2 ratio shifted and cyanobacterial Rubisco was forced to change due to a newly realized oxygenase reaction and the inefficiency that reaction involves. Somehow the Rubisco in the rhodophyte plastid was largely protected from this, possibly by carbon concentrating mechanisms that are common, but not yet understood, in modern rhodophytes and chromophytes (Raven 1991, Hobson 1994), or by oxygen scavenging mechanisms that very likely preceded oxygen evolution. Nevertheless, any evolution of Rubisco subsequent to this proposed endosymbiosis appears to have been in a different direction from what the cyanobacteria and chlorophytes undertook. At some later point, following Rubisco evolution, a now essentially modern cyanobacterium may have been endosymbiotically transformed to a chlorophyte plastid. Thus, the rhodophyte/chromophyte line possesses a Rubisco of a proto-cyanobacterium that closely resembles the ancestral purple bacterial Rubisco, and the chlorophyte line possesses a Rubisco from a modern cyanobacterium that is quite unlike purple bacteria. This hypothesis depends on a slow evolutionary rate for Rubisco, except for the transition period when oxygen levels rose and CO_2 levels dropped. Also any evolution of rhodophyte/chromophyte Rubisco during this transition may have been less extensive than expressed by cyanobacteria due to sequestering of the protein and the genes in the rhodophyte/chromophyte plastid. This is analogous to the hypothesis of a punctuated equilibrium for evolutionary rates of organisms (Eldredge and Gould 1972) applied to molecular evolution.

Support for the hypothesis of two primary endosymbioses

Timing

While compelling evidence for existence of eukaryotes goes back as far as 1.4 $\times 10^9$ years ago, some microfossils of large size (0.3 - 0.9 mm) found in rock dating 1.9 - 1.7 $\times 10^9$ years old have been identified as possibly rhodophytan (Hofmann and Schopf 1983). This corresponds to the time when oxygen levels were beginning to rise. Possibly the endosymbiosis proved to be a mechanism allowing protection of Rubisco from rising oxygen. Possibly the host had evolved mechanisms superior to those of the endosymbiont to protect itself from the generally toxic effects of oxygen. The endosymbiont may have found advantage in this, thus also protecting its Rubisco from evolutionary pressures. Possibly the endosymbiosis occurred at an earlier time, and merely for the advantage to the host of endogenous carbon fixation in the face of increasing competition for exogenous reduced carbon. Lack of a better fossil record of this event may lie in lack of fossilizable cell walls of the rhodophyte. Also, the host may have been unicellular and planktonic, factors that contrast with the benthic filaments of stromatolites that dominate the scant fossil record of this very early time.

This proposal of an endosymbiosis before a stable high oxygen level is not necessarily incompatible with the definition of a eukaryote. It has been shown that a group of anaerobic prokaryotes, including *Giardia,* that lack mitochondria are placed at the base of the eukaryote branch of the 16/18S rRNA tree, nearest the prokaryotes (Sogin *et al.* 1989). These findings indicate they have always been anaerobic and lacking mitochondria (Kabnick and Peattie 1991). The polyphyletic origin of mitochondria implied by the establishment of lines of eukaryotes without mitochondria is supported by data that shows that land plants have had a different source of mitochondria than green algae (Gray *et al.* 1989). Thus it is possible that establishment of chloroplasts preceded mitochondria and the timing proposed is not unreasonable.

Rubisco

Oxygen evolution occurs in close proximity to Rubisco and their incompatibility may have been felt before oxygen began to accumulate in the atmosphere. Even before oxygenic photosynthesis it is almost certain that organisms had developed mechanisms for dealing with oxygen. Small amounts of oxygen were produced in the early environment by non-biological means such as photolysis of water (Chapman and Schopf 1983). While there may have been spatially and temporally localized rises in oxygen due to oxygenic photosynthesis, it is likely that proto-cyanobacteria also had these mechanisms of oxygen handling. Further, it is possible that the endosymbiont was taken up by a unicellular planktonic urkaryote, and this habit helped diffuse oxygen away from the cell, in contrast to a larger build-up of oxygen that can occur in benthic mats.

There is other support for this hypothesis in the Rubisco analyses. The associations of prokaryotes and eukaryotes has a different pattern in the two main branches (Figure 9). The cyanobacteria are intermixed with the chlorophytes, whereas the purple bacteria constitute a separate sub-branch of the chromophyte branch. This could reflect the time since the endosymbiosis, and possibly the different evolutionary pressures exerted over those times. It may be that the purple bacteria, with a long time since separation from the proto-cyanobacteria, have since differentiated themselves, and diverged into their subclasses. Under the hypothesis advanced here, the cyanobacteria, on the other hand, have had less opportunity to diverge since the recent endosymbiosis and are less distinguishable from the chlorophyte chloroplasts. The branch lengths of these two main groups also differ, with the chromophyte/purple bacterial branches longer than those of the chlorophyte/cyanobacteria group. This also may reflect the time since endosymbiosis. These trends are also observable in the Rubisco sequence tree (Martin *et al.* 1991).

While the above considerations support the hypothesis, a characteristic of Rubisco kinetics does not. Rubisco efficiency can be measured and expressed in terms of a substrate specificity factor.

$$\tau = {}^{V_c\,K_o}/_{V_o\,K_c},$$

where V_c and V_o are the maximal velocities of the carboxylase and oxygenase reactions respectively, and K_c and K_o are the half saturation constants for these reactions. It can be seen that higher values of τ represent a higher effectiveness of Rubisco for carboxylation. In support of the hypothesis there is found to be a general trend of increasing specificity factor with evolution (Jordan and Ogren 1981, Kent and Tomany 1984, Lee *et al.* 1993) (Table 17). The hypothesis would predict that the chromophyte lineage would possess a Rubisco intermediate between that of purple bacteria and that of cyanobacteria. The one instance of data presently available (Read and Tabita 1992) does not support this hypothesis; the diatom has the highest specificity factor. However, if the specificity factor is partitioned between maximal velocity and half-saturation constants a different pattern for the diatom can be observed, compared to any of the other groups (Table 17). Clearly, more information about these kinetic parameters for chromophytes is required before any conclusion is possible.

Other data

There is support for this hypothesis in the contrasting distributions of cytochrome *c versus* plastocyanin in photosynthetic electron transport in algae and bacteria (Raven *et al.* 1989). Copper was likely unavailable before oxygen levels rose due to its insolubility in the early reduced environment. Copper metalloproteins are not seen in modern strict anaerobes (Chapman and Schopf 1983). Also, rhodophyte and chromophyte plastids have only cytochrome *c*, compatible with the suggestion of a more

Table 17. Rubisco specificity and other kinetic factors for several taxa. V_c, V_o) maximal velocity for carboxylase and oxygenase reactions of Rubisco. K_c, K_o) half-saturation constants of same. Specificity factor, $\tau = V_c K_o / V_o K_c$.

Taxon		Specificity factor, τ	V_c/ V_o	K_c	K_o/ K_c	ref.[1]
C3 plants						
Glycine max		82	1.7	9	48	1
Tetragonium expansa		81	1.8	13	46	1
Spinacea oleracea		80	2.3	14	34	1
Lolium perenne		80	2.6	16	31	1
Nicotiana tabacum		77	1.3	11	59	1
C4 plants						
Amaranthus hybridus		82	2.0	16	40	1
Zea mays		78	3.3	34	24	1
Chlorophyte algae						
Scenedesmus obliquus		63	3.6	38	17	1
Chlamydomonas reinhardii		61	3.7	29	17	1
Euglena gracilis		54	3.3	25	16	1
Diatom						
Cylindrotheca strain N1		107	5.5	33	19	2
Cyanobacteria						
Aphanizomenon flos-aquae		48	5.1	105	9.4	1
Cocochloris peniocystis		47	4.7	121	10	1
Synechococcus PCC 6301		41	8.9	142	4.7	2
Purple bacteria						
Rhodospirillum rubrum	type II	15	3.3	89	4.6	1
Rhodobacter sphaeroides	type II	9	3.3	80	2.8	1
" "	type I	62	2.7	36	23	1

[1] References:

1. Jordan and Ogren. 1981. *Nature* 291:513-515.
2. Read and Tabita. 1992. *Biochemistry* 31:5553-5560.

anaerobic history. The copper-containing protein, plastocyanin occurs in chlorophytes and modern cyanobacteria, indicating a condition that arose after oxygen levels were elevated.

Other molecular markers such as 16S rRNA may not have been under the same evolutionary pressure by the level of free oxygen and thus did not undergo rapid evolution at any time. Thus, it may be that they have undergone little evolutionary change through time or have evolved at about the same rate so as to obscure the distinction between the two endosymbioses. Slow evolution may be related to the close association of rRNA with ribosomal proteins and the tight constraints this relationship may impose (Martin *et al.* 1992).

On the other hand, molecular markers such as the D1 protein of PSII, may have undergone a rapid evolution before either endosymbiosis. It could be argued that this protein would have experienced evolutionary pressure in the process of developing water splitting activity, because it is so close to the active site of the OEC, and may even form a part of it. The distinction of D1 from its counterparts in purple bacteria is reflected in their having only 25% homology (Bryant 1986), but the conservation of D1 among oxygenic photoautotrophs is reflected in its 85% homology among cyanobacteria, rhodophytes, and chlorophytes (Morden *et al.* 1992). The changes involved in development of water splitting would have occurred before the first endosymbiosis proposed for the rhodophyte plastid. Subsequent evolution of this protein may have been more gradual.

