INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600
TWO STUDIES IN EVOLUTIONARY HISTORY: THE ORIGIN OF THE NONPHOTOSYNTHETIC ALGAE POLYTOMA AND THE REPRODUCTIVE MECHANISM OF ACANTHAMOEBA

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
The Degree Doctor of Philosophy in the Graduate School of the Ohio State University

By
Robert Rumpf, M. S.

The Ohio State University
1997

Dissertation Committee: 
Dr. C. William Birky, Jr., Adviser
Dr. Tom Byers
Dr. Daniel Crawford
Dr. Amanda Simcox

Approved By
Advisor
Department of Molecular Genetics
ABSTRACT

The evolutionary history of two disparate organisms, each with their own unique evolutionary question, was studied using phylogenetic analysis of DNA sequences. *Polytoma*, a non-photosynthetic alga of the family Chlamydomonadaeae, is most closely related to the photosynthetic genus *Chlamydomonas* based on morphological evidence. Complete sequences of the *Rrn18* genes were obtained from 13 strains of *Polytoma*. Phylogenetic analysis showed that these strains formed two clades. One clade shows modest sequence diversity but is represented by strains collected at widely dispersed sites in Europe and America. The other clade consists of a single isolate from the Canary Islands. Both clades lie well within the extended clade that includes all species of *Chlamydomonas* for which sequence data are available. The two *Polytoma* clades are separated from each other by several green species, indicating that the extant nonphotosynthetic Chlamydomonadaeae arose from photosynthetic ancestors at least twice. These results further suggest that nonphotosynthetic mutants are capable of establishing lineages that can spread widely but have a higher probability of extinction than their photosynthetic congeners.

*Acanthamoeba castellanii* is the agent responsible for *Acanthamoeba* keratitis, a potentially vision-threatening opportunistic parasitic infestation of the cornea in humans. It has generally been assumed that *Acanthamoeba* reproduces asexually, primarily from lack of any evidence indicating otherwise.
Since the method of reproduction can affect the etiology and treatment of the infection, we decided to use molecular sequence data to determine the method of reproduction. Ancient asexual lineages will accumulate allelic sequence divergence (ASD), and unless this divergence is masked by frequent convergence events such as mitotic recombination, gene conversion, or ploidy cycles, measurement of ASD can be used diagnostically to determine the mode of reproduction of an organism. The TBP gene was PCR amplified, cloned, and sequenced from \textit{Acanthamoeba castellanii} Neff and \textit{Acanthamoeba griffini} S7. Divergences between alleles within a clonal culture of each species were within the expected range for sexually reproducing species; alternately \textit{Acanthamoeba} might be an asexual species with a high frequency of sequence convergence events. A phylogenetic tree constructed from the data failed to show the allele-specific topology expected of an obligate asexual species in the absence of convergence. Southern blotting showed TBP to be a single-copy gene; thus the minimum ploidy of \textit{Acanthamoeba} can be determined by the number of unique alleles found. \textit{Acanthamoeba griffini} S7 had eleven unique alleles, suggesting a minimum ploidy for \textit{Acanthamoeba} of eleven. ASD was greater in \textit{A. griffinii} than in \textit{A. castellanii}, implying that there are significant differences in the effective population size, frequency of sex, or the frequency of convergence events between the two species. Additional study is required to resolve this issue.
Dedicated to my wife, Deborah, in thanks for her support, understanding, and patience...
ACKNOWLEDGMENTS

First and foremost, I'd like to thank my adviser, Dr. C. William Birky, for all of his help, support, encouragement, and for just plain being a nice guy. Without his scientific and intellectual input, my work would not be as complete as it is; it certainly wouldn't have been as much fun.

I'd like to thank the members of the Birky lab who helped me learn techniques and tricks, helped with analysis and discussion, and kept the alpha-geek in his place: Dawne Vernon, David Schreiber, Julie Gordon, Doug Byers, Pam MacKowski, and Stacey Seibert-Harley.

I'd like to thank the Fuerst lab and the Byers labs for their contributions of data, primers, DNA, and the occasional chuckle.

I'd also like to thank my wife, Deborah, who has beyond a doubt earned her application to sainthood. Long roads, Deb, but we're traveling together...

I'd like to thank my parents, who instilled their values of hard work and an education in me, and were always there for me.

And to all the rest who provided me with support, hope, help, inspiration, and sanity checks: The Haven Gaming group (Steve Schwartz, Carole McCallister, Chris Nickel, Kathy Hamilton, and Bruce Coulson), The Wednesday Gaming Group (Will Ray, Jim Emrick, Dan Schreiber, Chris Myers), Gary Grumbling, and the kids, Merlin and Mi-ke Rumpf. Special thanks to Sting and Vangelis for setting the mood, and Apple Computer for making the Macintosh.
VITA

October 17, 1966 Born - Dayton, Ohio

1988 B.S. Biology, University of Dayton

1990 M. S. Biology, University of Dayton

1990 - present Graduate Teaching and Research Associate,
The Ohio State University Department of
Molecular Genetics

PUBLICATIONS

Research Publication

1. Rumpf, R., Vernon, D., Schreiber, D, and Birky, C. W. Jr, "Evolutionary
Consequences of the Loss of Photosynthesis in Chlamydomonadaceae:
Phylogenetic Analysis of Rm18 (18S rDNA) in 13 Polytoma strains

"Sequences of the RRN18 Genes of Chlamydomonas humicola and C.
dysosmos are identical, in agreement with their combination in the species C.

FIELDS OF STUDY

Major Field: Molecular Genetics

vi
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Introduction</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Introduction to <em>Polytoma</em></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Goals of the <em>Polytoma</em> project</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Introduction to <em>Acanthamoeba</em></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Goals of the <em>Acanthamoeba</em> Project</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Introduction to Phylogenetic Methods</td>
<td>8</td>
</tr>
<tr>
<td>2.</td>
<td>Published Paper: Evolutionary Consequences of the Loss of Photosynthesis in Chlamydomonadaceae: Phylogenetic Analysis of <em>Rrn18</em> (18S rDNA) in 13 <em>Polytoma</em> strains (Chlorophyta)</td>
<td>10</td>
</tr>
<tr>
<td>3.</td>
<td>Allelic Sequence Divergence, Reproduction, and Ploidy in <em>Acanthamoeba</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Introduction</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Materials &amp; Methods</td>
<td>22</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>1</td>
<td>List of <em>Polytoma</em> stocks.</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>List of complete <em>Rm18</em> sequences and their sources</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>Numbers of substitution and indel differences between complete sequences of <em>Rm18</em> genes from pairs of strains</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>Organisms used in TBP alignment for primer design</td>
<td>24</td>
</tr>
<tr>
<td>5</td>
<td>Table of TBP Clones.</td>
<td>30</td>
</tr>
<tr>
<td>6</td>
<td>Distance matrix of TBP clones.</td>
<td>32</td>
</tr>
<tr>
<td>7</td>
<td>Distance matrix of TBP clones (introns excluded)</td>
<td>33</td>
</tr>
<tr>
<td>Figures</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Phylogram of complete <em>Rm18</em> sequences.</td>
<td>13</td>
</tr>
<tr>
<td>2</td>
<td>Cladograms based on partial <em>Rm18</em> sequences.</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>Schematic tree depicting the implications of ASD accumulation</td>
<td>21</td>
</tr>
<tr>
<td>4</td>
<td>Alignment of TBP used in designing degenerate primers</td>
<td>25</td>
</tr>
<tr>
<td>5</td>
<td>Map of the <em>A. castellanii</em> TBP gene</td>
<td>28</td>
</tr>
<tr>
<td>6</td>
<td>Neighbor-joining distance tree of <em>Acanthamoeba</em> TBP alleles</td>
<td>34</td>
</tr>
<tr>
<td>7</td>
<td>Southern blots of <em>Acanthamoeba</em> genomic DNA probed with</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>TBP</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Alignment of <em>Acanthamoeba</em> TBP clones.</td>
<td>39</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

**Polytoma**

The family Chlamydomonadaceae consists primarily of photosynthetic green algae. Eight genera within the Chlamydomonadaceae, however, are nongreen obligate heterotrophs which lack the ability to photosynthesize (Pringsheim 1963). One of the best characterized of these is *Polytoma*, which is believed to be closely related to *Chlamydomonas* based on morphological data. *Polytoma* are ovoid biflagellated unicells similar to *Chlamydomonas* except that they are salmon pink rather than green, as they produce carotenoids but not chlorophyll (Links et al. 1963). Over a dozen *Polytoma* isolates are readily available from culture collections, making the genus amenable to study.

Nonphotosynthetic chlorophytes can arise by mutation from photosynthetic ancestors; this occurs under laboratory conditions and is assumed to occur in the wild (Harris 1989). The nonphotosynthetic genus *Polytoma* is capable of metabolizing acetate as its primary carbon and energy source, whereas many chlamydomonads are facultative heterotrophs capable of acetate metabolism, photosynthesis, or both. The growth rate of *Polytoma* is comparable to that of *Chlamydomonas* (Boynton et al., 1972), suggesting that loss of photosynthesis may be selectively neutral, or nearly neutral, especially in a nutrient-rich environment.

A *Polytoma* species can be postulated to arise from a nonphotosynthetic *Chlamydomonas* mutant in a minimum of three steps: (1) the mutation is fixed
in the population, or establishes a new population in a new habitat; (2) the heterotrophic population becomes isolated from the photosynthetic members of the same species by geographical and/or reproductive isolation; and (3) additional mutations in the photosynthetic pathway accumulate in the absence of selective pressure, making the loss of photosynthesis irreversible (Rumpf et al., 1996). After chlorophyll synthesis is lost, the species would be classified as a *Polytoma*.

Loss of photosynthesis removes selective pressure from the majority of plastid genes involved in photosynthesis and protein synthesis (Wolfe et al. 1992a & 1992b), which will increase the probability of fixation of additional mutations in genes required for photosynthesis, making the loss of photosynthesis irreversible. Additionally, these nonphotosynthetic species are less versatile than the facultative heterotrophs from which they evolved, and presumably are more limited in the habitats they can occupy. It seems logical that nonphotosynthetic lineages would become established and speciate infrequently at best, and have a higher probability of extinction than their photosynthetic congener. Chapter 2 of this work details my investigations into the evolutionary relationship of *Polytoma* and its nearest photosynthetic relative, *Chlamydomonas*, in attempts to characterize its origin and extract information regarding the frequency with which nonphotosynthetic lineages arise, how long they survive, and the extent to which they speciate.

**Project Goal: Polytoma**

The goal of the first half of my project has been to investigate the evolutionary relationship between 13 *Polytoma* species and their photosynthetic relatives among the *Chlamydomonads* using the sequence of
the gene encoding the nuclear small subunit RNA of the ribosome (Rrn18). The Rrn18 gene is useful in this analysis because it is essential for protein synthesis and therefore ubiquitous and highly conserved. The rRNA product forms a complex secondary structure containing loops and stems, both of which contribute to the information content of the molecule. The gene has been used extensively in phylogenetic studies (e.g. Fernholm et al., 1989), providing a large database of sequences and phylogenetic trees for comparison.

