J Mol Evol (1995) 41:238–246

JOURNAL OF MOLECULAR EVOLUTION

Phylogenetic Position of *Dictyostelium* Inferred from Multiple Protein Data Sets

Kei-ichi Kuma, Naruo Nikoh, Naoyuki Iwabe, Takashi Miyata

Department of Biophysics, Faculty of Science, Kyoto University, Kyoto 606, Japan

Received: 8 June 1994 / Accepted: 14 October 1994

Abstract. The phylogenetic position of *Dictyostelium* inferred from 18S rRNA data contradicts that from protein data. Protein trees always show the close affinity of *Dictyostelium* with animals, fungi, and plants, whereas in 18S rRNA trees the branching of *Dictyostelium* is placed at a position before the massive radiation of protist groups including the divergence of the three kingdoms. To settle this controversial issue and to determine the correct position of *Dictyostelium,* we inferred the phylogenetic relationship among *Dictyostelium* and the three kingdoms Animalia, Fungi, and Plantae by a maximum-likelihood method using 19 different protein data sets. It was shown at the significance level of 1 SE that the branching of *Dictyostelium* antedates the divergence of Animalia and Fungi, and Plantae is an outgroup of the Animalia-Fungi-*Dictyostelium* clade.

Key words: Cellular slime molds — Animals — Fungi — Plantae — Maximum-likelihood method — Evolution

Introduction

The taxonomy of the cellular slime molds is the arena of a long-standing controversy among biologists: The cellular slime molds have features characteristic of animals, plants, and fungi. According to the five-kingdom system of Margulis and Schwartz (1988), the cellular slime molds belong to neither Animalia nor Plantae but to an independent phylum Acrasiomycota of the kingdom Protoctista. Zoologists called this group mycetozoa and classified them protozoa, while microbiologists classified them a phylum of Fungi called myxomycetes (e.g., Margulis and Schwartz 1988).

Furthermore, the phylogenetic position of *Dictyostelium* inferred from molecular data is currently controversial: Molecular phylogenetic trees inferred from 18S rRNAs show that the branching of *Dictyostelium* antedates the massive radiation of protist groups including the divergence of Animalia, Fungi, and Plantae (McCarroll et al. 1983; Hasegawa et al. 1985; Sogin et al. 1986, 1989; Hendriks et al. 1991; Douglas et al. 1991; Cavalier-Smith 1993). In sharp contrast, all protein data examined to date favor the close affinity of *Dictyostelium* with the three kingdoms (Simmer et al. 1990; Loomis and Smith 1990; Hasegawa et al. 1993).

Generally there may be several reasons for the discrepancy between 18S rRNA trees and protein trees. In rRNA trees, unusual G+C contents in certain lineages have serious effects on the whole tree topology, which often misleads molecular phylogenetic trees (e.g., Hashimoto et al., 1993). On the other hand, protein trees always involve a risk of paralogous comparison. In the two protein data sets out of four analyzed by Loomis and Smith (1990), for example, yeast proteins are probably paralogous. Thus their conclusion may be erroneous at least in the two protein cases.

Even by orthologous comparison, the tree topologies often differ for different proteins used, as recently dem-

Correspondence to: T. Miyata

Table 1. Protein data sets used in the present analysis[a]

Proteins	*Dictyostelium* Acc.	Animalia		Fungi		Plantae		Outgroup	
		Species	Acc.	Species	Acc.	Species	Acc.	Species	Acc.
1. EF2	M26017	HS	X51466	Sc	M59370	Ck	M68064	Hh	X17148
		Dm	X15805						
2 hsp70	S65739	Hs	M11717	Sc	J05637	At	X74604	Hs GRP78*	
		Dm	L01500			Gm	X62799		M19645
3. EF-1α	X55973	Hs	X03558	Sc	M15666	At	X16430	Hm	X16677
		Dm	X06870	Tr	Z23012	Ta	M90077		
4. Acin	X03281	Hs	X04098	Sc	L00026	At	M20016	Hs ARP*	
		Dm	K00670	Ca	X16377	Vc	M33963		Z14978
5. pol-II β′	S52651	Hs	X63564	Sc	X03128	At	X52494	Sc pol-III β′*	
		Dm	M27431	Sp	X56564				X03129
6. hmg	L19350	Hs	M11058	Sc	M22002	At	L19261	Hv	M83531
		Dm	M21329			Rs	X68652		
7. L3	L08391	Hs	X73460	Sp	X57734	At	M32654	Hm	J05222
		Mm	Y00225			Os	D12630		
8. L10	X56194	Hs	M17885	Sc	M26506	Cru	X15206	Hc	X15078
		Mm	X15267						
9. CK-II	L05535	Hs	M55265	Sc	M22473	At	D10247	Gf CDK2*	
		Dm	M16534			Zm	X61387		S40289
10. cdc2	M80808	Hs	X05360	Sc	X00257	At	X57840	Hs p58*	
		Dm	X57485	Sp	M12912	Zm	M60526		M37712
11. ATC	X14634	Hs	M38561	Sc	M27174	Le	X74072	Ec	K01472
		Dm	X04813						
12. L8	X15710	Rr	X62145	Sp	X16392	Le	X64562	Hm	J05222
		Aa	M99055			Nt	X62500		
13. ran	L09720	Hs	M31469	Sc	X71945	Vf	Z24678	Rr rab7p*	
		Gg	X66906						X12535
14. rab7p	U02928	Cf	M35522	Sc	X68144	Gm	L14930	Hs ran*	
		Rr	X12535			Vc	L08131		M31469
15. rab1A	L21009	R^+	X13905	Sp	X52099	At	D01027	Dd rab1B*	
		Ls	X72688	Nc	S51252	Vc	M93438		L21010
16. NDK	J05457	Hs	M36981	Sc	S64016	At	X69373	Bs	M80245
		Dm	X13107			Ps	X71388		
17. eIF-4D	X14970	Hs	M23419	Sc	M63542	Ms	X59441	Sa	X63132
		Gg	M99499	Nc	U02638	Nt	X63543		
18. Profilin	X61581	Dm	M84528	Sc	Y00469	Zm	X73279	Spu PRP*	
						Bv	M65179		S42185
19. Thioredoxin	M91383	Hs	X54539	Sc	M59169	At	Z14084	Ec	K02845
		Gg	J03882	Pc	X76120	Cre	S16090		

