João Marcelo Pereira Alves

**Caracterização e Filogenia Moleculares
de *Acanthamoeba***

Tese apresentada ao Instituto de
Ciências Biomédicas da Universidade
de São Paulo para obtenção do título
de Doutor em Ciências

Área de concentração: Biologia da
Relação Patógeno-Hospedeiro

Orientadora: Maria Heloiza
Trebilcock Affonso

São Paulo

2.001

DADOS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
Serviço de Biblioteca e Informação Biomédica do
Instituto de Ciências Biomédicas da Universidade de São Paulo

Alves, João Marcelo Pereira.

Caracterização e filogenia moleculares de *Acanthamoeba* / João Marcelo Pereira Alves. -- São Paulo, 2001.

Tese (Doutorado) -- Instituto de Ciências Biomédicas da Universidade de São Paulo. Departamento de Parasitologia.

Área de concentração: Biologia da Relação Patógeno-Hospedeiro.

Linha de pesquisa: Biologia Molecular de amebas de vida livre.

Orientador: Affonso, Maria Heloiza Trebilcock.

Versão do título para o inglês: Molecular characterization and phylogeny of *Acanthamoeba*.

Descritores: 1. *Acanthamoeba* 2. Filogenia 3. Identificação de organismo 4. RNA ribossômico 5. Sequenciamento de DNA

ICB/SBIB053/2001

*"We're off to the witch,
We may never, never, never come home.
But the magic that we'll feel
Is worth a lifetime."*

Ronnie James Dio e cols., 1984

Prefácio

“Esta tese foi elaborada de acordo com as Normas da CPG/ICB relativas a outras formas de elaboração de tese de doutorado que permitem a inclusão, como Anexos, de resultados já publicados em periódicos internacionais indexados em língua inglesa. Permite ainda que detalhes metodológicos e resultados obtidos sejam aqueles contidos nos artigos anexados ao corpo da tese.

Artigos que compõe o corpo da tese:

Anexo 1 – Random amplified polymorphic DNA profiles as a tool for the characterization of Brazilian keratitis isolates of the genus *Acanthamoeba*. Alves, J. M. P., Gusmão, C. X., Teixeira, M. M. G., Freitas, D., Foronda, A. S. & Affonso, H. T. (2.000) **Brazilian Journal of Medical and Biological Research**, **33**: 19-26.

Anexo 2 – Genus *Acanthamoeba*: parity among random amplified polymorphic DNA, total DNA and SSU rDNA restriction length polymorphism patterns for the grouping and characterization of Brazilian keratitis isolates. Alves, J. M. P., Saba, M. M. F., Teixeira, M. M. G. & Affonso, H. T. (Manuscrito em preparação).

Este trabalho teve o apoio financeiro de CNPq e FAPESP

Dedico este trabalho a

José Bento Monteiro Lobato (1882 – 1948) e **Carl Sagan** (1934 – 1996),

por terem me tornado, quase 20 anos atrás, um viciado em leitura e ciência.

Suas obras e idéias são responsáveis por boa parte do que sou.

Agradeço

Aos meus pais, João e Darci, sem os quais absolutamente nada em minha vida teria sido possível. Não lhes dedico meu trabalho por que a eles dedico toda minha vida;

Às minhas irmãs Geiza e Syntia, à minha avó Ernesta e ao Bendj por constituírem uma família perfeita, com todas suas qualidades e defeitos;

À Prof.a. Dra. Maria Heloiza Trebilcock Affonso, cujos orientação, amizade, paciência (*quase* sempre...), confiança, brigas, convivência e etc. tornaram os últimos sete anos muito mais construtivos e agradáveis;

À Prof.a. Dra. Marta Maria Geraldtes Teixeira, pela contínua colaboração e co-orientação em qualquer aspecto do trabalho e por nos acolher tão bem em seu laboratório;

Aos Profs. Drs. Luiz R. Nunes e Regina L. C. Oliveira, do Núcleo Integrado de Biotecnologia da Universidade de Mogi das Cruzes, cuja colaboração foi imprescindível para a finalização deste trabalho;

Aos vários professores (da Parasito, Bio, EPM, etc.), que muito ajudaram com ensinamentos, orientação, colaboração e amizade, em especial os Profs. Drs. Carlos E. Winter, Jeffrey J. Shaw e Lucile M. F. Winter (cujo laboratório foi o primeiro onde estivemos instalados, lá atrás em 1.992);

Ao Prof. Dr. Jeffrey J. Shaw por ter gentilmente cedido sua cópia do PAUP* para Windows (sem a qual boa parte das análises teria ficado apenas no nível hipotético) e por me proporcionar experiência, primeiro como aluno e, depois, como monitor informal em parte do seu curso de filogenia;

Aos Srs. Wolfgang Fischer, Heribert Pickart, Manoel A. Peres e à Sra. Márcia de Oliveira Rocha pelo valioso auxílio técnico prestado em diversas fases e aspectos deste trabalho;

Aos inúmeros colegas que passaram pelo laboratório de *Acanthamoeba* ao longo de todos estes anos, em especial aos amigos Cléber X. Gusmão, Lúcia M. Lopes e Dalva M. Hashimoto;

Aos amigos Myrna G. Serrano, Rogéria M. Ventura, Adriana R. Fuzato, Flávia M. Silva, Alane P. C. Zahar, Robson Ferreira, Carmen S. Takata e Marta Campaner pela fervilhante troca de idéias, por me fazer sentir sempre “em casa” em seu laboratório e por tornar meu cotidiano muito mais divertido; Ah, Myrna, valeu pela idéia que resultou no meu primeiro programinha de biocomputação! (pensou que eu iria esquecer a promessa?);

Às dezenas de alunos de vários laboratórios do departamento com quem convivi, responsáveis por tornar a Parasito um lugar a ser lembrado com carinho;

Ao pessoal da turma Bio-92 e “agregados”, muitos amigos para toda a vida. Em especial, mio caro fratello Luís F. Mucci (Luiggi);

À K., por ter me ensinado tanto em tão pouco tempo;

Aos diversos escritores, de ficção ou não, cujas influências sobre mim não podem ser subestimadas. Sem falar nas infinitas horas de diversão, claro;

Aos diversos músicos do rock, ídolos ou companheiros de banda (Rising Sun will never set!), sem os quais minha sobrevivência no mundo dos mentalmente sãos seria absolutamente impossível. E, especialmente, aos insuperáveis Johann S. Bach e Ludwig van Beethoven – nada pode ir mais alto;

Aos órgãos financiadores CNPq e FAPESP, por manter o laboratório, seus projetos e bolsistas durante todos esses anos;

A todos que ajudaram, de qualquer forma, mas não foram citados aqui por falha minha. E, principalmente, aos que não atrapalharam;

E, claro, às incontáveis amebas sacrificadas no decorrer de minhas atividades. “Meninas”, espero que não tenha sido em vão!

ÍNDICE

Abreviaturas	1
Resumo	4
Abstract	6
<hr/>	
1. Introdução	8
1.1 O gênero <i>Acanthamoeba</i>	8
1.1.1 Morfologia, ecologia e ciclo de vida	9
1.1.2 Patologias associadas	12
1.1.2.1 Encefalite amebiana granulomatosa (GAE)	12
1.1.2.2 Ceratite crônica	12
1.1.3 Genes ribossômicos em <i>Acanthamoeba</i> e sua utilização em sistemática	16
1.1.4 Taxonomia	18
1.2 Evolução, filogenia e dados moleculares	20
1.2.1 A natureza do pensamento evolutivo	20
1.2.2 A síntese evolutiva	24
1.2.3 Inferência filogenética	24
1.2.4 Métodos moleculares em sistemática	29
1.2.4.1 Métodos baseados em distâncias	33
1.2.4.2 Métodos baseados no princípio da parcimônia	37
1.2.4.3 Máxima verossimilhança e modelos de evolução molecular	39
1.2.5 Exame da robustez e confiabilidade das filogenias	45
1.2.6 Os conceitos de espécie	46
1.3 Objetivos	48
<hr/>	
2. Material e Métodos	49
2.1 Organismos e cultivo	49
2.1.1 Cepas referência da ATCC	49
2.1.2 Cepas isoladas de casos de úlcera de córnea humana	49
2.2 Extração de DNA genômico total	50
2.2.1 Reações de amplificação aleatória de DNA polimórfico	50
2.2.2 Clivagem do DNA genômico total por endonucleases de restrição	51
2.3 Construção dos oligonucleotídeos para PCR e seqüenciamento	52
2.4 Reações de amplificação por PCR e purificação dos produtos	55
2.5 Clivagem por enzimas de restrição dos produtos de PCR	56
2.6 Cálculo das estimativas de distâncias genéticas entre os isolados	57
2.7 Reações de seqüenciamento do SSU rDNA em termociclador	58

2.8 Alinhamento das seqüências do SSU rDNA de <i>Acanthamoeba</i>	58
2.9 Análise filogenética do SSU rDNA de <i>Acanthamoeba</i>	59
2.9.1 Análises de distância	59
2.9.2 Análises de parcimônia	61
2.9.3 Análises de máxima verossimilhança	62
3. Resultados e Discussão	68
3.1 Amplificação aleatória de DNA polimórfico (RAPD)	68
3.2 RFLP de DNA genômico total	69
3.3 RFLP do SSU rDNA	72
3.4 Determinação da seqüências do SSU rDNA	76
3.5 Análises filogenéticas	78
3.5.1 Alinhamento e determinação do tipo de seqüência do SSU rDNA	78
3.5.2 Conceitos de espécie	123
3.5.2.1 Conceitos reprodutivos	123
3.5.2.2 Conceito ecológico	123
3.5.2.3 Conceito evolutivo	125
3.5.2.4 Conceito filogenético ou cladístico	126
3.5.2.5 Conceito de coesão	127
3.5.2.6 A situação das espécies em <i>Acanthamoeba</i>	128
3.5.3 Determinação do modelo evolutivo disponível mais adequado para o alinhamento do SSU rDNA	129
3.5.4 Construção das árvores filogenéticas	136
3.5.4.1 Análises de distância	139
3.5.4.2 Procura de árvores ótimas por parcimônia	144
3.5.4.3 Análises de verossimilhança das árvores obtidas e procura da árvore de ML para um alinhamento parcial	150
4. Conclusões	157
5. Referências bibliográficas	159
Anexo 1 – Trabalho publicado	
Anexo 2 – Manuscrito em preparação	

ÍNDICE DE FIGURAS E TABELAS

Fig. 1	10
Fig. 2	11
Fig. 3	13
Fig. 4	17
Fig. 5	25
Fig. 6	28
Fig. 7	35
Fig. 8	36
Fig. 9	39
Fig. 10	41
Fig. 11	54
Fig. 12	70
Fig. 13	71
Fig. 14	74
Fig. 15	75
Fig. 16	83
Fig. 17	87
Fig. 18	135
Fig. 19	138
Fig. 20	141
Fig. 21	142
Fig. 22	143
Fig. 23	149
Fig. 24	152
Fig. 25	155

Tab. 1	65
Tab. 2	144
Tab. 3	153

ABREVIATURAS

$^{\circ}\text{C}$ – graus Celsius

A. astr. – *Acanthamoeba astronyxis*

A. cast. – *Acanthamoeba castellanii*

A. poly. – *Acanthamoeba polyphaga*

A. royr. – *Acanthamoeba royreba*

AMD – Advanced Micro Devices, Inc.