By reconstructing the phylogeny of Polytoma, I hope to answer several questions: (1) how many times has this nonphotosynthetic lineage arisen from photosynthetic ancestors; (2) how speciose is each of these lineages; and (3) how old is each lineage.

Nonphotosynthetic mutants frequently arise spontaneously under laboratory conditions, and, by extrapolation, in nature. These mutants can and do establish species of colorless algae. Over time, additional mutations arise in the pathway responsible for photosynthesis, which is no longer under selection, making reversion unlikely. Why hasn’t this happened to all photosynthetic lineages, rendering them incapable of photosynthesis? There are two approaches to answering this question: individual selection and species selection. Selection working at the level of the individual would prevent the nonphotosynthetic lineage from being established, whereas selection at the level of the species would lead to the extinction of the nonphotosynthetic lineage after it had already been established.

It is conceivable that the nonphotosynthetic mutants are infrequently fixed in the population, as a direct result of the mutation itself. Losing a major nutritional pathway would seem intuitively to be detrimental. It has been shown, however, that the growth rates of photosynthetic and nonphotosynthetic species
is similar in nutrient-rich environments (Boynton et al. 1972), so at least under these conditions the loss of photosynthesis is not detrimental, and individual selection would not come into play.

It is also likely that nonphotosynthetic species become extinct more readily or speciate less frequently, or both. Again this sounds reasonable as these species would be less adaptable to changing environmental conditions than their facultative heterotrophic competitors, and would thus occupy a narrower niche.

Phylogenetic trees of photosynthetic and nonphotosynthetic lineages could shed light on this by way of topology: if the nonphotosynthetic lineages were few in number but deeply rooted, this would suggest a low speciation rate; alternately if the nonphotosynthetic lineages were abundant but short-branched, this would suggest a high rate of extinction. Both of these would indicate that selection was working at the level of the species. A third alternative, in which all lineages were approximately equal in branch length, with an origin near the time of one of the major extinctions, such as the Cretaceous-Tertiary (KT) boundary, would suggest strong individual selection against nonphotosynthetic mutants except at times of low light and high nutrient availability, both of which were presumed present at the times of the major extinctions.

**Acanthamoeba**

*Acanthamoeba* is the causative agent of granulomatous amebic encephalitis, a generally fatal opportunistic infection of the central nervous system common in immunocompromised individuals (Visvesvara et al., 1983; Ma et al., 1990; John, 1993), and *Acanthamoeba* keratitis, a potentially vision-
threatening parasitic invasion of the cornea in contact lens wearers (Dart, 1988). This genus of free living protozoans has been divided into three morphological groups based on cyst morphology (Pussard and Pons, 1977). *A. castellanii* and *A. griffini*, the objects of this study, fall into Group II, which is characterized by cysts of < 18 μm in diameter, and stellate, polygonal, triangular, oval, or round endocysts (Visvesvara, 1991; Rivasi et al., 1995). *Acanthamoeba* alternates between trophozoite and cyst stages in its life cycle, making treatment of infections difficult. The cyst stage is highly resistant to a wide variety of conditions and antimicrobial compounds (Osato et al., 1991; Hay et al., 1994; Elder et al., 1994), allowing the organism to survive in this state until treatment is discontinued.

It is believed that *Acanthamoeba* reproduces asexually; this is due primarily to the lack of convincing evidence for genetic recombination (Akins, 1981). The observation of identical mitochondrial 16S genes in association with differing nuclear 18S genes suggests that reproduction in *Acanthamoeba castellanii* may not be entirely clonal (Ledee 1995). The method of reproduction affects the etiology of the diseases caused by an infectious parasite; an asexual organism makes an excellent rapid colonizer, whereas sexual reproduction of the infectious agent facilitates the quick and efficient dissemination of drug resistance genes throughout the population. The resulting diseases should be treated accordingly.

Another question in *Acanthamoeba* genetics is that of ploidy. Measurement of nuclear DNA in *A. castellanii* Neff suggests a total nuclear DNA content at log phase of $10^9$ bp (King, 1980). Kinetic complexity measurements suggest a haploid genome size of $4-5 \times 10^7$ bp (Bohnert & Heermann, 1974), whereas pulse field electrophoresis has suggested that the genome size is on
the order of $2.3-3.5 \times 10^7$ bp (Rimm et al., 1988). Based on these data, the ploidy may range anywhere from 10 to 22. Chapter 3 of this work details my investigation into both the method of reproduction and the ploidy of *Acanthamoeba*.

**Project Goal: Acanthamoeba**

The primary goal of my project regarding *Acanthamoeba* is to determine the method of reproduction of two species of *Acanthamoeba*, *A. castellanii* and *A. griffini*. To accomplish this goal, I utilized the measurement of allelic sequence divergence (ASD), a technique pioneered by Dr. Matthew Meselson, and for which the theory was fully developed by Birky (1996).

In a sexual species, the two alleles of any gene will have a most recent common ancestor, referred to as the coalescent of the alleles, within a relatively short span of time (an average of $2N_e$ generations where $N_e$ is equal to the effective population size). The expected neutral sequence divergence between two different alleles in a sexual population equals $4N_e\mu$, where $\mu$ is the mutation rate. In a lineage where the capability for sexual reproduction has been lost, however, the distance to the coalescent will be significantly greater; not only will the alleles have the existing $2N_e$ generations of divergence time from when the organism was sexual, but time $t$ (the time since the loss of sexual reproduction) will allow for additional divergence of some pairs of alleles, most notably those in one individual or clone. Thus the expected neutral sequence divergence between two different alleles in an asexual population equals $4N_e\mu + 2t\mu$, where $t =$ the time since the loss of sexual reproduction. This effect was clearly seen in the bdelloid rotifers, which had an ASD as high as 40% (M. Meselson and D. Welch, personal communication), which is significantly higher than the
maximum value found for sexual species (Avise, 1994; Moriyama and Powell, 1996).

Allelic sequence divergence will be limited by selection; in a protein coding gene, however, synonymous substitutions will be subject to weak or no selection. Thus using a protein coding gene and concentrating on divergence at the third codon position will largely avoid the problem of selection.

It is possible that, although a species is asexual, sequence divergence can be reduced in one of several ways; this is called sequence convergence. Mitotic recombination, a ploidy cycle, or rare sex (in which case the description of the species as asexual is a matter of semantics) can reduce or eliminate allelic sequence divergence. In such cases, measurement of sequence divergence could lead to an asexual species being mistaken for a sexually reproducing one. This effect can be partially countered by sequencing the alleles of several genes or the alleles of one gene in a polyploid; only a ploidy cycle will affect all genes in an organism or all genes on a single chromosome. Another interesting result of the loss of sexual reproduction is that the alleles within an organism could be more divergent to one another than to alleles in another organism; this was also seen in the bdelloid rotifers (M. Meselson and D. Welch, personal communication). This results in allele-dependent phylogenetic trees, in which different sets of alleles have different tree topologies (See Chapter 3 and Figure 3).

By amplifying, cloning, and sequencing all possible alleles of a given protein coding gene from *Acanthamoeba*, I will be able to measure the allelic sequence divergence, and possibly determine whether or not *Acanthamoeba* reproduces sexually. Additionally, by counting the number of unique alleles cloned I can determine a minimum ploidy number for the organism.
**Phylogenetic Methods**

A phylogenetic tree based on any data, be it sequence data or morphological characteristics, is at best an approximation of the true state of the universe. The level of agreement between the inferred tree and the true tree depends on the nature and quality of the data as well as the method of phylogenetic reconstruction employed (Gray, 1992). Strong correlation between trees derived from the same data by different reconstruction methods based on different evolutionary models suggests that the inferred trees are close to the true tree (Woese et al., 1980). Three reconstruction methods were used for my analysis of the *Rm18* gene in *Polytoma*: neighbor-joining, parsimony, and maximum likelihood.

Distance methods employ a specific evolutionary model to first calculate the number of substitutions between sequences in a pairwise fashion, generating a distance value for all possible pairs of taxa. Distances are generally corrected for multiple hits using one of a variety of methods. These values are used to construct a tree based on the neighbor-joining algorithm, which optimizes the tree based on observed distances (seen after tree construction) and actual distances (from the pairwise calculations). By adjusting the separation of pairs of nodes based on their divergence from all other nodes, the neighbor-joining algorithm counteracts the effects of differing rates of evolution in the various taxa (Saitou and Nei, 1987). The neighbor-joining trees in this work were constructed using the PHYLIP package (Felsenstein, 1993 version 3.5).

Parsimony algorithms construct trees so that the minimum possible number of base substitutions are required for the resulting tree topology. No correction is made for multiple hits, and only informative sites are taken into
consideration. Parsimony trees in this work were constructed using PAUP 3.5 (Swofford 1993).

Maximum-likelihood algorithms construct trees by recursively calculating the likelihood of node topologies for each position within the alignment, beginning at the top of the tree and moving downwards one node at a time, testing the likelihood of the topology against a specific evolutionary model. The likelihood calculations derived at each step are used in subsequent steps to find the likelihood of the next node. This process is repeated until all nodes are calculated; the overall likelihood is then calculated by summing the products of the root likelihoods with the prior likelihoods at each node; the best tree is the one with the highest likelihood. Maximum likelihood assumes that each site evolves independently following either the Poisson or some other probability process; thus the likelihood of a phylogeny can be calculated separately for each site (Felsenstein, 1981).

All trees in this work were bootstrapped. Bootstrapping is a resampling technique in which a dataset of $n$ positions is resampled $n$ times at a random position within the array, creating a new data set of the same size as the original. Since the resampling is random, any given site within the array may be represented not at all, once, or in multiple. The bootstrap values which are placed at nodes of the resulting phylogenetic trees represent the proportion of bootstrap datasets which supported that particular node (Felsenstein, 1988). These values, although not statistical measures themselves, can be used to infer which nodes are more strongly supported by the given dataset.
EVOLUTIONARY CONSEQUENCES OF THE LOSS OF PHOTOSYNTHESIS IN CHLAMYDOMONADACEAE: PHYLOGENETIC ANALYSIS OF Rrn18 (18S rDNA) IN 13 POLYTOMA STRAINS (CHLOROPHYTA)


Department of Molecular Genetics, The Ohio State University, 484 West Twelfth Avenue, Columbus, Ohio 43210

ABSTRACT

Complete sequences of the Rrn18 genes were obtained from 13 strains of the nonphotosynthetic algal genus Polytoma. Phylogenetic analyses showed that these strains formed two clades. One clade shows only modest sequence divergence but is represented by strains collected at widely dispersed sites in Europe and America. The other clade consists of a single isolate from the Canary Islands. Both clades lie well within the extended clade that includes all species of Chlamydomonas for which sequence data are available. The two Polytoma clades are separated from each other by several green species, suggesting that the extant nonphotosynthetic Chlamydomonadaeae arose from photosynthetic ancestors at least twice. These results suggest that nonphotosynthetic mutants are capable of establishing lineages that can spread widely but have a higher probability of extinction than their photosynthetic congeners.