[a] Acc., accession number; *, paralogous sequence. Abbreviations of proteins: EF2, elongation factor 2; hsp70, 70-kd heat-shock protein; GRP78, 78-kd glucose-regulated protein; EF1α, elongation factor 1α; ARP, actin-related protein; pol-II β′, RNA polymerase II β′ subunit; pol-III β′, RNA polymerase III β′ subunit; hmg, hydroxymethylglutaryl CoA reductase; L3, ribosomal protein large subunit L3; L10, ribosomal protein large subunit L10; CK-II, casein kinase II; ATC, aspartate transcarbamoylase; L8, ribosomal protein large subunit L8, ran, ras-like protein ran; NDK, nucleoside diphosphate kinase; eIF-4D, eukaryotic initiation factor 4D; PRP, profilin-related protein. Abbreviations of organisms: Hs, *Homo sapiens;* Dm, *Drosophila melanogaster;* Sc, *Saccharomyces cerevisiae;* Ck, *Chlorella kessleri;* Hh, *Halobacterium halobium;* At, *Arabidopsis thaliana;* Gm, *Glycine max;* Tr, *Trichoderma reesei;* Ta, *Triticum aestivum;* Hm, *Halobacterium marismortui;* Ca, *Candida albicans;* Vc, *Volvox carteri;* Sp, *Schizosaccharomyces pombe;* Rs, *Raphanus sativus; Hv, Haloferax volcanii;* Mm, *Mus musculus;* Os, *Oryza sativa;* Cru, *Chenopodium rubrum;* Hc, *Halobacterium cutirubrum;* Zm, *Zea mays;* Gf, goldfish (Unclassified); Le, *Lycopersicon esculentum;* Ec, *Escherichia coli;* Rr, *Rattus rattus;* Aa, *Aedes albopictus;* Nt, *Nicotiana tabacum;* Gg, *Gallus gallus;* Vf, *Vicia faba;* Cf, *Canis familiaris;* R^+, *rattus species;* Ls, *Lymnaea stagnalis;* Nc, *Neurospora crassa;* Ps, *Pisum sativum;* Bs, *Bacillus subtilis;* Ms, *Medicago sativa;* Sa, *Sulfolobus acidocaldarius;* Bv, *Betula verrucosa;* Spu, *Strongylocentrotus purpuratus;* Pc, *Penicillium chrysogenum;* Cre, *Chlamydomonas reinhardtii*

onstrated by 23 protein data sets in inferring phylogenetic relationships among Animalia, Fungi, and Plantae (Nikoh et al. 1994). It is therefore required for inferring reliable tree topologies to use a large number of protein data sets, but not a single protein data set, and to synthesize all the results obtained from different data sets, based on a statistically solid background. The extended version of the maximum-likelihood method recently developed by Hasegawa's group (Kishino and Hasegawa 1989; Kishino et al. 1990; Adachi and Hasegawa 1992) may have an advantage for this purpose. Using 23 protein data sets, we recently showed the close relatedness

Table 2. The difference Δli of log-likelihood li of tree i (i = 1–15) from that $l$max of the maximum-likelihood tree and its standard error ± SE and bootstrap probability p_i calculated for each of 19 different protein data sets[a]