The *tufA* analysis placed a primitive cyanobacterium, *Gloeobacter,* close to the rhodophytes and chromophytes, suggesting that this organism may be the progenitor of plastids in these groups. 16S rRNA analysis placed *Gloeobacter* at the base of the cluster of cyanobacteria and chloroplasts, closest to the purple bacteria (Douglas and Turner 1991). This cyanobacterium differs from other cyanobacteria in having no thylakoids; photosynthesis is carried out on the plasma membrane. This arrangement is similar to that for several purple bacteria (Glazer and Melis 1987). These data suggest that among modern cyanobacteria, *Gloeobacter* may be most like the proto-cyanobacterium that gave rise to the rhodophyte plastid. Its lack of thylakoids suggests it may even predate that stage. It would be interesting to see if Rubisco from this organism were typically cyanobacterial or more like purple bacteria.

Whence the dinoflagellates?

The results for the dinoflagellate Rubisco deserve separate attention. The numerous unique features of dinoflagellates, outlined in the Introduction, separate them from other protists, at a greater distance than between most other groups. Whether this separation represents their splitting from the eukaryote lineage at a very early stage, or represents a faster rate of evolution compared to other eukaryotes is not known yet. In either case, the highly divergent Rubisco indicated here reinforces the distinctiveness of dinoflagellates.

These Rubisco results could be taken to indicate a very different endosymbiont for dinoflagellate chloroplasts, a now lost photosynthetic progenitor. However, conclusions involving missing ancestors are not very tenable. Alternatively, the progenitor of dinoflagellate chloroplasts may have been part of the rhodophyte and chromophyte lineage, as indicated by possession of chlorophyll *c,* but divergence may have taken Rubisco in a very different path. If a different path were the case, it might be related to the utilization of heterotrophy by many dinoflagellates. Martin *et al.* (1992) have suggested that the Rubisco of a heterotrophic euglenoid, *Astasia,* is evolving more rapidly than for others because it is no longer used for photosynthesis. Perhaps this could also be argued for dinoflagellates. However, heterotrophy is also widespread among some classes of chrysophytes (Smith and Hobson 1994), albeit only by uptake of organic molecules and never by capture of living particle as is found in the dinoflagellates.

A case for facultative or evolutionary switching between heterotrophy and autotrophy for dinoflagellates has points both for and against it. Dodge (1983) has proposed that dinoflagellates are essentially heterotrophic, with photoautotrophy being acquired at various stages in its evolutionary history. Further, only half of the dinoflagellate species have photosynthetic capability, and many of the heterotrophs have no sign of ever possessing plastids (Dodge 1983, Taylor 1990). However, eyespots found in some species have been speculated to represent relic chloroplasts (Dodge 1983), thus possibly indicating a switch from photosynthesis back to a heterotrophic mode. Also, several photosynthetic species have heterotrophic capabilities. *Amphidinium carterae* is photosynthetic and possesses the distinctive carotenoid, peridinin, that is taken to indicate an endosymbiosis early in dinoflagellate evolution. Thus, the heterotrophic argument would not seem to apply. However, growth of *A. carterae* was enhanced by glycerol (Gaines and Elbrächter 1987) and other *Amphidinium* species are nonphotosynthetic, but it is not known whether heterotrophy represents an ancestral or derived condition for members of this genus. Clearly, more work needs to be done to determine a mechanism that may have resulted in a different Rubisco in dinoflagellates.

The net conclusion that can be drawn regarding *Amphidinium* is that the chloroplasts of, at least some dinoflagellates have diverged sufficiently from their ancestral endosymbiotic condition as to obscure reliable evidence of their origins. There is weak evidence that they may have arisen from the chromophyte lineage. If so, however, the greater depth of differences between these two groups indicates that the dinoflagellates have taken a radically different evolutionary path. There may have been two eras of endosymbiosis of dinoflagellate chloroplasts (Whatley 1989). *Amphidinium,* with its very different Rubisco, may represent that earlier era. Members of the more recent era would likely be placed differently among the Rubisco and other data. Possession of peridinin may be correlated to that earlier endosymbiosis.

Conclusion

Analyses of Rubisco LSU amino acid composition provided plausible estimates for phylogeny of chloroplasts. The two techniques used to analyze these data gave slightly different results. For example, the SΔn distance technique did not keep the chrysophytes as a homogeneous group, and because of this the other technique, principal components analysis, may be more reliable. There were also some minor differences between these and the phylogeny predicted by analysis of Rubisco sequences, as reported in the literature. These differences may be related to uncertainties in the compositional data.

Still, all analyses of Rubisco LSU gave the same result in finding two major lineages of plastids, the chlorophyte and the chromophyte. Inclusion of amino acid compositions and sequences reported in the literature with these data allowed inferences to be made about origins of these plastids from prokaryotic endosymbionts. The Rubisco analyses indicated that plastids of the chlorophyte lineage originated from cyanobacteria, and those of the chromophyte lineage from purple bacteria. Analyses of other molecules predicted a different origin - all plastids derived from cyanobacteria. These conflicting results demanded a resolution.

A hypothesis based on an origin of cyanobacteria from purple bacteria was explored. The hypothesis indicated that two endosymbioses of cyanobacteria may have occurred, separated in time. During this time interval, Rubisco was hypothesized to have undergone a radical evolution, such that it changed from a Rubisco characteristic of the ancestral purple bacteria to that resembling Rubisco of cyanobacteria. The proposed driving force for this change was a switch from an anoxic to an oxic environment, and the disadvantage to the organisms of Rubisco oxygenase activity. Several lines of evidence supported this hypothesis. However, a kinetic characteristic of Rubisco did not support the placement of chromophytes between purple bacteria and cyanobacteria. More importantly this kinetic characteristic concerns the Rubisco oxygenase function and brought into question the proposed mechanism for change. Clearly, there are not enough data to make a final decision on the validity of this hypothesis, especially since so little is known about the physiology of chromophytes, particularly Rubisco and other aspects of carbon acquisition.

Also a dinoflagellate was found to have a quite different Rubisco, indicating a different evolutionary history for their plastids. This finding fit with the general uniqueness of this group, but does not allow the postulation of a phylogenetic history for them.

Some chromophyte physiology was found to have greater similarities to that of purple bacteria. The purple bacteria have not previously been considered as models for chromophyte physiology. This suggestion of ancestry now allows consideration of an expanded set of models with which to examine chromophyte physiology. From this consideration may come clues to explain some of the puzzles of algal physiology and ecology.

References

Aitken, A. 1988. Protein sequences as taxonomic probes of cyanobacteria. *Methods in Enzymology* 167:145-157.

Akazawa, T., Takabe, T., and Kobayashi, H. 1984. Molecular evolution of ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO). *Trends in Biochemical Sciences* 9:380-383.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. D. 1989. *Molecular Biology of the Cell. 2nd ed.* Garland, New York. 1262 pp.

Aldrich, J., Cherney, B., Merlin, E., and Palmer, J. 1986a. Sequence of the *rbc* L gene for the large subunit of ribulose bisphosphate carboxylase-oxygenase from petunia. *Nucleic Acids Research* 14:9534-9534.

Aldrich, J., Cherney, B., Merlin, E., and Palmer, J. 1986b. Sequence of the *rbc* L gene for the large subunit of ribulose bisphosphate carboxylase-oxygenase from alfalfa. *Nucleic Acids Research* 14:9535-9535.

Altekar, W. and Rajagopalan, R. 1990. Ribulose bisphosphate carboxylase activity in halophilic *Archaebacteria. Archives of Microbiology* 153:169-174.

Amábile-Cuevas, C. F. and Chicurel, M. E. 1993. Horizontal gene transfer. *American Scientist* 81:332-341.

Amesz, J. 1989. Energy transfer and electron transport in *Heliobacterium chlorum. Photosynthetica* 23:403-410.

Amiri, I., Salnikow, J. and Vater, J. 1984. Amino acid sequence of the large subunit of D-ribulosebisphosphate carboxylase/oxygenase from *Nicotiana tabacum. Biochimica et Biophysica Acta* 784:116-123.

Andersen, K. and Caton, J. 1987. Sequence analysis of the *Alcaligenes eutrophus* chromosomally encoded ribulose bisphosphate carboxylase large and small subunit genes and their gene products. *Journal of Bacteriology* 169:4547-4588.

Assali, N.-E., Martin, W. F., Sommerville, C. C., and Loiseaux-de Goër. 1991. Evolution of the Rubisco operon from prokaryotes to algae: Structure and analysis of the *rbcS* gene of the brown alga *Pylaiella littoralis. Plant Molecular Biology* 17:853-863.

Berry-Lowe, S. L., McKnight, T. D., Shah, D. M., and Meagher, R. B. 1982. The nucleotide sequence, expression and evolution of one member of a multigene family encoding the small subunit of ribulose-1,5-bisphosphate carboxylase in soybean. *Journal of Molecular and Applied Genetics* 1:483-498.

Black, J. A. and Harkins, R. N. 1977. Amino acid compositions and evolutionary relationships with protein families. *Journal of Theoretical Biology* 66:281-295.

Boczar, B. A., Delaney, T. P., and Cattolico, R. A. 1989. Gene for the ribulose-1,5-bisphosphate carboxylase small subunit protein of the marine chromophyte *Olisthodiscus luteus* is similar to that of a chemoautotrophic bacterium. *Proceedings of the National Academy of Science, U. S. A.* 86:4996-4999.

Bold, H. C. and Wynne, M. J. 1985. *Introduction to the Algae, Structure and Reproduction.* 2nd ed. Prentice-Hall, Englewood Cliffs, N.J., 720 pp.

Bourrelly, P. 1968. *Les Algues d'Eau Douce. Initiation à la systématique. III. Les algues jaunes et brunes, Chrysophycées, Phéophycées, Xanthophycées et Diatomées.* N. Boubée, Paris. 512 pp.

Bourrelly, P. 1970. *Les Algues d'Eau Douce. Initiation à la systématique. III. Les algues bleues et rouges, les Eugléniens, Péridiniens et Cryptomonadines.* N. Boubée, Paris. 512 pp.