Key index words: 18S rRNA; Chlorophyta; Chlamydomonas; heterotrophic algae; loss of photosynthesis; molecular evolution; Polytoma; phylogenetics; Rrn18 gene

Most chlorophyte algae are photosynthetic and contain chlorophyll in their plastids, as suggested by the common appellation "green algae." However, a number of nongreen heterotrophic species are also classified in the Chlorophyta on the basis of morphological and biochemical similarities to various photosynthetic groups. One of the best-known genera of heterotrophic chlorophytes is Polytoma, oval to elongated biflagellate unicells with a single large plastid (leucoplast) partly surrounding the centrally located nucleus (Lang 1963, Pringsheim 1963). The leucoplast contains ribosomes and DNA (Scherbel et al. 1974, Siu et al. 1975a, b, c, Vernon-Kipp et al. 1989). Polytoma is believed to be closely related to Chlamydomonas on morphological grounds (Pringsheim 1963). The genus Chlamydomonas is diverse in terms of morphology, habitat, and metabolism and has been classified into over 450 species on the basis of morphology and sensitivity to cell wall autolysins (Ettl 1976, Schlösser 1984, Ettl and Schlösser 1995). We used DNA sequence data to verify the close relationship between one isolate of Polytoma and Chlamydomonas reinhardtii (Vernon-Kipp et al. 1989); here we show that 12 additional isolates of Polytoma are also members of the Chlamydomonas clade.

It is generally assumed that nonphotosynthetic chlorophytes arose from photosynthetic ancestors. Polytoma species are obligate heterotrophs capable of utilizing acetate as their sole carbon and energy source, while many chlamydomonads are facultative heterotrophs that can utilize acetate, photosynthesis, or both. Nonphotosynthetic mutants are easily obtained in C. reinhardtii and can result from mutations in any of a number of different genes in the chloroplast or nucleus (Harris 1989). Many of them are easily maintained on acetate in the light or dark, but detailed comparisons of growth rates to wild-type cells have been made for only two mutants (Boynton et al. 1972). One of these, ac-20, grows only slightly slower than wild type under heterotrophic conditions with no light and acetate, in spite of having a reduced rate of plastid protein synthesis. Thus, it is reasonable to suppose that some nonphotosynthetic mutants (perhaps a small minority) might be selectively neutral or nearly neutral in nature, especially in an environment rich in organic nutrients and under less intense and discontinuous illumination. Such a mutant could give rise to a Polytoma species by the following steps: 1) the mutation is fixed in the population (probably by random drift), or establishes a new population in a new habitat; 2) the heterotrophic population becomes isolated from photosynthetic members of the same species by geography and/or the acquisition of genetic isolation mechanisms; and 3) additional mutations in the photosynthetic pathway accumulate in the absence of selective pressure, making the loss of photosynthesis irreversible. The population will be classified as Polytoma when chlorophyll synthesis is lost, due to either the initial mutation or, more often, a subsequent mutation.

The loss of photosynthesis is likely to have important consequences for evolution of the plastid genome. These include the loss of most or all genes coding for proteins involved in photosynthesis (Wolfe et al. 1992b, but see Siemeister and Hachtel 1990) and the accelerated evolution of genes coding for ribosomal RNAs (rRNAs) and proteins involved in protein synthesis (Wolfe et al. 1992a). In addition, it is likely that selection operating at the level of species or populations favors photosynthetic over nonphotosynthetic organisms. Although an obligate
heterotroph might reproduce as rapidly as a photosynthesizer in an environment rich in organic nutrients, it will be unable to occupy the nutrient-poor environments that are open to photosynthesizers.

As part of our investigations of the evolutionary consequences of the loss of photosynthesis in algae, we determined the evolutionary relationships of a large sample of the available strains of Polytoma to species of Chlamydomonas. The nuclear Rml8 gene and the small subunit rRNA for which it codes have been widely used to reconstruct evolutionary histories. The sequences of these genes have been used to clarify the phylogeny of 28 Chlamydomonas species and 12 species from closely related genera (Jupe et al. 1988, Buchheim et al. 1990, Buchheim and Chapman 1991, 1992, Chapman and Buchheim 1991, Lar sen et al. 1992). The maximum sequence divergence within this group is about 10%, greater than that between gymnosperms and angiosperms. Parsimony trees generated with these data consistently group these organisms into six or seven clades. To which, if any, of the members of this group are the various species of Polytoma related?

We have now sequenced the complete Rml8 gene from 15 Polytoma strains and two strains of Chlamydomonas. We constructed phylogenetic trees using our sequences and those of other species of Chlamydomonas and closely related green algae provided by other researchers. Our goals were to determine 1) the number of distinct heterotrophic lineages, i.e. the number of times extant nonphotosynthetic species arose from photosynthetic ancestors; 2) the age of the heterotrophic lineages; and 3) the extent of speciation and dispersion of the nonphotosynthetic lineages. Our results show only two origins, one represented by a single isolate, the other by 12 widely distributed isolates that probably represent a single species.

MATERIALS AND METHODS

All strains were obtained from stock centers or other laboratories (Table 1) and subcloned (with the exception of SAG 62-2C, which we were unable to subclone successfully). Some strains have never been assigned specific names, whereas others were given provisional names that appear to have been dropped subsequently; therefore, we will refer to them by their strain numbers. Stocks were grown in 10 mL liquid PM media (2.0 g sodium acetate (hydrate), 1.0 g yeast extract, 1.0 g tryptone, doubly distilled water to 1 L) and transferred monthly using a 1:10 dilution of existing stock into fresh medium.

**DNA preparation.** Cells were grown to log phase in 1-2 L PM. Cells were harvested by centrifugation (Beckman JAIO, 14 x g, 10 min), and the resulting pellet was frozen in liquid nitrogen. The resulting lyssate was extracted once with 24:1 chloroform:isoamyl alcohol, and the DNA was precipitated with ethanol and resuspended in 1 mL TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0).

**Polymerase chain reaction amplifications.** The Rml8 gene was amplified from the crude lysates with primers located at the ends of the gene. The 5' and 3' primers were, respectively, SSUIA = 5' - TGGTTGACTCCTGCCAGTAG-3' (sense strand sequence) and SSU2 = 3'CATTGCAGCTTCTTTAGT-5' (antisense strand sequence) and SSU2 = 3'CATTGCAGCTTCTTTAGT-5' (antisense strand sequence). The optimal amplification conditions for each lyssate were determined empirically. Then multiple amplifications under these conditions were performed and pooled for use as template in subsequent sequencing reactions. The pooling of multiple individual amplifications minimizes sequencing errors due to misincorporation of nucleotides by Taq polymerase during individual amplifications. The pooled DNA was purified via the GeneClean kit (Bio 101).

Accidental amplification of DNA from the wrong organism was extremely unlikely because 1) signal DNA molecules greatly outnumbered any contaminating DNA molecules because multiple amplifications were done, each using about 1 µL of the crude lyssate from a 2-L culture, and 2) all sequences clearly belong to the Chlamydomonas clade, of which the only cultures in the lab were the Polytoma isolates, Chlamydomonas reinhardtii, C. tumulosa,

### Table 1. List of Polytoma stocks.

<table>
<thead>
<tr>
<th>Stock number</th>
<th>Species*</th>
<th>Where collected</th>
<th>Collector</th>
<th>Accession N°</th>
</tr>
</thead>
<tbody>
<tr>
<td>UTEX 19</td>
<td>P. wellii Ehrenberg</td>
<td>?</td>
<td>E. G. Pringsheim</td>
<td>U22940</td>
</tr>
<tr>
<td>UTEX 954</td>
<td>P. wellii Ehrenberg</td>
<td>Woods Hole, Mass., USA</td>
<td>F. Moewus</td>
<td>U22943</td>
</tr>
<tr>
<td>ATCC 20963</td>
<td>P. obtusum</td>
<td>?</td>
<td>?</td>
<td>U22958</td>
</tr>
<tr>
<td>SAG 195-80</td>
<td>(Tetraedrales sp.)</td>
<td>South Africa</td>
<td>E. G. Pringsheim</td>
<td>U22957</td>
</tr>
<tr>
<td>CCAP 62/2M</td>
<td>P. wellii Ehrenberg</td>
<td>Leiden, Holland</td>
<td>E. G. Pringsheim</td>
<td>U22942</td>
</tr>
<tr>
<td>SAG 62-2c</td>
<td>P. wellii Ehrenberg</td>
<td>Brixen, Tirol, Italy</td>
<td>E. G. Pringsheim</td>
<td>U22941</td>
</tr>
<tr>
<td>SAG 62-15</td>
<td>(P. mirum Pringsheim)</td>
<td>Solingen, Germany</td>
<td>E. G. Pringsheim</td>
<td>U22954</td>
</tr>
<tr>
<td>SAG 62-16</td>
<td>(P. difficultis Pringsheim)</td>
<td>Heidelberg, Germany</td>
<td>E. G. Pringsheim</td>
<td>U22932</td>
</tr>
<tr>
<td>SAG 62-18</td>
<td>(P. elipticus Pringsheim)</td>
<td>Klaasen-Brixen, Tirol, Italy</td>
<td>E. G. Pringsheim</td>
<td>U22953</td>
</tr>
<tr>
<td>SAG 62-20</td>
<td>Polytoma sp.</td>
<td>Prutz, Tirol, Italy</td>
<td>E. G. Pringsheim</td>
<td>U22959</td>
</tr>
<tr>
<td>SAG 62-21</td>
<td>(P. anomale Pringsheim)</td>
<td>California, USA</td>
<td>E. G. Pringsheim</td>
<td>U22951</td>
</tr>
<tr>
<td>SAG 62-27</td>
<td>(P. anomale Pringsheim)</td>
<td>Tenerife, Spain</td>
<td>E. G. Pringsheim</td>
<td>U22956</td>
</tr>
<tr>
<td>DH1</td>
<td>P. obtusum</td>
<td>?</td>
<td>?</td>
<td>U22955</td>
</tr>
</tbody>
</table>

* ATCC = American Type Culture Collection, CCAP = Culture Collection of Algae and Protozoa, SAC = Sammlung von Algenkulturen Göttingen, UTEX = University of Texas Culture Collection of Algae; DH1 was supplied by David Herrin and came originally from the culture collection. L. Provosoli at Yale University; strain designation is that of the authors.

* Names in parentheses are provisional or former names of strains now designated Polytoma sp.

* GenBank Accession number for Rml8 (188) gene sequence.
Molecular phylogenetics of Polytoma

Table 2. List of complete rRNA sequences and their sources.

