Curr Genet (1997) 31: 214–227

ORIGINAL PAPER

Michael Liss · David L. Kirk · Kurt Beyser · Stefan Fabry

Intron sequences provide a tool for high-resolution phylogenetic analysis of volvocine algae

Received: 7 October / 25 November 1996

Abstract Three nuclear spliceosomal introns in conserved locations were amplified and sequenced from 28 strains representing 14 species and 4 genera of volvocalean green algae. Data derived from the three different introns yielded congruent results in nearly all cases. In pairwise comparisons, a spectrum of taxon-specific sequence differences ranging from complete identity to no significant similarity was observed, with the most distantly related organisms lacking any conserved elements apart from exon-intron boundaries and a pyrimidine-rich stretch near the 3′ splice site. A metric (SI_{50}), providing a measure of the degree of similarity of any pair of intron sequences, was defined and used to calculate phylogenetic distances between organisms whose introns displayed statistically significant similarities. The rate of sequences divergence in the introns was great enough to provide useful information about relationships among different geographical isolates of a single species, but in most cases was too great to provide reliable guides to relationships above the species level. A substitution rate of approximately 3×10^{-8} per intron position per year was estimated, which is about 150-fold higher than in nuclear genes encoding rRNA and about 10-fold higher than the synonymous substitution rate in protein-coding regions. Thus, these homologous introns not only provide useful information about intraspecific phylogenetic relationships, but also illustrate the concept that different parts of a gene may be subject to extremely different intensities of selection. The intron data generated here (1) reliably resolve for the first time the relationships among the five most extensively studied strains of *Volvox*, (2) reveal that two other *Volvox* species may be more closely related than had previously been suspected, (3) confirm prior evidence that particular isolates of *Eudorina elegans* and *Pleodorina illinoisensis* appear to be sibling taxa, and (4) contribute to the resolution of several hitherto unsettled issues in *Chlamydomonas* taxonomy.

Key words Evolution · Intron sequences · Molecular clock · *Chlamydomonas* · *Volvox* · Volvocales

M. Liss · K. Beyser · S. Fabry (✉)
Genetik, Universität Regensburg,
Universitätsstraße 31,
D-93040 Regensburg, Germany

D.L. Kirk
Department of Biology,
Washington University,
St. Louis MO 63130, USA

Communicated by J.-D. Rochaix

Introduction

Ever since the pioneering work of Zuckerkandl and Pauling (1965), it has been widely recognized that molecular data provide potentially powerful tools for analyzing relationships among organisms and reconstructing their phylogenetic histories. It is equally well recognized, however, that ''different types of molecular assay provide genetic information ideally suited to different subsets of (the evolutionary) hierarchy, and a continuing challenge is to develop and utilize molecular methods appropriate for a particular biological problem at hand'' (Avise 1994). Comparative analysis of the sequences of rRNAs (Wilson et al. 1987; Olson and Woese 1993), and highly conserved proteins (such as actin, tubulin and certain glycolytic enzymes; Doolittle 1995; Doolittle et al. 1996) have greatly improved our understanding of many relationships among and within major evolutionary lineages. rRNA molecules have proven to be particularly useful for a wide range of phylogenetic studies, because they contain regions with markedly different degrees of sequence conservation (Larson and Wilson 1989). However, even the most

variable regions of rRNA are seldom sufficiently informative to resolve relationships among closely related species or subspecies. At the other end of the spectrum, markers such as "minisatellite," or "microsatellite" DNAs, that can be used for identifying individuals or determining their parentage, are much too variable to be useful for establishing relationships among sibling populations or species (Avise 1994). Thus, resolution of the branching patterns at the level of subspecies and sibling species clearly requires the use of molecules with a rate of change intermediate between these two extremes.

Certain intergenic and non-coding intragenic regions of DNA have the potential to serve this purpose because they accumulate changes more rapidly than adjacent coding segments, but less rapidly than satellite DNAs. Intergenic regions that are flanked by well-conserved sequence elements can be readily amplified from very small samples by PCR, permitting rapid analysis of many specimens. Thus, the internal transcribed spacer of rDNA and the intergenic segments of chloroplast DNA have been shown to permit phylogenetic resolution down to the species level, at least for certain groups of organisms. Although mitochondrial DNA has also proven very powerful for high-resolution analysis of vertebrate phylogenies (Cann et al. 1986), its strikingly different properties in non-vertebrates (Avise 1994) has made extensions of this method to green organisms rare (e.g. Hiesel et al. 1994).

In principle, spliceosomal introns of nuclear genes also have the potential to provide high-resolution phylogenetic information, but to date this potential has not been widely exploited (e.g. Tominaga and Narise 1995). Among the reasons for this may be the perceived difficulty of establishing true homology between introns in the same locations in two or more organisms. In addition, Li and Graur (1991) calculated (from a very limited data set) that average intron substitution rates did not exceed those of pseudogenes or degenerate codon sites (approximately 3.3×10^{-9} substitutions per site per year), which suggested that there would be no advantage to using introns rather than pseudogenes or codon replacements for high-resolution phylogenetic analysis.

Recently, however, we uncovered evidence which suggested that this conclusion should be re-evaluated (Dietmaier and Fabry 1994). We found that in each related set of organisms studied, genes encoding the same isotype of small G proteins possessed introns at identical positions, which we took as an indication that the introns in each such taxonomic group were related by descent. In strong contrast to their conserved locations, however, no sequence similarities could be detected between homologous intron pairs in species as closely related as two green algae, *Chlamydomonas reinhardtii* and *Volvox carteri* in the order Volvocales, which are thought to have shared a common ancestor $\geq$ 75 million years ago (MYA) (Rausch et al. 1989). We concluded that these introns must have been diverging at a rate far in excess of the rate at which silent substitutions had been accumulated in protein-coding genes of these two species (Rausch et al. 1989; Schmitt et al. 1992) and might, therefore, be suitable for high-resolution phylogenetic analysis (Dietmaier and Fabry 1994).

The volvocine algae seemed an ideal group with which to test this idea, and it also seemed possible that the resulting intron-sequence data might contribute to resolving certain questions of interest regarding the phylogenetic history of *Volvox* and its relatives. Volvocine algae (green flagellates in the order Volvocales, including unicellular *Chlamydomonas* plus certain multicellular organisms in the family Volvocaceae) have frequently been used as an example of the way that multicellular organisms may have arisen from unicellular ancestors (Kochert 1973; Schmitt et al. 1992). However, although rRNA sequence comparisons (Larson et al. 1992) have indicated that members of the family Volvocaceae are closely related to one another and to *C. reinhardtii*, they do not support the simplistic notion that the path leading from *Chlamydomonas* to *Volvox* involved a simple linear progression in organismic size and complexity (such as *Chlamydomonas* $\rightarrow$ *Gonium* $\rightarrow$ *Pandorina* $\rightarrow$ *Eudorina* $\rightarrow$ *Pleodorina* $\rightarrow$ *Volvox*). Rather, they suggest a highly branched family tree in which similar transitions between different grades of organismic complexity may have occurred independently in several different lineages, leading to a polyphyletic genus *Volvox* (Larson et al. 1992; plus unpublished studies). Although useful for suggesting the general form of the volvocine family tree, such rRNA-based studies have raised many questions at lower taxonomic levels for which rRNA sequence data lack the resolving power. For example, one such study indicated that several other species of *Chlamydomonas* may be more closely related to *Volvox* and other volvocaceans than *C. reinhardtii* is, but left the branching pattern among six unicellular and multicellular taxa unresolved (Buchheim and Chapman 1991). Moreover, the need for additional methods of high-resolution analysis of this group has been underscored by the realization that certain strains now available in culture collections may have been misdiagnosed and/or mis-catalogued in the past (Harris 1989; Starr and Zeikus 1993), and that the genus *Chlamydomonas* is in serious need of revision (Ettl and Schlösser 1992).

In this study we present evidence that selected spliceosomal introns of nuclear genes provide tools for analysis of relationships at the sub-species to sibling-species level. In addition to deciphering some new species and strain relationships, we provide a rough estimate of the evolutionary time span over which the events that can be resolved in intron sequences may have occurred.

Materials and methods

Sources and cultivation of organisms. The species and strains used, their sites of origin, and culture-collection accession numbers (or relevant references) are given in Table 1. With the exception of certain *Chlamydomonas* isolates that were grown in TAP medium (Harris 1989), all cultures were maintained in standard *Volvox* medium (SVM, Starr 1969; Kirk and Kirk 1983) under a 16-h light–8-h dark regime at constant temperature (24°, 28° or 32°C, depending on the strain).