ATCC – American Type Culture Collection

AVL – amebas de vida livre

BandB – “branch-and-bound”

dATP – trifosfato de desoxiadenosina

dCTP – trifosfato de desoxicitosina

dGTP – trifosfato de desoxiguanosina

DNA – ácido desoxiribonucleico

dTTP – trifosfato de desoxitimidina

EDTA – ácido etileno diamino tetra-acético

EPM – Escola Paulista de Medicina

F81 – modelo Felsenstein, 1981

Fig. – figura

FMUSP – Faculdade de Medicina da Universidade de São Paulo

g – aceleração gravitacional

gl – graus de liberdade do teste χ^2

GTR – “general time reversible”

HC – Hospital das Clínicas

HKY – modelo Hasegawa, Kishino & Yano, 1985

IGS – “intergenic spacer”

ITS – “internal transcribed spacer”

JC – modelo Jukes & Cantor, 1969

K2P – modelo Kimura 2 parâmetros, 1980

kpb – kilo pb (1.000 pares de base)

lnL – logaritmo neperiano da verossimilhança

LSU – subunidade maior (“large subunit”)

M – molar

Mb – megabytes

MHz – megahertz
 µg/mL – microgramas por mililitro
 µL – microlitro(s)
 µm – micrômetro(s)
 min – minuto(s)
 ML – máxima verossimilhança (“maximum likelihood”)
 mM – milimolar
 MS - Microsoft
 NaCl – cloreto de sódio
 ng – nanograma(s)
 NJ – “neighbor joining”
 NNI – “nearest-neighbor interchange”
 p.l. – parâmetros livres
 PBS – salina tamponada com fosfato (“phosphate buffered saline”)
 PCR – reação em cadeia da polimerase (“polymerase chain reaction”)
 pb – pares de base
 pM – picomol
 pmoles – picomoles
 RAM – “random access memory”
 RAPD – random amplified polymorphic DNA
 rDNA – gene do RNA ribossômico
 RFLP – polimorfismo no comprimento dos fragmentos de restrição (“restriction fragment length polymorphism”)
 RNA – ácido ribonucleico
 rRNA – RNA ribossômico
 s – segundo(s)
 SDS – dodecil sulfato de sódio
 SE – 0,1 M de NaCl, 100 mM de EDTA
 SPR – “subtree pruning-regrafting”
 SSU – subunidade menor (“small subunit”)
 T₁₀E₁₀ – Tris 10 mM + EDTA 10 mM
 TBR – “tree bisection-reconnection”
 TE – Tris 10 mM + EDTA 1 mM
 Tris – tris [hidroximetil] aminometano
 TrN – modelo Tamura & Nei, 1993

ts – transição

tv – transversão

U – unidade(s)

U/E – USP/EPM

U/HC – USP/HC

U/Oft – USP/Oft

UNIFESP – Universidade Federal de São Paulo

UPGMA – método de agrupamento de pares por média aritmética (“unweighted pair group method with arithmetic mean”)

USP – Universidade de São Paulo

v. – versão

RESUMO

Amebas de vida livre do gênero *Acanthamoeba* podem causar dois tipos de patologias graves: uma ceratite muito dolorosa, que pode levar à perda da visão, e encefalite amebiana granulomatosa, que é sempre fatal. Apesar da utilização intensiva de isolados de *Acanthamoeba* em estudos clássicos de biologia celular e molecular, a identificação, taxonomia, mecanismos de patogenicidade e a filogenia do gênero permanecem confusos. Para investigar o polimorfismo genético de 14 isolados brasileiros de ceratite e oito provenientes da ATCC (4 de ceratite), foram utilizados neste estudo diversos métodos moleculares: a) amplificação aleatória de DNA polimórfico (RAPD); b) polimorfismo de comprimento de fragmentos de restrição (RFLP) de DNA genômico total; c) RFLP do gene do RNA da subunidade menor do ribossomo (SSU rDNA) amplificado por PCR. Posteriormente, para inferir as relações filogenéticas entre estes isolados, as seqüências do SSU rDNA foram utilizadas para análises taxonômicas e filogenéticas.

Apesar do extenso polimorfismo revelado pelos métodos de RAPD e RFLP de DNA genômico e de SSU rDNA, foram identificados vários grupos de isolados mais geneticamente relacionados após a comparação dos fenogramas construídos utilizando distâncias de Jaccard e o agrupamento por UPGMA.

As análises baseadas em seqüências do SSU rDNA são as mais promissoras para o estabelecimento da filogenia e para a classificação de espécies no gênero *Acanthamoeba*. Doze tipos de seqüências (com a proposta de que cada um corresponda a uma espécie) foram definidos por Stothard e cols. (1998) baseados em medidas da divergência não corrigida entre seqüências (pelo menos 5% de diferença). Nossos resultados corroboram os desse trabalho, mostrando forte correlação entre os isolados tipo T4 e ceratite. Entretanto, maiores estudos envolvendo isolados de ambiente são necessários para se

verificar se esta correlação não seria devida simplesmente a uma maior prevalência ecológica das amebas apresentando seqüências do SSU rDNA tipo T4.

Nossos dados de seqüenciamento do SSU rDNA confirmaram aqueles previamente obtidos por RAPD e RFLP. Além disso, três de nossos isolados apresentaram divergência suficiente, quando comparados aos doze tipos previamente estabelecidos, para justificar a criação de três novos tipos de seqüência, incluindo um representado por um isolado de ceratite brasileiro (U/HC1). Desse modo, identificamos uma nova espécie causando ceratite, o que não seria inesperado considerando-se que nosso estudo é o primeiro envolvendo tipagem do SSU rDNA de isolados de *Acanthamoeba* do hemisfério sul.

ABSTRACT

Two severe pathologies of humans are caused by free-living amoebae of the *Acanthamoeba* genus: a painful sight-threatening keratitis and fatal granulomatous amebic encephalitis. In spite of the intensive use of *Acanthamoeba* species in classical molecular and cellular studies, the identification, taxonomy, pathogenicity mechanisms and phylogeny of this genus remain confused. In order to assess the genetic polymorphism among 14 Brazilian keratitis isolates and 8 ATCC isolates (4 from keratitis) we used various molecular methods: a) random amplified polymorphic DNA (RAPD); b) total genomic DNA restriction fragment length polymorphism (RFLP); c) PCR amplified small-subunit ribosomal DNA (SSU rDNA) RFLP analyses. Subsequently, to infer phylogenetic relationships among these isolates, sequences of SSU rDNA were used for taxonomic and phylogenetic analyses.

Despite the extensive polymorphism revealed by RAPD, total genomic DNA and SSU rDNA RFLP analyses, several groupings of more genetically related isolates could be disclosed after comparison of phenograms constructed using Jaccard distances and the UPGMA clustering algorithm.

Methods dealing with SSU rDNA sequences are the most promising in terms of phylogeny and species assignment in this genus. Twelve sequence types (each proposed to be an *Acanthamoeba* species) were defined by Stothard et al., 1998, based on measurements of uncorrected sequence divergence (at least 5% difference). Our results agree with those of that work, showing a strong correlation between type T4 isolates and keratitis. However, more extensive studies of environment isolates are needed in order to verify whether this correlation is not merely due to a greater ecological prevalence of amoebae presenting T4 sequence type.

Our data based on SSU rDNA sequencing corroborated those previously found by RAPD and RFLP approaches. Besides, three of our isolates showed divergence enough when compared with the twelve previously established sequence types to create three new sequence types, including one represented by a Brazilian keratitis isolate (U/HC1). Therefore, we identified a new species causing keratitis, what would not be surprising since our study is the first one of *Acanthamoeba* SSU rDNA sequence typing from Southern hemisphere isolates.

Introdução

1. INTRODUÇÃO

1.1 – O gênero *Acanthamoeba*

Desde sua descoberta, em meados do século XVIII⁽¹⁰⁵⁾, as amebas têm sido organismos bastante estudados. Isso se deve, em grande parte, à existência de isolados responsáveis por importantes patologias. Entre as amebas mais bem estudadas destaca-se *Entamoeba histolytica*, o protozoário patogênico que apresenta, talvez, a maior dispersão mundial, sendo responsável por um grande número de infecções, muitas vezes fatais^(81, 103). Além disso, a diversidade de formas e a facilidade de isolamento e cultivo destes organismos, muito abundantes na natureza, os tornam modelos biológicos interessantes para estudos de bioquímica, biologia celular e molecular.

Até recentemente, *E. histolytica* era o único gênero de amebas considerado patogênico para o homem. Em 1958, ao inocular intranasalmente animais de laboratório com uma vacina contaminada, Culbertson descobriu o potencial patogênico de uma outra ameba, que até então era considerada saprofítica e não patogênica⁽¹⁶⁾. Os animais haviam morrido de encefalite, e a investigação do agente etiológico revelou a presença de uma ameba como contaminante da vacina. Essa ameba, identificada como pertencente ao gênero *Acanthamoeba*, faz parte do grupo das Amebas de Vida Livre (AVL)⁽⁶⁴⁾.

Os gêneros de AVL considerados potencialmente patogênicos são *Acanthamoeba*, *Naegleria* e *Balamuthia*. Organismos pertencentes a esses gêneros são responsáveis por quadros de encefalites bastante graves, na maioria das vezes fatais⁽⁷⁹⁾. Além dessa patologia, organismos do gênero *Acanthamoeba* são também os agentes causadores de uma úlcera de córnea bastante severa, conhecida como ceratite crônica, descrita pela primeira vez no início da década de setenta do século passado^(30, 60, 91, 129).

Apesar da grande importância que as AVL vêm assumindo, principalmente pelo número crescente de casos clínicos em todo mundo^(55, 79, 80, 110, 130), muitas questões referentes principalmente aos fatores de patogenicidade desse grupo de amebas permanecem sem resposta. Ao contrário de *E. histolytica*, cujos mecanismos de patogenicidade já são bem conhecidos^(13, 81), ainda há muita incerteza quanto à importância dos diversos fatores envolvidos na capacidade de adesão, invasão e multiplicação das AVL nos tecidos que parasitam.

Um ponto bastante interessante a ser esclarecido é se todas as AVL, encontradas normalmente no ambiente, são potencialmente patogênicas, ou se só alguns desses organismos seriam capazes de provocar infecções. Tais estudos seriam de grande interesse em saúde pública, uma vez que foi mostrada recentemente a presença de *Acanthamoeba* em mais da metade das piscinas analisadas na Alemanha, assim como em quase 100% das amostras de solo da Turquia⁽⁸⁷⁾, além da observação feita no Japão de que cerca de metade dos indivíduos examinados apresentou anticorpos contra essas amebas⁽¹²⁴⁾.

Muitas destas questões podem ser investigadas com o auxílio de técnicas de biologia molecular, através da comparação de diversos tipos de seqüências gênicas de isolados de ambiente e de casos clínicos.

1.1.1 Morfologia, ecologia e ciclo de vida

Os organismos do gênero *Acanthamoeba*, isolados pela primeira vez em 1930 por Castellani⁽¹⁰⁾ e inicialmente classificados como *Hartmannella*⁽⁸⁰⁾, são encontrados sob duas formas: trofozoítos, que são as formas vegetativas da célula (Fig. 1), e cistos, que representam as formas de resistência (figura 1 do Anexo 2 e esquematizados na Fig. 2). Pode-se obter, em laboratório, qualquer uma das formas, dependendo do meio de cultura utilizado^(102, 80, 97). O gênero é

facilmente distinguível de outras amebas devido às formas características dos cistos e trofozoítos.

Os trofozoítos, forma em que as amebas se locomovem, se alimentam e se reproduzem, são células de formato variável, com 20 a 40 μm de tamanho, e portadoras de um núcleo bem definido e nucléolo grande. O citoplasma apresenta diversos vacúolos, com destaque para um vacúolo contrátil, envolvido no controle osmótico da célula. Os pseudópodes se formam constantemente, sendo responsáveis pela locomoção. Já os acantopódios, projeções aciculiformes da membrana celular típicas de *Acanthamoeba*, não têm função claramente definida.

Fig. 1 - Trofozoítos de *Acanthamoeba* sp. em microscopia de contraste de fase. Micrografia realizada pela Profa. Dra. Annette S. Foronda.

Os cistos, resistentes por longos períodos de tempo à dessecação, extremos de temperatura e produtos químicos diversos, tipicamente variam de diâmetro entre 13 e 20 μm . São compostos por celulose e diversas proteínas, e possuem duas paredes: o endocisto e o ectocisto. Apresentam um opérculo, um ponto determinado do

cisto por onde a ameba sai ao desencistar, assim que as condições ambientais forem mais favoráveis. De forma geral, o ectocisto é mais esférico, enquanto o endocisto apresenta diversas formas poligonais que dependem do isolado, variando de triangular a estrelado, com diversas variações. A distância entre as paredes do cisto e o número de pontos em que o endocisto se comunica com o ectocisto (braços do endocisto) também são variáveis. Essas características dos cistos, especialmente tamanho e distância entre paredes, foram utilizadas

inicialmente para a identificação e taxonomia do gênero⁽⁹⁷⁾, com a divisão de *Acanthamoeba* em três grupos morfológicos (I, II e III). A divisão em três grupos é um tanto arbitrária, uma vez que os grupos I e III representam extremos bem definidos, enquanto o grupo II contém todos os intermediários.