<table>
<thead>
<tr>
<th>Taxon</th>
<th>GenBank accession</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asteromastix gracilis</td>
<td>M95614</td>
<td>Unpublished</td>
</tr>
<tr>
<td>Chlamydomonas hemiclora*</td>
<td>U15884</td>
<td>Gordon et al. 1995</td>
</tr>
<tr>
<td>Chlamydomonas reinhardii</td>
<td></td>
<td>Gunderson et al. 1987</td>
</tr>
<tr>
<td>Chlorella vulgaris</td>
<td>X15588</td>
<td>Huang and Sogin 1989</td>
</tr>
<tr>
<td>Dunaliella salina</td>
<td>M56208</td>
<td>Published</td>
</tr>
<tr>
<td>Dunaliella tertiolecta</td>
<td>M84320</td>
<td>Wilcox et al. 1992</td>
</tr>
<tr>
<td>Glycera max</td>
<td>M20017</td>
<td>Eckertode et al. 1985</td>
</tr>
<tr>
<td>Oruga sinia</td>
<td>X00755</td>
<td>Takeiwa et al. 1984</td>
</tr>
<tr>
<td>Saccharomyces cerevisiae</td>
<td>M27607</td>
<td>Mankin et al. 1986</td>
</tr>
<tr>
<td>Valvata carteri</td>
<td>X53504</td>
<td>Raush et al. 1989</td>
</tr>
<tr>
<td>Zamia pumila</td>
<td>M20017</td>
<td>Nairn and Perl 1988</td>
</tr>
</tbody>
</table>

*Accession numbers.

**Combined with other strains under C. appananae by Exl and Schlösser (1992).

C. dynoos, and three isolates of Polytoma that have different sequences from those shown here (unpubl.).

Direct sequencing of double-stranded DNA. The pooled polymerase chain reaction (PCR) products were directly sequenced using the double-stranded DNA Cycle Sequencing kit and protocols (BRL), with two modifications: 200–500 ng of PCR product and 1.5–2 µL of 10 µM-1 TAQ DNA polymerase were used for each reaction. Primers for both strands were constructed for conserved sites at approximately 200-bp intervals along the gene and used for consecutive sequencing reactions. Each strain was sequenced at least twice on one strand; several were sequenced twice on both strands. The high level of conservation of this gene allowed for an additional proofreading step: sequences from two closely related strains were compared and sites of divergence were checked on the autoradiographs to positively identify differences, which were then sequenced again for verification. All sequence autoradiograms were read independently by at least two people.

Other DNA sequence. The complete rRNA sequences of other algae, yeast, and plants were obtained from GenBank; accession numbers and references are given in Table 2. Partial sequences of 18S rRNAs from additional Chlamydomonadaceae species shown in Figure 2 were provided by Mark A. Buchheim (University of Tulsa).

Sequence alignment. DNA sequences were entered directly in the SeqApp computer application package (Gilbert 1992) running on Macintosh computers. SeqApp, through links to CAP (Comig Assembly Program, Huang 1992) and ClustalV (Higgins and Sharp 1988, 1989, Higgins et al. 1995), was also used to assemble the sequencing fragments of each strain into a contig and then to produce the initial alignment of six green Chlamydomonadaceae and 15 Polytoma strains. We refined this alignment by hand, influenced by primary sequences from three land plants (Zamia pumila, Glycera max, and Oruga sinia) and a yeast (Saccharomyces cerevisiae). We further refined the alignment using conserved patterns in the secondary structures for this gene in one land plant (Zea mays, Gutell et al. 1985), three green algae (Chlorella vulgaris, Huw and Sogin 1989; Valvata carteri, Raush et al. 1989; Chlamydomonas reinhardii, Gutell et al. 1985), and S. cerevisiae (Gutell et al. 1985).

We were able to align 98.9% of the rRNA gene from 12 of the 15 Polytoma strains (the strains in the P. wollei clade), leaving out only the first 4 3' positions and the last 16 5' positions for lack of data in all species. For the entire array of complete algal sequences for the taxa shown on the accompanying table, we were able to align 96.6% of the gene, omitting only two sets of 6 and 55 bases where the alignment was ambiguous, as well as the short sections at the 5' and 3' ends for lack of data; this left 1775 base pairs.

For a second dataset of partial rRNA sequences from 29 additional Chlamydomonadaceae and 30 other chlorophytes (kindly provided by Mark Buchheim), our complete sequences were truncated to match the partial sequences and aligned as earlier. This data set had 963 total sites.

Phylogenetic analysis. Three types of phylogenetic analysis programs were run on Macintosh and Iris Indigo computers. The DNAboot, DNADist, and neighbor-joining algorithms of Phylip v. 3.56 (Felsenstein 1989, 1993) were used to construct neighbor-joining trees; the Kimura two-parameter method with a transition/transversion ratio of 2.0 was used to correct for multiple hits. PAUP v. 4.02d5 (Swolldorf 1995, results published with permission of the author) was used to generate maximum parsimony trees with three kinds of searches: exhaustive and branch-and-bound heuristic searches, which give the most parsimonious tree, and a full heuristic search, from which the 50% majority consensus tree was used. All trees (except the one from the exhaustive search maximum parsimony) were bootstraped 100 times. Branch swapping by tree-bisection-reconnection (TB) was used with accelerated transformation (ACCTRAN); 10 random sequence addition replicates were used for each of the 100 bootstrap replicates in the heuristic search, for a total of 1000 replicates. There were 91 informative sites in the complete sequences from six green Chlamydomonadaceae, 15 Polytoma strains, and Chlorarella, and 2) a subset of the partial sequences including 24 green Chlamydomonadaceae, Polytoma strains 964 and 62-27, and Chlorarella. Chlorarella vulgaris was used as an outgroup because Chlorarella species are consistently separated from Chlamydomonas species in trees based on 5S and 18S rRNA sequences (Chapman and Buchheim 1991). Distance matrices (Table 3) were constructed with the aid of The InDel/Substitution Machine, a Hypercard-based program (R. R., unpubl.). The program constructs matrices of pairwise base substitutions and indels (gaps due to either insertions or deletions).

Results and discussion

Choice of genes. Several factors make the rRNA gene an excellent candidate for phylogenetic analysis. The gene product (185 rRNA) is essential for protein synthesis and therefore can be found in all organisms. The rRNA product forms a secondary structure rich in stems and loops; the structure is sufficiently conserved that it can be used to improve alignments. Finally, this gene has already been used to determine the phylogenetic relationships of numerous organisms (e.g. Farnholm et al. 1989), and a large database of 18S sequences and phylogenetic trees is available for comparison (Maidak et al. 1994). For the photosynthetic Chlamydomonadaceae in particular, complete sequences of the rRNA gene are available for 8 species, and partial sequences of the 18S rRNA are available for 18 additional species (Vernon-Kapp et al. 1989, Buchheim et al. 1990, Buchheim and Chapman 1991, 1992, Chapman and Buchheim 1991; Mark A. Buchheim, pers. commun.).

Alignments. The secondary structure of the 18S rRNA is highly conserved, in part by compensating base substitutions (Hillis et al. 1994). This conservation can be used to verify the accuracy of alignments: if a region in one sequence has a particular secondary structure, the same region in another se-
Table 3. Numbers of substitution and indel differences between complete sequences of \( \text{RmI8} \) genes from pairs of strains. Indels are above the diagonal, substitutions below.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>110</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>14</td>
<td>15</td>
<td>15</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>19</td>
<td>19</td>
<td>15</td>
</tr>
<tr>
<td>105</td>
<td>14</td>
<td>15</td>
<td>14</td>
<td>18</td>
<td>15</td>
<td>15</td>
<td>13</td>
<td>15</td>
<td>20</td>
<td>14</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>88</td>
<td>88</td>
<td>81</td>
<td>0</td>
<td>5</td>
<td>8</td>
<td>4</td>
<td>12</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>90</td>
<td>90</td>
<td>84</td>
<td>12</td>
<td>3</td>
<td>8</td>
<td>4</td>
<td>12</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>106</td>
<td>102</td>
<td>94</td>
<td>40</td>
<td>43</td>
<td>11</td>
<td>7</td>
<td>15</td>
<td>10</td>
<td>11</td>
<td>9</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>83</td>
<td>105</td>
<td>14</td>
<td>15</td>
<td>18</td>
<td>14</td>
<td>15</td>
<td>13</td>
<td>15</td>
<td>20</td>
<td>14</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>88</td>
<td>88</td>
<td>81</td>
<td>0</td>
<td>5</td>
<td>8</td>
<td>4</td>
<td>12</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>90</td>
<td>90</td>
<td>84</td>
<td>12</td>
<td>3</td>
<td>8</td>
<td>4</td>
<td>12</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>106</td>
<td>102</td>
<td>94</td>
<td>40</td>
<td>43</td>
<td>11</td>
<td>7</td>
<td>15</td>
<td>10</td>
<td>11</td>
<td>9</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>83</td>
<td>105</td>
<td>14</td>
<td>15</td>
<td>18</td>
<td>14</td>
<td>15</td>
<td>13</td>
<td>15</td>
<td>20</td>
<td>14</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>88</td>
<td>88</td>
<td>81</td>
<td>0</td>
<td>5</td>
<td>8</td>
<td>4</td>
<td>12</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>90</td>
<td>90</td>
<td>84</td>
<td>12</td>
<td>3</td>
<td>8</td>
<td>4</td>
<td>12</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>106</td>
<td>102</td>
<td>94</td>
<td>40</td>
<td>43</td>
<td>11</td>
<td>7</td>
<td>15</td>
<td>10</td>
<td>11</td>
<td>9</td>
<td>9</td>
<td>12</td>
</tr>
</tbody>
</table>

Phylogenetic trees. Neighbor-joining, maximum parsimony, and maximum likelihood algorithms were used to reconstruct the phylogeny of these species. Each of these methods has distinct advantages and disadvantages and is prone to different sources of error (Hillis and Moritz 1990, Miyamoto and Cracraft 1991, Stewart 1993, Hillis et al. 1994), so that the strongest conclusions can be drawn when all methods produce the same tree topology. For the complete-sequence data set, all three methods produced the same cladogram (Fig. 1). For these trees we used the default weighting of 2 transitions : 1 transversion. We determined that the actual ratio in this data set was 1.726:1; when we used this ratio, the same parsimony and maximum likelihood trees were obtained (not shown). Twelve of the \( \text{Polytoma} \) species form a clade that we refer to as the \( \text{P. uweila} \) clade. For the Neighbor-joining algorithm, the 50% majority consensus tree obtained with a full heuristic search, the maximum likelihood tree, and the number of bootstrap replicas (out of 100) containing the corresponding clade; from top to bottom, these are from the neighbor-joining, maximum parsimony, and maximum likelihood algorithms. The most parsimonious tree had a branch length of 265, a consistency index of 0.8528, and a rescaled consistency index of 0.6456.
Molecular phylogenetics of Polytoma

Fig. 2. Cladograms based on partial Rn18 sequences. The cladograms were obtained with the neighbor-joining (left) or maximum parsimony (right) method. Numbers in boxes at the nodes are numbers of bootstrap replicas (out of 100) containing the clade. Clades with nodes marked by dark triangles on the nodal boxes were supported by the maximum likelihood tree.

have at least one green lineage between Polytoma 62-27 and the P. uvella clade. For the maximum parsimony trees, this includes all of the most parsimonious trees produced by bootstrap resampling, plus the 58 trees with lengths no more than 10 changes longer than that of the most parsimonious tree, i.e. from 265 to 275. In the complete sequence tree, a monophyletic origin of all Polytoma species from a single heterotrophic mutant would require a minimum of two reversions of the loss of photosynthesis (the partial sequence trees would require three or four reversions). We can assess the probability of two reversions from the branch lengths on the maximum parsimony tree of complete sequences. Between the original loss of photosynthesis and the second of these hypothetical reversions (on the C. humiola branch), the Rn18 gene accumulated a minimum of 18 base substitutions (data not shown) and six insertions or deletions (estimated by multiplying the number of base substitutions times the ratio of substitutions to insertions or deletions between strain 62-27 and C. humiola). During this time, each of the photosynthetic genes, which would not be subject to selection after photosynthesis was first lost, must have incurred many additional substitutions, insertions, and deletions. Reactivating photosynthesis would then require the loss of multiple base substitutions, insertions, and deletions in one lineage, which is extremely unlikely.