DNA extraction and amplification, and sequencing of introns. Total DNA was prepared in one of three ways. (1) In some cases existing DNA samples that had been extracted on a large scale and purified by CsCl centrifugation (Harper et al. 1987) were used. (2) In most cases, the following intermediate-scale, intermediate-purity cetyltrimethylammonium bromide (CTAB) extraction and precipitation procedure was employed (unless otherwise specified, all reagents and steps were at room temperature, buffers were at pH 8.0, and centrifugation was at 16 000 g in a microfuge): organisms from 300-ml cultures were collected (approximately 2 000 rpm in a tabletop centrifuge), vortexed repeatedly in the presence of 0.5-mm glass beads and a buffer containing 25 mM Tris, 5 mM EDTA and 5% sarcosyl, and subjected to five freeze-thaw cycles (at −20°C and 37°C). After separation from the glass beads, the fluid phase was centrifuged for 10 min. The resulting supernatant was mixed with an equal volume of a 65°C solution containing 2% CTAB, 100 mM Tris, 20 mM EDTA, 1.4 M NaCl, and 1% polyvinyl pyrrolidone and extracted with chloroform. The aqueous phase was mixed with 1/10 vol of 10% CTAB/0.7 M NaCl (pre-warmed to 65°C), vortexed, allowed to sit for 20 min, extracted with chloroform again, mixed with an equal volume of 1% CTAB/50 mM Tris/10 mM EDTA, allowed to sit for about 45 min and centrifuged for 2 min. The pellet was carefully drained and thoroughly re-suspended in 10 mM Tris/1 mM EDTA/1 M NaCl, then mixed with 2.5 vol of ethanol, held at −20°C for 1 h, and centrifuged for 10 min; the resulting pellet was dried *in vacuo* and dissolved in 10 mM Tris/1 mM EDTA. (3) Finally, in a few cases a quick-extraction, 'mini-prep' method was used: Cells from 10-ml cultures were harvested by centrifugation and re-suspended in 1 ml H_2O, and 1/10 vol M NaCl was added. The suspension was incubated at 95°C for 5 min, centrifuged at 14 000 rpm in a microfuge, and the

Table 1 Volvocalean strains employed

Genus, species	Forma and/or strain (synonym)	Geographic origin	Accession # (reference)	Abbrev.
Volvox:				
V. aureus		Michigan, USA	UTEX[a] 1899	Va
V. carteri	f. *nagariensis*/HK10	Japan	UTEX 1885	Vch
V. carteri	f. *nagariensis*/NIES	Japan	NIES[b] 397	Vcn
V. carteri	f. *nagariensis*/Poona	India	(Adams et al. 1990)[c]	Vcp
V. carteri	f. *kawasakiensis*/KK3	Japan	(Nozaki, 1988)[d]	Vck
V. carteri	f. *weismannia*	Nebraska, USA	UTEX 1874	Vcw
V. dissipatrix		Australia	UTEX 1871	Vd
V. obversus		Australia	UTEX 1866	Vo
V. powersii		Nebraska, USA	UTEX 1864	Vp
V. spermatosphaera		South Africa	UTEX 2273	Vs
V. tertius		England	UTEX 132	Vt
Pleodorina:				
P. californica		Tennesee, USA	UTEX 198	Pc
P. illinoisensis		Minnesota, USA	UTEX 808	Pi
Eudorina:				
E. elegans	12	Sweden	UTEX 12	E2
E. elegans	1193	*Indiana, USA*	UTEX 1193	E3
Chlamydomonas:				
C. debaryana		not available	CCAP[e] 11/56a	Cda
C. debaryana		not available	CCAP 11/56b	Cdb
C. debaryana		Mexico	UTEX 344	Cd3
C. debaryana	(*C. komma*)	Japan	UTEX 579	Cd5
C. incerta		Cuba	CC[f]-1870	Ci
C. reinhardtii	nit1 cw15	Massachusetts, USA	CC-2424	Cr
C. reinhardtii	Jarvik 6	Pennsylvania, USA	CC-2342	C6
C. reinhardtii	Jarvik 124	Florida, USA	CC-2343	C1
C. reinhardtii	Jarvik 356	Pennsylvania, USA	CC-2344	C3
C. reinhardtii	Harris 6	North Carolina, USA	CC-2931	Ch6
C. reinhardtii	(*C. grossii*)	Minnesota, USA	CC-1952	Cg
C. reinhardtii	(*C. smithii* mt^+)	Massachusetts, USA	CC-1373	Cs
C. zebra	(*C. texensis*)	Texas, USA	UTEX 1904	Cz

[a]UTEX: The Culture Collection of Algae at the University of Texas at Austin (Starr and Zeikus 1993)
[b]NIES: Microbial Culture Collection, National Institute for Environmental Studies, P.O. 16-2 Onogawa, Tsukuba-shi, Ibaraki 305, Japan
[c]Provided by Richard C. Starr, Dept. of Botany, Univ. of Texas, Austin, TX 78713, USA
[d]Provided by Hisayoshi Nozaki, now at the National Insitute for Environmental Studies (address above)
[e]CCAP: Culture Collection of Algae and Protozoa; Freshwater Biological Association, Ambleside, Cambria, UK
[f]CC: Catalogue of *Chlamydomonas* strains of the *Chlamydomonas* Genetic Center, Duke University, Durham, NC 27706, USA

supernatant was extracted once with phenol, and once with chloroform. DNA was precipitated with 1 vol isopropanol and re-suspended in 50 µl H_2O; 1–5 µl of DNA was used for each PCR. Primers for use in PCR were designed on the basis of known sequences of the relevant *V. carteri f. nagariensis* and *C. reinhardtii ypt4* and actin genes (Cresnar et al. 1990; Fabry et al. 1993; Dietmaier et al. 1995; Sugase et al. 1996) and known algal codon biases (Schmitt et al. 1992). The primer pairs used for amplification of intron VI of the *ypt4* gene were (4/6-5′) AGCACVGCNCA-CAACGTNGAGGA and (4/6-3′) AGTCTCCTTNGCBGTGT-TRATGAA, and for intron VII (4/7-5′) AGCTACAAGAAGA-TYCAGGAYGG and (4/7-3′) AGTANCCVACCTTRATNC-CRTA. For actin intron IX, the primer pairs were (A/9-5′) AGCATGAARTGYGAYGTNGAYAT and (A/9-3′) GA-CACNACYTTDATYTTCATNGC (see Fig. 1A). The standard PCR reaction was performed in 50 or 100 µl with 200 nM/µl of each primer. The cycling conditions for *ypt4* introns were 95°C/1 min, 57°C (ypt4-VI) or 52°C (ypt4-VII)/1 min, and 72°C/1 min (35 repetitions). For intron act-IX, a two-step, hot start PCR was performed: 3 cycles at 95°C/1 min, 53°C/1 min, and 72°C/1 min were followed by 35 cycles at 95°C/1 min, 55°C/l min, and 72°C/1 min. Each intron from each strain was amplified at least twice in independent PCR reactions. PCR products were separated on 2% agarose by electrophoresis, cut from the gel, and purified with silica particles according to Boyle and Lew (1995). In early phases of this study, PCR products were cloned in plasmid pUCBM20 (Boehringer Mannheim) using the DI/TRISEC method (Dietmaier et al. 1993), and recombinant plasmids were sequenced using T7 polymerase according to the manufacturer's description (Pharmacia). Later, gel-purified PCR products were sequenced directly by PCR, using the ^{33}P-labelled PCR primers, in 7.5-µl reaction mixtures, with 40 repetitions of 95°C/30 s, 55°C/30 s, and 72°C/1 min. In certain cases, in which the sequence obtained with the original PCR primers was incomplete (because of intron length) or ambiguous in some region, additional primers were designed and used to complete the sequencing.

Sequence analysis and phylogenetic calculations. Pairwise alignments of all introns within a group (ypt4-VI, ypt4-VII or act-IX) were performed by the GAP program of the GCG software package (Genetics Computer Group Inc., Madison, WIS., USA; version 7) with a gap penalty of 1.0, and no gap-length penalty. A calculation of the statistical significance of the similarities observed was performed by calculating a similarity index (SI_{50}). This was obtained by dividing the computer-generated Q value (number of matches minus number of gaps) for the aligned sequences by the mean Q value of 50 additional analyses of the same sequence pair under identical alignment conditions, but with one sequence repeatedly randomized. SI_{50} values from different experiments are directly comparable, as long as the alignment parameters are not changed. In theory, full dissimilarity between two sequences should result in $SI_{50} = 1.0$. In practice, however, deviations from this value occur that vary with the base composition and sizes of the two sequences being compared. Simulations showed that with short sequences, SI_{50} values about 1.10 (and exceptionally even as high as 1.15) can occasionally be obtained with unrelated partners (see also intron act-IX; Table 3c). Therefore, $SI_{50} \geq 1.15$ was set as the lower limit of significance. Two identical sequences usually generated a SI_{50} value > 2.15. For the analysis of the mutation type (transition/transversion/gap) at each position in aligned sequences, and for detection of consensus motifs within a set of similar sequences, the Turbo Pascal 6.0 programs MUTATION and NEWSITE, respectively, were written. These programs and their description are available from M.L. on request. To minimize interference from 'multiple hits', only those intron pairs exhibiting high levels of similarity (≥ 80% sequence identity) were used to compute transition/transversion/gap ratios. For each intron set, a multiple alignment based on successive pairwise alignments according to Needleman and Wunsch (1970) was established with the program PileUp of the GCG software package, version 7 (1991), on a SUN SparcStation IPX. The penalty parameters for alignments were again 1.0 for a gap, and 0 for gap length. Using the method of Kimura (1980), distance values were calculated from the multiple alignment for all pairwise comparisons with $SI_{50} > 1.15$, resulting in three independent distance matrices for each group of similar sequences. These values were converted into phylogenetic trees according to Fitch and Margoliash (1967), treating the branch lengths as adjustable parameters, and considering the best tree (based on least-square analysis) as the correct one (Woese 1987). The root was inferred by using the remaining sequences as outgroups. Phylogenetic distances were calculated with the program DNADIST; transitions and transversions were weighted equally. To estimate the significance of all branchpoints, 100 additional data sets for bootstrap analysis were generated with the program DNABOOT (Felsenstein 1985), and bootstrap analyses and cladograms were calculated with the program KITSCH using the least-square and Fitch-Margoliash algorithms. An rRNA tree for the taxa used in this study was calculated from data in Larson et al. (1992), using the DNADIST, DNABOOT, DNAMLK and KITSCH programs in the PHYLIP software package, version 3.2, of Felsenstein (1989).