Bowien, B. and Mayer, F. 1978. Further studies on the quaternary structure of D-ribulose-1,5-bisphosphate carboxylase from *Alcaligenes eutrophus. European Journal of Biochemistry* 88:97-107.

Brok, M., Vasmel, H., Horikx, J. T. G., and Hoff, A. J. 1986. Electron transport components of *Heliobacterium chlorum* investigated by EPR spectroscopy at 9 and 35 GHz. *FEBS Letters* 194:322-326.

Bryant, D. A. 1986. The cyanobacterial photosynthetic apparatus: comparisons to those of higher plants and photosynthetic bacteria. *In* Platt, T. and Li, W. K. W. (Eds.) *Photosynthetic Picoplankton. Canadian Bulletin of Fisheries and Aquatic Sciences* 214:423-500.

Bujak, J. P. and Williams, B. L. 1981. The evolution of dinoflagellates. *Canadian Journal of Botany* 59:2077-2087.

Cavalier-Smith, T. 1982. The origin of plastids. *Biological Journal of the Linnean Society* 17:289-306.

Cavalier-Smith, T. 1989. The kingdom Chromista. *In* Green, J. C., Leadbeater, B. S. C., and Diver, W. L. (Eds.) *The Chromophyte Algae: Problems and Perspectives* , Systematics Association Special Volume No. 38. Clarendon, Oxford. pp. 381-407.

Chapman, D. J. and Schopf, J. W. 1983. Biological and biochemical effects of the development of an aerobic environment. *In* Schopf, J. W. (Ed.) *Earth's Earliest Biosphere* Princeton University Press, Princeton NJ. pp. 302-320.

Chapman, V. J. and Chapman, D. J. 1973. *The Algae. 2nd Ed.* MacMillan, New York. 497 pp.

Chesnick, J. M. and Cattolico, R. A. 1989. The small subunit gene is chloroplast encoded in two binucleate dinoflagellate species. *In Proceedings of the Fourth Northwest Algal Symposium 1989.* Walla Walla College, Wash.

Chory, J., Muller, E. D., and Kaplan, S. 1985. DNA-Directed in vitro synthesis and assembly of the Form II D-ribulose-1,5-bisphosphate carboxylase/oxygenase from *Rhodomonas sphaeroides. Journal of Bacteriology* 161:307-313.

Christensen, T. A. 1962. *In* Böcher, T. W., Lange, M., and Sørensen, T. (Eds.) *Systematisk Botanik, Vol. 2, No. 2* Munksgaard, Copenhagen. 178 pp.

Clayton, M. N. 1989. Brown algae and chromophyte phylogeny. *In* Green, J. C., Leadbeater, B. S. C., and Diver, W. L. (Eds.) *The Chromophyte Algae: Problems and Perspectives.* Systematics Association Special Volume No. 38, Clarendon, Oxford. pp. 229-253.

Cornish-Bowden, A. 1977. Assessment of protein sequence identity from amino acid composition data. *Journal of Theoretical Biology* 65:735-742.

Cornish-Bowden, A. 1979. How reliably do amino acid composition comparisons predict sequence similarities between proteins? *Journal of Theoretical Biology* 76:369-386.

Cornish-Bowden, A. 1980. Critical values for testing significance of amino acid composition indexes. *Analytical Biochemistry* 105:233-238.

Cornish-Bowden, A. 1983. Relating proteins by amino acid composition. *Methods in Enzymology* 91:60-75.

Curtis, S. E. and Haselkorn, R. 1983. Isolation and sequence of the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase from the cyanobacterium *Anabaena* 7120. *Proceedings of the National Academy of Science U. S. A.* 80:1835-1839.

Dayhoff, M. O. and Hunt, L. T. 1972. Composition of Proteins. *In* Dayhoff, M. O. (Ed.) *Atlas of Protein Sequence and Structure. Vol. 5.* National Biomedical Research Foundation, Washington, D. C. p. D355-D359.

Dayhoff, M. O., Schwartz, R. M., and Orcutt, B. C. 1979. *In* Dayhoff, M. O. (Ed.) *Atlas of Protein Sequence and Structure. Vol. 5, Suppl. 3.* National Biomedical Research Foundation, Washington, D. C. p. 345-352.

Dickerson, R. E. 1980. Evolution and gene transfer in purple photosynthetic bacteria. *Nature* 283:210-212.

Dodge, J. D. 1983. Dinoflagellates: investigation and phylogenetic speculation. *British Phycological Journal* 18:335-356.

Dodge, J. D. 1987. Dinoflagellate ultrastructure and complex organelles: A. General ultrastructure. *In* Taylor, F. J. R. (Ed.) *The Biology of the Dinoflagellates* Botanical Monographs Vol. 21 Blackwell Scientific, Oxford. pp. 92-119

Dodge, J. D. 1989. Phylogenetic relationships of dinoflagellates and their plastids. *In* Green, J. C., Leadbeater, B. S. C., and Diver, W. L. (Eds.) *The Chromophyte Algae: Problems and Perspectives.* Systematics Association Special Volume No. 38, Clarendon, Oxford. pp. 207-227.

Doolittle, R. F. 1987. *Of Urfs and Orfs: A Primer on How to Analyze Derived Amino Acid Sequences.* University Science Books, Mill Valley CA. 103 pp.

Dörhöfer, G. and Davies E. H. 1980. Evolution of the archeopyle and tabulation in rhaetogonyaulacinean dinoflagellate cysts. *Royal Ontario Museum Life Sciences Miscellaneous Publications.* pp. 1-91.

Douglas, S. E. and Turner, S. 1991. Molecular evidence for the origin of plastids from a cyanobacterium-like ancestor. *Journal of Molecular Evolution* 33:267-273.

Douglas, S. E., Durnford, D. G., and Morden, C. W. 1990. Nucleotide sequence of the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from *Cryptomonas* Φ: evidence supporting the polyphyletic origin of plastids. *Journal of Phycology* 26:500-508.

Douglas, S. E., Murphy, C. A., Spencer, D. F., and Gray, M. W. 1991. Cryptomonad algae are evolutionary chimeras of two phylogenetically distinct unicellular eukaryotes. *Nature* 350:148-151.

Dron, M., Rahire, M. and Rochaix, J. D. 1982. Sequence of the chloroplast DNA region of *Chlamydomonas reinhardii* containing the gene of the large subunit of ribulose bisphosphate carboxylase and part of its flanking genes. *Journal of Molecular Biology* 162:775-93.

Edmondson, D. L., Kane, H. J., and Andrews, T. J. 1990. Substrate isomerization inhibits Rubisco during catalysis. *FEBS Letters* 260:62-66.

Eldredge, N. and Gould, S. J. 1972. Punctuated equilibria: an alternative to phyletic gradualism. *In* Schopf, T. J. M. and Thomas, J. M. (Eds.) *Models in Paleobiology* Freeman, Cooper, San Francisco. pp. 82-115.

Ellis, R. J. and van der Vies, S.M. 1988. The Rubisco subunit binding protein. *Photosynthesis Research* 16:101-115.

Fritsch, F. E. 1935. *The Structure and Reproduction of the Algae. Vol. 1.* Cambridge University Press, Cambridge. 791 pp.

Gabellini, N. 1988. Organization and structure of the genes for the cytochrome b/c_1 complex in purple photosynthetic bacteria. A phylogenetic study describing the homology of the b/c_1 subunits between prokaryotes, mitochondria, and chloroplasts. *Journal of Bioenergetics and Biomembranes* 20:59-83.

Gabrielson, P. W., Garbary, D. J. and Scagel, R. F. 1985. The nature of the ancestral red alga: inferences from a cladistic analysis. *BioSystems* 18:335-346.

Gabrielson, P. W., Garbary, D. J., Sommerfeld, M. R., Townsend, R. A., and Tyler, P. L. 1990. Phylum Rhodophyta. *In* Margulis, L., Corliss, J. O., Melkonian, M., and Chapman, D. J. (Eds.) *Handbook of Protoctista* pp. 102-118.

Gaines, G. and Elbrächter, M. 1987. Heterotrophic nutrition. *In* Taylor, F. J. R. *The Biology of Dinoflagellates. Botanical Monographs Vol. 1.* Blackwell Scientific, Oxford. pp. 224-268.

Gatenby, A. A. 1991. Protein folding and chaperonins. *Plant Molecular Biology* 19:677-687.

Gayral, P. and Billard, C. 1986. A survey of the marine Chrysophyceae with special reference to the Sarcinochrysidales. *In* Kristiansen, J. and Andersen, R. A. (Eds.) *Chrysophytes: Aspects and Problems.* Cambridge University Press, New York. pp. 37-48.

Gest, H. and Schopf, J. W. 1983. Biochemical evolution of anaerobic energy conversion: the transition from fermentation to anoxygenic photosynthesis. *In* Schopf, J. W. (Ed.) *Earth's Earliest Biosphere: its Origin and Evolution.* Princeton University Press, Princeton NJ. pp. 135-148.

Ghanotakis, D. F. and Yocum, C. F. 1990. Photosystem II and the oxygen-evolving complex. *Annual Review of Plant Physiology and Molecular Biology* 41:255-276.

Gibbs, S. P. 1978. The chloroplasts of *Euglena* may have evolved from symbiotic green algae. *Canadian Journal of Botany* 56:2883-2889.

Gibson, J. L. and Tabita, F. R. 1977. Different molecular forms of D-ribulose 1,5-bisphosphate carboxylase from *Rhodomonas sphaeroides*. *Journal of Biological Chemistry* 252:943-949.

Gibson, J. L., Falcone, D. L., and Tabita, F. R. 1991. Nucleotide sequence, transcriptional analysis, and expression of genes encoded within the form I CO_2 fixation operon of *Rhodobacter sphaeroides*. *Journal of Biological Chemistry* 266:14646-14653.