The P. uvella clade may represent a single species or several closely related species. The topology of the phylogenetic trees demonstrates that Polytoma species arose at least twice. However, it does not eliminate the possibility that the P. uvella clade is polyphyletic, i.e. that there are more than two origins overall. That this is unlikely is indicated by the low genetic variability within this clade. Within the P. uvella clade, 62-3 and 62-18 are identical. The pairwise sequence divergence within the remaining P. uvella clade is less than 0.17% base substitutions (0-3 substitutions/1775 bp). This can be compared to the divergence of the Rn18 genes of Chlamydomonas moewusii Gerloff (UTEX 97) and C. eugametos Moewus (UTEX 9), which are considered to be conspecific on the basis of their ability to interbreed in the laboratory (Gowans 1963) and of their similar morphology (Ettl and Schlösser 1992). Partial sequences of the small-subunit rRNA of C. moewusii and C. eugametos, provided by Mark Buchheim (pers. comm.), differ by 1 base substitution out of 737 base pairs (approximately 0.5% of the gene); this corresponds to a sequence divergence of 0.13% base substitutions. All the substitution differences within the P. uvella clade occur at only four positions, none of which are phylogenetically informative. Thus, it is not surprising that the clade was not convincingly resolved by any of the three computer algorithms.

There is more variation within the P. uvella clade with respect to indels than base substitutions. Members of the clade differ by 0-12 indels, compared to
2 in the partial sequences from C. moewusii and C. eugametos. All are insertions or deletions of a single base pair. The pairwise comparisons show no apparent correlation between the number of base substitutions and the number of indels between a pair of species. Although there are many more indels than substitutions within this clade, all the indels occur at just 25 sites, of which only 8 are phylogenetically informative. Parsimony analysis of the P. uvella clade using indels placed P. uvella strain 964 with P. obtusum DH1 in a separate clade from the other strains, which form a polytomy (not shown). This agrees with sequence data from the leucoplast rrm16 gene (Dawne Vernon, unpubl.)

Estimating ages of Polytoma clades. The loss of photosynthesis that gave rise to the P. uvella clade must have occurred after the common ancestor of the P. uvella clade and C. humicola, but before the last common ancestor of all members of the P. uvella clade. The first point can be estimated from the minimum divergence between a member of the P. uvella clade and C. humicola, which is approximately 1% (18 base substitutions/1775). The second point can be estimated from the maximum divergence between two members of the P. uvella clade, which is 0.17%. Wilson et al. (1987) presented evidence indicating that the small-subunit rRNA gene evolves at an approximately constant rate of 10⁻¹⁴ base pair substitutions per site per year in organisms ranging from bacteria to vertebrates. This would place the loss of photosynthesis in the lineage leading to the P. uvella clade somewhere between 50 and 8.5 mya. For 62-27, we can only estimate the maximum time since the loss of photosynthesis, because this nonphotosynthetic lineage has only one member. The average divergence between 62-27 and members of the C. humicola-P. uvella clade is approximately 3.1%, which gives us a maximum divergence time of roughly 155 million years between 62-27 and its nearest green relatives. Photosynthesis could have been lost at any time since then.

CONCLUSIONS

Our data show only two origins for the 13 Polytoma strains we examined. One of these gave rise to the P. uvella clade, which may represent a single species. It may have arisen recently, as indicated by the limited genetic divergence within the clone. Nevertheless, it spread widely across both continents from which samples have been taken. The example of P. uvella shows that a heterotrophic clade can be very successful for at least several million years. This conclusion assumes that the collection sites in Table 1 are correct, but some may be erroneous due to contamination or other mistakes in the long maintenance of the strains in culture. The second Polytoma clade is represented by a single strain (62-27). It might have arisen more recently and had less time to spread, or it might be the remnant of an ancient heterotrophic lineage that is dying out. Of course we cannot rule out the possibility that it represents a clade that is successful in nature but is rarely collected by Pringsheim's methods or does not thrive in stock collections.

These observations are compatible with the hypothesis that heterotrophic mutants rarely give rise to heterotrophic species, presumably because the mutants are at a selective disadvantage relative to their photosynthetic cohorts. The data are also compatible with the hypothesis that heterotrophic species are at a disadvantage relative to photosynthetic species. This might be manifested as a higher extinction rate, lower speciation rate, or both. This disadvantage could be expected on the basis that the heterotrophs are more limited in the range of environments they can occupy than are their photosynthetic congeners, which are also facultative heterotrophs. They may also be at a disadvantage in environments rich in organic nutrients where both can grow. Normal photosynthetic strains of both Chlamydomonas reinhardti and C. humicola show higher growth rates with mixotrophic growth, where they have both sunlight and acetate as carbon and energy sources, than with strictly heterotrophic growth (on acetate in the dark) (Boynton et al. 1972, Laliberte and de la Noue 1993). Similarly, photosynthetic C. reinhardtii grow faster under mixotrophic conditions than do nonphotosynthetic mutants that are obligately heterotrophic. Only in the dark do some mutants grow at nearly the same rate as the wild type (Boynton et al. 1972). A higher extinction rate and/or lower speciation rate of nonphotosynthetic lineages is an attractive explanation for the failure of nonphotosynthetic species of Chlamydomonadaceae to replace the photosynthetic forms.

If heterotrophic mutants frequently gave rise to heterotrophic species, we would expect to see many independent origins of Polytoma, not just two. And if there were no species-level selection favoring photosynthetic over heterotrophic species, we would expect to see more heterotrophic lineages branching from deep within the inclusive Chlamydomonas clade, some of which might show substantial divergence within the lineage. However, conclusive evidence for or against selection at the individual and species level would require the phylogenetic analysis of a much larger sample of sequences, preferably from a random sample of recent isolates of photosynthetic and nonphotosynthetic species, preferably coupled with breeding analysis to identify biological species and competition experiments under as nearly natural conditions as possible.

The earliest possible origin of the P. uvella heterotrophic clade, about 50 mya, falls close to the time of the KT (Cretaceous–Tertiary) boundary. The time of the KT boundary coincides with the time of mass extinctions, probably brought about by an extraterrestrial impact (Alvarez et al. 1980).
The large quantities of dust placed in the atmosphere by such an event might significantly reduce the available sunlight. It is interesting to speculate that a nonphotosynthetic mutant might be least disadvantaged relative to its green relatives and most likely to become established as an independent lineage, under these conditions. Genes from more strain of _P. weda_ would have to be sequenced in order to test this hypothesis by accurately dating the origin of this nonphotosynthetic lineage.

In addition to illuminating the evolutionary history of nonphotosynthetic algae, our data provide the necessary historical background for studies of the effects of the loss of photosynthesis on the molecular evolution of plastid genes. These studies show that the sequence evolution of several plastid genes is accelerated in _P. weda_ but not in _Polystoma_ strain 62-27 (Dawne Vernon, unpubl.). Accelerated evolutionary rates also have been reported for plastid rRNA and ribosomal protein genes in the nonphotosynthetic angiosperm _Epifagus_ (Wolfe et al. 1992a). The lack of detectable acceleration in strain 62-27 suggests that photosynthesis may have been lost more recently in this lineage than in the _P. weda_ lineage.

We thank Mark A. Buchheim for providing us with his rRNA sequence data; Paul Fuerst and his collaborators for primers; and Mark A. Buchheim, Russell L. Chapman, Nicholas W. Gillham, and Jeffrey A. Palmer for helpful suggestions with the manuscript. This research was supported by NSF research grant BSR-9107069 and an REU supplement. Authors may be contacted by e-mail as follows: C. W. Birky, Jr.: birky.2@osu.edu; R. Rumpf: rumpf1@osu.edu; D. Schreiber: fswa/sr@schreiber/ ou=irsi2105a@msahs.atmail.com; and D. Vernon: vernon.16@osu.edu.


17
CHAPTER 3

ALLELIC SEQUENCE DIVERGENCE, REPRODUCTION, AND PLOIDY IN ACANTHAMOEBA

Introduction

Acanthamoeba is a genus of ubiquitous free-living amoebae which are the causative agent of Acanthamoeba keratitis, a potentially vision-threatening invasion of the cornea, and granulomatous amebic encephalitis, a potentially fatal invasion of the central nervous system (Ma et al. 1990). It is believed that Acanthamoeba reproduces exclusively asexually, primarily because there is no convincing evidence for genetic recombination (Akins, 1981); however, the observation of identical mitochondrial 16S genes in association with differing nuclear 18S genes suggests that reproduction in Acanthamoeba castellanii may not be entirely clonal (Ledee 1995). It is possible that Acanthamoeba has rare or cryptic sex; for example, exchange of genetic information between individuals may occur only every hundred or every thousand generations, or in a manner which we are unable to detect. The method of reproduction affects the etiology of the diseases caused by an infectious parasite; an asexual organism is an excellent rapid colonizer, whereas sexual reproduction of the infectious agent facilitates the quick and efficient dissemination of drug resistance genes throughout the population. Consequently it is important to
know the mode of reproduction for the prognosis, treatment, and epidemiology of the disease.

Another unanswered question in *Acanthamoeba* genetics is that of ploidy. Measurement of nuclear DNA in *A. castellanii* Neff suggests a total nuclear DNA content at log phase of $10^9$ bp (King, 1980). Kinetic complexity measurements suggest a haploid genome size of $4-5 \times 10^7$ bp (Bohnert & Heermann, 1974), whereas pulse field electrophoresis has suggested that the genome size is on the order of $2.3-3.5 \times 10^7$ bp (Rimm et al., 1988). Based on these data, the ploidy may range anywhere from 10 to 22.

In order to address the questions of sexuality and ploidy in *Acanthamoeba*, I used a technique pioneered by Dr. Matthew Meselson and theory developed by Birky (1996) that uses allelic sequence divergence (ASD) to identify ancient asexual lineages. Allelic sequence divergence refers to the sequence differences among alleles of a given gene within an individual. In organisms that reproduce sexually, the alleles of a gene are shared among the members of the species and random drift eliminates most mutations one at a time. In an asexual species, however, the different copies of a gene within a lineage will accumulate different mutations from the moment that sexual reproduction is lost; thus in an asexual lineage the amount of ASD can be much higher than that found in a sexual species. This has been seen in two species of the asexual bdelloid rotifers, which have ASD values ranging from 10% to 45%, with an average of 29%, in 4 genes from 3 species (Meselson, personal communication). In contrast, in sexual species ASD ranges from 0% to 5.7% with an average of less than 1% (Avise, 1994; Moriyama and Powell, 1996).
Another interesting result of the loss of sexual reproduction is that the structure of phylogenetic trees can differ depending on which specific alleles are used to represent each taxonomic unit. In particular, two alleles in one individual or clone may differ more from each other than either does to an allele in another strain or species (see Figure 3); this was also seen in the bdelloid rotifers (Meselson, unpublished). By constructing a phylogenetic tree of allele sequences between two or more strains of an organism, it is possible to test whether or not the species are asexual by analyzing the resulting tree topology.