Results

Amplification and sequencing of introns

Using appropriate exon-specific primer pairs, introns VI and VII of the *ypt4* gene were amplified from genomic DNA of 28 selected volvocalean species and strains (Table 1). In order to obtain a third, completely independent, source of sequence information, we also amplified one intron (IX) from the actin gene of most of the same strains. Figure 1 A shows a map of these genes, introns and primer positions in *C. reinhardtii* strain CC 2424. Actin and *ypt* genes are present in multiple copies in some organisms (e.g. ten actin genes in *Arabidopsis*, and three isoforms of *YPT5* in yeast; Singer-Krüger et al. 1994; McDowell et al. 1996); but it has been shown that actin and *ypt4* are single-copy genes in the haploid algae *C. reinhardtii* and *V. carteri f. nagariensis* HK10 (Cresnar et al. 1990; Dietmaier et al. 1995; Sugase et al. 1996), ruling out any complications from multiple loci or alleles. Accordingly, we obtained single bands in PCR and unique DNA sequences of the PCR products for every alga analyzed.

To minimize PCR artifacts, we determined the complete sequence of each intron from two independent PCR products, and then, if these sequences were not identical, analyzed additional PCR products to resolve discrepancies. With one exception (discussed below), each PCR product that was recovered revealed an intron (flanked by canonical consensus splice sites) that interrupted the coding sequence at the expected, conserved position in a conserved exon. In general, intron ypt4-VI was the largest, and act-IX the smallest of the introns analyzed, but sizes varied considerably in different taxa (Table 2).

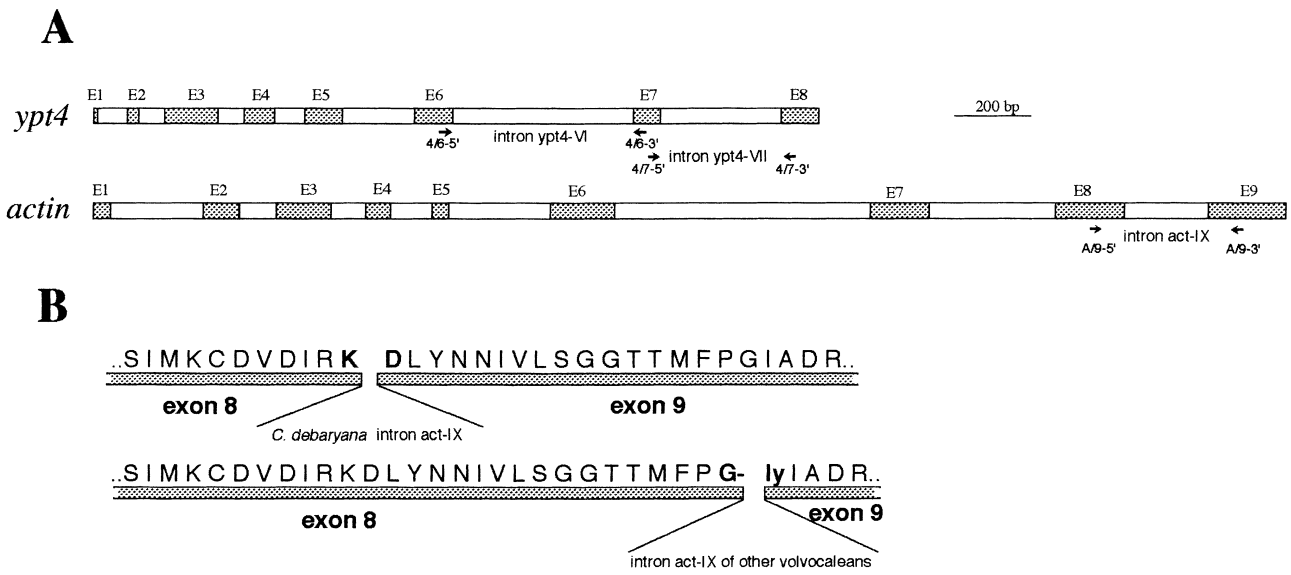

Fig. 1A, B Locations of introns analyzed in this study. **A** organization of the *ypt4* and the actin genes of *C. reinhardtii* with respect to exons (*shaded boxes*) and introns (*open boxes*). The location and orientation of the primers used for amplification of introns ypt4-VI, ypt4-VII and act-IX in *C. reinhardtii* (and the homologous introns of other algae) are indicated below the adjacent exons. **B** the location of intron act-IX in *C. debaryana* (above), compared to its location in all other algae that were studied (below). Amino acids encoded by the adjacent exons are given in the single-letter code, except for one Gly residue, the codon for which is interrupted by an intron between the first and second nucleotide

In all four *C. debaryana* isolates analyzed, intron act-IX was found at an unexpected position, shifted 49 bp in the 5′ direction from its location in all other organisms studied (Fig. 1 B). Although this displacement casts doubt on the homology of these introns to the act-IX introns of the other algae, there can be little doubt that they are homologous to one another, and thus can still be used in conjunction with the ypt4-VI introns to estimate the degree of relationship among *C. debaryana* strains. Furthermore, *C. zebra* produced no PCR product with primer pairs 4/7-5′ and 4/7-3′. This could be either because it lacks intron ypt4-VII, because the flanking exons are too different to allow primer hybridization, or because this intron is also shifted – but too far to be captured by the primers used. However, intron ypt4-VI could be amplified at the conserved position in this strain. A short (about 50 bp) central part of intron ypt4-VI from *P. californica* was unreadable for unknown reasons. Intron sequences were deposited in the GenBank database (accession numbers in Table 2).

Sequence divergence in homologous introns

To obtain an overview of the sequence divergence range, pairwise comparisons between all introns within a homologous set were performed. We obtained a spectrum ranging from identity (*V. carteri* f. *nagariensis* strains HK10 and NIES) to absence of any discernible similarity (many cases). To create a common basis for comparison, a similarity index (SI_{50}) was established for each sequence pair (see Materials and methods for details). Identical sequences were defined by $SI_{50} \geq 2.14$, and sequence pairs without significant similarity by $SI_{50} \leq 1.15$ (see Fig. 3, upper-right sectors). The term "closely related" was applied to organisms only when they displayed a SI_{50} score ≥ 1.15 for at least two different introns. Using this stringent dual criterion, five groups of closely related organisms, comprising 20 of the 29 isolates studied, could be readily identified. These groups were: (1) five *V. carteri* isolates, (2) seven *C. reinhardtii* isolates, (3) the *V. tertius*/*V. spermatosphaera* pair, (4) the *E. elegans* UTEX 1193/*P. illinoisensis* pair, and (5) four *C. debaryana* strains. The relationship between *C. incerta* and the seven members of the *C. reinhardtii* clade was found to be at the borderline of resolution by this method: significant similarities were detected for intron ypt4-VI (average $SI_{50} = 1.20$, range = 1.18–1.22), but similarity values at or below the significance threshold were observed for the other two introns (ypt4-VII: average $SI_{50} = 1.14$, range = 1.13–1.16; intron act-IX: average $SI_{50} = 1.15$, range = 1.12–1.16). All other pairwise strain comparisons yielded no significant similarities by SI_{50} values.