Gillott, M. A. and Gibbs, S. P. 1980. The cryptomonad nucleomorph: its ultrastructure and evolutionary significance. *Journal of Phycology* 16:558-568.

Gingrich, J. C. and Hallick, R. B. 1985. The *Euglena gracilis* chloroplast ribulose-1,5-bisphosphate carboxylase gene. II. The spliced mRNA and its product. *Journal of Biological Chemistry* 260:16162-8.

Glazer, A. N., Cohen-Bazire, G., and Stanier, R. Y. 1971. Comparative immunology of the algal biliproteins. *Proceedings of the National Academy of Science, U. S. A.* 68:3005-3008.

Glazer, A. N. and Melis, A. 1987. Photochemical reaction centers: structure, organization, and function. *Annual Review of Plant Physiology* 38:11-45.

Goodman, D. K. 1987. Dinoflagellate cysts in ancient and modern sediments. *In* Taylor, F. J. R. (Ed.) 1987. *The Biology of the Dinoflagellates* Blackwell, Oxford, pp. 649-722.

Graham, L. E. 1985. The origin of the life cycle of land plants. *American Scientist* 73:178-186.

Gray, M. W. 1989. The evolutionary origins of organelles. *Trends in Genetics* 5:294-299.

Gray, M. W., Cedergren, R., Abel, Y., and Sankoff, D. 1989. On the evolutionary origin of the plant mitochondrion and its genome. *Proceedings of the National Academy of Sciences. U. S. A.* 86:2267-2271.

Gray, J. C. and Kekwick, R. G. O. 1974. An immunological investigation of the structure and function of ribulose 1,5-bisphosphate carboxylase. *European Journal of Biochemistry* 44:481-489.

Griffith, O. M. 1986. *Techniques of Preparative, Zonal, and Continuous Flow Ultracentrifugation.* Beckman, Palo Alto, CA. 50 pp.

Guillard, R. R. L. and Ryther, J. H. 1962. Studies of marine planktonic diatoms. I. *Cyclotella nana* Hustedt and *Detonula confervacea* (Cleve) Gran. *Canadian Journal of Microbiology* 8:229-239.

Hardison, L. K. and Cattolico, R. A. 1989. Characterization and regulation of Rubisco genes in *Olisthodiscus luteus. In Proceedings of the Fourth Northwest Algal Symposium 1989.* Walla Walla College, Wash.

Hardison, L. K., Boczar, B. A., Reynolds, A. E., and Cattolico, R. A. 1992. A description of the Rubisco large subunit gene and its transcript in *Olisthodiscus luteus. Plant Molecular Biology* 18:595-599.

Harrison, P. J., Waters, R. E., and Taylor, F. J. R. 1980. A broad spectrum artificial seawater medium for coastal and open ocean phytoplankton. *Journal of Phycology* 16:28-35.

Hartman, F. C., Stringer, C. D., Omnass, J., Donnelley, M. I. and Fraij, B. 1982. Purification and sequencing of cyanogen bromide fragments from ribulosebisphosphate carboxylase/oxygenase from *Rhodospirillum rubrum. Archives of Biochemistry and Biophysics* 219:422-437.

Hobson, L. A. 1974. Effects of interactions of irradiance, daylength, and temperature on division rates of three species of marine unicellular algae. *Journal of the Fisheries Research Board of Canada* 31:391-395.

Hobson, L. A., Morris, W. J., and Guest, K. P. 1985. Varying photoperiod, ribulose 1,5-bisphosphate carboxylase/oxygenase and CO_2 uptake in *Thalassiosira fluviatilis* (Bacillariophyceae). *Plant Physiology* 79, 833-837.

Hobson, L. A. *in press.* Carbon concentrating mechanisms in photosynthetic cyanobacteria and protista - a review. *Current Topics in Plant Physiology.*

Hofmann, H. J. and Schopf, J. W. 1983. Early Proterozoic microfossils. *In* Schopf, J. W. (Ed.) *Earth's Earliest Biosphere: its Origin and Evolution.* Princeton University Press, Princeton NJ. pp.321-360.

Hori, H. and Osawa, S. 1987. Origin and evolution of organisms as deduced from 5S ribosomal RNA sequences. *Molecular Biology and Evolution* 4:445-472.

Houtz, R. L., Stults, J. T., Mulligan, R. M., and Tolbert, N. E. 1989. Post-translational modifications in the large subunit of ribulose bisphosphate carboxylase/oxygenase. *Proceedings of the National Academy of Science, USA* 86:1855-1859.

Houtz, R. L., Poneleit, L., Jones, S. B., Royer, M. and Stults, J. T. 1992. Posttranslational modifications in the amino-terminal region of the large subunit of Ribulose-1,5-bisphosphate carboxylase/oxygenase from several plant species. *Plant Physiology* 98:1170-1174.

Hunt, L. T., George, D. G., and Barker, W. C. 1985. The prokaryote-eukaryote interface. *BioSystems* 18:223-240.

Hurt, E. C. and Hauska, G. 1984. Purification of membrane-bound cytochromes and a photoactive P840 complex of the green sulfur bacterium *Chlorobium limicola* f. *thiosulfatophilum. FEBS Letters* 168:149-154.

Hwang, S.-R. and Tabita, F. R. 1989. Cloning and expression of the chloroplast-encoded *rbc* L and *rbc* S genes from the marine diatom *Cylindrotheca* sp. strain N1. *Plant Molecular Biology* 13:69-79.

Jones, C. W. and Poole, R. K. 1985. The analysis of cytochromes. *In* Gottschalk, G. (Ed.) *Methods in Microbiology* 18:285-328.

Jordan, D. B. and Ogren, W. L. 1981. Species variation in the specificity of ribulose bisphosphate carboxylase/oxygenase. *Nature* 291:513-515.

Jordan, D. B. and Chollet, R. 1983. Inhibition of ribulose-1,5-bisphosphate carboxylase by substrate ribulose-1,5-bisphosphate. *Journal of Biological Chemistry* 258:13752-13758.

Kabnick, K. S. and Peattie, D. A. 1991. *Giardia:* a missing link between prokaryotes and eukaryotes. *American Scientist* 79:34:43.

Kawashima, N. 1969. Comparative studies on fraction 1 protein from spinach and tobacco leaves. *Plant and Cell Physiology* 10:31-40.

Kent, S. S. and Tomany, M. J. 1984. Kinetic variance of ribulose-1,5-bisphosphate carboxylase/oxygenase isolated from diverse taxonomic sources. II. Analysis by two dual label methods. *Plant Physiology* 75:645-650.

Kies, L. 1979. Zur systematischen Einordnung von *Cyanophora paradoxa, Gloeochaete wittrockiana* und *Glaucocystis nostochinearum. Berichte der Deutschen Botanischen Geselleschaft.* 92:445-454.

Klein, K. M., and Cronquist, A. 1967. A consideration of the evolutionary and taxonomic significance of some biochemical, morphological and physiological characters in the Thallophytes. *Quarterly Review of Biology* 42:105-296.

Kondratieva, E. N., Pfennig, N., and Trüper, H. G. 1991. The phototrophic prokaryotes. *In* Balows, A., Trüper, H. G., Dworkin, M., Harder, W., and Schleifer, K.-H. (Eds.) *The Prokaryotes 2nd ed.* Springer-Verlag, New York. pp. 312-330.

Kostrzewa, M., Valentin, K., Maid, U., Radetzky, R., and Zetsche, K. 1990. Structure of the rubisco operon from the multicellular red alga *Antithamnion spec. Current Genetics* 18:465-469.

Kuehn, G, D. and McFadden, B. A. 1969. Ribulose 1,5-diphosphate carboxylase from *Hydrogenomonas eutropha* and *Hydrogenomonas facilis.* II. Molecular weight, subunits, composition, and sulfhydryl groups. *Biochemistry* 8:2403-2408.

Larsen, J. 1988. An ultrastructural study of *Amphidinium poecilochroum* (Dinophyceae), a phagotrophic dinoflagellate feeding on small species of cryptophytes. *Phycologia* 27:366-377.

Lee, G. J., Kostov, R. V., and McFadden, B. A. 1993. A facile method to determine the CO_2 / O_2 specificity factor for ribulose bisphosphate carboxylase/oxygenase. *Photosynthesis Research* 37:81-86.

Lee, R. E. 1980. *Phycology* Cambridge University Press, Cambridge.

Li, W.-H. and Graur, D. 1991. *Fundamentals of Molecular Evolution.* Sinauer, Sunderland Mass. 284 pp.

Lockhart, P. J., Howe, C. J., Bryant, D. A., Beanland, T. J., and Larkum, A. W. D. 1992. Substitutional bias confounds inference of cyanelle origins from sequence data. *Journal of Molecular Evolution* 34:153-162.

Loeblich, A. R. III 1976. Dinoflagellate evolution: speculation and evidence. *Journal of Protozoology* 23:13-28.

Loeblich, A. R. III 1984. Dinoflagellate evolution. *In* Spector, D. L. (Ed.) *Dinoflagellates* Academic Press, London, pp. 481-522.

Marchalonis, J. J. and Weltman, J. K. 1971. Relatedness among proteins: a new method of estimation and its application to immunoglobulins. *Comparative Biochemistry and Physiology.* 38:609-625.

Markowicz, Y. and Loiseaux-de Goër,S. 1991. Plastid genomes of the Rhodophyta and Chromophyta constitute a distinct lineage which differs from that of the Chlorophyta and have a composite phylogenetic origin, perhaps like that of the Euglenophyta. *Current Genetics* 20:427-430.

Martin, W., Somerville, C. C., and Loiseaux-de Goër, S. 1992. Molecular phylogenies of plastid origins and algal evolution. *Journal of Molecular Evolution* 35:385-404.

Mayr, E. 1965. Numerical phenetics and taxonomic theory. *Systematic Zoology.* 14:73-97.