Fig. 2 – Esquema da estrutura da parede do cisto de *Acanthamoeba*. b: braço do endocisto; ex: ectocisto; en: endocisto; op: opérculo; os: ostíolo; p: poro. Modificado de Pussard & Pons, 1977.

Organismos do gênero *Acanthamoeba* apresentam ampla distribuição na natureza, sendo encontrados no mundo inteiro e nos mais diversos ambientes, como solo, água de rios, lagos e mares, esgotos, piscinas, sistemas de condicionamento de ar, água de torneira, vegetais, cultura de

células, em peixes, répteis, pássaros, mamíferos e até mesmo no ar^(5, 22, 55, 64, 66). Águas termais ou termicamente poluídas costumam conter mais amebas patogênicas do que águas mais frias⁽⁸⁰⁾. Além de ser a ameba mais freqüentemente isolada, estima-se que *Acanthamoeba* possa ser o protozoário de vida livre mais comum⁽⁹⁷⁾, o que confere ao gênero uma grande importância ecológica, uma vez que essas amebas se alimentam de bactérias, fungos, cianobactérias e protozoários⁽⁶⁴⁾. Além disso, *Acanthamoeba* é o hospedeiro natural de diversas bactérias, como várias Chlamydiales⁽³³⁾ e, especialmente, *Legionella pneumophila*, causadora da doença dos legionários em humanos^(101, 106, 126), o que aumenta ainda mais sua importância tanto ecológica como em saúde pública. Além disso, *Acanthamoeba* é freqüentemente isolada do trato respiratório superior de indivíduos assintomáticos^(55, 124), sugerindo que o contato entre essas amebas e seus hospedeiros potenciais é bastante intenso.

1.1.2 Patologias associadas

Embora já tenham sido encontrados como oportunistas em infecções em uma grande variedade de tecidos, como pele, ouvido, fígado, pâncreas, baço, próstata, tireóide, trato urogenital, úlcera péptica, etc., os organismos do gênero *Acanthamoeba* podem ser responsabilizados por apenas dois tipos de patologia em humanos:

1.1.2.1 Encefalite Amebiana Granulomatosa (GAE) - é uma infecção considerada oportunista, pois quase todos os casos relatados até o momento se referem a indivíduos imunologicamente debilitados, especialmente aqueles que sofreram algum tipo de terapia ou situação imunossupressora, como alcoolismo, gravidez, quimioterapia ou uso de antibióticos de amplo espectro^(30, 79). Mais recentemente, vêm sendo descritos casos de infecção por *Acanthamoeba* em pacientes com AIDS^(9, 35, 111, 130). Presumivelmente, a invasão ocorre pela pele (ulcerações) ou pelo trato respiratório (neuroepitélio), seguindo então por via sangüínea até o sistema nervoso central. O período de incubação pode variar de semanas a meses. Após esse período, vários sintomas aparecem: dor de cabeça acompanhada de febre, náuseas, vômitos, letargia, mudança da personalidade e loucura^(78, 79, 133). Como o desenvolvimento da doença é fulminante, geralmente o diagnóstico correto só é obtido na necropsia, por análise de fluido cefalorraquidiano ou do próprio tecido cerebral.

1.1.2.2 Ceratite crônica - os primeiros casos de ceratite identificados foram descritos em 1973 e a contaminação estava associada a lesões na córnea^(60, 91, 129). O relato de casos de ceratite por *Acanthamoeba* vem aumentando significativamente nos últimos anos, paralelamente à popularização do uso de lentes de contato, principalmente hidrofílicas^(55, 76, 79, 122, 130). Ao contrário da

encefalite, a ceratite pode ocorrer em indivíduos sem nenhum sinal de comprometimento imunológico.

A contaminação ocorre principalmente quando o indivíduo utiliza práticas inadequadas de desinfecção das lentes (Fig. 3). Um caso bastante ilustrativo desse ponto é o estudo de Ledee e cols., 1996, utilizando métodos moleculares, onde foi demonstrada a presença de um mesmo isolado de *Acanthamoeba griffini* na caixa d'água, pia e na caixa de acondicionamento das lentes de contato de um paciente com ceratite.

Fig. 3 – O tipo de higiene das lentes de contato associado a alto risco de contaminação por *Acanthamoeba*. Reproduzido com permissão do autor.

Acredita-se que *Acanthamoeba* não tenha a capacidade de invadir a córnea íntegra, havendo a necessidade da presença de uma lesão prévia, o que explicaria a grande correlação entre a doença e o uso de lentes de contato⁽⁴⁵⁾. Em contato com a córnea, o próximo passo para o estabelecimento da infecção é a aderência e penetração dos trofozoítos no epitélio⁽⁹⁰⁾. Foi observado que a aderência é mediada por uma proteína da membrana do trofozoíto, que se liga à manose de glicoproteínas do epitélio da córnea⁽¹⁴⁴⁾. Além disso, alguns estudos sugerem que enzimas como neuraminidase, elastase e várias proteases podem ser fatores que determinam a infecção por *Acanthamoeba*^(29, 88, 99). Após a

penetração, seguem-se ulcerações, infiltrado em forma de anel (exceto *A. griffini*, que produz infiltrados pontuais)⁽⁶⁹⁾ e inflamações.

Ainda não foi determinado o tempo de incubação para o estabelecimento dos primeiros sintomas da ceratite por *Acanthamoeba*. A irritação devida ao trauma inicial pode se confundir com os primeiros sintomas da infecção mas, em alguns casos mais raros, poderia haver um período de cerca de duas semanas entre o trauma e o aparecimento dos primeiros sintomas⁽⁵⁵⁾.

Os sintomas mais comuns da ceratite por *Acanthamoeba* são fotofobia, opacidade da córnea, dores intensas (normalmente desproporcionais à gravidade aparente das lesões) e diminuição da visão^(55, 64, 78).

Um dos problemas mais sérios e preocupantes nas infecções oculares por *Acanthamoeba* é a deficiência de um diagnóstico rápido e correto. A porcentagem de falso diagnóstico costuma ser bastante alta. Na maioria dos casos, o diagnóstico efetuado refere-se à presença de herpes, fungos e bactérias. A presença de *Acanthamoeba* muitas vezes só é detectada através de raspado ou biópsia corneana e, muitas vezes, somente após a remoção da córnea^(55, 110). Em muitos desses casos, foi realizada ceratoplastia, por vezes sem sucesso devido à recidiva da infecção, levando à necessidade de remoção do globo ocular. Mesmo após um diagnóstico correto, um outro problema é determinar um tratamento eficaz, uma vez que as amebas desse gênero, especialmente quando na forma de cistos, mostram-se bastante resistentes a uma série de quimioterápicos^(31, 59, 74, 110).

Embora o número de relatos de ceratite por *Acanthamoeba* venha aumentando, ainda assim existe uma desproporção evidente entre o número de casos e a dispersão ambiental desses protozoários. Esse fato pode ser devido à existência de organismos com diferentes graus de patogenicidade, à resistência do hospedeiro e/ou ao desconhecimento da classe médica em relação às AVL,

com a conseqüente ocorrência de inúmeros casos não diagnosticados, mas principalmente à inexistência de métodos precisos de diagnóstico⁽⁵⁵⁾.

Pode-se perceber, portanto, a necessidade cada vez maior da utilização precoce de métodos rápidos e precisos para o diagnóstico correto de organismos do gênero *Acanthamoeba*, principalmente nos casos de úlcera de córnea. Estudos recentes têm utilizado a técnica de PCR para o diagnóstico desses casos com relativo sucesso, superando a eficiência e sensibilidade dos métodos tradicionais⁽⁷¹⁾. A identificação correta desses protozoários pode favorecer também o desenvolvimento de métodos alternativos de tratamento. Tradicionalmente, o diagnóstico é feito a partir de cultura de material proveniente de raspados de córnea. Além de ser demorado e pouco sensível, especialmente em casos com baixa quantidade de amebas, esse método não permite uma identificação adequada do isolado envolvido, devido à falha dos dados morfológicos em distinguir a maioria das espécies no gênero.

A determinação correta da taxonomia dentro do gênero é um fator essencial para que se possa chegar a generalizações e conclusões acerca da ecologia, patogenicidade e epidemiologia relacionadas a qualquer patógeno, uma vez que a falta de linhagens bem definidas causa grandes equívocos na comparação de diversos casos estudados. Assim, várias "espécies" do gênero já foram responsabilizadas pelas infecções^(80, 130). Entretanto, conforme detalhado adiante (Taxonomia), vem sendo demonstrado, ao longo dos anos, que muitas das espécies de *Acanthamoeba*, especialmente aquelas pertencentes ao grupo morfológico II, são agrupamentos de organismos geneticamente muito diversos, enquanto outros organismos praticamente idênticos estão colocados em espécies diferentes. Mais recentemente, foi estabelecido um novo critério de identificação e taxonomia para *Acanthamoeba*: o uso da seqüência do gene que codifica o RNA da subunidade menor do ribossomo (SSU rDNA), dividindo o gênero em

doze grupos, T1 a T12, de acordo com a divergência não corrigida entre as seqüências⁽¹¹⁷⁾. Em relação à patologia, o resultado mais interessante dessa nova abordagem foi o encontro de uma forte correlação entre o grupo T4 e a ocorrência de úlcera de córnea, uma vez que somente uma exceção foi observada mas, mesmo assim, causando uma patologia de aparência diferente do habitual. Embora seja uma associação forte (24 em 25 casos de ceratite causados por organismos do grupo T4), ainda assim resta saber se tal correlação não se deve apenas, por exemplo, a uma maior presença do tipo T4 no ambiente, como os próprios autores reconhecem. Estudos mais aprofundados de isolados provenientes do ambiente poderão ajudar a resolver a questão.

*1.1.3 Genes ribossômicos em **Acanthamoeba** e sua utilização em sistemática*

Entre as diversas seqüências de DNA que vêm sendo utilizadas para a identificação e taxonomia de uma série de organismos, aquelas que codificam os RNAs ribossômicos são as que vêm apresentando resultados mais promissores^(47, 95, 96, 113, 137). Suas grandes vantagens são:

- tamanho (cerca de 2.000 pares de bases no SSU rDNA), que geralmente possibilita a obtenção de dados suficientes para a observação de sinal filogenético;
- grande quantidade de cópias no genoma, quase sempre evoluindo em conjunto, o que facilita o isolamento da molécula;
- ubiquidade, pois estão presentes em todos organismos conhecidos, à exceção dos vírus, permitindo a construção de filogenias universais, impossíveis de serem determinadas por métodos morfológicos e mesmo métodos moleculares utilizando outras seqüências alvo, por exemplo;
- grande significado biológico, que impõe uma grande pressão seletiva e assegura homologia entre todos os seres vivos, uma vez que se acredita que os

genes ribossômicos tenham surgido uma única vez e sejam, portanto, compartilhados por ancestralidade comum;

- presença de regiões mais e menos variáveis na mesma molécula, possibilitando a análise tanto de organismos proximamente relacionados (utilizando regiões de alta taxa de substituição) quanto pertencentes a reinos diferentes (utilizando agora as regiões de evolução mais lenta na molécula);

- grande quantidade de dados já gerados sobre essas seqüências em diversos grupos de organismos, facilitando a análise comparativa.

Assim como para outros protozoários, os genes ribossômicos de alguns gêneros de AVL vêm sendo estudados, na tentativa de isolar seqüências que possam ser utilizadas na identificação e taxonomia desse grupo de organismos. No gênero *Acanthamoeba*, já foi determinado que a unidade de repetição do cistron ribossômico é de cerca de 12.000 pares de bases (pb). O número de unidades de repetição parece ser em torno de 600 cópias por genoma⁽⁸⁾, e estão organizadas conforme esquematizado na figura 4.

Fig. 4 – Esquema do cistron ribossômico em *Acanthamoeba*.

A subunidade maior (LSU rDNA), da mesma maneira que ocorre, por exemplo, em tripanosomatídeos^(11, 42), está subdividida em dois segmentos de 2.400 pb (LSU a) e 2.000 pb (LSU b), separados por um espaçador interno transcrito (ITS) de cerca de 200 pb, que é removido do transcrito primário⁽¹¹⁶⁾.