The numbers of transitions, transversions, and gaps between pairs of aligned sequences were determined for the organisms displaying the greatest similarities (≥ 80% sequence identity). When all pairwise comparisons within these groups were averaged, the mean transition-to-transversion-to-gap ratios obtained were 1:0.67:0.38 for ypt4-VI, 1:0.75:0.39 for ypt4-VII, and 1:0.83:0.38 for actin-IX, indicating that, within these closely related introns, transitions are somewhat more

Table 2 Products of PCR amplification of volvocalean introns

Strain[a]	Intron					
	ypt4-VI		ypt4-VII		act-IX	
	bp[b]	GB #[c]	bp	GB #	bp	GB #
Va	583	U55913	226	U55914	n.d.[d]	–
Vch	568	L08130	247	L08130	282	M33963
Vcn	568	U55917	247	U55918	282	U70573
Vcp	572	U55919	245	U55920	286	U70574
Vck	499	U55915	215	U55916	281	U70572
Vcw	676	U55921	224	U55922	292	U70575
Vd	455	U55923	265	U55924	157	U70576
Vo	483	U55925	284	U55926	263	U70577
Vp	637	U55927	248	U55928	n.d.	–
Vs	748	U55929	242	U55930	103	U70578
Vt	601	U55931	238	U55932	111	U70579
Pc	>808[e]	U55907	193	U55908	n.d.	–
Pi	796	U55909	235	U55910	209	U70570
E2	852	U55899	254	U55900	n.d.	–
E3	1261	U55897	241	U55898	414	U70567
Cda	612	p.d.[f]	250	p.d.	161	p.d.
Cdb	612	p.d.	250	p.d.	n.d.	–
Cd3	612	U55886	n.d.	–	161	U70564
Cd5	586	U55887	n.d.	–	138	U70565
Ci	451	U55895	349	U55996	200	U70566
Cr	485	U13167	325	U13167	222	D50838
Cg	480	U55893	338	U55894	361	U70563
C6	483	U55905	327	U55906	215	U70569
Cl	485	U55889	338	U55890	315	U70561
C3	484	U55891	327	U55892	272	U70562
Ch6	483	U55901	327	U55902	217	U70568
Cs	485	U55911	338	U55912	215	U70571
Cz	634	U55888	n.d.	–	n.d.	–

[a] For abbreviations see Table 1
[b] Length of amplified intron, in base pairs
[c] GenBank accession number
[d] n.d. = not determined
[e] Approximately 50 bp not readable
[f] p.d. = preliminary sequence data (not yet submitted)

frequent than transversions, and insertions/deletions are surprisingly abundant. When these data were disaggregated to consider individually the groups with the most similar sequences (i.e. the seven *C. reinhardtii* isolates, two *C. debaryana* isolates, four *V. carteri* isolates [excluding f. *weismannia*], or the *P. illinoisensis*/*E. elegans* 1193 pair), the number of transversions per transition ranged between 0.5 and 1.7, and the ratio of gaps per transition ranged between 0.0 and 1.25. In the light of this variability in the ratio of transitions to transversions in different introns and different taxa, these two types of nucleotide replacement were weighted equally in the subsequent calculations of distance values.

Most changes resulting in gaps were small. An exception occurred in intron ypt4-VI of *P. illinoisensis*, where a 463-bp insert was found at position 297 relative to the otherwise extremely similar intron of *E. elegans* strain 1193. A database search did not detect any similarity of this insert to any known mobile genetic element. Thus, this difference does not appear to resemble a situation recently detected in intron III of the *Adh* gene of the genus *Brachyscome* (Asteraceae), where one main reason for sequence variability was attributed to insertions of mobile elements (Denda et al. 1995).

Most insertions/deletions occurred at different, apparently random, locations in different taxa. A curious exception to this generalization was in intron act-XI of *C. reinhardtii*, where the introns of all seven strains that were analyzed differed from one another in one particular region. At this location each strain possesses one or more copies of an asymmetric element that comprises a 7–15-bp highly variable region flanked by 33-bp and 8-bp conserved regions (Fig. 2). This element is present only once (in slightly different forms) in four of the *C. reinhardtii* isolates, but the remaining three isolates possess two, three, or four copies of it. The significance of this curious feature is unknown. In calculating relationships among these isolates, each extra copy of this element was (like other insertion/deletion gaps) treated as a unitary mutational event.

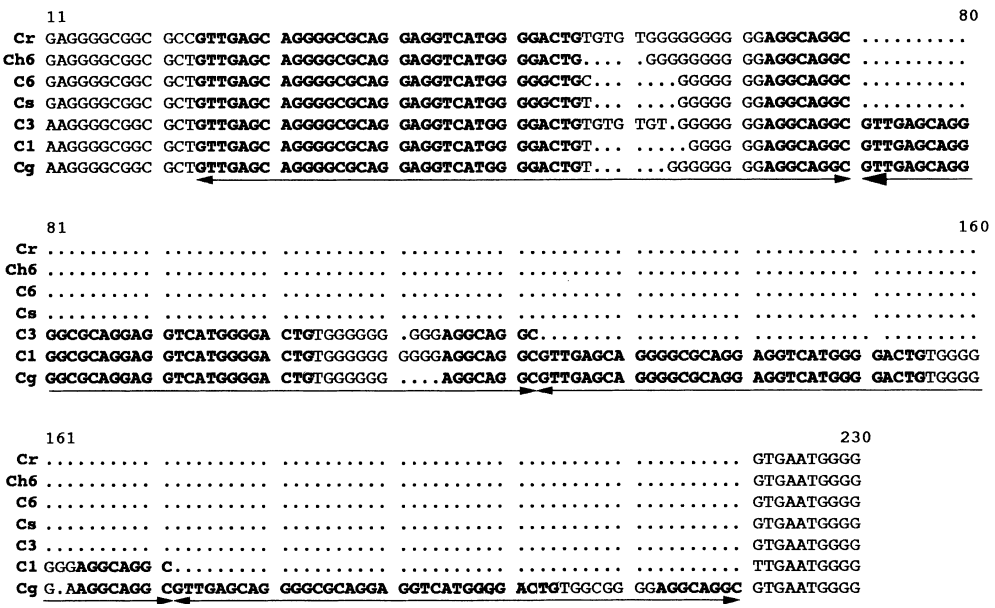

Fig. 2 Aligned sequences of part of intron act-IX in seven different *C. reinhardtii* strains (identified as in Table 1) in the region where a similar element is repeated different numbers of times in various strains. *Double-headed arrows* indicate the extent of each such element, and *boldface letters* indicate conserved bases within that element

A search for possible internal splicing consensus motifs (e.g. lariat consensus sequences) in these introns was not successful. A program written to find the "best" common motif in a given set of sequences (NEWSITE) readily detected the *S. cerevisiae* lariat consensus (RCTRAC) within a set of 14 arbitrarily chosen yeast introns, but failed to find any related elements within any of the following four algal sequence groups: (1) 26 homologous ypt4-VI introns, (2) 23 homologous ypt4-VII introns, (3) 116 different introns of *C. reinhardtii* CC-2424 for which sequences were available, or (4) 78 different introns of *V. carteri* f. *nagariensis* HK10 for which sequences were available (data not shown). The same sets of intron sequences, however, gave pyrimidine contents of 66%, 69%, 71%, and 65%, respectively, within a 30-bp segment near the 3′ splice site (between positions −35 and −5), whereas the overall average pyrimidine contents of *C. reinhardtii* and *V. carteri* f. *nagariensis* introns are 50% and 53%, respectively. This confirms earlier observations that, apart from the conserved GT-AG borders and a pyrimidine-rich tail, neither consensus motifs nor common folding structures can be detected in volvocalean introns (Dietmaier and Fabry 1994). Also (apart from the one exception discussed above) no significant 'hotspots' for mutation were found when comparing many different, homologous sequence pairs. The substitution rate in volvocalean introns seems, therefore, to follow a nearly random and unbiased ("neutral") pattern.

Phylogenetic analyses

Independent of the pairwise comparisons discussed above, a multiple alignment of all members of each intron group was established, and a distance matrix was derived for all sequences that exhibited significant similarities. The resulting distance values (Table 3, lower sector) were then used to establish the dendrograms presented in Fig. 4. The following taxa did not exhibit significant intron sequence similarities to any other taxon studied here, and therefore are not included in Figs. 3 or 4: *V. aureus, V. dissipatrix, V. obversus, V. powersii, P. californica, E. elegans* strain 12, and *C. zebra*.

With one exception to be discussed shortly, dendrograms derived from all three sets of intron data for the remaining taxa were fully consistent. For example, all three sets of data produced precisely the same branching pattern for the five members of the *V. carteri* complex that were examined, with branches of similar length and all branchpoints supported by high bootstrap values (≥ 93%; Fig. 4 A, B, C). A similarly consistent pattern of relationship was observed for *C. debaryana*. Of particular interest are the consistent relationships deduced for the members of the *Eudorina-Pleodorina* group that were examined. All three data sets supported a close relationship between *E. elegans* strain 1193 and *P. illinoisensis,* while at the same time failing to detect any significant similarities in intron sequences between *E. elegans* strains 12 and 1193, or between *P. californica* and *P. illinoisensis.* It was also of interest that in each case the intron data indicated a very close affiliation between *V. spermatosphaera* and *V. tertius.*

Only the results obtained with the *C. reinhardtii* group were not fully consistent. All three sets of intron data indicated very close relationships among the seven