Allelic sequence divergence can be reduced or eliminated by any of several phenomena, collectively called allelic sequence convergence; these are (1) mitotic crossing-over and gene conversion; (2) a ploidy cycle; or (3) rare sex (in which case the description of the species as asexual is a matter of semantics) (Birky, 1996).

If two homologous chromosomes have a crossover during G2, 50% of the time the loci distal to the point of crossover will be homozygous after cell division. If two homologs undergo mitotic gene conversion during G1, the region involved in the conversion will become homozygous unless the event is reciprocal; if it occurs during G2 half of the progeny will be homozygous. In any of these cases ASD will be reduced in segments of chromosomes. Only small segments on the order of a few kilobases will be involved in a single gene conversion event, while mitotic recombination can affect segments as large as an entire chromosome arm. Mechanisms by which the ploidy of an organism is decreased and subsequently restored (ploidy cycle) have been described (Kondrashov 1994). A ploidy cycle will reduce ASD in each chromosome that is involved; for example if ploidy is reduced to one, ASD will be reduced to zero and will begin to accumulate again after the ploidy is increased. If the ploidy
Figure 3. Diagramatic representation of the consequences of the loss of sexual reproduction.

(a) A phylogenetic tree representing a sexual lineage which has speciated (left) and a lineage which has lost the ability to reproduce sexually (right). The alleles of the gene of interest begin to diverge following the loss of sex, and the lineage speciates; alleles 1 and 3 are in one species; alleles 2 and 4 in the second. By tracing backwards to the coalescent of the individual alleles, it can be seen that alleles within either species are less closely related to one another than they are to alleles in the other species.

(b) Phylogenetic tree from (a) showing allelic relationships in the asexual lineage.
were cycled via aneuploid intermediates, convergence would affect all loci on a chromosome to the same extent, and different chromosomes would have different levels of ASD.

If sequence convergence is frequent, measurement of sequence divergence could lead to an asexual species being mistaken for a sexually reproducing one. This effect can be partially countered by sequencing the alleles of several genes; only a ploidy cycle affecting all chromosomes at the same time will reduce the ASD to the same extent in all genes in the organism.

Selecting a target gene is crucial for the measurement of ASD. Detrimental mutations will accumulate more slowly than neutral mutations and reduce the observed value of ASD. This problem can be largely avoided by using a protein coding gene and concentrating on divergence at the third codon position, where most mutations are neutral or nearly so. Additionally, the gene must be unique and not a member of a multi-gene family; otherwise it is possible that an allele of a related gene might be mistakenly identified as an allele of the gene of interest. This would inflate the resulting ASD values and could lead to an unfounded conclusion of asexuality in the species in question.

We selected the TATA-binding protein (TBP) gene because it is a protein coding gene for which the sequence of an allele from *Acanthamoeba castellanii* was available from GenBank, and Southern hybridization indicates that TBP is a single-copy gene in *Acanthamoeba castellanii* Neff (Wong et al. 1992).

**Materials and Methods**

**Strains and DNA Acquisition**

Purified whole-cell DNA from *Acanthamoeba castellanii* strain Neff and *Acanthamoeba griffini* strain S7 was obtained from Jill Schreoder-Diedrich in the lab of Dr. Tom Byers at The Ohio State University.
PCR Amplification

Non-degenerate primers for the 5' and 3' ends of the GenBank A. castellanii TBP sequence were designated ACTFIID5' (sequence: gga tcc cat ctt cga cca gc) and ACTFIID3' (sequence: aac tga tca gct gat gat gt) respectively. Initial PCR amplification using nondegenerate primers resulted in clones which, when sequenced, were identical. To ensure that all possible alleles of TBP were amplified, including those which might have diverged significantly from the GenBank TBP sequence of Acanthamoeba castellanii Neff, degenerate primers were used for all subsequent PCR amplifications. Degenerate primers were designed using an alignment of TBP genes from the organisms listed in Table 4. Conserved regions of the gene were chosen for primer locations (Figure 4; see also Figure 5). Degenerate primers for the 5' and 3' ends of the amplified region of TBP were designated dACTFIID5' (sequence: cac caa gca ccc xtc xgg xa) and dACTFIID3' (sequence: xcg xac xtt xgc xcc c) respectively. PCR amplification used one μl of DNA (~1 μg/μl) in a PCR reaction mix containing 50 pmol of each primer, 1.25 mM concentrations of dATP, dCTP, dGTP, and dTTP, 10 μl 10X TAQ buffer containing 15mM magnesium (Perkin-Elmer Cetus), 10 units TAQ polymerase (Life Technologies), and 10 μl triple-distilled water. Amplifications involving degenerate primers differed only in using 1000 pmol of primer. Amplifications used a cycle of 95°C for 4 minutes, 55°C for 1 minute, and 72°C for 1 minute for the first cycle, and 95°C for 1 minute, 55°C for 1 minute, and 72°C for 1 minute for an additional 29 cycles. The degenerate amplification yielded a smear in all lanes. The amplification products of approximately 700 to 800 bp in size were band-isolated, purified by the QIAquick PCR purification protocol (Qiagen), and used as the template for a second degenerate amplification,
<table>
<thead>
<tr>
<th>Organism</th>
<th>Genbank Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Saccharomyces cerevisiae</em></td>
<td>M27135</td>
</tr>
<tr>
<td><em>Schizosaccharomyces pombe</em></td>
<td>X53415</td>
</tr>
<tr>
<td><em>Dictyostelium discoideum</em></td>
<td>M64861</td>
</tr>
<tr>
<td><em>Triticum aestivum</em></td>
<td>X59874</td>
</tr>
<tr>
<td><em>Acetabularia cliftonii</em></td>
<td>Z28644</td>
</tr>
<tr>
<td><em>Acanthamoeba castellanii</em></td>
<td>M93340</td>
</tr>
</tbody>
</table>

Table 4. Organisms used in TBP alignment for primer design.
Figure 4. Alignment of TBP used in designing degenerate primers. DNA sequences were obtained from Genbank and aligned using ClustalW. Conserved regions indicated by the brackets at the 5' and 3' ends of the gene were used as a basis for designing degenerate primers.
Figure 5. Map of the *A. castellanii* TBP gene. Arrows above sequence map indicate amplification primer positions.
which routinely produced a distinct band of the proper size. Degenerate amplification products from *A. castellanii* were 681 bases in length, as expected based on the GenBank sequence; degenerate amplification products from *A. griffini* were 795 bases in length. The size difference between the two species is the result of additional bases in the introns in *A. griffini.*

**Cloning of PCR Products**

The TBP amplification products were either cloned into pUC18 (Life Technologies) or cloned by the TA cloning technique (Invitrogen). These clones were sequenced either by direct double stranded cycle sequencing (Life Technologies) in our laboratory or by automated sequencing via an Applied Biosystems Model 373A version 1.2.1 automated sequencer at the Molecular Genetics Sequencing Facility at the University of Georgia. All clones were sequenced on both strands; sequences will be deposited into GenBank. Table 5 gives the amplification method, vector, and sequencing technique used for each clone.

**Southern Blotting**

Whole cell DNA was digested overnight with *Bst*EII, *Xba*I, *Kpn*I, and *Hind*III (all enzymes from Life Technologies), run out on an a 0.7% agarose gel, and transferred to a nitrocellulose membrane for Southern analysis (Maniatis, 1989). The membrane was prehybridized at 65° C for 1 hour with a prehybridization solution (30 ml 20×SSC, 5 ml 100× Denhardts, 5 ml 10% SDS, and 1 ml of 10 mg/ml salmon sperm DNA, boiled for 10 minutes prior to use). The membrane was then hybridized at 68° C for 12 hours after adding the probe solution consisting of 150ul water, full-length TBP amplification fragment from *A. castellanii* Neff labelled with Redi-Prime (Amersham), and 0.5 ml salmon sperm DNA, boiled for 10 minutes prior to use. After hybridization the
<table>
<thead>
<tr>
<th>Clone #</th>
<th>Species</th>
<th>Cloned By:</th>
<th>Sequenced By:</th>
<th>Vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBP1</td>
<td>A. castellanii Neff</td>
<td>Standard</td>
<td>Manual</td>
<td>pUC18</td>
</tr>
<tr>
<td>TBP2</td>
<td>A. castellanii Neff</td>
<td>Standard</td>
<td>Manual</td>
<td>pUC18</td>
</tr>
<tr>
<td>TBP3</td>
<td>A. castellanii Neff</td>
<td>Standard</td>
<td>Manual</td>
<td>pUC18</td>
</tr>
<tr>
<td>TBP4</td>
<td>A. castellanii Neff</td>
<td>Standard</td>
<td>Manual</td>
<td>pUC18</td>
</tr>
<tr>
<td>TBP5</td>
<td>A. castellanii Neff</td>
<td>Standard</td>
<td>Manual</td>
<td>pUC18</td>
</tr>
<tr>
<td>TBP6</td>
<td>A. castellanii Neff</td>
<td>Standard</td>
<td>Manual</td>
<td>pUC18</td>
</tr>
<tr>
<td>dTBP1</td>
<td>A. castellanii Neff</td>
<td>CloneAmp</td>
<td>Automated</td>
<td>pAMP10</td>
</tr>
<tr>
<td>dTBP2</td>
<td>A. castellanii Neff</td>
<td>CloneAmp</td>
<td>Automated</td>
<td>pAMP10</td>
</tr>
<tr>
<td>dTBP3</td>
<td>A. castellanii Neff</td>
<td>CloneAmp</td>
<td>Automated</td>
<td>pAMP10</td>
</tr>
<tr>
<td>dTBP4</td>
<td>A. castellanii Neff</td>
<td>CloneAmp</td>
<td>Automated</td>
<td>pAMP10</td>
</tr>
<tr>
<td>dTBP5</td>
<td>A. castellanii Neff</td>
<td>CloneAmp</td>
<td>Automated</td>
<td>pAMP10</td>
</tr>
<tr>
<td>dTBP6</td>
<td>A. castellanii Neff</td>
<td>CloneAmp</td>
<td>Automated</td>
<td>pAMP10</td>
</tr>
<tr>
<td>dTBP7</td>
<td>A. castellanii Neff</td>
<td>CloneAmp</td>
<td>Automated</td>
<td>pAMP10</td>
</tr>
<tr>
<td>dTBP8</td>
<td>A. castellanii Neff</td>
<td>CloneAmp</td>
<td>Automated</td>
<td>pAMP10</td>
</tr>
<tr>
<td>ACgTBP1</td>
<td>A. griffini S7</td>
<td>TA</td>
<td>Automated</td>
<td>pCR1 2.1</td>
</tr>
<tr>
<td>ACgTBP2</td>
<td>A. griffini S7</td>
<td>TA</td>
<td>Automated</td>
<td>pCR1 2.1</td>
</tr>
<tr>
<td>ACgTBP3</td>
<td>A. griffini S7</td>
<td>TA</td>
<td>Automated</td>
<td>pCR1 2.1</td>
</tr>
<tr>
<td>ACgTBP4</td>
<td>A. griffini S7</td>
<td>TA</td>
<td>Automated</td>
<td>pCR1 2.1</td>
</tr>
<tr>
<td>ACgTBP5</td>
<td>A. griffini S7</td>
<td>TA</td>
<td>Automated</td>
<td>pCR1 2.1</td>
</tr>
<tr>
<td>ACgTBP6</td>
<td>A. griffini S7</td>
<td>TA</td>
<td>Automated</td>
<td>pCR1 2.1</td>
</tr>
<tr>
<td>ACgTBP7</td>
<td>A. griffini S7</td>
<td>TA</td>
<td>Automated</td>
<td>pCR1 2.1</td>
</tr>
<tr>
<td>ACgTBP8</td>
<td>A. griffini S7</td>
<td>TA</td>
<td>Automated</td>
<td>pCR1 2.1</td>
</tr>
<tr>
<td>ACgTBP9</td>
<td>A. griffini S7</td>
<td>TA</td>
<td>Automated</td>
<td>pCR1 2.1</td>
</tr>
<tr>
<td>ACgTBP10</td>
<td>A. griffini S7</td>
<td>TA</td>
<td>Automated</td>
<td>pCR1 2.1</td>
</tr>
<tr>
<td>ACgTBP11</td>
<td>A. griffini S7</td>
<td>TA</td>
<td>Automated</td>
<td>pCR1 2.1</td>
</tr>
</tbody>
</table>