McFadden, B. A., Torres-Ruiz, J., Daniell, H., and Sarjoni, G. 1986. Interaction, functional relations and evolution of large and small subunits in Rubisco from Prokaryota and Eukaryota. *Philosophical Transactions of the Royal Society of London* Series B 313:247-358.

McFadden, B. A. and Small, C. L. 1988. Cloning, expression and directed mutagenesis of the genes for ribulose bisphosphate carboxylase/oxygenase. *Photosynthesis Research* 18:245-260.

McIntosh, L., Poulsen, C. and Bogorad, L. 1980. Chloroplast gene sequence for the large subunit of ribulose bisphosphate oxygenase of maize. *Nature* (Lond.) 288:556-560.

McNeill, J. 1979. Purposeful phenetics. *Systematic Zoology* 28:465-482.

Meijer, W. G., Arnberg, A. C., Enequist, H. G., Terpstra, P., Lidstrom, M. E., and Dijkhuizen, L. 1991. Identification and organization of carbon dioxide fixation genes in *Xanthobacter flavus* H4-14. *Molecular and General Genetics* 225:320-330.

Michel, H. and Deisenhofer, J. 1988. Relevance of the photosynthetic reaction center from purple bacteria to the structure of photosystem II. *Biochemistry* 27:1-7.

Miziorko, H. M. and Lorimer, G. H. 1983. Ribulose-1,5-bisphosphate carboxylase-oxygenase. *Annual Review of Biochemistry* 52:507-535.

Moon, E., Kao, T.-H., and Wu, R. 1987. Rice chloroplast DNA molecules are heterogeneous as revealed by DNA sequences of a cluster of genes. *Nucleic Acids Research* 15:611-630.

Moon, K. E. and Thompson, E. O. P. 1969. Subunits from reduced and S-carboxymethylated ribulose diphosphate carboxylase (fraction I protein). *Australian Journal of Biological Sciences* 22:463-470.

Morden, C. W. and Golden, S. S. 1991. Sequence analysis and phylogenetic reconstruction of the genes encoding the large and small subunits of ribulose-1,5-bisphosphate carboxylase/oxygenase from the chlorophyll *b*- containing prokaryote *Prochlorothrix hollandica. Journal of Molecular Evolution* 32:379-395.

Morden, C. W., Delwiche, C. F., Kuhsel, M., and Palmer, J. D. 1992. Gene phylogenies and the endosymbiotic origin of plastids. *BioSystems* 28:75-90.

Nagy, F., Fluhr, R., Morelli, G., Kuhlemeier, C., Poulsen, C., Keith, B., Boutry, M., and Chua, N.-H. 1986. The Rubisco small subunit gene as a paradigm for studies on differential gene expression during plant development. *Philosophical Transactions of the Royal Society of London.* Series B 313:409-417.

Nargang, F., McIntosh, L., and Sommerville, C. 1984. Nucleotide sequence of the ribulosebisphosphate carboxylase gene from *Rhodospirillum rubrum. Molecular and General Genetics* 193:220-224.

Newman, S. M., Derocher, J., and Cattolico, R. A. 1989. Analysis of chromophytic and rhodophytic ribulose-1,5-bisphosphate carboxylase indicates extensive structural and functional similarities among evolutionarily diverse algae. *Plant Physiology* 91:939-946.

Newman, S. M. and Cattolico, R. A. 1987. Structural, functional, and evolutionary analysis of ribulose-1,5-bisphosphate carboxylase from the chromophytic alga *Olisthodiscus luteus. Plant Physiology* 84:483-490.

Nierzwicki-Bauer, S. A., Curtis, S. E., and Haselkorn, R. 1984. Cotranscription of genes encoding the small and large subunits of ribulose-1,5-bisphosphate carboxylase in the cyanobacterium *Anabaena* 7120. *Proceedings of the National Academy of Sciences, USA* 81:5961-5965.

Ogren, W. L., Salvucci, M. E., and Portis, A. R. Jr. 1986. The regulation of Rubisco activity. *Philosophical Transactions of the Royal Society of London.* Series B 313:337-346.

Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H. and Ozeki, H. 1986. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature* (Lond.) 322:572-4.

Olson, J. M. and Pierson, B. K. 1987. Evolution of reaction centers in photosynthetic prokaryotes. *International Review of Cytology* 108:209-248.

Palenik, B. and Haselkorn, R. 1992. Multiple evolutionary origins of prochlorophytes, the chlorophyll *b* -containing prokaryotes. *Nature* 355:265-267.

Palmer, J., Jansen, R. K., Michaels, H. J., Chase, M. W., and Manhart, J. R. 1988. Chloroplast DNA variation and plant phylogeny. *Annals of the Missouri Botanical Garden* 75:1180-1206.

Parry, M. A. J., Keys, A. J., Foyer, C. H., Furbank, R. T., and Walker, D. A. 1988. Regulation of ribulose-1,5-bisphosphate carboxylase activity by the activase system in lysed spinach chloroplasts. *Plant Physiology* 87:558-561.

Pickett-Heaps, J. D. 1975. *Green Algae: Structure, Reproduction and Evolution in Selected Genera.* Sinauer. Sunderland MA. 606 pp.

Plumley, F. G., Kirchman, D. L., Hodson, R. E., and Schmidt, G. W. 1986. Ribulose bisphosphate carboxylase from three chlorophyll *c*-containing algae: physical and immunological characterizations. *Plant Physiology* 80:685-691.

Portis, A. R. Jr. 1990. Rubisco activase. *Biochimica et Biophysica Acta* 1015:12-28.

Raven, J. A. 1991. Implications of inorganic carbon utilization: ecology, evolution, and geochemistry. *Canadian Journal of Botany* 69:908-924.

Raven, J. A., Johnston, A. M., and bin Surif, M. 1989. The photosynthetic apparatus as phyletic character. *In* Green, J. C., Leadbeater, B. S. C., and Diver, W. L. (Eds.). *The Chromophyte Algae: Problems and Perspectives.* Clarendon, Oxford. pp. 63-84.

Read, B. A. and Tabita, F. R. 1992. A hybrid ribulosebisphosphate carboxylase/oxygenase enzyme exhibiting a substantial increase in specificity factor. *Biochemistry* 31:5553-5560.

Reith, M. E. and Cattolico, R. A. 1985. In vivo chloroplast protein synthesis by the chromophytic alga *Olisthodiscus luteus. Biochemistry* 24:2556-2561.

Robinson, S. P. and Portis A. R. Jr. 1988. Release of the nocturnal inhibitor of ribulose-1,5-bisphosphate carboxylase/oxygenase by Rubisco activase. *FEBS Letters* 233:413-416.

Rohlf, F. J. 1988. *NTSYS-pc Numerical Taxonomy and Multivariate Analysis System. version 1.50* Exeter Publishing Ltd. Setauket NY 186 pp.

Rothschild, L. J. and Heywood, P. 1987. Protistan phylogeny and chloroplast evolution: conflicts and congruence. *Progress in Protistology* 2:1-68.

Round 1977. Cited but not referenced in Bold and Wynne 1985.

Round, F. E., Crawford, R. M., and Mann, D. G. 1990. *The Diatoms: Biology and Morphology of the Genera.* Cambridge University Press. 747 pp.

Roy, H., Chaudhari, P., and Cannon, S. 1988. Incorporation of large subunits into ribulose bisphosphate carboxylase in chloroplast extracts. *Plant Physiology* 86:0044-0049.

Rutner, A. C. and Lane, M. D. 1967. Nonidentical subunits of ribulose diphosphate carboxylase. *Biochemical and Biophysical Research Communications* 28:531-537.

Salvucci, M. E., Portis, A. R. Jr., and Ogren, W. L. 1985. A soluble chloroplast protein catalyzes ribulose bisphosphate carboxylase / oxygenase activation in vivo. *Photosynthesis Research* 7:193-201.

Scheller, H. V. and Møller, B. L. 1990. Photosystem I polypeptides. *Physiologia Plantarum* 78:484-494.

Schidlowski, M., Hayes, J. M., and Kaplan, I. R. 1983. Prebiotic organic syntheses and the origin of life. *In* Schopf, J. W. (Ed.) *Earth's Earliest Biosphere: its Origin and Evolution.* Princeton University Press, Princeton, NJ. pp. 149-186.

Schloss, J. V., Phares, E. F., Long, M. V., Norton, I. L., Stringer, C. D., and Hartman, F. C. 1979. Isolation, characterization, and crystallization of ribulosebisphosphate carboxylase from autotrophically grown *Rhodospirillum rubrum. Journal of Bacteriology* 137:490-501.

Schopf, J. W. 1993. Microfossils of the early Archean Apex chert: new evidence of the antiquity of life. *Science* 260:640-646.

Schwartz, R. M. and Dayhoff, M. O. 1978. Origins of prokaryotes, eukaryotes, mitochondria, and chloroplasts. *Science* 199:395-403.

Seemann, J. R., Kobza, J., and Moore, B. d. 1990. Metabolism of 2-carboxyarabinitol 1-phosphate and regulation of ribulose-1,5-bisphosphate carboxylase activity. *Photosynthesis Research* 23:119-130.

Shinozaki, K. and Sugiura, M. 1985. Genes for the large and small subunits of ribulose-1,5-bisphosphate carboxylase/oxygenase constitute a single operon in a cyanobacterium *Anacystis nidulans* 6301. *Molecular and General Genetics* 200:27-32.

Shinozaki, K. and Sugiura, M. 1982. The nucleotide sequence of the tobacco chloroplast gene of the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase. *Gene* 20:91-102.

Shinozaki, K., Yamada, C., Takahata, N. and Sugiura, M. 1983. Molecular cloning and sequence analysis of the cyanobacterial gene for the large subunit of ribulose-1,-bisphosphate carboxylase/oxygenase. *Proceedings of the National Academy of Science, U.S.A.* 80:4050-4054.