A: intron ypt4-VI

SI_{50} values

	Vch	Ven	Vep	Vek	Vew	Vt	Vs	Pi	E3	Cr	Cg	C6	C1	Cr356	Ch6	Cs	Ci	Cd3	Cd5
Vch		2.17	2.10	1.50	1.27	1.01	1.06	1.01	1.02	1.03	1.02	1.03	1.02	1.02	1.02	1.03	1.05	1.04	1.03
Ven	0.000		2.11	1.50	1.28	1.01	1.06	1.01	1.02	1.03	1.02	1.03	1.02	1.02	1.02	1.03	1.05	1.04	1.03
Vep	0.023	0.021		1.52	1.29	1.02	1.05	1.02	1.01	1.02	1.02	1.03	1.02	1.03	1.02	1.02	1.05	1.05	1.02
Vek	0.215	0.212	0.205		1.35	1.03	1.03	1.01	1.00	1.02	1.01	1.02	1.02	1.02	1.02	1.02	1.03	1.05	1.04
Vew	0.286	0.283	0.281	0.300		1.00	1.01	1.02	1.00	1.00	1.00	1.00	1.00	1.00	1.02	1.00	1.02	1.05	1.03
Vt							1.23	1.00	1.01	1.02	1.02	1.03	1.02	1.03	1.03	1.01	1.02	1.01	1.01
Vs						0.322		1.03	1.01	1.02	1.04	1.03	1.03	1.03	1.03	1.02	1.04	1.01	1.02
Pi									1.63	1.02	1.03	1.04	1.02	1.03	1.03	1.02	1.03	1.05	1.03
E3								0.074		1.01	1.02	1.01	1.01	1.01	1.01	1.01	1.01	1.03	1.02
Cr											2.02	2.02	2.10	2.02	2.03	2.11	1.20	1.04	1.03
Cg										0.034		2.01	2.03	2.02	2.01	2.03	1.22	1.03	1.02
C6										0.023	0.041		2.03	2.10	2.12	2.02	1.18	1.04	1.04
C1										0.006	0.036	0.030		2.01	2.02	2.10	1.20	1.03	1.04
C3										0.028	0.041	0.008	0.034		2.11	2.02	1.19	1.04	1.03
Ch6										0.021	0.039	0.002	0.028	0.006		2.03	1.19	1.04	1.03
Cs										0.008	0.034	0.028	0.010	0.032	0.026		1.21	1.04	1.03
Ci										0.351	0.348	0.353	0.351	0.352	0.348	0.356		1.02	1.04
Cd3																			1.55
Cd5																		0.204	

B: intron ypt4-VII

	Vch	Ven	Vep	Vek	Vew	Vt	Vs	Pi	E3	Cr	Cg	C6	C1	C3	Ch6	Cs	Ci
Vch		2.25	2.12	1.62	1.36	1.04	1.04	1.04	1.07	1.04	1.06	1.06	1.07	1.08	1.07	1.06	1.04
Ven	0.000		2.12	1.62	1.36	1.04	1.04	1.04	1.07	1.04	1.06	1.06	1.07	1.08	1.07	1.06	1.04
Vep	0.055	0.055		1.59	1.34	1.05	1.03	1.06	1.07	1.06	1.07	1.07	1.07	1.06	1.07	1.08	1.02
Vek	0.232	0.232	0.246		1.48	1.12	1.08	1.09	1.07	1.08	1.09	1.11	1.10	1.12	1.10	1.09	1.04
Vew	0.361	0.361	0.386	0.272		1.05	1.02	1.07	1.05	1.01	1.02	1.04	1.02	1.04	1.11	1.02	1.04
Vt							1.52	1.07	1.05	1.07	1.08	1.09	1.08	1.08	1.08	1.08	1.04
Vs						0.366		1.06	1.05	1.09	1.08	1.09	1.10	1.09	1.09	1.09	1.04
Pi									2.04	1.04	1.04	1.07	1.05	1.06	1.05	1.05	1.05
E3								0.081		1.05	1.06	1.06	1.05	1.07	1.07	1.06	1.04
Cr											1.98	1.96	2.02	1.97	1.97	2.05	1.16
Cg										0.038		1.99	2.02	2.00	1.99	2.03	1.15
C6										0.026	0.038		2.01	2.10	2.11	2.03	1.14
C1										0.016	0.040	0.022		2.03	2.03	2.10	1.15
C3										0.019	0.031	0.006	0.016		2.11	2.05	1.14
Ch6										0.023	0.035	0.003	0.019	0.003		2.02	1.13
Cs										0.009	0.033	0.016	0.006	0.009	0.012		1.14
Ci										0.496	0.529	0.476	0.489	0.469	0.469	0.487	

C: intron act-IX

	Vch	Ven	Vep	Vek	Vew	Vt	Vs	Pi	E3	Cr	Cg	C6	C1	C3	Ch6	Cs	Ci	Cd3	Cd5
Vch		2.20	2.15	1.53	1.33	1.06	1.05	1.07	1.02	1.08	1.06	1.09	1.06	1.08	1.07	1.07	1.02	(1.02)	(1.01)
Ven	0.000		2.15	1.53	1.33	1.06	1.05	1.07	1.02	1.08	1.06	1.09	1.05	1.09	1.07	1.07	1.02	(1.02)	(1.01)
Vep	0.011	0.011		1.52	1.32	1.05	1.04	1.07	1.03	1.06	1.04	1.08	1.04	1.08	1.07	1.05	1.02	(1.01)	(1.00)
Vek	0.260	0.260	0.262		1.37	1.06	1.08	1.02	1.01	1.00	1.04	1.05	1.03	1.01	1.00	1.06	0.99	(1.00)	(1.00)
Vew	0.421	0.421	0.430	0.404		1.10	1.05	1.02	1.02	1.00	1.05	1.05	1.09	1.05	1.00	1.06	0.99	(1.00)	(1.00)
Vt							1.18	1.00	1.00	1.12	1.10	1.10	1.16	1.12	1.09	1.11	1.05	(1.11)	(1.13)
Vs						0.481		1.00	1.02	1.11	1.09	1.07	1.12	1.11	1.09	1.06	1.00	(1.04)	(1.05)
Pi									1.54	1.09	1.15	1.14	1.16	1.14	1.12	1.12	1.07	(1.00)	(1.09)
E3								0.102		1.07	1.07	1.08	1.06	1.08	1.07	1.08	1.03	(1.06)	(1.09)
Cr											2.00	2.03	2.00	2.05	2.07	2.04	1.15	(1.09)	(1.12)
Cg										0.048		2.00	2.07	2.04	2.03	2.00	1.16	(1.08)	(1.09)
C6										0.038	0.054		2.01	2.00	2.09	2.12	1.16	(1.06)	(1.13)
C1										0.048	0.038	0.044		2.05	2.03	2.03	1.16	(1.07)	(1.09)
C3										0.042	0.028	0.048	0.029		2.02	2.02	1.15	(1.08)	(1.08)
Ch6										0.023	0.043	0.014	0.034	0.038		2.08	1.14	(1.06)	(1.13)
Cs										0.038	0.054	0.014	0.044	0.043	0.014		1.12	(1.07)	(1.11)
Ci										0.491	0.492	0.512	0.495	0.504	0.492	0.512		(1.10)	(1.14)
Cd3																			1.74
Cd5																		0.185	

distance values

Fig. 3A–C Similarity indices (upper right) and distance values (lower left) calculated for the three different sets of intron sequences. Only the taxa exhibiting significant similarity values ($SI_{50} > 1.15$) are included in the figure, and sets of significant SI_{50} values are enclosed in *heavy lines*. Values in *parentheses in* **C** refer to introns that are shifted in position in *C. debaryana* relative to their locations in all other taxa. For further details see text

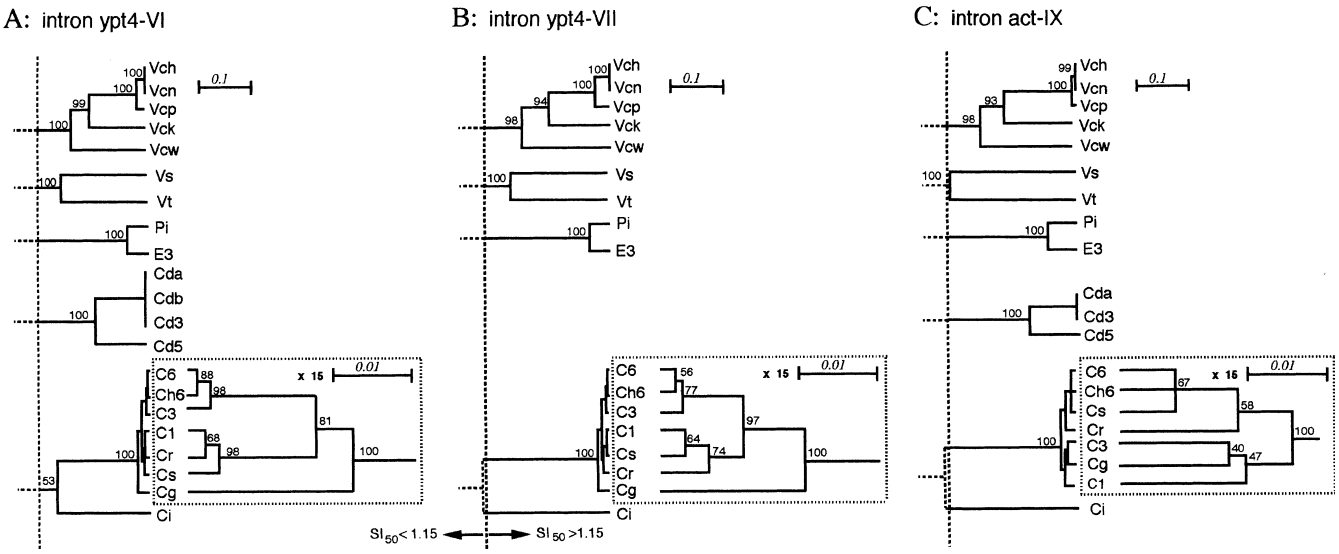

Fig. 4A–C Dendrograms of volvocalean relationships inferred from the data in Fig. 3. Strain identifications are as in Table 1, branch lengths are proportional to distance values given in Fig. 3 (see scales), and *vertical dashed lines* indicate thresholds for statistical significance ($SI_{50} \geq 1.15$). Bootstrap confidence levels (from 100 replicates) are given next to all branchpoints. For the *C. reinhardtii* clade, a 15-fold expansion for each dendrogram is given at the right

C. reinhardtii isolates (including the one previously catalogued as *C. smithii* mt^+), and set them clearly apart from all other *Chlamydomonas* isolates examined with the exception of *C. incerta*. But, whereas the two ypt4 introns yielded *C. reinhardtii* trees with nearly identical topologies for the *C. reinhardtii* clade (Fig. 4 A, B), data from the actin intron (Fig. 4C) yielded a different topology and failed to resolve the branching pattern for three of the isolates. Moreover, whereas all three sets of intron data identified *C. incerta* as a probable sister taxon on the *C. reinhardtii* clade, in two cases this relationship fell at or below the threshold of statistical significance.