O RNA da subunidade menor de *Acanthamoeba* tem um tamanho de cerca de 2.300 a 2.700 pares de bases, estando entre os maiores conhecidos⁽¹²⁾. Recentemente, descobriu-se que nem todas as cópias do SSU rDNA são

obrigatoriamente idênticas, uma vez que sete entre 53 isolados estudados apresentaram mais de um alelo do gene⁽¹¹⁷⁾. Não foi demonstrado, entretanto, se as diferentes variantes são todas expressas ou se algumas delas podem ser pseudogenes. Os alelos identificados em cada isolado se mostraram muito semelhantes entre si (1 a 4 diferenças por 1000 bases), sendo que a diferença entre eles foi menor do que entre eles e outras seqüências de isolados pertencentes ao mesmo tipo de SSU rDNA.

1.1.4 Taxonomia

Conforme citado anteriormente, a diferenciação dos gêneros das AVL normalmente é feita através da morfologia dos cistos e trofozoítos, sendo relativamente simples a identificação do gênero *Acanthamoeba*. Entretanto, a taxonomia do gênero ainda é bastante indeterminada no nível específico.

Diversas espécies foram definidas pelos parâmetros morfológicos clássicos, sendo depois agrupadas e separadas diversas vezes conforme novas técnicas demonstravam a pouca consistência dos agrupamentos iniciais^(117, 131). Desse modo, ainda hoje existem referências a diversas “espécies” do gênero *Acanthamoeba*, como pode ser verificado na listagem da página de Taxonomia do National Center for Biotechnology Information, baseada nos registros do GenBank (<http://www.ncbi.nlm.nih.gov/Taxonomy/>): *Acanthamoeba astronyxis*, *A. castellanii*, *A. comandoni*, *A. culbertsoni*, *A. griffini*, *A. hatchetti*, *A. healyi*, *A. lenticulata*, *A. lugdunensis*, *A. palestinensis*, *A. pearcei*, *A. polyphaga*, *A. pustulosa*, *A. rhysodes*, *A. royreba*, *A. stevensoni* e *A. tubiashi*. Diversos outros binômios, como *A. divionensis*, *A. gigantea*, *A. mauritaniensis*, *A. quina* e *A. triangularis*, por exemplo, já foram encontrados na literatura⁽⁹⁷⁾, mas são hoje considerados sinônimos para outras espécies. Os diversos métodos subsequentes de investigação genética demonstraram que muitas das espécies ainda

reconhecidas são agrupamentos de organismos muito diversos e, muitas vezes, mais proximamente aparentados com integrantes de outras espécies do gênero.

Até meados da década de 1990, os parâmetros para a classificação de novos isolados tinham se baseado principalmente na morfologia dos cistos^(80, 97, 102), características antigênicas⁽¹⁷⁾, composição enzimática⁽¹⁵⁾, aglutinação com lectinas⁽⁸²⁾ e padrão de isoenzimas^(18, 57). Mais recentemente, novos métodos têm sido propostos, utilizando técnicas de Biologia Molecular, tais como padrões de restrição de DNA total^(62, 63), padrões de restrição de DNA mitocondrial^(6, 7, 14, 38, 138, 139), seqüências de DNA ribossômico 5S e 5,8S⁽⁸⁵⁾, amplificação de seqüências específicas por PCR^(71, 132), cariotipagem por eletroforese em campo pulsado (PFGE)⁽⁸³⁾, seqüências de rDNA mitocondrial⁽⁷⁷⁾ e do gene nuclear da subunidade menor do ribossomo (SSU rDNA)^(37, 43, 58, 67, 69, 114, 117).

O critério mais promissor parece ser a utilização da seqüência do SSU rDNA⁽¹¹⁷⁾, que permitiu a divisão de *Acanthamoeba* em doze tipos de seqüência (T1 a T12). Os autores desse estudo sugeriram que cada tipo de seqüência poderia ser considerado como uma espécie separada (discutido mais detalhadamente no item 3.5.2.6 de Resultados e Discussão). Os tipos se baseiam na divergência não corrigida entre as seqüências (mais de 5% para pertencer a um grupo diferente), além da análise filogenética a partir dos alinhamentos, sem a qual a simples divergência não tem muito significado. Esse critério é, até o momento, o mais objetivo e prático para a taxonomia de *Acanthamoeba*, uma vez que permite facilmente a comparação de isolados utilizados em diferentes estudos, além de colocar os dados relevantes (seqüências e alinhamento) à disposição de outros pesquisadores por meio de bancos de dados. Essas vantagens não são possíveis ao se utilizar métodos morfológicos, ou mesmo moleculares de vários outros tipos.

O estudo dos tipos de SSU rDNA demonstrou que os organismos colocados no grupo morfológico II, especialmente *A. castellanii* e *A. polyphaga*, são os que apresentam maior variabilidade, o que poderia ser devido a uma possível maior prevalência ambiental. Assim, vários isolados agrupados nessas espécies (devido à alta similaridade morfológica) foram separados em linhagens diferentes, enquanto isolados anteriormente considerados como pertencentes a espécies diferentes foram reunidos em um mesmo tipo de sequência.

Desse modo, parece claro que o conhecimento da organização de determinadas seqüências de DNA desses organismos poderá levar ao desenvolvimento de métodos mais precisos para sua classificação e taxonomia, contribuindo assim para um diagnóstico preciso desse grupo de patógenos, o que terá grande importância tanto do ponto de vista clínico como epidemiológico.

Tendo em vista a importância da aplicação de novas estratégias para a classificação e taxonomia do gênero *Acanthamoeba*, é apresentada, a seguir, uma explanação sobre a utilização de dados moleculares no estudo de evolução e filogenia, assim como sobre os vários métodos atualmente empregados.

1.2 – Evolução, filogenia e dados moleculares

1.2.1 A natureza do pensamento evolutivo

Ao contrário do que comumente se pensa, Charles Darwin e sua obra "On the Origin of Species by Means of Natural Selection", de 1859, não foram o início da história do evolucionismo^(84, 104). Desde fins do século XVII, pensadores como Buffon, Cuvier, diversos outros da França e da Alemanha, além de pesquisadores da Geologia, especialmente Charles Lyell, vinham paulatinamente coletando evidências e desenvolvendo a noção de que o mundo (tanto vivo quanto físico) não poderia ser tão jovem e imutável quanto afirmavam as interpretações da natureza à luz de textos religiosos (teologia natural). Além disso, não se pode

esquecer a contribuição essencial de Lamarck, primeiro evolucionista verdadeiro. Embora muitos desses pensadores fossem abertamente antievolucionistas, ainda assim cada qual, com sua contribuição, foi preparando o ambiente intelectual e compondo o arcabouço de dados necessário aos desenvolvimentos conceituais posteriores, essenciais à formação do pensamento evolutivo de Darwin e Alfred R. Wallace. Suas idéias, observações e grande quantidade (e qualidade) de dados foram, portanto, importantes precursores na mudança de visão que demandou muita discussão e que levou praticamente o século XIX inteiro e parte do XX para se consolidar: não só é possível explicar os fenômenos naturais sem recorrer ao sobrenatural, como também se deve abandonar a noção do ser humano como uma entidade separada da natureza, incluindo-o como mais um integrante da filogenia universal. Tal filogenia não é uma "scala naturae", uma progressão de formas em direção à perfeição, e não se deve aplicar o pensamento teleológico às questões evolutivas, ou seja, não existem finalidades ou objetivos no processo de modificação dos organismos. O que haveria, no caso, seria a sobrevivência diferencial das variedades melhor adaptadas às condições prevalentes no momento. Além disso, foi iniciada a mudança (incrivelmente ainda em curso) de um pensamento tipológico, herança da filosofia platônica, para um pensamento populacional. Isso significa que os indivíduos de uma espécie não eram mais encarados como cópias imperfeitas de uma essência (o tipo divinamente criado), mas variações naturais cujas características eram passadas à progênie. Ainda hoje, no cotidiano da sociedade e, muitas vezes, até mesmo no meio científico, pode-se observar que o pensamento, em geral, ainda costuma resistir de forma inconsciente a essas idéias, apesar do enorme peso das evidências em seu favor.

O que Darwin trouxe de importante, portanto, não foi o conceito de mudança nas linhagens (fato que mesmo os criadores de animais domésticos, muito

consultados por Darwin, há muito já conheciam), mas o mecanismo pelo qual poderia ocorrer: a seleção natural. Tudo seria consequência, basicamente, da existência de variabilidade hereditária nas populações, da luta entre os organismos pela sobrevivência em um mundo de recursos limitados (conceito herdado das idéias de Malthus) e das contingências ambientais imprevisíveis, de natureza muitas vezes estocástica. Esses fatores iriam determinar quais das variedades, se alguma, se saíam melhor na “luta pela sobrevivência”, deixando mais descendentes. Curiosamente, o próprio termo “evolução” não é sequer citado em “Origin of Species” de Darwin, e a única menção que se aproxima disso é exatamente a última palavra do livro, na frase “...whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, **evolved**.”⁽²⁰⁾

A partir de então e, aos poucos, anatomistas e morfologistas, que antes encaravam as semelhanças entre os organismos como evidência de um criador e de “Baupläne”, agora utilizavam as análises comparativas como evidência de parentesco e da existência de ancestrais comuns. Os paleontólogos passaram a compreender a verdadeira natureza dos fósseis: como ancestrais dos *taxa* vivos ou outras linhagens que não deixaram descendência, e não mais como evidência de um ou mais dilúvios e catástrofes de origem divina. Esses fósseis agora serviriam como importantes componentes da inferência filogenética, ao fornecer “elos perdidos” (como *Ichtyostega* ou *Archaeopteryx*) e informações sobre a homologia de caracteres. As mudanças no mundo vivo deixam, portanto, de ter a forma de uma escada, de uma progressão de formas em direção à perfeição predefinida (que se acreditava que fosse o ser humano), e passa a se organizar com uma árvore em que todas as linhagens contemporâneas correm

paralelamente, tendo se originado de linhagens extintas e, potencialmente, podendo dar origem a outras linhagens no futuro.

Entretanto, embora algumas das teorias de Charles Darwin tenham sido razoavelmente bem sucedidas em sua época (especialmente o conceito de evolução em si e, como consequência, a existência de um ancestral comum a todos os seres vivos), ainda assim alguns conceitos-chave envolvidos não alcançaram ampla aceitação por cerca de 80 anos. Isso se deveu, em grande parte, a uma deficiência não resolvida: os críticos argumentavam, com razão, que o fator mais importante não estava sendo explicado – de onde surge a variação e, principalmente, como se dá a herança das características observadas. Dependendo do modo de herança, a seleção natural poderia ser implausível.

As (re)descobertas da Genética, por volta de 1900, não foram imediatamente encaradas como a solução para o impasse darwiniano. Ao contrário, serviram inclusive como críticas ao gradualismo que caracterizava as idéias de Darwin. Nessa época, havia basicamente duas correntes evolucionistas: a dos mendelianos e a dos naturalistas ou biométricos. Os mendelianos achavam que a evolução deveria ocorrer por saltos maiores do que as mudanças graduais e contínuas, postuladas por Darwin. Isso se devia à visão propiciada pela natureza da maioria dos experimentos executados por esses cientistas, que analisavam mutações bastante conspícuas em organismos como drosófilas. Já os naturalistas tinham a visão oposta, de que as variações realmente ocorreriam de forma gradual, e concluíam que assim poderia ocorrer também na evolução, uma vez que suas observações de medidas de características morfológicas, como altura ou peso, sempre indicavam variação de um indivíduo para outro dentro das populações estudadas, resultando sempre em uma curva. A evolução seria, portanto, um deslocamento nas médias dessas curvas, devido à seleção direcional. Os problemas enfrentados pelos biométricos eram o fato de suas

observações não se encaixarem às predições mendelianas e a tendência em considerar o transformismo de Lamarck, ao invés da seleção natural, como mecanismo para a evolução.