Table 5. Cloning methods and vectors used in cloning TBP from A. castellanii and A. griffini. Standard cloning was performed as in Sambrook (1989). CloneAmp cloning indicates Life Technologies CloneAmp System. TA cloning was done with the Invitrogen TA PCR Cloning System. pAMP and pCR1 2.1 were obtained from Life Technologies and Invitrogen respectively.
membrane was washed as per Sambrook (1989) and visualized via phoshoimager.

**Sequence Alignment and Analysis**

Sequences were aligned using ClustalW (Higgins et al. 1992) on an Iris Indigo computer and SeqApp (Gilbert 1992) on a Power Macintosh. Of the 795 bases sequenced, 472 were in exons and could be unambiguously aligned; the remainder were in introns and, although highly conserved in the majority of clones, could not be unambiguously aligned in all clones. Pairwise sequence differences were determined from the alignment using the DNADIST algorithm of Phylip 3.5 (Felsenstein 1989, 1993) and corrected for multiple hits by the Nei 2-parameter method (Tables 6 and 7). These distances were used to construct a Neighbor-Joining distance tree (Figure 6) with the DNADIST algorithm of Phylip 3.5.

**Results & Discussion**

Southern blotting was performed on *A. castellanii* and *A. griffinii* to verify that the TBP gene is unique and not a member of a multi-gene family. The Southern blot, probed at low stringency with the TBP product amplified from *A. castellanii*, showed that TBP homology was limited to a single restriction fragment for most enzymes in both species. The fact that the TBP probe from *A. castellanii* hybridized to *A. griffini* showed that the hybridization conditions were of sufficiently low stringency to detect sequences differing by as much as 25%. We conclude that the TBP gene is unique in the *A. castellanii* and *A. griffinii* genomes (Figure 7), in agreement with the results of Wong et al. (1992) for *A. castellanii*. Consequently the different clones of TBP we identified probably
Table 6. Distance Matrix of *Acanthamoeba* TBP clones. Values above the diagonal represent distances for the entire cloned region, corrected for multiple hits via the Jukes-Cantor method. Values below the diagonal represent the absolute number of substitutions found.
Table 7. Distance Matrix of *Acanthamoeba* TBP clones. Values above the diagonal represent distances for only exons within the cloned region, corrected for multiple hits via the Jukes-Cantor method. Values below the diagonal represent the absolute number of substitutions found.
Figure 6. Neighbor-joining tree of *Acanthamoeba* TBP alleles. Tree was constructed using the DNABoot (100 bootstrap replicates), DNADist (Nei 2-parameter correction for multiple hits), and Neighbor algorithms of Phylip 3.5p. Bootstrap values at nodes represent the proportion of trees supporting that particular node topology. Bootstrap values of 100 strongly support the two clades, one containing the *A. castellanii* clones (TBP 1-6 and dTBP 1-7) and another containing the *A. griffini* clones (ACgTBP 1-11).
Figure 7. Southern blot showing copy-number of TBP genes in *A. castellanii* and *A. griffini* genomes. Genomic DNA from both species was digested with *BstII, HindIII, KpnI,* and *XbaI.* Digested DNA was run out on a 0.7% agarose gel and probed with the TBP PCR product. Unlabelled probe was used as a control. Presence of a single band in three of the four *A. castellanii* lanes and all four of the *A. griffini* lanes confirms that TBP is single-copy in both species.
occupy the same locus on different chromosomes, and thus represent true alleles.

It is of interest that when degenerate amplification was used to derive cloning inserts, no duplicate clones (as determined by sequence) were obtained. This suggests that the population of different clones has not been exhausted and that additional alleles exist.

The sequence divergence data for *Acanthamoeba castellanii* show that the clones are extremely similar to one another. The minimum divergence was 0 substitutions out of 455 bases in exons; the maximum divergence was 10. This translates to a maximum divergence of approximately 0.0219 or 2.2%. When the introns are included, the divergence reaches a maximum of 15 substitutions out of 681 bases, which is equal to 0.022 or 2.2% divergence. This is well within the range of 0% - 5.7% diversity found between alleles of a sexually reproducing species (Avise 1994; Moriyama & Powell 1996), and far short of the 10-45% divergence Meselson and Welch found in the asexual bdelloid rotifers. Of the 9 polymorphic sites within *A. castellani*i's TBP coding region, 5 were in the first or second codon position, and 4 were in the third codon position.

The sequence divergence data for *A. griffinii* shows similar results; the minimum divergence was 1 substitution out of 455 coding bases; the maximum divergence was 13. This translates to a maximum distance of approximately 0.0285 or 2.9% divergence. When the introns are included, the divergence reaches a maximum of 55 substitutions out of 795 bases, which is equal to 0.0691 or 6.9% divergence. Again, this is well within the range expected for sexually reproducing species. Of the 49 polymorphic sites within *A. griffinii*i's TBP coding region, 6 were in the first or second codon position, and 43 were in
the third codon position. Table 6 shows the distance calculations and raw
substitution numbers for all of the cloned TBP alleles as well as the GenBank A.
castellanii TBP sequence.

Nucleotide diversity (\(\pi\)) was also calculated for each set of clones. \(\pi\), a
measure of polymorphism in a population, is the average proportion of
nucleotide differences per site between any two randomly chosen sequences
and is calculated using the formula

\[
\pi = \sum_{ij} x_i x_j \pi_{ij}
\]

where \(x_i\) and \(x_j\) are the frequencies of the \(i\)th and \(j\)th type of DNA sequences
and \(\pi_{ij}\) is the proportion of different nucleotides between these sequences (Li
and Grauer 1991). \(\pi\) is usually calculated by sampling alleles from a population
rather than from an individual as in our case; if we assume that Acanthamoeba
is reproducing sexually under conditions of random mating, then the alleles
within an individual will be as random as if we had taken them from multiple
individuals in a population. We can then compare our estimates of \(\pi\) to known
values of \(\pi\) for sexual species to test whether or not Acanthamoeba could be
truly sexual. In cases of identical clone sequences (i.e. A. castellanii TBP 2-6),
where it was unclear whether we had cloned separate alleles that happened to
be identical in sequence, or had cloned the same allele multiple times, only one
of the duplicate sequences was used in determining \(\pi\). This could lead to a
slight overestimation of the true value of \(\pi\) in the population if the duplicates are
in fact separate alleles. The clones from A. castellanii have a \(\pi\) of 2.3 x 10^{-5},
and the clones from A. griffini have a \(\pi\) of 9.1 x 10^{-5}. These values are lower
than the maximum \(\pi\) of 9.7 x 10^{-3} seen for several species of Drosophila

37
(Moriyama and Powell 1996) and the average $\pi$ of $4.4 \times 10^{-4}$ seen for humans (Li and Sadler 1991). Since $\pi$ can be equated with ASD in a randomly mating population, this suggests that *Acanthamoeba* may be reproducing sexually at least part of the time.

An alignment made using the cloned TBP alleles from *Acanthamoeba castellanii*, *Acanthamoeba griffini*, the GenBank *A. castellanii* TBP sequence, and the *S. cerevisiae* TBP sequence is shown in Figure 8. This alignment was used to construct the neighbor-joining (distance) tree shown in Figure 6. This tree clearly separates the *A. castellanii* clones and the *A. griffini* clones into separate clades. Thus there is no indication that any allele in one species is more closely related to any allele in the other species, as would be expected if *Acanthamoeba* has been reproducing strictly asexually since before the divergence of these two species and if their convergence is rare. We conclude that either *Acanthamoeba* reproduces sexually occasionally or it is undergoing frequent sequence convergence.

The TBP alleles in *A. griffini* fall into two clades, as shown on the tree: one containing alleles 2, 4, and 9; and the other containing the rest of the alleles. Within these clades the divergences are comparable to those seen in *A. castellanii*. The branches leading to the genes in *A. griffini* are longer than those leading to the *A. castellanii* genes, suggesting that the rate of evolution may be greater in the lineage leading to *A. griffini*. An increase in the rate of base pair substitution between two lineages must be due to a higher mutation rate, or a higher probability of fixation (decreased selection), or both. If mutation rates in these two species are similar, other factors might contribute to the observed difference in divergence. The effective population size of the two strains may vary; since this affects fixation of alleles, it can contribute to
Figure 8. Alignment of TBP Clones from A. castellanii (AcTBP1-6 and dTBPI-7) and A. griffini (ACgTBPI-11). The Genbank sequence (AcTBPgb) is included as a reference.
Figure 8. (continued)
Figure 8. (continued)
Figure 8. (continued)
### Figure 8. (continued)