Smayda, T. J. 1973. The growth of *Skeletonema costatum* during a winter-spring bloom in Naragansett Bay, Rhode Island. *Norwegian Journal of Botany* 20:219-247.

Smith, E. L., Landon, M., Piszkiewicz, D., Brattin, W. J., Langley, T. J., and Melamed, M. D. 1970. Bovine liver glutamate dehydrogenase: tentative amino acid sequence; identification of a reactive lysine; nitration of a specific tyrosine and loss of allosteric inhibition by guanosine triphosphate. *Proceedings of the National Academy of Sciences, U. S. A.* 67:742-730.

Smith, P. J. and Hobson, L. A. *in press.* Temporal variations in the taxonomic composition of flagellated nanoplankton in a temperate fjord. *Journal of Phycology* 30:.

Sneath, P. H. A. and Sokal, R. R.. 1973. *Numerical Taxonomy.* Freeman, San Francisco, 573 pp.

Sogin, M. L., Gunderson, J. H., Elwood, H. J., Rogelio, A. A., and Peattie, D. A. 1989. Phylogenetic meaning of the kingdom concept: and unusual ribosomal RNA from *Giardia lamblia. Science* 243:75-77.

Stewart, K. D. and Mattox, K. R. 1978. Structural evolution in the flagellated cells of green algae and land plants. *BioSystems* 10:145-152.

Stone, K. L. and Williams, K. R. 1986. High-performance liquid chromatographic peptide mapping and amino acid analysis in the sub-nanomole range. *Journal of Chromatography* 359:203-212.

Streusand, V. J. and Portis, A. R. Jr. 1987. Rubisco activase mediates ATP-dependent activation of ribulose bisphosphate carboxylase. *Plant Physiology* 85:152-154.

Sugiyama, T. and Akazawa, T. 1970. Subunit structure of spinach leaf ribulose-1,5-diphosphate carboxylase. *Biochemistry* 9:4499-4504.

Sugiyama, T., Ito, T., and Akazawa, T. 1971. Subunit structure of ribulose 1,5-diphosphate carboxylase from *Chlorella ellipsoidea. Biochemistry* 10:3406-3411.

Tabita, F. R. and McFadden, B. A. 1974. D-Ribulose 1,5-bisphosphate carboxylase from *Rhodospirillum rubrum. Journal of Biological Chemistry* 249:3453-3464.

Takabe, T. And Akazawa, T. 1975. The role of sulfhydryl groups in the ribulose-1,5-bisphosphate carboxylase and oxygenase reactions. *Archives of Biochemistry and Biophysics* 169:686-694.

Taylor, F. J. R. 1980. On dinoflagellate evolution. *BioSystems* 13:65-108.

Taylor, F. J. R. 1987. Appendix: taxonomy and classification. *In* Taylor, F. J. R. (Ed.) *The Biology of the Dinoflagellates* Blackwell, Oxford, pp. 723-731.

Taylor, F. J. R. 1990. Phylum Dinoflagellata. *In* Margulis, L., Corliss, J. O., Melkonian, M., and Chapman D. J. (Eds.) *Handbook of Protoctista* Jones and Bartlett, Boston. pp. 419-437.

Toman, D. P., Lynch, J. J. Jr., and Schmidt, R. R. 1985. Purification and partial characterization of ribulose bisphosphate carboxylase holoenzyme and its subunits from *Chlorella sorokiniana* and use of its antigen affinity-purified antibodies in specific immunoprecipitation and immunoadsorption procedures. *Plant Physiology* 79:806-814.

Tomas, R. N. and Cox, E. R. 1973. Observations on the symbiosis of *Peridinium balticum* and its intracellular alga. I. Ultrastructure. *Journal of Phycology* 9:304-323.

Towbin, H., Staehelin, T., and Gordon, J. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. *Proceedings of the National Academy of Sciences U. S. A.* 76:4350-5354.

Trench, R. K. 1982. Physiology, biochemistry, and ultrastructure of cyanellae. *In* Round, F. E. and Chapman D. J. (Eds.) *Progress in Phycological Research, Vol 1.* Elsevier, Amsterdam. pp. 257-288.

Trüper, H. G. and Pfennig, N. 1991. The Family Chlorobiaceae. *In* Balows, A., Trüper, H. G., Dworkin, M., Harder, W., and Schleifer, K.-H. (Eds.) *The Prokaryotes* (2nd Ed.) Springer-Verlag, New York. pp. 3583-3592.

Turner, S., Burger-Wiersma, T., Giovannoni, S. J., Mur, L. R., and Pace, N. R. 1989. The relationship of a prochlorophyte *Prochlorothrix hollandica* to green chloroplasts. *Nature* 337:380-382.

Urbach, E., Robertson, D. L., and Chisholm, S. W. 1992. Multiple evolutionary origins of prochlorophytes within the cyanobacterial radiation. *Nature* 355:267-270.

Valentin, K. and Zetsche, K. 1989. The genes of both subunits of ribulose-1,5-bisphosphate carboxylase constitute an operon on the plastome of a red alga. *Current Genetics* 16:203-209.

Valentin, K. and Zetsche, K. 1990a. Structure of the Rubisco operon from the unicellular red alga *Cyanidium caldarium* : evidence for a polyphyletic origin of the plastids. *Molecular and General Genetics* 222:425-430.

Valentin, K. and Zetsche, K. 1990b. Nucleotide sequence of the gene for the large subunit of Rubisco from *Cyanophora paradoxa* - phylogenetic implications. *Current Genetics* 18:199-202.

van den Hoek. C. and Jahns, H. M. 1978. *Algen. Einführung indie Phykologie*. G. Thieme, Stuttgart.

van den Hoek, C. 1981. Chlorophyta: morphology and classification. *In* Lobban, C. S. and Wynne, M. J. (Eds.) *The Biology of Seaweeds.* Blackwell, Berkeley and Los Angeles, pp. 86-132.

Viale, A. M., Kobayashi, H., and Akazawa, T. 1989. Expressed gene for the plant-type ribulose 1,5-bisphosphate carboxylase/oxygenase in the photosynthetic bacterium *Chromatium vinosum,* which possesses two complete sets of the genes. *Journal of Bacteriology* 171:2391-2400.

Wagner, S. J., Stevens, S. E. Jr., Nixon, B. T., Lambert, D. H., Quivey, R. G. Jr., and Tabita, F. R. 1988. Nucleotide sequence and deduced amino acid sequence of the *Rhodobacter sphaeroides* gene encoding form II ribulose-1,5-bisphosphate carboxylase/oxygenase and comparison with other deduced form I and II sequences. *FEMS Microbiology Letters* 55:217-222.

Walker, J. C. G., Klein, C., Schidlowski, M., Schopf, J. W., Stevenson, D. J., and Walter, M. R. 1983. Environmental evolution of the Archaean-early Proterozoic earth. *In* Schopf, J. W. (Ed.) *Earth's Earliest Biosphere* Princeton. pp. 260-290.

Wang, Z. Y. and Portis, A. R. Jr. 1992. Dissociation of ribulose-1,5-bisphosphate bound to ribulose-1,5-bisphosphate carboxylase/oxygenase and its enhancement by ribulose-1,5-bisphosphate carboxylase/oxygenase activase-mediated hydrolysis of ATP. *Plant Physiology* 99:1348-1353.

Watanabe, M. M., Suda, S., Inouye, I., Sawaguchi, T., and Chihara, M. 1990. *Lepidodinium viride* gen. et sp. nov. (Gymnodiniales, Dinophyta), a green dinoflagellate with a chlorophyll *a-* and *b-* containing endosymbiont. *Journal of Phycology* 26:741-751.

Werneke, J. ., Chatfield, J. .,and Ogren, W. . 1988. Catalysis of ribulosebisphosphate carboxylase / oxygenase activation in the product of a Rubisco activase cDNA clone expressed in *Escherichia coli*. *Plant Physiology* 87:917-920.

Whatley, J. M. and Whatley, F. R. 1981. Chloroplast evolution. *New Phytologist* 87:233-247.

Whatley, J. M. 1989. Chromophyte chloroplasts - a polyphyletic origin?. *In* Green, J. C., Leadbeater, B. S. C., and Diver, W. L. (Eds.) *The Chromophyte Algae: Problems and Perspectives.* Systematics Association Special Volume No. 38, Clarendon, Oxford. pp. 125-144

Woese, C. R. 1987. Bacterial evolution. *Microbiological Reviews* 51:221-271.

Wolfe, K. H., Li, W.-H., and Sharp, P. M. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNA. *Proceedings of the National Academy of Sciences U. S. A.* 84:9054-9058.

Yoshinaga, K., Ohta, Y., Suzuki, Y. and Sugiura, M. 1988. *Chlorella* chloroplast DNA sequence containing a gene for the large subunit of ribulose1-,5-bisphosphate/carboxylase/oxygenase and a part of a possible gene for the ß' subunit of RNA polymerase. *Plant Molecular Biology* 10:245-250.

Zurawski, G., Perrot, B., Bottomley, W. and Whitfield, P. R. 1981. The structure of the gene for the large subunit of ribulose 1,5-bisphosphate carboxylase from spinach chloroplast DNA. *Nucleic Acids Research* 9:3251-70.

Zurawski, G., Whitfield, P. R., and Bottomley, W. 1986. Sequence of the gene for the large subunit of ribulose 1,5-bisphosphate carboxylase from pea chloroplasts. *Nucleic Acids Research* 14:3975-3975.

Appendix 1. Amino acid compositions (this work). Cysteine is shown but not included in calculations.