1.2.2 A síntese evolutiva

Somente entre as décadas de 20 e 40 do século passado é que se observou o desenvolvimento do conceito que agora é conhecido como neodarwinismo ou síntese evolutiva, ou seja, a combinação da teoria evolutiva de Darwin com a Genética mendeliana, formando uma teoria muito coesa e desde então pouco modificada em suas características básicas. No início dos anos 20, R. A. Fisher, J. B. S. Haldane, S. Wright e outros lançaram as bases teóricas necessárias, ao mostrar que as características medidas pelos biométricos não eram de modo algum incompatíveis com a genética estudada pelos mendelianos e que, portanto, os dois lados poderiam estar com parte da razão. A teoria da evolução, como a conhecemos atualmente, foi sendo refinada e acabou de ser definida por volta de 1940 com diversos trabalhos, em especial aqueles de Theodosius Dobzhansky, Julian Huxley e Ernst Mayr, que sintetizaram o progresso da década de 30 na área e aplicaram os conceitos resultantes a diversos problemas evolutivos, contribuindo para a difusão e consolidação do evolucionismo até hoje vigente ^(84, 104).

1.2.3 Inferência Filogenética

Com as noções de que todos os seres vivos partilham um único ancestral comum e que a estrutura dos organismos de hoje traz gravada em si parte de sua história evolutiva, surgiu então um grande interesse no estudo das relações de parentesco entre os organismos, especialmente na época de Darwin. Isso pode ser observado nos trabalhos de Ernst Haeckel, que inventou o termo

filogenia e já desenhava árvores relacionando os grupos conhecidos na década de 1860 (Fig. 5).

Qualquer filogenia arbitrária, por mais absurda que possa parecer, pode explicar os dados obtidos, sempre incompletos. Para tanto, bastaria apenas que escolhêssemos algum cenário evolutivo em que tal árvore fosse produzida. Portanto, são necessários métodos de análise filogenética consistentes com a realidade biológica para selecionar uma (ou mais) árvore(s) que seja(m) favorecida(s) por nossos critérios dentre o enorme número de possibilidades à disposição.

Fig. 5 – Filogenias “universais” feitas por Ernst Haeckel. Retirado de Swofford, 1996 e Gould, 1977.

Assim, várias abordagens para inferir as relações filogenéticas foram sendo desenvolvidas ao longo do século XX, resultando em diversas controvérsias acerca tanto de tais relações quanto dos métodos e tipos de dados utilizados para se chegar a elas⁽¹⁾, várias das quais continuam até hoje de muitas maneiras. Inicialmente, foram utilizadas principalmente as características morfológicas para a definição de filogenias. Uma autoridade, alguém com bastante experiência no estudo do grupo em questão, utilizava seus conhecimentos da morfologia, paleontologia, ecologia, ontogenia e outras áreas para decidir, de maneira hoje considerada excessivamente subjetiva, quais seriam as relações filogenéticas envolvidas. Essa é a chamada sistemática evolutiva, onde não há um método bem definido de como são feitas a escolha e a análise das características utilizadas, o que tira muito da objetividade das conclusões alcançadas.

Com o tempo, críticos dessa falta de método na inferência filogenética propuseram o que hoje chamamos de taxonomia numérica ou sistemática fenética. Nessa abordagem, o que importa é a quantidade de similaridade geral entre os organismos estudados: quanto maior a semelhança, maior a proximidade filogenética, uma vez que houve menos tempo para o acúmulo de diferenças entre eles. Assim, compilando-se as semelhanças em diversas características, seria possível construir matrizes que possibilitariam o cálculo de uma quantidade muito mais objetiva, a similaridade, interpretada como distância evolutiva entre os organismos medidos.

Apesar do avanço obtido, para a sistemática, com a adoção de métodos objetivos bem definidos e repetíveis, a fenética também encontrou muitas críticas aos seus procedimentos, que poderiam muitas vezes levar a conclusões errôneas devido, principalmente, a convergências. Nas décadas de 50 e 60, paralelamente ao estabelecimento da taxonomia numérica, foi desenvolvida uma

nova metodologia: a sistemática cladística. Essa nova abordagem, introduzida principalmente por Walter Zimmermann e Willi Hennig, considera válidas para análise apenas as sinapomorfias, ou seja, as características homólogas derivadas compartilhadas, ao contrário do que ocorria anteriormente, com a utilização também de plesiomorfias (características primitivas) e homoplasias (convergências, reversões ou paralelismos). Além disso, os únicos agrupamentos aceitos por essa escola de sistemática são os monofiléticos, aqueles em que o ancestral e todos seus descendentes pertencem ao grupo definido (Fig. 6A)⁽¹⁰⁴⁾.

O componente essencial da cladística para seu sucesso na inferência das relações evolutivas foi a exigência de homologia derivada entre as características comparadas, ou seja, a característica deve ser compartilhada por ter vindo de um ancestral comum próximo. Características homólogas, mas primitivas no grupo estudado (plesiomorfias), não serviriam para inferência filogenética, uma vez que grupos definidos segundo essas características estariam excluindo organismos mais derivados (ou seja, que tiveram muitas mudanças), mas descendentes do mesmo ancestral. Tais grupos, chamados parafiléticos, são considerados inválidos pela cladística, uma vez que excluem linhagens irmãs. O exemplo clássico de um grupo parafilético são os répteis, pois esse grupo exclui seus descendentes aves e mamíferos (Fig. 6B). Grupos definidos por homoplasias costumam possuir vários ancestrais diferentes e não proximamente relacionados; tais grupos são chamados polifiléticos, e também não são aceitos pelos cladistas (Fig. 6C). Seria o caso extremo, por exemplo, de um grupo que reunisse cetáceos e peixes, excluindo todos os outros organismos, baseando-se na forma geral do corpo (nesse caso, claramente uma convergência)⁽¹⁰⁴⁾.

Fig. 6 – Árvores exemplificando grupos mono, para e polifiléticos, e o tipo de características que os definem. Modificado de Ridley, 1996.

O problema agora não seria tanto o método de classificação e suas bases filosóficas, mas preocupações de ordem técnica; em especial, a preocupação na definição da homologia entre as características observadas. Tal definição é complexa por ter embutido em si, de antemão, o conceito de ancestralidade comum: como definir se a característica é homóloga, ou seja, vinda do ancestral comum mais próximo, se ainda não conhecemos a filogenia e, portanto, não sabemos se duas linhagens em questão possuem mesmo um ancestral comum exclusivo de onde teriam herdado a característica analisada? Não seria mais objetivo utilizar todas as características indistintamente, como fazem os feneticistas, sem decidir previamente quais teriam mais valor? A análise não estaria sofrendo de circularidade? Críticas como essas, feitas pelos defensores da fenética, deram origem a muito debate nas primeiras décadas da cladística^(54, 70), e tiveram como resultado a definição de critérios cujas interpretações e aplicações fossem independentes da filogenia, como a comparação paleontológica, a observação de grupos externos e, especialmente, a análise ontogenética das características em estudo.

Paralelamente ao avanço dos métodos de inferência filogenética, as técnicas moleculares foram sendo desenvolvidas. A utilização de dados moleculares em estudos evolutivos foi se tornando cada vez mais comum, conforme o desenvolvimento técnico foi aumentando a facilidade de obtenção de tais dados, indo do estudo indireto dos genes a partir dos efeitos fenotípicos presentes em mutantes, até a obtenção de genomas completos que se tem atualmente.

1.2.4 Métodos moleculares em sistemática

Embora os métodos morfológicos sejam muito úteis em uma grande variedade de situações, especialmente na análise de grupos de Metazoa, há muito mais instâncias em que eles não podem ser aplicados, uma vez que não existem características morfológicas suficientes compartilhadas entre organismos mais distantemente relacionados⁽¹¹⁵⁾. Não é possível fazer comparações significativas entre bactérias, plantas e animais, ou entre os diversos protozoários, para citar alguns exemplos. Mesmo entre organismos mais aparentados, como os mamíferos, existe muita dificuldade em se resolver certos aspectos, como bem demonstram, por exemplo, as diversas controvérsias nas filogenias de cetáceos e proboscídeos^(34, 94).

Conforme postulado por Zuckerkandl & Pauling, 1965, *"em qualquer nível de organização, a quantidade de história preservada será tanto maior quanto maior for a complexidade dos elementos naquele nível, e menores as partes destes elementos que tem de ser afetadas para causar uma mudança significativa"*. Por essa definição, dentre as moléculas presentes em organismos vivos, DNA, RNA e proteínas (as chamadas "moléculas semantoforéticas", ou carregadoras de significado) seriam as mais adequadas para utilização em inferência filogenética. Desse modo, as macromoléculas biológicas teriam grande potencial em refletir a história evolutiva dos organismos aos quais pertencem, uma vez que são

passadas do ancestral ao descendente ou, no caso das proteínas, refletem o que está codificado nas moléculas herdadas.

Entretanto, devido a limitações técnicas e conceituais, muito tempo se passou até que a utilização desse tipo de dados se tornasse possível na prática. Os primeiros estudos moleculares envolvendo filogenia, no começo do século, envolviam reações cruzadas de anticorpos (distância imunológica), e sugeriam que a intensidade dessas reações deveria ser diretamente proporcional ao grau de parentesco entre os organismos. Outro método que também resulta diretamente em distâncias, a utilização de curvas de hibridação, foi desenvolvido mais tarde, permitindo uma análise menos indireta do genoma. Esses métodos têm limitações severas devido a dificuldades técnicas e à falta de resolução inerente, e são atualmente muito pouco utilizados para a inferência filogenética⁽¹²¹⁾.

Métodos bioquímicos cada vez mais refinados passaram a permitir estudos mais aprofundados das macromoléculas biológicas. As proteínas receberam, historicamente, maior investimento dos pesquisadores, uma vez que se suspeitou, até a década de 40, que fossem as responsáveis pelo armazenamento e transmissão da informação genética. Assim, os métodos utilizados para se trabalhar com peptídeos foram desenvolvidos bem antes daqueles necessários para a manipulação de ácidos nucléicos e isso, naturalmente, fez com que as proteínas fossem as primeiras moléculas diretamente utilizadas para inferência filogenética. Inicialmente, embora bastante trabalhoso, o seqüenciamento de proteínas era mais simples que o de ácidos nucléicos. A comparação de padrões de isoenzimas ainda hoje é uma técnica bastante útil para vários propósitos em sistemática, embora já não seja tão utilizada para filogenia como costumava ser há alguns anos, devido à disponibilidade de técnicas muito mais confiáveis e refinadas que envolvem seqüências de DNA ou RNA.

Após os trabalhos de Avery e cols., 1944, estabelecendo definitivamente que a hereditariedade se deve à transmissão de ácidos nucleicos, e de Watson e Crick, 1953, sugerindo o modelo de dupla hélice de fitas complementares e antiparalelas para a estrutura do DNA, seguiu-se um maior interesse no estudo dessas moléculas, levando ao aperfeiçoamento das técnicas de manipulação de ácidos nucléicos, como a utilização de enzimas de restrição, técnicas de clonagem, hibridação e seqüenciamento, entre muitas outras. Logo tornou-se claro que a investigação direta das informações genéticas seria sempre mais informativa do que a utilização de meios indiretos⁽¹⁴⁶⁾. Afinal, por definição, toda característica utilizada em inferência filogenética deve ter base genética, uma vez que só assim será herdada e poderá refletir as relações de parentesco entre os organismos. Além disso, os dados moleculares tendem a colocar uma quantidade bem maior de características à disposição do investigador do que os dados morfológicos, pois geralmente se considera que cada nucleotídeo ou aminoácido seja um caráter independente na análise, o que faz com que o estudo de um gene típico, por exemplo, envolva cerca de 1.000 características. Mesmo levando-se em consideração o fato de que, dependendo do método de análise dessas seqüências, muitas das posições podem nem ser utilizadas, ainda assim a quantidade de informação disponível atualmente é muito maior do que no passado^(72, 121).

Assim como a natureza dos dados moleculares é bastante diversa daquela das características morfológicas, os métodos de análise filogenética utilizando seqüências também tiveram de ser adaptados a partir dos métodos tradicionais, e novas abordagens foram criadas. São utilizadas três abordagens na análise de dados moleculares para a inferência filogenética: métodos baseados em análise de distância, máxima parcimônia e máxima verossimilhança^(27, 121). A definição de quais métodos seriam “fenéticos” ou “cladísticos” é muito mais sutil quando se

trata dos dados moleculares do que quando são considerados os dados morfológicos. Isso porque, excetuando-se casos extremos como, por exemplo, UPGMA ("Unweighted Pair Group Method with Arithmetic Mean"), aplicado a dados sem indicação de homologia, decididamente fenético, é muito mais difícil a aplicação desses conceitos à maioria dos outros métodos, sendo que talvez nem mesmo seja muito relevante tal tipo de classificação no caso das análises moleculares. Exceto pela comparação com um grupo externo, a definição de quais substituições em um alinhamento seriam sinapomorfias não pode ser feita *a priori*, uma vez que é praticamente impossível analisar as seqüências de fósseis (exceto em alguns poucos casos), e a ontogenia, nesse caso, obviamente não tem nenhuma utilidade. Entretanto, a incerteza ainda é grande, mesmo na comparação com um grupo externo, uma vez que a probabilidade de uma homoplasia estar sendo observada depende muito da natureza da molécula em questão, e a maioria dos métodos de análise filogenética não leva fatores probabilísticos em consideração.