		Proteins no.								
		Total	1	2	3	4	5	6	7	8
		No. of sites compared								
		5,462	700	597	421	370	360	343	330	296
		$l$max								
		−53,072.9	−6,801.5	−4889.9	−3,715.8	−2,494.8	−3,740.5	−3,649.3	−3,148.3	−3,201.6
Tree 1	Δl_1	−35.8	−4.0	−5.5	0.0	−4.2	0.0	−1.2	−1.7	−0.9
	SE	21.3	6.5	9.9	0.0	3.6	0.0	4.0	3.0	1.6
((AF)(PD))	p_1	0.004	0.071	0.061	0.447	0.028	0.689	0.168	0.028	0.069
Tree 2	Δl_2	0.0	0.0	0.0	−4.1	0.0	−8.1	0.0	−1.3	0.0
	SE	0.0	0.0	0.0	4.7	0.0	8.9	0.0	6.1	0.0
(((AF)D)P)	p_2	0.859	0.417	0.343	0.082	0.513	0.171	0.386	0.169	0.392
Tree 3	Δl_3	−46.7	−7.1	−10.1	−2.3	−4.2	−15.2	−1.4	−2.4	−1.0
	SE	19.7	5.4	8.4	5.7	3.6	7.3	3.9	5.7	1.5
(((AF)P)D)	p_3	0.006	0.024	0.009	0.284	0.009	0.000	0.205	0.102	0.049
Tree 4	Δl_4	−187.6	−19.2	−14.4	−30.8	−14.7	−61.9	−12.6	−9.3	−4.2
	SE	37.5	14.7	13.6	13.2	8.7	17.1	7.1	8.0	5.1
((AP)(FD))	p_4	0.000	0.000	0.006	0.00	0.000	0.000	0.001	0.004	0.010
Tree 5	Δl_5	−156.2	−20.5	−4.4	−30.7	−9.8	−58.0	−10.1	−10.0	−2.2
	SE	39.2	14.5	15.6	13.2	9.8	16.4	7.5	8.0	5.7
(((AP)D)F)	p_5	0.000	0.003	0.186	0.000	0.010	0.000	0.020	0.003	0.159
Tree 6	Δ	−163.6	−21.2	−14.8	−23.0	−14.6	−60.1	−9.9	−8.4	−3.9
	SE	37.1	13.7	14.0	13.2	8.4	17.0	7.2	8.2	4.9
(((AP)F)D)	p_6	0.000	0.004	0.004	0.009	0.000	0.000	0.018	0.012	0.030
Tree 7	Δl_7	−202.2	−28.9	−24.7	−30.4	−8.8	−64.7	−11.3	−10.1	−5.9
	SE	35.5	12.8	12.2	13.3	10.0	16.5	7.2	8.3	4.3
((AD)(PF))	p_7	0.000	0.000	0.000	0.000	0.006	0.000	0.003	0.002	0.001
Tree 8	Δl_8	−159.5	−24.6	−13.2	−30.2	−5.8	−60.6	−9.6	−9.1	−3.6
	SE	37.3	13.5	13.8	13.2	10.6	15.8	7.6	8.4	5.2
(((AD)P)F)	p_8	0.000	0.003	0.002	0.000	0.190	0.000	0.028	0.016	0.018
Tree 9	Δl_9	−128.5	−17.8	−11.6	−28.9	−6.0	−55.2	−8.5	−6.7	−4.0
	SE	29.4	10.9	7.6	13.1	9.2	17.4	6.4	8.7	3.4
(((AD)F)P)	p_9	0.000	0.014	0.008	0.002	0.107	0.000	0.036	0.046	0.021
Tree 10	Δl_{10}	−210.0	−31.1	−22.6	−31.3	−13.7	−62.8	−13.8	−10.0	−6.0
	SE	36.0	12.8	12.8	13.0	8.9	16.4	6.7	6.7	4.3
(((FP)D)A)	p_{10}	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.001	0.002
Tree 11	Δl_{11}	−196.8	−18.1	−17.4	−33.8	−14.7	−61.4	−13.7	−7.2	−6.0
	SE	36.3	14.9	12.3	12.2	8.7	16.9	6.7	7.5	4.3
((FD)P)A))	p_{11}	0.000	0.004	0.003	0.000	0.000	0.000	0.000	0.042	0.000
Tree 12	Δl_{12}	−62.6	−8.2	−11.2	−3.2	−9.5	−4.0	−5.9	0.0	−3.6
	SE	28.3	9.4	12.5	3.0	5.4	3.7	6.1	0.0	3.2
(((PD)F)A)	p_{12}	0.000	0.024	0.008	0.041	0.002	0.122	0.005	0.335	0.010
Tree 13	Δl_{13}	−184.7	−29.1	−21.5	−23.3	−13.5	−62.5	−10.7	−10.0	−5.6
	SE	35.3	12.3	12.6	13.1	8.6	16.5	6.9	7.9	4.1
(((FP)A)D)	p_{13}	0.000	0.000	0.001	0.009	0.000	0.000	0.008	0.005	0.002