Discussion

Introns as sources of phylogenetic information

Although spliceosomal introns generally experience more rapid evolution than exons, their use for high-resolution molecular phylogeny has been very limited so far (Li and Graur 1991). A continuing dispute regarding the origins and antiquity of introns (e.g. Hurst 1994; Stolzfuss 1994; Kwiatowski et al. 1995; Long et al. 1995) may be one reason they remain largely unexploited. Here we show, however, that once introns of unambiguous homology are identified, they do provide a potentially valuable tool for resolving phylogenetic relationships within groups of closely related strains and species.

The choice of the actin and *ypt* genes as molecular objects of this study was based on knowledge that these are both single-copy genes with highly conserved exons encoding ubiquitous eukaryotic proteins in the algal group examined. Actin is one of the most slowly evolving proteins known (Doolittle 1995), and (at least within the volvocine algae) the coding region of the *ypt4* gene is equally conservative; in sharp contrast to their introns, the exons of both of these genes exhibit about 85% nucleotide identity between *V. carteri* strain HK 10 and *C. reinhardtii* strain CC-2424 (Cresnar et al. 1990; Fabry et al. 1993; Dietmaier et al. 1995; Sugase et al. 1996). The present data indicate that the introns of these two classes of genes are evolving at similarly rapid rates, since the two *ypt4* introns and the actin intron yielded comparable data (Fig. 3), with intergenic differences (*ypt* vs actin) being not significantly greater than intragenic ones (ypt4-VI vs ypt4-VII). Thus, for many purposes, the results from the three introns data sets can be discussed together.

Most molecular phylogenetic studies employ sequences that exhibit extensive similarity within all members of a study group, and which therefore can be aligned relatively easily. In contrast, this study dealt with sequences for which alignment was often difficult at best because, within this rather limited group of related organisms, they ranged from perfect identity to complete dissimilarity, and often exhibited substantial differences in length. Therefore, we recognized the need for some simple metric that could be applied to all intron pairs to distinguish those that retained enough sequence similarity to make further analysis meaningful from those that did not. For this purpose, the similarity index "SI_{50}" was calculated. We further restricted our phylogenetic

analysis to cases in which at least two of the introns of a pair of strains being compared exhibited SI_{50} values of 1.15 or greater. This dual criterion excluded from our analysis a few pairs of organisms that had marginal similarities in a single intron (data not shown), while permitting us to infer a probable relationship between two taxa (*C. incerta* and *C. reinhardtii*) for which sequence similarities were at the borderline of significance (see below).

In contrast to introns in organisms that have been more extensively studied, no conserved splicing signals could be detected in the volvocalean introns, apart from their GT…AG intron borders and a modest overrepresentation of pyrimidine residues near the 3′ splice site (65–71%), somewhat reminiscent of the polypyrimidine tracts of animal introns (Wiebauer et al. 1988). Moreover, no common folding motifs could be detected. Thus, if green algal introns contain any additional splicing information within their sequences, the operative motifs are very weakly conserved and apparently do not constrain local substitution rates, because we could not detect any regions of higher or lower mutability within introns that could be unambiguously aligned. It thus appears that these algal introns can be treated as neutrally mutating DNA segments that are nearly ideal for the documentation of rapid, unconstrained nucleotide substitutions.

Phylogenetic conclusions

The analyses reported here clustered 20 of the 28 taxa that we examined into five groups that share high SI_{50} values, and which therefore appear to represent clades of very closely related organisms:

(1) The five isolates of *V. carteri* that were analyzed in this study are the representatives of the genus *Volvox* that have been studied most extensively over the years by ourselves and others (reviewed in Kirk and Harper 1986; Schmitt et al. 1992; Kirk 1996). Previous comparative studies of rRNA sequences had clearly indicated that these five taxa are closely related, but rRNA sequences lacked the resolving power to establish a statistically robust set of relationships among these strains and certain other closely related species of *Volvox* (Larson et al. 1992; Kirk et al., unpublished). In marked contrast, all three intron data sets employed here led to dendrograms for these isolates that are identical in topology, similar in branch lengths, and supported by high bootstrap values (≥ 93%). Because the intron-derived relationships among these five strains are fully consistent with all past genetic, molecular and morphological observations (Schmitt et al. 1992; Kirk 1996), the topology of this consensus tree might well have been predicted – whereas its branch lengths could not have been. Therefore, the uniformity and robustness of the branch lengths in the trees that were derived here appear not only to provide the first reliable reconstruction of the phylogenetic history of these five *V. carteri* isolates of interest, but also provide clear evidence of the potential power of comparative studies of intron sequences as a tool for resolving relationships among very closely related taxa.[1]

(2) The second set of relationships suggested by our intron analysis was completely unexpected. All three data sets indicated that *V. spermatosphaera* and *V. tertius* have shared a common ancestor about as recently as have members of the *V. carteri* clade. Although *V. spermatosphaera* and *V. tertius* (like *V. obversus* and *V. powersii,* which were also studied here) are classified in the same section of the genus as *V. carteri* (Smith 1944), they are readily distinguished morphologically, and there have been no prior suggestions that they might be particularly closely related. This novel finding therefore deserves further study with other methods.

(3) Another particularly interesting result of the present analysis was the set of relationships it revealed among the *Eudorina* and *Pleodorina* strains that were sampled. One of the most surprising aspects of a preliminary rRNA-based molecular phylogeny of volvocaceans that was published earlier (Larson et al. 1992) had been that it placed representatives of *Eudorina* and *Pleodorina* on widely separate branches of the family tree. The reason this was surprising was that Goldstein (1964) had shown that certain representatives of these two genera could mate and produce viable hybrid progeny, and had therefore proposed that the genus *Pleodorina* should be included in the genus *Eudorina.* However, we recognized that the two isolates that had originally been subjected to rRNA sequencing (*E. elegans* strain 12 and *P. californica*) were not the ones that Goldstein had succeeded in hybridizing (*E. elegans* strain 1193 and *P. illinoisensis*). Therefore, we included all four of these isolates in the present study. All three introns indicated that *E. elegans* strain 1193 and *P. illinoisensis* are extremely close relatives of one another (Figs. 3, 4), whereas each is so distantly related to the strain with which it shares a generic or specific name that no discernible sequence similarities are retained within introns. This result is fully consistent with the observations of Goldstein (1964), who succeeded in mating *E. elegans* strain 1193 with *P. illinoisensis*, but failed in repeated attempts to mate it with *E. elegans* strain 12.

[1] Although it is not obvious from the data included in the figures, the intron data also permitted us to reject certain phylogenetic hypotheses constructed on other grounds. For example, previous rRNA sequence data (A. Larson and D. Kirk, unpublished) had suggested the possibility that *V. obversus* might be more closely related to certain *V. carteri* isolates than the latter were to other *V. carteri* isolates, but did not have the resolving power to provide a statistically significant test of this hypothesis. In contrast, the present study appears to permit a secure rejection of this hypothesis, since no significant sequence similarities were detected between *V. obversus* and *V. carteri* introns

Furthermore, this result has now been corroborated by additional rRNA sequencing, which also indicates that *E. elegans* strain 1193 is an extremely close relative of *P. illinoisensis*, but that *E. elegans* strain 12 is more closely related to members of five other volvocacean genera than it is to *E. elegans* strain 1193 (A. Larson and D. Kirk, unpublished). Together these results strongly support the developing view that many genera in the family Volvocaceae are polyphyletic, and that the morphological criteria used for classification of volvocaceans identify grades of organizational complexity, not clades of closely related organisms (see also Larson et al. 1992).

(4) A clade of particular interest identified also in this study includes seven strains of *C. reinhardtii*. One of these (CC-1952) was once called *C. grossii*, and another (CC-1373) has long been listed in catalogues as *C. smithii* mt^+ (Harris 1989; Starr and Zeikus 1993), but our data provide strong support for placing both of them within *C. reinhardtii*. Such a reassignment had already been recommended by others, based on the knowledge that both are fully interfertile with laboratory strains of *C. reinhardtii* (E. Harris, personal communication) and also share extremely similar ITS sequences with them (A. Coleman, personal communication; ms. in preparation). The present data reinforce those recommendations.