Um método pode também ser algorítmico ou de critério de optimalidade. Um método algorítmico é aquele que segue uma seqüência de passos para a construção da árvore filogenética, chegando a uma resposta única. Já nos métodos que utilizam um critério de optimalidade, analisam-se árvores filogenéticas arbitrárias de modo a encontrar aquela(s) que minimize(m) ou maximize(m) uma certa grandeza, por exemplo, número de substituições ou verossimilhança, entre outras. Assim, os métodos de optimalidade são de implementação computacional muito mais complexa do que os algorítmicos, uma vez que precisam analisar o maior número de árvores possível, sendo que o total de árvores pode ser tão grande que impossibilite totalmente a análise de cada uma delas, conforme detalhado no item 3.5.4.2 de Resultados e Discussão. Assim, tornam-se necessários métodos de exploração do espaço de árvores

possíveis, de modo a tentar encontrar a “melhor”, sem necessidade de se examinar todas elas. Diversos métodos de procura e avaliação de filogenias foram desenvolvidos nos últimos anos, chegando a aproximações bastante satisfatórias^(28, 121). Ainda assim, métodos de computação intensa, como máxima verossimilhança, ainda não podem ser plenamente utilizados, mesmo para quantidades modestas de *taxa*, ao se comparar com as possibilidades atuais de obtenção de dados. Uma exposição mais detalhada dos métodos de exploração está presente no item 3.5.4.2 de Resultados e Discussão.

1.2.4.1 Métodos baseados em distâncias

Entre os métodos de inferência filogenética, aqueles baseados na análise de distâncias entre seqüências são os de computação mais veloz, o que os torna especialmente úteis na análise de grandes quantidades de *taxa*. Um estudo recente, por exemplo, analisou 500 seqüências na tentativa de construir uma filogenia dos grupos eucarióticos⁽¹²⁸⁾ utilizando, para tanto, um método de cálculo de distâncias que leva em conta a heterogeneidade de taxa de substituição ao longo da seqüência (no caso, SSU rDNA), seguido pela construção de uma árvore por “neighbor joining”⁽¹⁰⁷⁾, um algoritmo que agrupa as seqüências par a par. Tal análise seria impraticável sem a utilização de um método algorítmico. Mesmo assim, a computação das taxas de substituição para 750 seqüências levou cerca de 20 horas, utilizando o software TREECON⁽¹²⁷⁾, em um computador Pentium 200 MHz, com 64 Mb de RAM e sob Windows 95.

A análise de um alinhamento por distâncias envolve basicamente dois passos independentes: computação das distâncias entre as seqüências e construção (nos métodos algorítmicos) ou avaliação (nos métodos de optimalidade) de uma árvore. Caso seja utilizado um método algorítmico, o resultado será uma única

árvore, enquanto os métodos de optimalidade podem encontrar diferentes árvores que se adequem igualmente bem à matriz de distâncias.

A computação das distâncias sempre se dá par a par, resultando em uma matriz triangular com $N(N-1)/2$ valores, sendo N o número de *taxa* no alinhamento. Existem diversas fórmulas para o cálculo da distância entre duas seqüências, baseadas em diferentes modelos do processo evolutivo. Cada modelo assume diferentes premissas, o que pode levar a conclusões errôneas no caso de se utilizar um certo modelo para analisar seqüências que evoluíram de um modo diferente. Além disso, tais fórmulas contêm correções matemáticas para o fato de que, conforme a divergência entre duas seqüências aumenta, vai aumentando a probabilidade de que uma posição tenha sofrido mais de uma mutação (eventos superpostos). Isso levaria a uma subestimação da verdadeira distância evolutiva entre tais seqüências, já que o número de substituições observadas é menor do que o que realmente ocorreu, como pode ser visto no exemplo hipotético da figura 7. Os principais modelos de evolução de nucleotídeos serão detalhados mais adiante, na seção 2.9.3, sobre máxima verossimilhança, do item Material e Métodos.

Os diversos métodos baseados em critério de optimalidade para construção de árvores a partir de matrizes de distância são, em geral, do tipo “mínimos quadrados”, ou seja, eles procuram a(s) árvore(s) que minimize(m) a soma dos quadrados das diferenças entre as distâncias observadas e aquelas computadas na árvore analisada⁽¹²¹⁾.

Fig. 7 – Simulação da evolução de duas seqüências a partir de um ancestral, mostrando todas as mutações ocorridas entre a divergência e o tempo atual e os nomes dos tipos de substituição envolvidos. Modificado de Li & Graur, 1991.

Deve-se notar que os métodos de optimalidade não fornecem as árvores a serem analisadas conforme seu critério, mas apenas resultam em um valor que relaciona cada uma das árvores à matriz em questão, guardando aquelas que melhor satisfaçam ao critério de optimalidade envolvido. Desse modo, esses métodos, embora utilizem distâncias, não são de computação tão veloz quanto os métodos algorítmicos, uma vez que precisariam analisar todas as árvores possíveis para encontrar, com certeza, aquela que minimiza (ou maximiza) a quantidade procurada. Conforme citado anteriormente, essa tarefa se torna impossível na prática mesmo para pequenos números de *taxa*, o que torna a eficiência desses métodos dependente do modo como será feita a exploração do espaço de árvores.

Já os métodos algorítmicos de construção de filogenias não precisam analisar nenhuma árvore, uma vez que foram desenvolvidos de modo a executar uma série de passos bem definidos que vão juntando os *taxa* e calculando o comprimento dos ramos até chegar a uma só árvore. Assim, pode-se perceber que tais métodos trazem embutidos, em um mesmo processo (o algoritmo), tanto a construção da árvore quanto à definição de como ela deve ser. Entre os métodos algorítmicos mais utilizados estão o UPGMA⁽¹¹²⁾ e o “neighbor joining”⁽¹⁰⁷⁾. O UPGMA (“Unweighted Pair Group Method with Arithmetic Mean”) assume que todas as linhagens evoluíram com a mesma velocidade, ou seja, obedecendo a um relógio molecular. As distâncias, em um caso como esse, são chamadas ultramétricas, e o aspecto de uma árvore ultramétrica (aquela em que a distância entre a raiz e qualquer dos *taxa* é sempre a mesma) pode ser visto na figura 8. O “neighbor joining”, por outro lado, permite que cada linhagem tenha uma taxa de evolução diferente das outras, resultando em uma árvore aditiva e não-ultramétrica, cujas distâncias entre a raiz (se houver) e os *taxa* terminais pode, portanto, ser variável (Fig. 8). Ou seja, as árvores ultramétricas são casos especiais de árvores aditivas.

Fig. 8 – Exemplos de árvore ultramétrica e aditiva e suas propriedades. Modificado de Li & Graur, 1991.

A maior crítica aos métodos de distância é que ocorreria uma grande perda de informação ao se calcular a distância entre duas seqüências, uma vez que todo o

padrão de diferenças seria reduzido a um único número. É evidente que se observarmos três seqüências alinhadas A, B e C, sendo que a distância entre A e B (D_{AB}) seja igual a 30%, $D_{AC}=30\%$ e $D_{BC}=10\%$, isso não levará à conclusão de que B é igual a C simplesmente pelo fato de estarem igualmente distantes de A, uma vez que a natureza das diferenças entre A, B e C foi perdida no cálculo das distâncias. Ou seja, os métodos de construção das árvores estariam se baseando em uma versão muito resumida dos dados (a matriz de distâncias), o que diminuiria muito o poder de resolução desses métodos na inferência filogenética. É quase o mesmo tipo de crítica que os cladistas fazem aos métodos fenéticos puros, de que a natureza das diferenças, essencial para a inferência, estaria sendo perdida ao se misturar todas as características como igualmente válidas para a definição do parentesco.

1.2.4.2 Métodos baseados no princípio da parcimônia

Filosoficamente, a parcimônia se baseia nas idéias do filósofo britânico do século XIII William de Ockham, cuja “navalha de Ockham” é o princípio de que hipóteses mais simples têm preferência sobre as mais complexas, quando ambas explicarem igualmente bem uma dada situação⁽²⁷⁾.

Basicamente, as diversas variações de métodos baseados em parcimônia tomam como premissa a evolução lenta das características analisadas. Esses métodos sempre procuram a árvore que minimize o número de passos (em estudos moleculares, substituições de nucleotídeos ou aminoácidos) necessários para explicar os padrões observados nos dados⁽¹²¹⁾. Devido ao conceito de simples compreensão e ao grande número de programas de computador disponíveis para esse tipo de análise, os métodos de parcimônia são os mais utilizados em filogenia, para qualquer tipo de dados.

Os diferentes métodos de parcimônia diferem entre si pelos tipos de substituição permitidos e pelos pesos que serão dados a cada tipo, resultando em várias abordagens diferentes, como as parcimônias de Fitch (todas as mudanças são iguais), Wagner (as mudanças têm polaridade), Dollo (apomorfias devem surgir uma única vez), Camin-Sokal (evolução irreversível) e de transversão (peso maior para as mudanças que envolvem transversões). De modo geral, todos esses modelos podem ser resumidos na parcimônia generalizada⁽¹²¹⁾, onde se utiliza uma matriz auxiliar que estipula o peso que cada tipo de mudança deve ter, baseada em evidências independentes da frequência de cada tipo possível.

A principal crítica aos métodos de parcimônia se deve a sua principal premissa: de que a evolução procederia lentamente ou, pelo menos, com igual velocidade em todas as linhagens. Já foi fartamente demonstrado que este não é o caso para muitas das moléculas analisadas^(72, 121, 143), levando à inconsistência do método⁽²⁴⁾. Ocorreria, portanto, o fato paradoxal de que, quanto mais dados fossem analisados, mais certeza haveria de que a árvore obtida (incorreta) é a árvore certa. Esse tipo de problema ocorre com grande frequência na chamada “zona de Felsenstein” do espaço de árvores, conforme demonstrado por diversos estudos de simulação^(50, 52, 121). Essa é a parte desse espaço em que a árvore real (de quatro *taxa*) possui dois organismos de evolução muito rápida e dois de evolução mais lenta, sendo que um dos organismos rápidos é mais aparentado a um dos mais lentos, e o ramo interno da árvore também é curto (Fig. 9). Tal tipo de situação é comum em radiações adaptativas em que novas *taxa* se diferenciam rapidamente dos grupos irmãos devido à exploração de novos nichos, sofrendo grandes pressões de seleção⁽¹⁰⁴⁾. A árvore estimada pela parcimônia será aquela que reúne os dois organismos de evolução rápida. Tal fenômeno é conhecido como atração dos ramos longos, e ocorre devido ao

grande número de substituições sofridas pelas moléculas desses organismos de evolução mais veloz, o que leva a um grande número de substituições convergentes e paralelas (como exemplificado na figura 7), que acabam por criar uma similaridade artificial entre eles.

Assim como ocorre com todo método de optimalidade, a parcimônia também depende do algoritmo de busca de árvores utilizado. Quanto melhor for a exploração do espaço de árvores, tanto maior será a probabilidade de ter realmente encontrado a árvore com o menor número de passos possível (árvore mais parcimoniosa).

Fig. 9 – Exemplo mostrando o potencial de inconsistência da parcimônia em casos de ocorrência de taxas de evolução muito diferentes entre organismos. A árvore **A** é a árvore correta (com os ramos proporcionais ao número de substituições ocorridas), enquanto a árvore **B** é aquela inferida por parcimônia. Devido ao grande número de substituições ocorrendo nos longos ramos que levam aos organismos x e y, é maior a chance de ocorrerem muitas homoplasias entre as seqüências destes taxa, tornando-as artificialmente semelhantes.