	A	B	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Z
Amphi1.A*	10.86	10.53		-	-	4.99	12.70	1.52	5.54	5.32	7.82	0.76	-	4.13	-	3.91	5.86	4.99	6.51	-	3.91	10.64
Amphi2.A*	10.53	11.05		-	-	4.74	13.05	1.58	5.26	5.05	7.47	0.63	-	4.21	-	4.00	5.89	5.26	6.32	-	3.79	11.16
Amphi3.A*	10.67	10.57		-	-	5.12	13.66	1.60	5.23	5.66	7.47	0.53	-	4.06	-	3.84	5.76	4.70	6.19	-	4.06	10.89
AmphiM.A*	10.69	10.72		-	-	4.95	13.14	1.57	5.34	5.34	7.59	0.64	-	4.13	-	3.92	5.84	4.98	6.34	-	3.92	10.90
Chaet1.A*	11.61	11.03	0.12	-	-	4.41	9.99	1.74	4.88	4.07	9.52	2.21	-	3.60	-	5.69	4.76	5.69	7.32	-	5.11	8.36
Chaet2.A*	12.27	10.55		-	-	4.42	10.06	1.72	4.91	3.93	9.94	2.09	-	3.80	-	5.28	4.42	5.77	7.48	-	5.28	8.10
ChaetM.A*	11.94	10.79		-	-	4.42	10.03	1.73	4.90	4.00	9.73	2.15	-	3.70	-	5.49	4.59	5.73	7.40	-	5.20	8.23
Chroo1.A*	10.42	11.35	0.10	-	-	4.06	11.25	1.77	5.10	4.79	8.33	2.92	-	3.54	-	6.25	4.17	5.94	5.94	-	4.48	9.69
Chroo2.A*	9.98	11.08	0.10	-	-	4.19	11.28	1.90	5.29	4.99	8.38	3.79	-	3.49	-	5.99	4.09	5.99	5.79	-	4.49	9.28
Chroo3.A*	10.36	10.98	0.10	-	-	4.15	11.50	1.76	5.08	4.87	8.29	4.25	-	3.52	-	5.80	4.15	5.91	5.80	-	4.35	9.22
Chroo4.A*	9.89	11.37	0.10	-	-	4.06	11.57	1.88	5.24	4.35	7.91	4.06	-	3.56	-	6.23	4.35	5.93	5.74	-	4.45	9.40
Chroo5.A*	10.82	11.47	0.11	-	-	3.57	12.34	1.84	4.76	3.57	7.58	4.22	-	3.79	-	6.06	4.44	6.06	5.74	-	4.11	9.63
ChrooM.A*	10.29	11.25		-	-	4.01	11.59	1.83	5.09	4.51	8.10	3.85	-	3.58	-	6.07	4.24	5.97	5.80	-	4.38	9.44
Iso1.A*	9.91	12.78	0.10	-	-	4.26	10.11	1.78	4.56	2.87	9.71	2.78	-	4.06	-	6.05	5.15	6.34	6.24	-	4.56	8.82
Iso2.A*	9.98	12.77	0.10	-	-	4.19	10.08	1.70	4.39	3.69	9.58	2.69	-	4.09	-	5.99	4.99	6.39	6.09	-	4.59	8.78
Iso3.A*	10.22	12.58	0.10	-	-	4.19	9.92	1.64	4.40	4.29	9.82	2.66	-	3.99	-	5.93	4.81	6.13	6.13	-	4.60	8.69
Iso4.A*	10.55	12.13	0.11	-	-	4.22	10.44	1.05	4.85	4.54	9.92	0.53	-	4.22	-	5.91	4.64	6.54	7.49	-	4.54	8.44
Iso5.A*	10.47	12.15	0.10	-	-	4.19	10.26	1.05	4.82	4.61	9.84	0.63	-	4.19	-	6.18	4.61	6.60	7.43	-	4.61	8.38
IsoM.A*	10.23	12.48		-	-	4.21	10.16	1.44	4.60	4.00	9.77	1.86	-	4.11	-	6.01	4.84	6.40	6.68	-	4.58	8.62
Skel1.A*	10.60	13.15		-	-	3.71	13.26	0.00	4.77	3.92	8.27	0.00	-	3.08	-	6.68	5.51	5.30	7.00	-	2.97	11.77
Skel2.A*	12.18	12.06		-	-	3.65	11.57	1.83	4.87	2.92	8.65	1.22	-	3.29	-	6.33	4.99	5.48	7.55	-	3.53	9.87
Skel3.A*	12.45	10.46		-	-	3.99	13.20	1.74	5.23	1.87	9.71	2.24	-	3.99	-	5.85	4.48	4.86	7.60	-	3.74	8.59
Skel4.A*	12.39	10.90		-	-	3.59	14.37	1.73	4.96	1.86	9.05	1.98	-	3.47	-	6.32	4.58	4.83	7.06	-	3.84	9.05
SkelM.A*	11.91	11.64		-	-	3.74	13.10	1.33	4.96	2.64	8.92	1.36	-	3.46	-	6.30	4.89	5.12	7.30	-	3.52	9.82
Spin1.A*	9.43	10.18	0.09	-	-	4.52	10.93	2.64	4.05	4.71	9.33	0.47	-	5.00	-	6.41	3.86	6.50	8.29	-	3.96	9.71
Spin2.A*	9.49	9.96	0.09	-	-	4.55	10.82	2.66	3.98	5.31	9.39	0.28	-	4.93	-	6.17	3.80	6.45	8.35	-	3.98	9.87
Spin3.A*	9.68	9.87	0.10	-	-	4.36	10.94	2.81	3.97	5.13	9.39	0.29	-	4.94	-	6.39	3.78	6.29	8.42	-	3.97	9.78
SpinM.A*	9.53	10.00		-	-	4.48	10.90	2.70	4.00	5.05	9.37	0.35	-	4.96	-	6.32	3.81	6.41	8.35	-	3.97	9.79
Thal1.A*	11.57	10.76	0.46	-	-	4.28	10.53	2.08	5.67	4.28	9.95	0.93	-	3.70	-	6.13	4.05	5.32	6.94	-	4.98	8.80
Thal2.A*	11.67	12.25	0.23	-	-	4.32	11.20	1.52	5.72	4.43	10.2	0.35	-	3.73	-	5.72	4.32	3.03	6.88	-	4.78	9.92
ThalM.A*	11.62	11.51		-	-	4.30	10.87	1.80	5.70	4.36	10.1	0.64	-	3.72	-	5.93	4.19	4.18	6.91	-	4.88	9.36

Appendix 2. Literature amino acid compositions. For codes and references see Table 1. Cysteine is shown but not included in calculations.

	A	B	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Z
Acido.A	9.62	11.93	2.68	-	-	4.62	9.25	2.59	4.99	3.70	8.09	3.14	-	4.99	-	6.56	4.62	5.78	7.30	-	3.19	9.62
Alcal1.A	10.35	11.28	2.18	-	-	4.44	9.95	2.67	4.04	4.35	7.95	2.84	-	4.89	-	6.26	5.15	6.71	7.55	-	3.47	8.09
Alcal2.A	10.54	10.54	2.98	-	-	4.32	8.73	2.42	4.41	3.80	7.78	3.28	-	5.57	-	6.91	5.01	6.53	7.09	-	3.54	9.51
Beta1.A	9.93	9.10		-	-	4.66	10.39	2.90	4.08	5.37	8.91	1.55	-	4.95	-	5.87	4.13	6.55	7.32	-	3.54	10.74
Beta2.A	9.89	9.29	1.62	-	-	4.64	10.31	2.88	4.09	5.34	8.87	1.53	-	4.92	-	5.85	4.32	6.55	7.29	-	3.53	10.68
Chlore.A	11.18	9.75		-	-	4.11	10.57	2.59	3.45	4.15	9.17	1.85	-	5.59	-	7.15	4.03	6.46	5.10	-	4.03	10.81
Chlors.A	11.05	10.06	1.17	-	-	4.14	11.44	2.56	3.75	4.73	8.48	1.38	-	5.52	-	6.31	3.75	6.51	5.92	-	4.14	10.26
ChromD.A	9.63	11.62	1.40	-	-	4.52	10.45	3.08	4.57	5.52	7.24	2.76	-	5.11	-	5.88	4.12	6.51	6.33	-	3.75	8.91
Nico.A	9.85	10.02		-	-	4.42	10.77	2.74	3.36	4.72	9.54	1.59	-	4.99	-	6.36	3.80	6.40	6.58	-	3.75	11.13
Oli.A	11.27	10.35	0.00	-	-	4.00	10.87	2.12	5.27	4.08	8.55	3.60	-	4.19	-	6.15	5.23	5.47	6.07	-	4.16	8.63
Phaseo.A	9.67	9.00		-	-	4.88	10.05	3.51	4.84	5.30	9.15	1.66	-	4.46	-	6.54	3.93	6.31	6.69	-	3.49	10.52
Rrub1.A	13.58	9.84	0.97	-	-	3.94	10.83	2.17	4.33	4.13	7.68	2.56	-	5.91	-	6.10	5.31	5.51	6.30	-	2.36	9.45
Rub2.A	13.49	9.92	0.98	-	-	4.96	10.71	2.58	4.17	4.17	7.74	2.98	-	4.96	-	6.15	4.17	5.56	5.95	-	3.57	8.93
Spin4.A	9.48	9.57		-	-	4.35	10.13	3.35	4.00	4.91	9.39	1.70	-	4.87	-	6.43	3.04	7.52	7.26	-	4.09	9.91
Spin5.A	9.57	9.66		-	-	4.43	9.30	2.97	3.90	5.23	9.26	1.86	-	5.00	-	6.38	3.59	7.75	7.25	-	4.12	9.74
Spin6.A	9.67	10.28		-	-	4.37	10.63	2.89	3.02	4.55	9.58	1.53	-	5.12	-	6.21	3.81	7.13	6.65	-	3.98	10.59

Appendix 3. Amino acid compositions derived from gene and protein sequences. Residue number above. Percent below, for only those amino acids in common with compositions. For codes and references see Table 1.