Table 2. Continued

Proteins no.										
9	10	11	12	13	14	15	16	17	18	19
No. of sites compared										
277	261	259	223	192	176	158	144	133	123	99
*l*max										
−2,583.0	−3,024.4	−3,101.5	−2,168.4	−1,571.3	−1,738.0	−1,041.6	−1,511.8	−1,520.2	−1,531.0	−1,578.7
−4.5 3.5 0.004	−7.5 6.6 0.006	−24.3 9.9 0.000	−5.3 4.2 0.003	−9.8 8.8 0.002	−4.3 3.8 0.002	−2.9 5.1 0.005	−2.8 9.6 0.176	−8.7 6.1 0.000	−4.8 5.2 0.009	−4.9 3.5 0.001
−3.8 3.9 0.101	−7.5 6.7 0.010	−7.7 5.2 0.002	−2.7 6.6 0.175	−5.5 6.7 0.120	−0.5 5.8 0.214	−4.7 4.2 0.001	−3.4 9.6 0.066	−4.7 5.1 0.053	−3.4 6.7 0.117	−4.0 3.9 0.003
0.0 0.0 0.635	−1.4 4.5 0.247	−25.2 9.7 0.000	−6.6 4.9 0.000	−11.6 7.9 0.000	−5.1 4.6 0.004	0.0 0.0 0.021	−3.2 9.6 0.099	−4.6 7.2 0.066	−4.3 6.4 0.083	−2.6 2.5 0.047
−10.9 8.2 0.001	−6.1 4.3 0.001	−14.9 7.9 0.005	−7.1 5.7 0.006	−12.4 7.4 0.002	−6.7 5.3 0.001	−4.0 3.3 0.001	−2.2 3.5 0.116	−10.0 6.2 0.000	−6.4 6.6 0.000	−1.4 2.0 0.034
−10.0 8.4 0.028	−3.8 5.2 0.100	−18.1 9.4 0.002	−7.2 5.6 0.005	−12.1 7.3 0.001	−8.3 5.5 0.000	−4.0 3.3 0.000	0.0 0.0 0.371	−7.0 6.1 0.012	−0.2 3.8 0.265	−1.3 2.1 0.076
−7.8 7.6 0.083	0.0 0.0 0.370	−19.4 9.2 0.000	−7.2 5.7 0.006	−14.0 7.2 0.000	−8.1 5.8 0.001	0.0 0.0 0.122	−2.3 3.4 0.078	−5.2 7.2 0.030	−5.1 6.9 0.040	0.0 0.0 0.464
−14.0 8.3 0.001	−8.1 6.2 0.000	−9.3 10.5 0.072	−7.2 5.2 0.001	−5.3 3.6 0.005	−5.2 6.4 0.000	−2.4 4.8 0.030	−11.2 6.9 0.000	−4.3 3.2 0.001	−8.8 5.8 0.000	−3.1 4.7 0.014
−13.5 8.3 0.002	−5.4 6.5 0.051	−12.0 9.7 0.001	−7.7 5.2 0.000	−4.6 4.1 0.071	−5.6 6.4 0.002	−2.4 4.8 0.057	−6.2 4.8 0.003	−2.6 3.8 0.080	−1.7 2.7 0.063	−2.4 4.7 0.152
−13.8 8.0 0.002	−8.4 6.4 0.002	−2.5 7.1 0.294	−3.3 6.8 0.051	0.0 0.0 0.681	−1.5 7.4 0.134	−2.4 4.9 0.140	−10.8 7.1 0.000	0.0 0.0 0.401	−5.7 7.0 0.007	−2.8 4.8 0.080
−13.8 8.4 0.001	−7.2 6.5 0.019	−14.0 11.1 0.040	−0.8 2.0 0.186	−8.0 4.5 0.000	−0.1 4.5 0.251	−4.7 4.2 0.000	−11.0 7.0 0.002	−6.2 3.6 0.000	−9.4 5.6 0.000	−4.9 3.5 0.000
−12.7 8.4 0.003	−8.1 5.7 0.004	−14.7 7.9 0.011	−1.1 1.8 0.060	−12.3 7.4 0.003	−2.7 3.5 0.017	−4.6 4.2 0.002	−8.4 6.6 0.003	−10.7 5.9 0.000	−7.6 6.6 0.000	−3.1 3.0 0.013
−6.3 4.8 0.033	−7.9 6.6 0.009	−21.4 9.9 0.000	0.0 0.0 0.417	−9.6 8.9 0.026	0.0 0.0 0.231	−2.9 5.1 0.001	−8.9 7.6 0.002	−10.6 5.8 0.000	−6.0 4.4 0.002	−5.0 3.4 0.000
−9.3 7.4 0.052	−2.3 4.0 0.105	−17.0 11.2 0.000	−7.1 5.3 0.002	−9.1 4.4 0.000	−5.0 6.6 0.022	0.0 0.0 0.450	−9.9 7.0 0.001	−0.5 5.3 0.334	−7.3 6.2 0.000	−2.6 2.6 0.040

Table 2. Continued

		Proteins no.								
		Total	1	2	3	4	5	6	7	8
		No. of sites compared								
		5,462	700	597	421	370	360	343	330	296
		*l*max								
		−53,072.9	−6,801.5	−4889.9	−3,715.8	−2,494.8	−3,740.5	−3,649.3	−3,148.3	−3,201.6
Tree 14	Δl_{14}	−124.7	−7.4	−7.6	−30.9	−10.4	−53.6	−9.8	−5.0	−3.7
	SE	29.8	13.0	8.7	12.4	7.8	17.7	5.8	8.7	3.6
(((FD)A)P)	p_{14}	0.000	0.240	0.058	0.000	0.003	0.000	0.006	0.106	0.021
Tree 15	Δl_{15}	−32.0	−4.9	−0.2	−3.0	−3.9	−4.6	−4.2	−1.3	−1.0
	SE	30.1	10.0	14.3	3.1	6.7	3.4	6.7	3.3	4.4
(((PD)A)F)	p_{15}	0.131	0.196	0.309	0.126	0.132	0.018	0.116	0.129	0.216

[a] The total values of $\Delta li \pm$ SE and p_i are also shown. The values of $\Delta li \pm$ SE and p_i of tree i are boxed in case of $|\Delta li| < 1$ SE

of Animalia and Fungi, and Plantae is an outgroup of the Animalia-Fungi clade (Nikoh et al. 1994).

Applying the same method to 19 different protein data sets, we here show with statistical confidence that *Dictyostelium* is closely related to the Animalia-Fungi clade and is distantly related to Plantae.

Materials and Methods

To know the phylogenetic position of *Dictyostelium*, the amino acid sequence was compared with those from animals, fungi, and plants, together with that of an outgroup for each of 19 different protein species. The data sets used in the present analysis were listed in Table 1. All the sequence data were taken from Genbank release 80.0.

Optimal alignments of sequences were obtained by the methods of Needleman and Wunsch (1970) and Berger and Munson (1991), together with manual inspections. The aligned sequences were applied to phylogenetic tree inferences for regions where unambiguous alignment is possible.