1.2.4.3 Máxima verossimilhança e modelos de evolução molecular

Os métodos mais consistentes, robustos e eficientes para a inferência filogenética são aqueles baseados na máxima verossimilhança, um critério de optimalidade derivado de uma grande quantidade de cálculos estatísticos, e que requer a utilização de modelos de evolução molecular completamente explícitos⁽²⁵⁾. Esses métodos procuram, idealmente, a árvore que tenha o maior valor de verossimilhança dentre todas as possíveis para um grupo de seqüências.

A verossimilhança de uma árvore pode ser definida como a probabilidade de que os dados observados (alinhamento) tivessem sido obtidos caso uma árvore (arbitrária) e um modelo de evolução das seqüências (arbitrário) tivessem ocorrido na história evolutiva das seqüências analisadas - $P(D|A,M)$. Como a verossimilhança é diretamente proporcional a essa probabilidade, pode-se concluir que, ao maximizar a verossimilhança, maximiza-se também a probabilidade de que os dados observados tivessem ocorrido caso a árvore e o modelo utilizados fossem os verdadeiros^(27, 121).

O método consiste em calcular a verossimilhança de cada uma das posições do alinhamento, e depois multiplicar todas as verossimilhanças para encontrar o valor geral:

$$L = \prod_{j=1}^N L_{(j)} \quad \text{ou} \quad \ln L = \sum_{j=1}^N \ln L_{(j)} \quad (1)$$

onde L é a verossimilhança, N é o número de posições analisadas, j é a posição atual e $\ln L$ é o logaritmo neperiano da verossimilhança. De modo a tornar possível a manipulação da enorme quantidade de multiplicações e números com muitas casas decimais envolvidos (que inviabilizariam a computação devido à arquitetura interna dos processadores), utiliza-se, na realidade, o logaritmo neperiano da verossimilhança ($\ln L$) para a expressão da verossimilhança de uma determinada árvore. Desse modo é possível expressar o produto das verossimilhanças como a soma dos $\ln L$.

A verossimilhança de cada posição ($L_{(j)}$ nas equações acima) é dada pela soma das probabilidades de cada cenário evolutivo ancestral que possa ter dado origem ao padrão presente no alinhamento, conforme pode ser visto na Fig. 10. Felizmente, existe uma maneira ("algoritmo de poda") de evitar a necessidade de se fazer todos estes cálculos, uma vez que o número de probabilidades a serem calculadas para cada posição da seqüência é de 2^{2n-2} (n = número de

taxa), rapidamente impossibilitando as análises^(25, 121). O cálculo dessas probabilidades depende do modelo evolutivo escolhido, e cada modelo leva em conta uma série de fatores que se acredita influenciarem o processo de divergência entre as seqüências.

Como todo método estatístico, a máxima verossimilhança funciona comparando os desvios entre os valores esperados e os estimados, a partir dos dados de certos parâmetros. Para isso, é necessário que se tenha modelos que predigam o que seria esperado em determinada situação. No caso das filogenias moleculares, os modelos devem especificar o modo como as seqüências evoluem. Desse modo, o método irá verificar o quanto o conjunto dados obtidos (alinhamento) mais hipótese (árvore e modelo) concorda com o que seria esperado, caso essas seqüências tivessem realmente evoluído segundo o modelo e árvore utilizados.

Fig. 10 – Esquema do cálculo de verossimilhança para cada posição de um alinhamento, para uma dada árvore. No momento observado, está sendo feito o cálculo para a posição j ($j=1..N$, N é o comprimento do alinhamento), que apresenta o padrão CCAG. São calculadas, de acordo com o modelo de evolução escolhido, as probabilidades das ocorrências dos diversos cenários ancestrais possíveis (ou $P(1,2)$, nucleotídeos nos nós 1 e 2 da árvore utilizada). A soma dessas probabilidades é a verossimilhança da posição j , e o produto das verossimilhanças de todas as posições dá a verossimilhança da árvore toda (equação 1 acima). Modificado de Swofford e cols., 1996.

Dos diversos modelos desenvolvidos para explicar a evolução de seqüências, os mais comumente empregados são aqueles de Jukes & Cantor, 1969 (JC), Kimura, 1980 (K2P), Felsenstein, 1981 (F81), Hasegawa, Kishino & Yano, 1985 (HKY) e Tamura & Nei, 1993 (TrN). Esses modelos diferem pelo número de tipos

de substituição consideradas e se todas as bases terão ou não a mesma frequência. Além disso, estão aninhados, em uma hierarquia de complexidade, o que significa que cada modelo mais simples é, na verdade, um caso especial de um modelo mais complexo. No caso dos modelos aqui descritos, a seqüência (do mais simples para o mais complexo) seria JC – K2P ou F81 – HKY – TrN. Deve-se notar que K2P e F81 derivam de JC por motivos diferentes (conforme detalhado na seção 2.9.3 do item Material e Métodos), e um não está subordinado ao outro.

Todos esses modelos, além de vários outros menos citados na literatura, são casos especiais do modelo conhecido como GTR (“general time reversible”), que considera tanto frequências de bases diferentes quanto seis tipos de substituição possíveis ($A \leftrightarrow T$, $A \leftrightarrow C$, $A \leftrightarrow G$, $G \leftrightarrow C$, $G \leftrightarrow T$ e $C \leftrightarrow T$)^(140, 149).

Para cada um dos modelos apresentados podem ser derivadas, a partir das probabilidades de substituição, fórmulas para cálculos de distância entre seqüências, permitindo o uso desses modelos também em análises de distância, como as descritas anteriormente.

Outros fatores relevantes do processo evolutivo podem ser adicionados aos modelos utilizados pelo método de máxima verossimilhança, como a determinação de diferentes taxas de evolução para diferentes posições da seqüência, muito importante de se considerar quando se utilizam genes como os ribossômicos, que possuem tanto regiões de evolução lenta quanto outras de evolução mais rápida. A heterogeneidade de taxa de substituição pode ser incorporada aos modelos utilizando uma distribuição conhecida como gama, contínua ou discreta^(121, 141, 142). Tal distribuição possui um parâmetro denominado α : quanto menor o valor de α , maior será a heterogeneidade de taxas de substituição nas seqüências analisadas. Do mesmo modo, um alto valor

de α significará que todas as posições da sequência evoluem na mesma taxa^(39, 121), conforme detalhado mais adiante no item 3.5.3 de Resultados e Discussão.

Além da heterogeneidade de taxa de substituição entre sítios, esses métodos também podem testar se as diferentes linhagens estão evoluindo segundo um relógio molecular único (todos com a mesma taxa de evolução), ou não. Caso as sequências em questão estejam evoluindo segundo um único relógio molecular, é possível fazer estimativas do tempo de divergência entre os organismos envolvidos, desde que se tenha algum tipo de dado independente para “calibrar o relógio” (estimar a relação entre o número estimado de substituições e o tempo de divergência).

Conforme vão sendo utilizados modelos mais complexos, com mais parâmetros para explicar a evolução das sequências, maior vai se tornando o valor de verossimilhança. Embora possa parecer interessante utilizar o modelo mais complexo disponível, de modo a maximizar o critério de optimalidade, isso não se constitui em vantagem se a complexidade adicional introduzida não resultar em uma explicação melhor do processo evolutivo em relação a modelos mais simples, uma vez que os parâmetros em excesso estarão introduzindo mais incerteza nas estimativas. Desse modo, é necessário que sejam utilizados critérios para avaliar até que ponto vale a pena continuar adicionando fatores aos modelos utilizados.

Vários estudos envolvendo simulações computadorizadas do processo evolutivo foram realizados para determinar os efeitos da utilização desses modelos sob diversas condições, de modo a desenvolver conceitos e testes que possibilitem a escolha do modelo mais adequado para análise filogenética em cada caso^(40, 145). A utilização de simulação de conjuntos de dados e de uma distribuição nula, e o teste do modelo utilizado em relação a essa distribuição é o modo mais correto, estatisticamente, de se fazer tais comparações e escolher o

modelo mais adequado⁽⁴⁰⁾. Entretanto, a complexidade computacional leva a tempos de análise impraticáveis para muitos *taxa*, uma vez que é necessária a simulação de muitos conjuntos de dados. Assim, para comparação de modelos aninhados (aqueles que são casos especiais uns dos outros, como os descritos anteriormente) costuma-se utilizar uma aproximação para a distribuição nula: a distribuição do χ^2 (Qui-quadrado), associada ao teste de razão de verossimilhanças (LRT)^(25, 53). Tal teste é de implementação bastante simples:

$$LRT = 2(\ln L_n - \ln L_a) \quad (2)$$

onde L_n é a verossimilhança obtida para o modelo nulo (mais simples) e L_a é a verossimilhança obtida para o modelo alternativo (mais complexo). O valor de LRT é então comparado ao valor crítico presente na distribuição do χ^2 , utilizando-se, como graus de liberdade, o número de parâmetros livres (ver tabela 1 no item Material e Métodos) do método alternativo menos o número de parâmetros livres do método nulo (como exemplo: em uma comparação entre HKY e K2P, seriam utilizados 3 graus de liberdade). Caso o LRT seja menor do que o valor crítico de χ^2 para uma certa significância estatística (geralmente 5%), então não se pode rejeitar o modelo mais simples, que passa a ser preferido por apresentar menor número de parâmetros a estimar e, portanto, menor variância nas estimativas.

Embora o conceito da verossimilhança em si seja razoavelmente simples, pode-se perceber que a implementação do método é bastante complexa, sendo impossível na prática sem a utilização de computadores velozes. Para cada árvore analisada, são necessárias quantidades enormes de cálculos envolvendo multiplicação de muitos números e somas de logaritmos, além da utilização de modelos de evolução bastante sofisticados e idealmente com diferentes graus de realismo (complexidade). Todos esses fatores contribuíram para a baixa

utilização desse método até pouco tempo atrás. Desse modo, apenas mais recentemente (década de 1980 para nucleotídeos e de 1990 para aminoácidos), esses métodos passaram a ser mais utilizados⁽¹²¹⁾. Ainda hoje, a máxima verossimilhança é de difícil aplicação quando se tem uma quantidade maior de *taxa*, uma vez que cada análise poderia demorar semanas nos computadores normalmente utilizados.

1.2.5 Exame da robustez e confiabilidade das filogenias

Deve-se manter em mente o fato de que a procura de uma filogenia é sempre uma estimativa que se faz sobre um processo passado ao qual não se tem mais acesso. Tal investigação é feita a partir de dados contemporâneos incompletos e que refletem o processo de maneiras muitas vezes incertas. É necessária, portanto, a utilização de meios para se avaliar a incerteza envolvida na atividade, de modo a tentar maximizar a exatidão das estimativas filogenéticas. Basicamente, há quatro formas de se chegar a esse objetivo: simulações do processo evolutivo seguidas de inferência filogenética utilizando os dados simulados⁽⁵²⁾, análise de filogenias que se conhece com certeza^(48, 51), testes estatísticos⁽⁷³⁾ e estudos de congruência (comparação de diferentes abordagens)⁽⁸⁹⁾. Essas abordagens tornam possível a avaliação das condições que maximizem consistência (habilidade de encontrar a árvore correta conforme os dados tendem ao infinito), eficiência (quanto menos dados necessários para se chegar à filogenia correta, maior a eficiência) e robustez (resistência à violação das premissas) dos métodos de inferência filogenética, permitindo a melhor compreensão dos fatores envolvidos no funcionamento de cada método e, portanto, a escolha daquele que melhor se adequa à análise em questão⁽⁵¹⁾.

Entre os vários métodos desenvolvidos para se investigar o nível de suporte para uma filogenia, o "bootstrap" não-paramétrico é um dos mais utilizados,

devido à facilidade de implementação e interpretação^(26, 49, 108). Esse método se baseia na perturbação aleatória do conjunto de dados um grande número de vezes (originando diversas subréplicas), seguida do cálculo de uma árvore para cada subréplica. O consenso de todas as árvores indicar quais ramos são melhor apoiados pelos dados em questão. O método se baseia na suposição de que pequenas modificações nos dados não diminuirão a capacidade de se encontrar os grupamentos se estes estiverem bem representados pelos dados. Assim, grupamentos com baixos valores de "bootstrap" são aqueles que deixaram de ser inferidos em muitas das subréplicas, o que sugere baixo suporte dos dados em questão para esses grupos. Entretanto, é bom observar que os valores obtidos não indicam a probabilidade de que os grupamentos sejam corretos filogeneticamente, uma vez que diferentes seqüências podem determinar diferentes grupos, dependendo do caso analisado (por motivos como paralogia ou transferência horizontal).