The *C. reinhardtii* clade is the only one for which the dendrograms generated from the three different intron data sets are not in full agreement. The two trees generated from *ypt4* intron sequences have very similar topology. It is particularly noteworthy that, with a bootstrap value of 100, both of them indicate that strain CC-1952 (which was isolated in the American midwest) is the basal member of the clade, and a sister taxon to the other six strains (all of which were isolated from the east coast region). These two *ypt*-based trees differ only with respect to whether strain CC-2424 or -1373 is identified as the closer relative of strain CC-2343, but in both cases these assignments have low bootstrap values (< 70), indicating that they are not robust. Similarly, although the act-IX-based *C. reinhardtii* tree has a topology quite different from the two *ypt4*-based trees, none of its branchpoints is supported by a high bootstrap value, and three *C. reinhardtii* strains have act-IX sequences so similar that they cannot be resolved by this criterion (Fig. 4 C). We believe that the reason that the act-IX intron does not provide either the same or as robust a *C. reinhardtii* dendrogram as the *ypt* introns is that this intron is significantly shorter, and much of the sequence variability within it is centered in and around its enigmatic repeated element (see Fig. 2). Although the *ypt* introns yield more nearly identical, and generally more robust, dendrograms of the *C. reinhardtii* clade than the act-IX intron does, even the *ypt* introns appear to have accumulated too few nucleotide differences distinguishing strains CC-1373, -2343 and -2424 to resolve the relationships among them unambiguously, presumably because these strains diverged so recently and at so nearly the same time. With a mean nucleotide-identity level of 99.1% (averaged across the two *ypt* introns), these three *C. reinhardtii* strains obviously define the upper limit of resolution of a phylogenetic analysis based on these *ypt* introns.

The lower limit of resolution of the method appears to be defined by the relationship of *C. incerta* to the *C. reinhardtii* clade. We included *C. incerta* in our study because we had learned that sequences of internal transcribed spacers [ITS] of rRNA-encoding nuclear genes of all *Chlamydomonas* species available in world culture collections indicate that *C. incerta* is the closest relative of *C. reinhardtii* (A. Coleman, personal communication; ms. in preparation). A similar conclusion has recently been reached from analysis of a rapidly evolving gene involved in mating-type specification (P. Ferris and U. Goodenough, personal communication). Our data indicate that the relationship between these two species is just above this threshold for statistical significance in the case of the ypt-VI intron, and just at or below this threshold in the case of the other two introns (Figs. 3, 4). Thus, with its average nucleotide identity of 48% to the seven members of the *C. reinhardtii* clade (averaged across all three introns), the relationship of *C. incerta* to *C. reinhardtii* defines the lower limit of resolution of the present method.

(5) The final clade analyzed in this study involved four *C. debaryana* strains, whose relationships to one another had previously been ambiguous. Our interest in this species was stimulated by a report that it appears to be more closely related to the Volvocaceae than *C. reinhardtii* is (Buchheim and Chapman 1991). UTEX strain 579 had originally been catalogued as *C. komma*, but was re-catalogued as *C. debaryana* in 1993, on the basis of morphological re-evaluation (Starr and Zeikus 1993). UTEX strain 344 has always been catalogued as *C. debaryana*, but with the appended suggestion that it is possibly identical to CCAP strains 11/56a and 11/56b (Starr and Zeikus 1993). When all four strains were analyzed here, 344, 11/56a and 11/56b were, indeed, found to be indistinguishable. Although distinguishable from these others, strain 579 was clearly similar enough to them in intron sequences to justify its inclusion in *C. debaryana* (Fig. 4). However, none of the *C. debaryana* strains exhibits any significant intron-sequence homologies to any of the volvocaceans analyzed yet.

Estimated rates of intron evolution

Is it possible to estimate the rate at which introns of the sort studied here accumulate mutations, and from that to estimate over what period of evolutionary diversification intron sequences may be expected to provide

phylogenetically useful information? Our attempt to answer this question (discussed below) produces only a rough first approximation, because it is based on a single reference date, combined with interpolations that rest on untested assumptions. Nevertheless, we believe that it provides a starting point for further analysis.

The reference point for this analysis is the conclusion (based on the extent of rRNA-sequence divergence and the abundance of silent substitution in tubulin genes) that *C. reinhardtii* and *V. carteri* f. *nagariensis* last shared a common ancestor not more than 75 MYA (Rausch et al. 1989). In Fig. 5 A we display a dendrogram based on 253 informative positions of 18S and 26S rRNA sequences for all of the volvocalean algae studied here for which published data were available; the branching pattern and branch lengths within this dendrogram are essentially identical to those obtained by Larson et al. (1992). If we set the branchpoint between the lineages leading to *C. reinhardtii* and the volvocaceans at 75 MYA, and assume a uniform replacement rate within volvocalean rRNAs since then, we can interpolate to estimate that the lineages leading to *V. carteri* f. *weismannia* and *V. carteri* f. *nagariensis* diverged approximately 10 MYA. This divergence, which is robustly resolved by all three intron data sets, is then used as the reference point to establish a time scale for a second dendrogram (Fig. 5 B) derived from the intron sequence data reported here (using mean distance values for each isolate). Once again assuming constant replacement rates, we can then interpolate the approximate divergence times within each of the clades resolved by the intron sequence data.

Because the relationship between *V. carteri* f. *weismannia* and *V. carteri* f. *nagariensis* is robustly reflected in the intron data, but the relationship of *V. obversus* to the *V. carteri* clade is not, we can estimate that the maximum divergence time that can be resolved by these intron sequences lies somewhere between these two branchpoints, that is between about 10 and 15 MYA (Fig. 5 A). Because the relationships among the three most closely related isolates of *C. reinhardtii* lie near the upper limit of resolution of the present method, we estimate from Fig.5 B that the shortest period of divergence that these introns are capable of resolving lies in the vicinity of 100 000 years.

The three introns of *V. carteri* f. *weismannia* and *V. carteri* f. *nagariensis* that were studied here exhibit together a mean value of 325 mutational events within an average total sequence length of 1120 bp. From this, and an estimated divergence time of 10 MY, a rate of 2.9×10^{-8} fixed mutations per site per year (3% sequence variation per MY) can be estimated for these introns. This is some 150-times greater than the mean replacement rate in 16/18S rRNA, and about 10-fold higher than the replacement rate that was previously calculated from a very limited data set of other spliceosomal introns (3.3×10^{-9} mutations per site per year; Li and

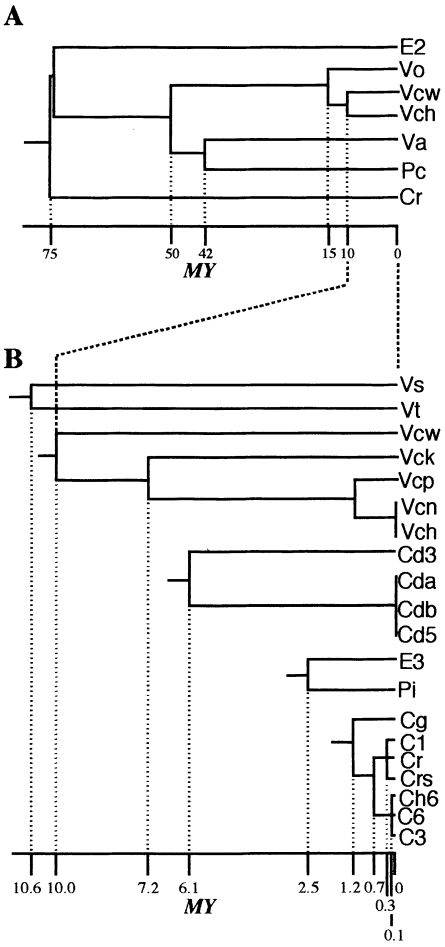

Fig. 5A, B Dendrograms used to estimate the time scale over which the intron sequences analyzed in this study provide phylogenetically useful information. **A** dendrogram based on 253 informative positions of 18S and 26S rRNA for all of the species used in this study for which rRNA data have been published (Larson et al. 1992), with branch lengths proportional to the number of informative interchanges between taxa. The time scale was generated by linear interpolation from the 75-MY time point, which was the estimate generated by Rausch et al. (1989) for the maximum period that has elapsed since *V. carteri* f. *nagariensis* and *C. reinhardtii* last shared a common ancestor. Assuming uniform rates of nucleotide substitutions in rRNAs of all lineages over this time period, this leads to an estimate of 15 MY since the *V. obversus* and *V. carteri* lineages diverged, and about 10 MY since diversification within the *V. carteri* clade began. The maximum period of time over which the introns we have analyzed retain phylogenetically useful data appears to lie in this interval. **B** dendrogram relating five groups of volvocalean algae resolved by *ypt4* intron analysis, with branch lengths indicating mean distance values calculated from the two *ypt4* introns. The time scale was established using the estimate of 10 MYA (obtained above) as the time when *V. carteri* f. *weismannia* diverged from the rest of the *V. carteri* clade; interpolations were again performed assuming that nucleotide replacement rates were uniform in both introns in all lineages. This leads to an estimated minimum time for accumulation of phylogenetically useful numbers of differences in these two introns of approximately 100 000 years (the interval during which the three most closely related *C. reinhardtii* strains appear to have been diverging)

Graur 1991). Had the previously calculated replacement rate of 3.3×10^{-9} been correct for algal introns, and if the divergence time for *V. carteri* and *C. reinhardtii* estimated by Rausch et al. (1989) is even approximately correct, the introns of these two species should presently exhibit 75% sequence identity, which is obviously not the case. Indeed, they exhibit *no* detectable sequence homologies. Only if Rausch et al. (1989) underestimated the time since the divergence of *V. carteri* and *C. reinhardtii* lineages by an order of magnitude, could a replacement rate of 3.3×10^{-9} have resulted in intron sequences as dissimilar as are observed for these taxa. We believe that such an underestimate is highly unlikely, but also realize that further analyses of several types will be required to provide a solid framework in which intron data can be used to calculate reliable time scales for recent phylogenetic events.