The method used in the present analysis is essentially identical to that by Nikoh et al. (1994). To determine an outgroup closest to animals, fungi, plants, and *Dictyostelium*, and to exclude a possibility of paralogous comparison, an unrooted tree was inferred by the neighbor-joining method (Saitou and Nei 1987) for each protein data set, including many sequences from a wide range of species available from database. On the basis of the unrooted tree, we determined an outgroup and selected one or two species for each kingdom as representatives, as shown in Table 1.

For each set of protein sequence data, the phylogenetic tree was inferred by the maximum-likelihood (ML) method of protein sequence (Kishino et al. 1990; Adachi and Hasegawa 1992) based on the JTT model (PROTML version 1.10 in Adachi and Hasegawa's program package MOLPHY). To evaluate the statistical significance of tree topologies inferred by the ML method, we calculated the difference Δli of log-likelihood of tree i from that of the ML tree and the standard error (SE) by the method of Kishino and Hasegawa (1989). A bootstrap probability for a particular tree being the highest-likelihood tree among the alternatives during bootstrap resamplings (Felsenstein 1985) was estimated approximately by the RELL (resampling estimated log-likelihood) method (Kishino et al. 1990). We also calculated the overall value of log-likelihoods of the 19 different protein data sets and that of bootstrap probabilities (Kishino et al. 1990).

Results

Based on the ML method of protein phylogeny developed by Kishino et al. (1990), the difference Δli (= li − $l$max) of log-likelihood li of a tree i (i = 1 − 15) from that $l$max of the ML tree and its bootstrap probability p_i were calculated for each of the 19 different protein data sets listed in Table 1. The results were summarized in Table 2. No data set suggested a unique tree that is significant at the level of 1SE; several alternative trees are possible within the confidence interval in all the cases examined here.

The ML method has advantages over other known tree-making methods in that it allows one to synthesize results on tree topologies inferred from different protein data sets: It is possible to estimate the total values of log-likelihoods and bootstrap probabilities of different data sets, and thus the reliability of a particular tree topology can be evaluated overall (Kishino et al. 1990). Furthermore, the reliability of inferred tree topologies can be evaluated on a solid statistical background (Kishino and Hasegawa 1989). The estimated total value of log-likelihoods and bootstrap probabilities of the 19 protein data sets were also shown in Table 2.

Judging from the total value of log-likelihood, the ML method strongly favors tree 2, representing the phylogenetic relationship (((Animalia, Fungi), *Dictyostelium*), Plantae). The total value of log-likelihoods of tree 2 is significantly higher than that of tree 15 (((*Dictyostelium*, Plantae), Animalia), Fungi), the second ML tree with $\Delta l_{15} = -32.0 \pm 30.1$ (Fig. 1). In the 14 cases out of 19 data sets, the values of Δl_2 of tree 2 are in the confidence interval, although tree 2 is the ML tree only in five cases (Table 2). In addition, tree 2 has the highest value (= 0.86) of total bootstrap probability, which is remarkably higher than that of the tree 15, the second largest (= 0.13). In the remaining 13 trees, the corresponding values are negligibly small. Furthermore, an analysis by

Table 2. Continued

Proteins no.										
9	10	11	12	13	14	15	16	17	18	19
No. of sites compared										
277	261	259	223	192	176	158	144	133	123	99
*l*max										
−2,583.0	−3,024.4	−3,101.5	−2,168.4	−1,571.3	−1,738.0	−1,041.6	−1,511.8	−1,520.2	−1,531.0	−1,578.7
−11.6	−8.5	0.0	−3.3	−7.4	−1.9	−4.6	−8.0	−5.5	−4.4	−2.6
8.2	5.7	0.0	6.9	6.2	6.8	4.3	6.8	4.8	7.7	3.5
0.012	0.001	0.573	0.069	0.034	0.113	0.007	0.009	0.019	0.085	0.072
−5.7	−5.2	−22.7	−5.2	−9.1	−4.3	−2.9	−4.2	−7.0	0.0	−4.0
4.8	6.7	10.2	4.2	8.8	3.7	5.1	5.9	5.8	0.0	3.6
0.042	0.075	0.000	0.019	0.055	0.008	0.163	0.074	0.004	0.329	0.004

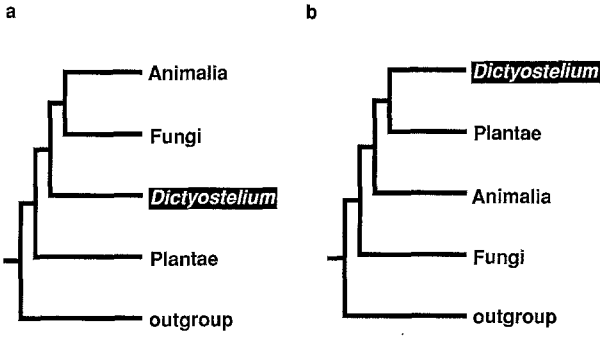

Fig. 1. The maximum-likelihood tree and an alternative tree inferred from 19 different protein data sets. **a** The ML tree with the maximum value of total log-likelihood (*l*max) of −53,072.9 and total bootstrap probability of 0.86. This tree corresponds to tree 2 of Table 2. **b** An alternative tree (tree 15 of Table 2) with the second-highest values for both the total log-likelihood ($\Delta l_{15} = l_{15} - l\max = -32.0 \pm 30.1$) and total bootstrap probability ($p_{15} = 0.13$). Note that the total log-likelihood is significantly higher in **a** than in **b** at the level of 1 SE.

maximum parsimony (MP) method (PROTPARS in Felsenstein's program package PHYLIP, version 3.5c) using the same data sets again favors tree 2. (The total bootstrap probability is 0.62.)