<table>
<thead>
<tr>
<th>405</th>
<th>415</th>
<th>425</th>
<th>435</th>
<th>445</th>
</tr>
</thead>
<tbody>
<tr>
<td>GGCCTGGC</td>
<td>AGTGCGGTT</td>
<td>GCCCTGCT</td>
<td>ACGGAGG</td>
<td>TCBGTTTC</td>
</tr>
<tr>
<td>TCAGCACAC</td>
<td>GGCCTGG</td>
<td>GCCCTGCT</td>
<td>ACGGAGG</td>
<td>TCBGTTTC</td>
</tr>
<tr>
<td>TCGTACCA</td>
<td>TCGTACCA</td>
<td>TCGTACCA</td>
<td>TCGTACCA</td>
<td>TCGTACCA</td>
</tr>
<tr>
<td>TCGTACCA</td>
<td>TCGTACCA</td>
<td>TCGTACCA</td>
<td>TCGTACCA</td>
<td>TCGTACCA</td>
</tr>
<tr>
<td>TCGTACCA</td>
<td>TCGTACCA</td>
<td>TCGTACCA</td>
<td>TCGTACCA</td>
<td>TCGTACCA</td>
</tr>
<tr>
<td>AGTACCTGC</td>
<td>AGTACCTGC</td>
<td>AGTACCTGC</td>
<td>AGTACCTGC</td>
<td>AGTACCTGC</td>
</tr>
<tr>
<td>AGTACCTGC</td>
<td>AGTACCTGC</td>
<td>AGTACCTGC</td>
<td>AGTACCTGC</td>
<td>AGTACCTGC</td>
</tr>
<tr>
<td>AGTACCTGC</td>
<td>AGTACCTGC</td>
<td>AGTACCTGC</td>
<td>AGTACCTGC</td>
<td>AGTACCTGC</td>
</tr>
<tr>
<td>AGTACCTGC</td>
<td>AGTACCTGC</td>
<td>AGTACCTGC</td>
<td>AGTACCTGC</td>
<td>AGTACCTGC</td>
</tr>
<tr>
<td>AGTACCTGC</td>
<td>AGTACCTGC</td>
<td>AGTACCTGC</td>
<td>AGTACCTGC</td>
<td>AGTACCTGC</td>
</tr>
<tr>
<td>CAGACATCG</td>
<td>CAGACATCG</td>
<td>CAGACATCG</td>
<td>CAGACATCG</td>
<td>CAGACATCG</td>
</tr>
<tr>
<td>CAGACATCG</td>
<td>CAGACATCG</td>
<td>CAGACATCG</td>
<td>CAGACATCG</td>
<td>CAGACATCG</td>
</tr>
<tr>
<td>CAGACATCG</td>
<td>CAGACATCG</td>
<td>CAGACATCG</td>
<td>CAGACATCG</td>
<td>CAGACATCG</td>
</tr>
<tr>
<td>CAGACATCG</td>
<td>CAGACATCG</td>
<td>CAGACATCG</td>
<td>CAGACATCG</td>
<td>CAGACATCG</td>
</tr>
<tr>
<td>CAGACATCG</td>
<td>CAGACATCG</td>
<td>CAGACATCG</td>
<td>CAGACATCG</td>
<td>CAGACATCG</td>
</tr>
<tr>
<td>CACGGAGA</td>
<td>CACGGAGA</td>
<td>CACGGAGA</td>
<td>CACGGAGA</td>
<td>CACGGAGA</td>
</tr>
<tr>
<td>CACGGAGA</td>
<td>CACGGAGA</td>
<td>CACGGAGA</td>
<td>CACGGAGA</td>
<td>CACGGAGA</td>
</tr>
<tr>
<td>CACGGAGA</td>
<td>CACGGAGA</td>
<td>CACGGAGA</td>
<td>CACGGAGA</td>
<td>CACGGAGA</td>
</tr>
<tr>
<td>CACGGAGA</td>
<td>CACGGAGA</td>
<td>CACGGAGA</td>
<td>CACGGAGA</td>
<td>CACGGAGA</td>
</tr>
<tr>
<td>CACGGAGA</td>
<td>CACGGAGA</td>
<td>CACGGAGA</td>
<td>CACGGAGA</td>
<td>CACGGAGA</td>
</tr>
<tr>
<td>GCCAGTTG</td>
<td>GCCAGTTG</td>
<td>GCCAGTTG</td>
<td>GCCAGTTG</td>
<td>GCCAGTTG</td>
</tr>
<tr>
<td>GCCAGTTG</td>
<td>GCCAGTTG</td>
<td>GCCAGTTG</td>
<td>GCCAGTTG</td>
<td>GCCAGTTG</td>
</tr>
<tr>
<td>GCCAGTTG</td>
<td>GCCAGTTG</td>
<td>GCCAGTTG</td>
<td>GCCAGTTG</td>
<td>GCCAGTTG</td>
</tr>
<tr>
<td>GCCAGTTG</td>
<td>GCCAGTTG</td>
<td>GCCAGTTG</td>
<td>GCCAGTTG</td>
<td>GCCAGTTG</td>
</tr>
<tr>
<td>GCCAGTTG</td>
<td>GCCAGTTG</td>
<td>GCCAGTTG</td>
<td>GCCAGTTG</td>
<td>GCCAGTTG</td>
</tr>
<tr>
<td>TACGGCTCGG</td>
<td>TACGGCTCGG</td>
<td>TACGGCTCGG</td>
<td>TACGGCTCGG</td>
<td>TACGGCTCGG</td>
</tr>
<tr>
<td>TACGGCTCGG</td>
<td>TACGGCTCGG</td>
<td>TACGGCTCGG</td>
<td>TACGGCTCGG</td>
<td>TACGGCTCGG</td>
</tr>
<tr>
<td>TACGGCTCGG</td>
<td>TACGGCTCGG</td>
<td>TACGGCTCGG</td>
<td>TACGGCTCGG</td>
<td>TACGGCTCGG</td>
</tr>
<tr>
<td>TACGGCTCGG</td>
<td>TACGGCTCGG</td>
<td>TACGGCTCGG</td>
<td>TACGGCTCGG</td>
<td>TACGGCTCGG</td>
</tr>
<tr>
<td>TACGGCTCGG</td>
<td>TACGGCTCGG</td>
<td>TACGGCTCGG</td>
<td>TACGGCTCGG</td>
<td>TACGGCTCGG</td>
</tr>
</tbody>
</table>

(Continued)
Figure 8. (continued)
Figure 8. (continued)
observed divergence of alleles. If *A. castellanii* is asexual, the time since the last convergence event (or the frequency of convergence) will affect the observed divergence. Alternately the time since the most recent sexual episode (or the frequency of sexual reproduction) may vary between the two strains, which would also contribute to the observed difference in divergence. Further study is required before this question can be conclusively answered.

The neighbor-joining tree also points out a peculiarity within the *A. griffini* clones. As with the *A. castellanii* clones, the majority of the *A. griffini* clones are similar in sequence to one another and have short branch lengths separating them from one another. *A. griffini* clones 2, 4, and 9, however, are not only set apart in a separate clade within the primary *A. griffini* clade, but have branch lengths and corresponding bootstrap values that distinctly separate them from one another and the other *A. griffini* clones. This suggests that *A. griffini* clones 2, 4, and 9 were not affected by a recent convergence event or events that affected the other 8 clones. Chromosomal rearrangements such as reciprocal translocations can render chromosomes immune to sequence convergence in several ways (Birky 1996):

1. Chromosomal rearrangements function as crossover suppressors in that they may prevent the necessary alignment of chromosomes from occurring.
2. Improper pairing of rearranged chromosomes during meiosis is likely to yield inviable progeny.
3. Chromosomal rearrangements may prevent ploidy reduction if the reduction causes the resulting cells to be homozygous for a deletion.
4. If a chromosome differs sufficiently from its homolog(s), then during sexual reproduction it will effectively function as a separate chromosome,
ensuring transmission of its particular allele to all of its offspring and eliminate sequence convergence by inbreeding.

The observation that the overwhelming majority of substitutions in the A. griffinii TBP clones occurs at synonymous sites suggests that the alleles are all under selection and functional. The number of polymorphic sites in the A. castellanii strain is too small to make any statistically significant conclusions regarding this, although the low divergence from the GenBank sequence suggests that all the A. castellanii alleles are probably functional. Additional evidence supporting the functionality of these genes can be derived from the probability of that no stop codons are introduced into the gene by substitution in the absence of selection. There are 35 sense codons within the A. castellanii clone dTBP5 which can be changed to a stop codon (a nonsense mutation) by the substitution of a single base (i.e. UCG to UAG), and 6 which can be converted to a stop by the substitution of one of two bases (i.e. UAC to UAA or UAG). The expected number of nonsense (stop) mutations can be determined by the formula:

\[ \frac{S_1 \mu}{3} + 2 \frac{S_2 \mu}{3} \]

where \( S_1 \) is the number of codons which could be made into a stop codon by one substitution, \( S_2 \) is the number of codons which could be made into a stop codon by one of two substitutions, and \( \mu \) is the expected number of synonymous substitutions per site in the allele in question if it became a pseudogene when it diverged. A minimum estimate of \( \mu \) can be obtained from the divergence between clone dTBP5 and A. griffinii clone AcgTBP7. In third codon positions, which are mostly synonymous, these two clones diverge by 0.216%. From this
it can be calculated that 3.38 nonsense mutations should have occurred in this clone. The probability that no nonsense mutations would occur is given by applying the first term of the Poisson distribution to the formula for the expected number of nonsense mutations:

\[ e^{-S_1(\mu)/3} + 2S_2(\mu)/3 \]

which in this case gives us 0.034 or 3.4%. The absence of stop codons in this clone is significant and strengthens the argument that these alleles are functional.

Since TBP is a single-copy gene in *Acanthamoeba*, we can use these data to determine the minimum ploidy number of *Acanthamoeba*. *A. castellanii* had 10 unique alleles, whereas *A. griffini* had 11 alleles. This suggests that the minimum ploidy for *A. castellanii* is 10, and the minimum ploidy for *A. griffini* is 11. These estimates are in line with other estimates of ploidy in *Acanthamoeba* (Byers 1986).

Conclusions

Allelic sequence divergences among TBP alleles in *A. castellanii* and *A. griffini* are well within the range for alleles in a sexual species. My data suggest that either *A. castellanii* is sexual, or it is undergoing frequent sequence convergence. The fact that sexual reproduction has never been observed in *Acanthamoeba* in spite of years of laboratory study suggests that the species is having sex infrequently if at all. Based on the number of unique alleles of TBP
cloned, the minimum ploidy of *A. castellanii* must be 10, and the minimum ploidy of *A. griffini* must be 11.
BIBLIOGRAPHY


Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package) version 3.5p. Distributed by the author, Department of Genetics, University of Washington.


King, L. E. 1980. Changes in nuclear and mitochondrial DNA levels and
metabolism during ageing and encystment in Acanthamoeba castellanii.
Ph.D. dissertation. Columbus: The Ohio State University.

Kondrashov, A. S., 1994. Mutation load under vegetative reproduction and

Chlamydomonas humicola (Chlorophyceae) on acetate. J. Phycol. 29:612-
20.

J. Protozool. 10:333-339.


LeDee, D. R. 1995. Interspecific mitochondrial rRNA and tRNA gene variation
in Acanthamoeba: new insights into phylogeny, taxonomy, RNA editing and

Associations, Inc. Sunderland, Massachusetts.

129:513-523.


Ma, P., Visvesvara, G. S., Martinez, A. J., Theodore, F. H., Daggett, P. M., and

Maidak, B. L., Larsen, N., McCaughey, M. J., Overbeek, r., Olsen, G. J., Fogel, K.,
Acids Res. 22:3485-3487.


54