	A	B	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Z	#
Alcal.G	49 10.36	51 10.78	5	33	21	21 4.44	45 9.51	11 2.33	22 4.65	21 4.44	35 7.40	17 3.59	18	22 4.65	15	34 7.19	20 4.23	32 6.77	38 8.03	9	19 4.02	36 7.61	487 473
Anab.G	48 10.43	45 9.78	7	28	31	19 4.13	42 9.13	13 2.83	25 5.43	25 5.43	41 8.91	8 1.74	17	20 4.35	15	30 6.52	15 3.26	34 7.39	30 6.52	9	19 4.13	46 10.00	476 460
Antith.G	46 9.66	50 10.50	4	31	27	18 3.78	43 9.03	9 1.89	35 7.35	24 5.04	39 8.19	18 3.78	19	19 3.99	14	26 5.46	22 4.62	31 6.51	34 7.14	8	21 4.41	41 8.61	488 476
Chlamy.G	45 9.89	44 9.67	12	30	30	19 4.18	49 10.77	13 2.86	21 4.62	23 5.05	38 8.35	13 2.86	14	21 4.62	10	31 6.81	15 3.30	29 6.37	35 7.69	8	19 4.18	40 8.79	475 455
Chromv.G	48 10.60	47 10.38	9	32	30	21 4.64	43 9.49	15 3.31	25 5.52	26 5.74	34 7.51	13 2.87	15	20 4.42	8	27 5.96	16 3.53	30 6.62	32 7.06	10	18 3.97	38 8.39	472 453
Chlore.G	47 10.26	43 9.39	9	29	31	19 4.15	47 10.26	13 2.84	22 4.80	22 4.80	41 8.95	10 2.18	14	23 5.02	13	31 6.77	17 3.71	32 6.99	28 6.11	8	19 4.15	44 9.61	475 458
Crypto.G	50 10.59	50 10.59	9	29	25	20 4.24	46 9.75	9 1.91	26 5.51	25 5.30	39 8.26	18 3.81	21	15 3.18	16	25 5.30	20 4.24	31 6.57	37 7.84	7	20 4.24	41 8.69	488 472
Cyanid.G	50 10.37	49 10.17	3	27	31	19 3.94	40 8.30	10 2.07	28 5.81	27 5.60	45 9.34	17 3.53	22	18 3.73	16	28 5.81	19 3.94	32 6.64	33 6.85	8	20 4.15	47 9.75	493 482
Cyanoph.G	45 9.80	43 9.37	8	27	31	21 4.58	46 10.02	14 3.05	26 5.66	21 4.58	35 7.63	13 2.83	16	23 5.01	12	33 7.19	21 4.58	33 7.19	26 5.66	8	16 3.49	43 9.37	475 459
Cylindro.G	51 10.74	48 10.11	7	27	25	18 3.79	43 9.05	10 2.11	26 5.47	21 4.42	42 8.84	16 3.37	21	17 3.58	14	28 5.89	27 5.68	32 6.74	35 7.37	8	22 4.63	39 8.21	490 475
Euglena.G	44 9.69	40 8.81	12	26	32	19 4.19	50 11.01	13 2.86	20 4.41	21 4.63	38 8.37	12 2.64	14	20 4.41	13	31 6.83	20 4.41	33 7.27	28 6.17	9	20 4.41	45 9.91	475 454
March.G	42 9.23	44 9.67	9	30	31	21 4.62	47 10.33	13 2.86	24 5.27	24 5.27	39 8.57	9 1.98	17	22 4.84	12	30 6.59	17 3.74	31 6.81	30 6.59	8	19 4.18	43 9.45	475 455
Nico.G	45 9.78	42 9.13	9	27	33	21 4.57	46 10.00	14 3.04	21 4.57	25 5.43	41 8.91	8 1.74	15	21 4.57	13	30 6.52	17 3.70	29 6.30	36 7.83	8	18 3.91	46 10.00	477 460
Nico.P	45 9.76	44 9.54	8	29	30	21 4.56	47 10.20	14 3.04	21 4.56	25 5.42	41 8.89	8 1.74	15	21 4.56	13	30 6.51	17 3.69	29 6.29	37 8.03	8	18 3.90	43 9.33	477 461

Appendix 3 cont.

	A	B	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Z	#
Oli.G	51 10.81	51 10.81	8	30	27	17 3.60	44 9.32	8 1.69	29 6.14	21 4.45	39 8.26	17 3.60	21	18 3.81	11	29 6.14	22 4.66	27 5.72	36 7.63	8	25 5.30	38 8.05	488 472
Oryza.G	46 10.02	43 9.37	10	28	34	22 4.79	46 10.02	15 3.27	22 4.79	25 5.45	36 7.84	12 2.61	15	23 5.01	10	29 6.32	18 3.92	30 6.54	31 6.75	8	17 3.70	44 9.59	477 459
Petunia.G	45 9.76	41 8.89	8	27	32	22 4.77	46 9.98	15 3.25	22 4.77	24 5.21	40 8.68	9 1.95	14	21 4.56	14	32 6.94	18 3.90	28 6.07	35 7.59	8	17 3.69	46 9.98	477 461
Pisum.G	44 9.61	42 9.17	9	27	31	21 4.59	46 10.04	14 3.06	22 4.80	26 5.68	41 8.95	9 1.97	15	23 5.02	12	29 6.33	16 3.49	31 6.77	32 6.99	8	19 4.15	43 9.39	475 458
Porphy.G	46 9.70	52 10.97	6	30	25	19 4.01	43 9.07	10 2.11	27 5.70	23 4.85	40 8.44	18 3.80	22	21 4.43	17	27 5.70	23 4.85	27 5.70	36 7.59	8	20 4.22	42 8.86	488 474
Prochl.G	39 8.67	41 9.11	12	28	36	23 5.11	45 10.00	15 3.33	23 5.11	21 4.67	43 9.56	11 2.44	13	22 4.89	13	30 6.67	13 2.89	31 6.89	30 6.67	8	14 3.11	49 10.89	470 450
Rbactf1.G	46 9.77	44 9.34	6	28	27	19 4.03	48 10.19	11 2.34	24 5.10	26 5.52	38 8.07	16 3.40	16	21 4.46	14	26 5.52	19 4.03	32 6.79	39 8.28	9	21 4.46	41 8.70	486 471
Rbactf2.G	54 12.00	44 9.78	4	31	26	23 5.11	48 10.67	15 3.33	20 4.44	18 4.00	33 7.33	20 4.44	13	19 4.22	13	29 6.44	18 4.00	28 6.22	23 5.11	5	19 4.22	39 8.67	459 450
Rrub.P	62 13.63	44 9.67	5	29	27	23 5.05	50 10.99	12 2.64	21 4.62	18 3.96	33 7.25	14 3.08	15	22 4.84	13	27 5.93	19 4.18	25 5.49	28 6.15	6	17 3.74	40 8.79	466 455
Spin.G	43 9.39	41 8.95	9	26	33	20 4.37	45 9.83	15 3.28	20 4.37	23 5.02	41 8.95	10 2.18	15	22 4.80	11	29 6.33	18 3.93	33 7.21	35 7.64	8	19 4.15	44 9.61	475 458
Synech.G	44 9.65	45 9.87	7 -	31 -	29 -	24 5.26	45 9.87	16 3.51	22 4.82	26 5.70	44 9.65	12 2.63	14	21 4.61	12	27 5.92	19 4.17	29 6.36	27 5.92	9	14 3.07	41 8.99	472 456
Xbactf1.G	50 10.62	47 9.98	6	33	27	18 3.82	50 10.62	10 2.12	23 4.88	25 5.31	36 7.64	15 3.18	14	19 4.03	17	28 5.94	16 3.40	31 6.58	40 8.49	11	19 4.03	44 9.34	488 471
Zea.G	45 9.87	45 9.87	11	31	29	19 4.17	48 10.53	14 3.07	23 5.04	25 5.48	38 8.33	12 2.63	14	21 4.61	12	27 5.92	19 4.17	30 6.58	30 6.58	8	19 4.17	41 8.99	475 456

VITA

Surname: Jeffs Given Names: Bruce Thomas

Place of Birth: Victoria B.C. Date of Birth: March 16, 1951

Educational Institutions Attended:

University of Victoria	1984-1991
Camosun College	1979-1984
Simon Fraser University	1972-1973
University of Victoria	1969-1970

Degrees Awarded:

B.Sc.	University of Victoria	1988

Honours and Awards

Melburn-Brenton Scholarship	1987
Lewis J. Clark Memorial Fellowship	1988
Lewis J. Clark Memorial Fellowship	1989
University of Victoria Fellowship	1988
NSERC Fellowship	1989

Publications:

Lucey, Wm. P., Moore, B., Jeffs, B. T., Barraclough, C. L., Congdon, K. E., and Austin, A. P. 1992. Impact on a pristine mountain river of domestic waste waters from remotely located, recreationally accelerated, development of a wilderness. Technical Report. *Canadian Journal of Fisheries and Aquatic Science. Technical Report No.* 1879:49-60.

Nickolichuk, K. J., Jeffs, B. T., Austin, A. P., Lucey, Wm. P., and Ridley-Thomas, C. I. 1989. Suitability of periphyton as a dietary ingredient in rainbow trout (*Oncorhynchus mykiss*) feeds. *European Aquaculture Society, EAS Special Publication No.* 10:189-192.

PARTIAL COPYRIGHT LICENSE

I hereby grant the right to lend my thesis to users of the University of Victoria Library, and to make single copies only for such users or in response to a request from the Library of any other university, or similar institution, on its behalf or one of its users. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by me or a member of the University designated by me. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Title of Thesis: Rubisco LSU amino acid composition as a phylogenetic indicator.

Author [redacted] [redacted]

Bruce T. Jeffs

940314