In summary, the present studies demonstrate that comparisons of selected spliceosomal nuclear intron sequences provide a means of estimating relationships within and among various volvocine algal groups with a hitherto unprecedented level of resolution. Stated differently, in the case of the volvocaleans at least, intron sequences constitute a ''second hand'' for the molecular clock. It remains to be seen whether they will be found to be of equal phylogenetic utility in the analysis of other groups of organisms.

Acknowledgements We are grateful to Patrick Ferris for providing several of the *Chlamydomonas* samples that were analyzed here, and to Annette Coleman for providing information about certain results of her studies of volvocine ITS sequences prior to publication. This work was supported by DFG grant Fa232/6-1 to S.F., and grant MCB-930447 from the National Science Foundation to D.K.

References

Adams CR, Stamer KA, Miller JK, McNally JG, Kirk MM, Kirk DL (1990) Patterns of organellar and nuclear inheritance among progeny of two geographically isolated strains of *V. carteri*. Curr Genet 18:141–153

Avise JC (1994) Molecular markers, natural history and evolution. Chapman and Hall, New York

Boyle JS, Lew AM (1995) An inexpensive alternative to glassmilk for DNA purification. Trends Genet 11:8

Buchheim MA, Chapman RL (1991) Phylogeny of the colonial green flagellates: a study of 18S and 26S rRNA sequence data. BioSystems 25:85–100

Cann RL, Stoneking M, Wilson AC (1986) Mitochondrial DNA and human evolution. Nature 325:31–36

Cresnar B, Mages W, Müller K, Salbaum JM, Schmitt R (1990) Structure and expression of a single actin gene in *Volvox carteri*. Curr Genet 18:337–346

Denda T, Kosuge K, Watanabe K, Ito M, Suziki Y, Short PS, Yahara T (1995) Intron length variation of the *Adh* gene in *Brachyscombe*, Asteraceae. Plant Mol Biol 28:1067–1073

Dietmaier W, Fabry S (1994) Analysis of the introns in genes encoding small G proteins. Curr Genet 26:497–505

Dietmaier W, Fabry S, Schmitt R (1993) *DISEC*/*TRISEC*: *di*- and *tri*nucleotide-*s*ticky-*e*nd *c*loning of PCR-amplified DNA. Nucleic Acids Res 21:3603–3604

Dietmaier W, Fabry S, Huber H, Schmitt R (1995) Analysis of a family of *ypt* genes and their products from *Chlamydomonas reinhardtii*. Gene 158:41–50

Doolittle RF (1995) The origins and evolution of eukaryotic proteins. Phil Trans R Soc Lond B 349:235–240

Doolittle RF, Feng D-F, Tsang S, Cho G, Little E (1996) Determining divergence times of the major kingdoms of living organisms with a protein clock. Science 271:470–477

Ettl H, Schlösser UG (1992) Towards a revision of the systematics of the genus *Chlamydomonas* (Chlorophyta). 1. *Chlamydomonas applanata* Pringsheim. Bot Acta 105:323–330

Fabry S. Jacobsen A, Huber H, Palme K, Schmitt R (1993) Structure, expression, and phylogenetic relationships of a family of *ypt* genes encoding small G-proteins in the green alga *Volvox carteri*. Curr Genet 24:229–240

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Felsenstein J (1989) PHYLIP – Phylogeny Inference Package, version 3.2. Cladistics 5:164–166

Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. Science 155:279–284

Goldstein M (1964) Speciation and mating behavior in *Eudorina*. J. Protozool 11:317–344

Harper JF, Huson KS, Kirk DL (1987) Use of repetitive sequences to identify DNA polymorphisms linked to *regA*, a developmentally important locus in *Volvox*. Genes Dev 1:573–584

Harris EH (1989) The *Chlamydomonas* sourcebook. A comprehensive guide to biology and laboratory use. Academic Press, San Diego, California

Hiesel R, von Haeseler A, Brennicke A (1994) Plant mitochondrial nucleic acid sequences as a tool for phylogenetic analysis. Proc Natl Acad Sci USA 91:634–638

Hurst LD (1994) The uncertain origin of introns. Nature 371:381–382

Kimura M (1980) A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Kirk DL (1996) *Volvox*. Molecular genetic origins of multicellularity and cellular differentiation. Cambridge University Press, Cambridge (in press)

Kirk DL, Harper JF (1986) Genetic, biochemical, and molecular approaches to *Volvox* development and evolution. Int Rev Cytol 99:217–293

Kirk DL, Kirk MM (1983) Protein synthetic patterns during the asexual life cycle of *Volvox carteri*. Dev Biol 96:493–506

Kochert G (1973) Colony differentiation in green algae. In : Coward RJ (ed) Developmental regulation: aspects of cell differentiation. Academic Press, New York; pp 155–167

Kwiatowski J, Krawczyk M, Kornacki M, Bailey K, Ayala FJ (1995) Evidence against the exon theory of genes derived from the triose-phosphate isomerase gene. Proc Natl Acad Sci USA 92:8503–8506

Larson A, Wilson AC (1989) Patterns of ribosomal RNA evolution in salamanders. Mol Biol Evol 6:131–154

Larson A, Kirk MM, Kirk DL (1992) Molecular phylogeny of the volvocine flagellates. Mol Biol Evol 9:85–105

Li W-H, Graur D (1991) Fundamentals of molecular evolution. Sinauer Associates, Sunderland, Massachusetts

Long M, Rosenberg C, Gilbert W (1995) Intron phase correlations and the evolution of the intron/exon structure of genes. Proc Natl Acad Sci USA 92:12495–12499

McDowell JM, Huang S, McKinney EC, An Y-Q, Meagher RB (1996) Structure and evolution of the actin gene family in *Arabidopsis thaliana*. Genetics 142:597–602

Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino-acid sequences of two proteins. J Mol Biol 48:443–453

Nozaki H (1988) Morphology, sexual reproduction, and taxonomy of *Volvox carteri* f. *kawasakiensis* f. nov. (Chlorophyta) from Japan. Phycologica 27:209–220

Olson GJ, Woese CR (1993) Ribosomal RNA: a key to phylogeny. FASEB J 7:113–123
Rausch H, Larsen N, Schmitt R (1989) Phylogenetic relationships of the green algae *Volvox carteri* deduced from small-subunit ribosomal RNA comparisons. J Mol Evol 29:255–265
Schitt R, Fabry S, Kirk DL (1992) In search of the molecular origins of cellular differentiation in *Volvox* and its relatives. Int Rev Cytol 139:189–265
Singer-Krüger B, Stenmark H, Düsterhöft A, Philippsen P, Yoo J-S, Gallwitz D, Zerial M (1994) Roles of three Rab5-like GTPases, Ypt51p, Ypt52p and Ypt53p, in the endocytic and vacuolar protein-sorting pathway of yeast. J Cell Biol 125:283–298
Smith GM (1944) A comparative study of the species of *Volvox*. Trans Am Micro Soc 63:265–310
Starr RC (1969) Structure, reproduction, and differentiation in *Volvox carteri* f. *nagariensis* Iyengar, strains HK9 and 10. Arch Protistenkd 111:204–222
Starr RC, Zeikus JA (1993) UTEX – the culture collection of algae at the University of Texas at Austin. 1993 list of cultures. J Phycol 29 Supplement:1–106
Stoltzfus A, Spencer DF, Zuker M, Logsdon JM Jr. Doolittle WF (1994) Testing the exon theory of genes: the evidence from protein structure. Science 265:202–207
Sugase Y, Hirono M, Kindle KL, Kamiya R (1996) Cloning and characterization of the actin-encoding gene of *Chlamydomonas reinhardtii*. Gene 168:117–121
Tominaga H, Narise S (1995) Sequence evolution of the Gpdh gene in the *Drosophila viridis* species group. Genetica 96:293–302
Wiebauer K, Herrero J-J, Filipowicz W (1988) Nuclear pre-mRNA processing in plants: distinct modes of 3′ splice site selection in plants and animals. Mol Cell Biol 8:2042–2051
Wilson AC, Ochman H, Prager EM (1987) Molecular time scale for evolution. Trends Genet 3:241–247
Woese C (1987) Bacterial evolution. Microbiol Rev 51:221–271
Zuckerkandl E, Pauling L (1965) Molecules as documents of evolutionary history. J Theoret Biol 8:357–366