In 18 rRNA trees reported to date, the branching of *Dictyostelium* antedates the divergence of Animalia, Fungi, and Plantae (McCarroll et al. 1983; Hasegawa et al. 1985; Sogin et al. 1986, 1989; Hendriks et al. 1991; Douglas et al. 1991; Cavalier-Smith 1993). This branching pattern of *Dictyostelium* is strongly excluded by the present analysis; the total bootstrap probabilities of three trees (trees 3, 6, and 13 of Table 2), all of which represent *Dictyostelium* as an outgroup of the three kingdoms, are very low, — 0.006, 0.0, and 0.0, respectively.

According to 18S rRNA trees, *Plasmodium falciparum* represents a closer affinity with Animalia, Fungi, and Plantae than *Dictyostelium* does (Sogin et al. 1989; Cavalier-Smith 1993). We have reexamined the phylogenetic relationships among Animalia, Fungi, Plantae, *Dictyostelium,* and *Plasmodium* by multiple protein sequences. Although only five protein data sets are available at present, the ML analysis strongly favors the earliest divergence of *Plasmodium* among the five groups at the confidence limit of 1 SE: The inferred ML tree among the five groups is ((((Animalia, Fungi), *Dictyostelium*), Plantae), *Plasmodium*) (Table 3).

Because distantly related sequences were used as outgroups in the present analysis, the phylogenetic relationships among Animalia, Fungi, Plantae, and *Dictyostelium* were also reexamined by using a *Plasmodium* sequence as an outgroup, based on the same data set shown in Table 3. As shown in Table 4a, the ML analysis confirmed the tree (((Animalia, Fungi), *Dictyostelium*), Plantae) at the level of 1 SE. The same result was also obtained, even when two sequences, a *Plasmodium* sequence and a sequence used as an outgroup in Table 3, were used as outgroups for each protein data set (Table 4b).

From these results we conclude that the branching of *Dictyostelium* antedates the divergence of the Animalia-Fungi clade, and Plantae is an outgroup of the Animalia-Fungi-*Dictyostelium* clade. This result is also consistent with our previous conclusion that Plantae is an outgroup of Animalia and Fungi (Nikoh et al. 1994).

Discussion

From an analysis of 19 different protein data sets by the ML method, together with that by the MP method, we here showed the closer affinity of *Dictyostelium* to the Animalia-Fungi clade than to Plantae. None of the protein data sets, however, gives any significant preference for this tree topology, and several alternative trees cannot be excluded at the significance level of 1 SE. This suggests the importance of analysis based on a large number

Table 3. The maximum-likelihood analysis for the phylogenetic relationships among *Dictyostelium*, Animalia, Plantae, and *Plasmodium*[a]

Proteins	No. of sites compared	Animalia	Plantae	Outgroup	*l*max	Δ*l*i		
						Tree 1 (((D,A),P),Pf)	Tree 2 ((A,(D,P)),Pf)	Tree 3 ((A,D),(P,Pf))
1 hsp70	607	Hs, Dm	At, Gm	Hs GRP78	−5,065.4	ML	−6.8 ± 10.3	−8.2 ± 8.8
2 EF-1α	421	Hs, Dm	At, Ta	Hm	−3,630.3	ML	−3.2 ± 11.5	−11.1 ± 7.2
3 pol-II β′	373	Hs, Dm	At	Sc pol-III β′	−3,594.7	−6.8 ± 8.5	−5.2 ± 7.3	−10.2 ± 8.0
4 Actin	371	Hs, Dm	At, Vc	Hs ARP	−2,581.2	ML	−15.1 ± 8.2	−7.1 ± 4.1
5 cdc2	260	Hs, Dm	At, Zm	Hs p58	−2,774.4	−4.5 ± 6.9	−2.2 ± 8.5	−5.8 ± 4.1
Total	2,032							
Δ*L*i					−17,657.4	ML	−21.1 ± 18.8	−31.0 ± 13.7
P_i						0.83	0.14	0.00

[a] Δ*l*i = *l*i − *l*max, where *l*i and *l*max are the log-likelihood of tree i and that of the maximum-likelihood tree, respectively. For each protein datum, the values of Δ*l*i and *l*max are shown only for the highest three trees out of 15 possible trees. ML, the maximum-likelihood tree with the highest log-likelihood value (i.e., Δ*l*i = 0.0). D, *Dictyostelium;* A, Animalia; P, Plantae; Pf, *Plasmodium falciparum.* In "Total" the total values of five data sets are shown; Δ*L*i = *L*i −*L*max, where *L*i = Σ*l*i, the total value of log-likelihoods of tree i over five data, and *L*max (= −17,657.4) is the total log-likelihood of ML tree; *pi*, total bootstrap probability. Abbreviations: EF-1α, elongation factor-1α; pol-II β′, RNA polymerase II β′ subunit; GRP78, 78-kd glucose-regulated protein; pol-III β′, RNA polymerase III β′ subunit; ARP, actin-related protein; p58, protein kinase p58; Hs, *Homo sapiens;* Dm, *Drosophila melanogaster,* At, *Arabidopsis thaliana;* Gm, *Glycine max;* Ta, *Triticum aestivum;* Hm, *Halobacterium maris-mortui;* Sc, *Saccharomyces cerevisiae;* Vc, *Volvox carteri;* Zm, *Zea mays*

Table 4. Phylogenetic relationships among Animalia, Fungi, Plantae, and *Dictyostelium* inferred by maximum-likelihood method using **(a)** a *Plasmodium* sequence and **(b)** a *Plasmodium* sequence and a distantly related sequence as outgroups, respectively[a]

Proteins	Outgroup	*l*max	Δ*l*i		
			Tree 1 (((A,F),D),P)	Tree 2 ((A,F),(D,P))	Tree 3 (((A,F),P),D)
a)					
hsp70	*Plasmodium*	−5,116.6	ML	−5.0 ± 11.5	−8.9 ± 10.2
pol-II β′	*Plasmodium*	−6,617.7	−9.1 ± 8.8	−4.3 ± 5.1	−10.5 ± 8.3
EF-1α	*Plasmodium*	−3,645.3	ML	−12.9 ± 13.1	−16.4 ± 12.1
Actin	*Plasmodium*	−2,272.9	ML	−9.2 ± 6.4	−8.6 ± 6.7
cdc2	*Plasmodium*	−2,865.1	ML	−4.9 ± 4.2	−2.2 ± 5.5
Total					
Δ*L*i		−20,526.6	ML	−27.3 ± 20.0	−37.5 ± 18.5
P_i			0.87	0.06	0.01
b)					
hsp70	*Plasmodium,* Hs GRP78	−5,484.2	ML	−13.6 ± 10.4	−15.7 ± 9.5
pol-II β′	*Plasmodium,* Sc pol-III β′	−4,147.6	ML	−1.1 ± 5.4	−4.0 ± 4.2
EF-1α	*Plasmodium,* Hm	−4,172.2	ML	−3.8 ± 12.7	−10.9 ± 11.2
Actin	*Plasmodium,* Hs ARP	−2,844.3	ML	−12.1 ± 7.6	−11.9 ± 7.7
cdc2	*Plasmodium,* Hs p58	−3,389.4	−3.3 ± 6.6	−2.4 ± 6.8	−0.1 ± 4.7
Total					
Δ*L*i		−20,040.9	ML	−29.7 ± 19.2	−39.4 ± 17.6
P_i			0.93	0.04	0.01

[a] Δ*l*i = *l*i − *l*max, where *l*i and *l*max are the log-likelihood of tree i (i = 1–15) and that of the maximum-likelihood tree, respectively. For each protein datum, the values of Δ*l*i and *l*max are shown only for the highest three trees among 15 possible trees. ML, the maximum-likelihood tree with the highest log-likelihood value (i.e., Δ*l*i = 0.0). In "Total" the total values of five data are shown; Δ*L*i = *L*i −*L*max, where *L*i = Σ*l*i, the total value of log-likelihoods of tree i over 5 data sets, and *L*max is the total log-likelihood of ML tree; P_i, total bootstrap probability. Sequence data for A (Animalia), F (Fungi), P (Plantae), and D (*Dictyostelium*) are the same as those used in Table 3. Abbreviations: GRP78, 78-kd glucose-regulated protein; pol-III β′, RNA polymerase III β′ subunit; ARP, actin-related protein; p58, protein kinase p58; Hs, *Homo sapiens;* Sc, *Saccharomyces cerevisiae;* Hm, *Halobacterium maris-mortui*

of protein data sets for the robust inference of phylogenetic tree.

In the present analysis, we used only one or two species as representatives of each kingdom. It may therefore be required to test the robustness of phylogenetic trees inferred from such small numbers of representatives. Recently we have inferred the phylogenetic relationship among vertebrates, echinoderms, arthropods, and mollusks from 11 mitochondrial DNA-coded proteins, using five species for vertebrates, three species for echinoderms, three species for arthropods, and three species for mollusks. We also carried out the same analysis using

two species for vertebrates, two species for echinoderms, two species for arthropods, and one species for mollusks, and for each tree topology the total values of log-likelihoods were compared between the two cases. A remarkable correlation was observed between the two cases (the correlation coefficient is 0.99), although the correlation was not always strong in each protein data set (Nikoh et al., manuscript in preparation). This suggests that even with such small numbers of representatives as one or two species, the robust inference of tree topology may be possible if a large body of protein data is used, although the result should be confirmed by many data before final conclusion.

Protein trees always involve a risk of paralogous comparison, and thus protein sequences from organisms should be chosen carefully. Yeast sequences for dihydroorotase and orotate phosphoribosyltransferase used by Loomis and Smith (1990) are probably paralogous, and thus their conclusion that *Dictyostelium* represents the closest association with animals may be erroneous at least in the two cases. In the present analysis, an unrooted tree based on a protein data set including many sequences from a variety of organisms was inferred by neighbor-joining method as a first step, by which paralogous sequences were excluded in the final comparisons.

The phylogenetic position of *Dictyostelium* revealed by the present analysis would provide a unique opportunity for understanding a possible relationship between evolution of multicellular organisms and diversification of genes associated with cell-cell communication. *Dictyostelium* is a model organism for cell-cell communication, cell growth, and differentiation in multicellular organisms. In *Dictyostelium,* a series of developmental processes is initiated by the secretion of cAMP, which attracts nearby cells, which leads to the formation of a multicellular organism. Aggregated cells respond by cAMP and by relaying the signal through receptor-mediated activation of a signal transduction system similar to those of higher animals (e.g., Johnson et al. 1992; Cubbit et al. 1992). The cAMP receptor has already been cloned from *Dictyostelium* and has been shown to be a member of the G protein–coupled receptor superfamily (Klein et al. 1988). A phylogenetic tree of the superfamily revealed an extensive diversification of the family members interacting with various ligands in the early evolution of metazoa after the separation from *Dictyostelium.* A similar pattern of divergence was also found in the G protein superfamily and phospholipase C superfamily (Iwabe et al., manuscript in preparation). Interestingly, in each of the superfamilies, the diversification of genes occurred independently in each lineage of *Dictyostelium* and metazoa from a single precursor that is shared between them. This strongly suggests a possible link between evolution of multicellular organisms and the diversification of genes with functions related to cell-cell interactions.

Acknowledgments. We thank Prof. M. Hasegawa for kindly providing us his program package of the maximum-likelihood method. We also thank Prof. M. Sogin for his critical reading of the manuscript. This work was supported by grants from the Ministry of Education, Science and Culture of Japan.

References

Adachi J, Hasegawa M (1992) Computer science monographs, No. 27, MOLPHY: programs for molecular phylogenetics I. PROTML: maximum likelihood inference of protein phylogeny. Institute of Statistical Mathematics, Tokyo

Berger MP, Munson PJ (1991) A novel randomized iterative strategy for aligning multiple protein sequences. CABIOS 7:479–484

Cavalier-Smith T (1993) Kingdom Protozoa and its 18 phyla. Microbiol Rev 57:953–994

Cubitt AB, Carrel F, Dharmawardhane S, Gaskins C, Hadwiger J, Howard P, Mann SKO, Okaichi K, Zhou K, Firtel RA (1992) Molecular genetic analysis of signal transduction pathways controlling multicelluar development in Dictyostelium. Cold Spring Harbor Symp Quanti Biol LVII:177–192

Douglas SE, Murphy CA, Spencer DF, Gray MW (1991) Cryptomonad algae are evolutionary chimaeras of two phylogenetically distinct unicellular eukaryotes. Nature 350:148–151

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

Hasegawa M, Iida Y, Yano T, Takaiwa F, Iwabuchi M (1985) Phylogenetic relationships among eukaryotic kingdoms inferred from ribosomal RNA sequences. J Mol Evol 22:32–38

Hasegawa M, Hashimoto T, Adachi J, Iwabe N, Miyata T (1993) Early branchings in the evolution of eukaryotes: ancient divergence of *Entamoeba* that lacks mitochondria revealed by protein sequence data. J Mol Evol 36:380–388

Hashimoto T, Nakamura Y, Nakamura F, Shirakura T, Adachi J, Goto N, Okamoto K, Hasegawa M (1994) Protein phylogeny gives a robust estimation for early divergences of eukaryotes: phylogenetic place of a mitochondria-lacking protozoan, *Giardia lamblia.* Mol Biol Evol 11:65–71

Hendriks L, De Baere R, Van de Peer Y, Neefs J, Goris A, De Wachter R (1991) The evolutionary position of the rhodophyte *Porphyra umbilicalis* and the basidiomycete *Leucosporidium scottii* among other eukaryotes as deduced from complete sequences of small ribosomal subunit RNA. J Mol Evol 32:167–177

Johnson RL, Gundersen R, Hereld D, Pitt GS, Tugendreich S, Saxe III CL, Kimmel AR, Devreotes PN (1992) G-protein-linked signaling pathways mediate development in Dictyostelium. Cold Spring Harbor Symp Quanti Biol LVII:169–176

Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J Mol Evol 29:170–179

Kishino H, Miyata T, Hasegawa M (1990) Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. J Mol Evol 30:151–160

Klein PS, Sun TJ, Saxe III CL, Kimmel AR, Johnson RL, Devreotes PN (1988) A chemoattractant receptor controls development in *Dictyostelium discoideum.* Science 241:1467–1472

Loomis WF, Smith DW (1990) Molecular phylogeny of *Dictyostelium discoideum* by protein sequence comparison. Proc Natl Acad Sci USA 87:9093-9097

Margulis L, Schwartz KV (1988) Five kingdoms: an illustrated guide to the phyla of life on earth, 2nd ed. WH Freeman, New York

McCarroll R, Olsen GJ, Stahl YD, Woese CR, Sogin ML (1983) Nucleotide sequence of the *Dictyostelium discoideum* small-subunit

ribosomal ribonucleic acid inferred from the gene sequence: evolutionary implications. Biochemistry 22:5858–5868

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

Nikoh N, Hayase N, Iwabe N, Kuma K, Miyata T (1994) Phylogenetic relationship of the kingdoms Animalia, Plantae, and Fungi inferred from twenty-three different protein species. Mol Biol Evol 11:762–768

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Simmer JP, Kelly RE, Rinker AG, Zimmermann BH, Scully JL, Kim H, Evans DR (1990) Mammalian dihydroorotaste: nucleotide sequence, peptide sequences, and evolution of the dihydroorotase domain of the multifunctional protein CAD. Proc Natl Acad Sci USA 87:174–178

Sogin ML, Elwood HJ, Gunderson JH (1986) Evolutionary diversity of eukaryotic small-subunit rRNA genes. Proc Natl Acad Sci USA 83:1383–1387

Sogin ML, Gunderson JH, Elwood HJ, Alonso RA, Peattie DA (1989) Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia.* Science 243:75-77