Métodos mais sofisticados e com melhores propriedades estatísticas como, por exemplo, simulações associadas a testes de máxima verossimilhança, tanto para comparações entre modelos como entre diferentes métodos^(39, 52, 145), normalmente exigem muito tempo de computação conforme o número de *taxa* vai aumentando. Tanto é assim que tais estudos sempre envolvem apenas quatro ou cinco seqüências, uma vez que números maiores tornariam impossível a análise do grande número (no mínimo 100, normalmente) de simulações necessárias para se alcançar a significância estatística mínima desejável.

1.2.6 Os conceitos de espécie

Ao longo da história, diversos conceitos de espécie foram sendo desenvolvidos e utilizados para auxiliar no estudo da diversidade orgânica, na tentativa de estabelecer critérios para a classificação dos diferentes organismos. Embora a

discussão dos conceitos de espécie seja quase invariavelmente considerada importante, dificilmente os inúmeros autores trabalhando em problemas taxonômicos explicitam quais seriam os critérios utilizados para a definição dessa categoria nos *taxa* com os quais trabalham. Os primeiros conceitos utilizados foram os morfológicos, com suas variantes fenéticas ou tipológicas. Tais conceitos, ainda implicitamente (e, talvez, inadvertidamente) muito utilizados na literatura de taxonomia, consistem em se delimitar um tipo básico, um modelo, ao qual outros espécimes serão comparados para se determinar sua identidade específica. Dependendo da variante em questão, o conceito permite maiores ou menores níveis de variabilidade intraespecífica (ou seja, divergência em relação ao tipo). Com o desenvolvimento de novos conceitos nas ciências biológicas, diversos outros conceitos foram sendo propostos para diferentes casos, e os principais e melhor representados na literatura estão discutidos no item 3.5.2 de Resultados e Discussão. Embora se possa concluir, de todas essas tentativas, que é muito difícil, talvez impossível, a tarefa de se estabelecer algum critério de espécie universal, ainda assim tais conceitos, sem dúvida, representaram progressos no entendimento da especiação e da natureza e realidade das espécies.

A situação corrente na definição e determinação das espécies do gênero *Acanthamoeba* será abordada no item 3.5.2.6 de Resultados e Discussão. Além disso, é bom enfatizar que, embora os organismos reunidos sob o termo “amebas” não formem um grupo natural, assim como todos aqueles chamados genericamente de protozoários, a praticidade e a ampla utilização desses termos são motivos para que se continue a empregá-los. Deve-se, entretanto, ter sempre em mente o fato de que não são grupos válidos do ponto de vista filogenético, uma vez que não são monofiléticos (definidos por relações de parentesco com um ancestral exclusivo).

1.3 – Objetivos

O presente trabalho teve por principais objetivos:

- caracterizar molecularmente e tentar inferir a similaridade genética entre os diversos isolados de *Acanthamoeba* - de úlcera de córnea brasileiros e os de referência provenientes da ATCC (American Type Culture Collection), por meio de RFLP (Restriction Fragment Length Polymorphism) do DNA total, RFLP do SSU rDNA e RAPD (Random Amplified Polymorphic DNA);
- determinar a seqüência do SSU rDNA dos isolados utilizados, a fim de verificar a que grupo de seqüência pertenceriam, ou se determinariam novos tipos de SSU rDNA;
- utilizar as seqüências do SSU rDNA obtidas, na tentativa de estimar a posição filogenética dos isolados estudados, tanto dentro do gênero *Acanthamoeba* quanto em relação a outros grupos eucarióticos, seja de protozoários ou de outros organismos.

Material e Métodos

2. MATERIAL E MÉTODOS

2.1. Organismos e cultivo

2.1.1. Cepas-referência da ATCC:

a) Isolados de *Acanthamoeba polyphaga* -

A. poly. #1 (ATCC 30873) - isolado de córnea humana

A. poly. #2 (ATCC 30461) - isolado de córnea humana

A. poly. #3 (ATCC 30871) - isolado de água

A. poly. #4 (ATCC 30872) - isolado de água

b) Isolados de *A. castellanii* -

A. cast. #1 (ATCC 50374) - isolado de cultura de levedura

A. cast. #2 (ATCC 30868) - isolado de córnea humana

c) *A. astronyxis* (ATCC 30901) - isolado de água

d) *A. royreba* (ATCC 30884) - isolado de células de carcinoma de córion humano

2.1.2. Cepas isoladas de casos de úlcera de córnea humana, provenientes da Escola Paulista de Medicina da UNIFESP, do Hospital das Clínicas da FMUSP e do Laboratório Oftalmolab: U/HC1; U/E1; U/E2; U/E3; U/E4; U/E5; U/E6; U/E7 (clones 1 e 2), U/E8R, U/E8L e U/E8X (isolados dos olhos direito, esquerdo e recidiva, respectivamente, do mesmo paciente), U/E10 e U/Oft1.

Os organismos foram rotineiramente mantidos em placas de ágar contendo 0,2% de extrato de soja. Para o estabelecimento de culturas axênicas, o crescimento das placas foi transferido para garrafas de cultura contendo meio Neff⁽⁹²⁾, acrescido de 100 µg/mL de ampicilina e/ou 80 µg/mL de gentamicina. Após a formação de uma monocamada, as células foram removidas com 10 mM de EDTA, lavadas uma vez em PBS e uma vez em T₁₀E₁₀ (10 mM de Tris e 10 mM

de EDTA), com centrifugações a 1.200 g por 15 min após cada lavagem. O sedimento final foi então armazenado a -20°C ou imediatamente submetido aos procedimentos de extração do DNA genômico total. Foi mantido, para experimentos posteriores, apenas um clone de cada caso clínico, com exceção de U/E7, em que foram mantidos dois clones, devido à grande diferença de tamanho do cisto observada entre eles (ver Anexo 1), e U/E8, um caso de infecção nos dois olhos (U/8R, olho direito e U/8L, olho esquerdo), com um novo isolamento devido à ocorrência de uma recidiva (U/8X).

2.2. Extração de DNA genômico total

O sedimento de células foi ressuspensão em SE (0,1 M de NaCl, 100 mM de EDTA). A lise das células foi realizada pela adição de 1% de SDS, na presença de 10 $\mu\text{g/ml}$ de RNase e de 50 $\mu\text{g/ml}$ de proteinase K. Após homogeneização suave, as preparações foram incubadas por 2 horas a 50°C e, em seguida, tratadas com igual volume de fenol durante alguns minutos (centrifugando então a 10.000 g por 10 min para separar as fases), e repetindo-se o mesmo processo duas vezes, utilizando então fenol/clorofórmio (1:1) e clorofórmio. O DNA foi precipitado pela adição de 0,3 M de acetato de sódio e 0,6 volumes de isopropanol, lavado em etanol a 70%, ressuspensão em TE (10 mM de Tris, 1 mM de EDTA), e mantido a 4°C .

2.2.1 Reações de amplificação aleatória de DNA polimórfico (RAPD)

O método de RAPD é um método multi-locus muito útil na caracterização molecular de organismos^(135, 136), e se baseia na amplificação de seqüências polimórficas de DNA, utilizando para tanto uma variante de PCR onde somente um oligonucleotídeo pequeno (cerca de 10 pb) é utilizado com baixa temperatura de associação (37°C).

Para a caracterização dos isolados de *Acanthamoeba* por RAPD (exceto U/E8X, U/E10 e U/Oft1, isolados em época posterior a esses experimentos), foram inicialmente investigados os padrões resultantes da utilização de diversos oligonucleotídeos arbitrários, a fim de selecionar aqueles que resultassem em padrões de bandas para todos os isolados estudados. Os três oligonucleotídeos escolhidos para a construção de fenogramas cujo consenso está apresentado no Anexo 1 foram os seguintes:

606: 5'- CGGTCGGCCA -3'

688: 5'- GCAGGAGCGT -3'

694: 5'- GGTTTGGAGG -3'

As reações de amplificação foram feitas em volume de 50 µL, contendo 100 ng de DNA molde, 250 µM de dNTPs, 100 pmoles de oligonucleotídeo, 2 U de *Taq* DNA polimerase (Gibco BRL, USA), e 1,5 mM de MgCl₂, em um termociclador Progene (Techne, UK). Os tempos e temperaturas utilizados foram: 94⁰ C por 3 min e 35 repetições de 94⁰ C por 1 min, 37⁰ C por 2 min e 72⁰ C por 2 min. O último ciclo teve a temperatura de 72⁰ C mantida por 10 min. Os produtos de amplificação foram fracionados por eletroforese em géis de agarose a 2%, corados com brometo de etídio e fotografados sob luz ultravioleta em um processador de imagem Eagle Eye II (Stratagene, USA).

As imagens dos géis foram utilizadas para o cálculo de distância genética entre os diversos isolados de *Acanthamoeba* utilizados, conforme os métodos descritos na seção 2.6.

2.2.2 Clivagem do DNA genômico total por endonucleases de restrição

O DNA genômico total de diversos isolados foi submetido à clivagem por diversas enzimas de restrição, a fim de selecionar aquelas que resultassem em padrões adequados, isto é, nem poucas nem muitas bandas, espaçadas de modo

a facilitar ao máximo a visualização para a comparação entre os isolados. As enzimas escolhidas para dar continuidade aos experimentos foram então utilizadas para a clivagem do DNA genômico total de todos os isolados, com exceção de U/E8X, U/E10 e U/Oft1.

A reações de clivagem por endonucleases de restrição foram feitas em volumes de 30 µL, contendo entre 0,5 e 1,0 µg de DNA genômico total, tampão de reação em quantidade recomendada pelos fabricantes e cerca de 1U de enzima de restrição (Pharmacia e New England Biolabs), nas condições indicadas para cada enzima. Os produtos das clivagens foram fracionados por eletroforese em géis de agarose a 0,8% corados com brometo de etídio e fotografados sob luz ultravioleta. O marcador de peso molecular utilizado foi o DNA de fago λ clivado por *Hind*III (λ /*Hind*III).

A partir dos padrões de bandas, foram construídas matrizes codificando a presença ou ausência, no padrão de cada isolado, de cada uma das bandas possíveis para uma determinada enzima. As matrizes foram então utilizadas para o cálculo da similaridade (porcentagem de bandas compartilhadas) de modo similar ao utilizado com os padrões de RAPD (ver seção 2.6 abaixo).

2.3. Construção dos oligonucleotídeos para PCR e seqüenciamento

A partir de seqüências disponíveis no GenBank, foi feito um alinhamento manual de várias seqüências que determinam o RNA da subunidade menor do ribossomo (SSU rDNA) de diversas espécies e isolados do gênero *Acanthamoeba*, com o auxílio do programa ESEE (Eyeball Sequence Editor, v.3.2), de modo a selecionar regiões do SSU rDNA que fossem conservadas em todos representantes do gênero. Desse modo, foram selecionados dois oligonucleotídeos de 18 bases de comprimento, denominados Ac18.4 e Ac18.5 (seqüências abaixo), que foram utilizados como iniciadores em reações de PCR e

nas reações de seqüenciamento. Esses oligonucleotídeos flanqueiam o gene completo, com tamanho esperado de cerca de 2.300 pb de comprimento (exceto para *A. astronyxis*, cujo tamanho esperado é de 2.800 pb)⁽¹¹⁷⁾.

Além dos oligonucleotídeos utilizados para a amplificação do DNA total de *Acanthamoeba*, foram construídos oito novos pares, utilizados para o seqüenciamento do SSU rDNA (em cada par, um baseado na fita "sense" e o outro na "anti-sense"). Esses oligonucleotídeos foram selecionados com base no alinhamento das seqüências descrito por Stothard e cols., 1998. Foram escolhidas regiões das seqüências que fossem totalmente conservadas entre todos os 53 isolados presentes no alinhamento (portanto, com alta probabilidade de estarem presentes também nos isolados utilizados em nossos estudos). O intervalo escolhido entre cada par de oligonucleotídeos foi de cerca de 300 pares de base, o que nos assegura a sobreposição das seqüências obtidas por cada oligonucleotídeo, possibilitando a confirmação entre as diferentes reações. Portanto, o seqüenciamento do SSU rDNA de nossos isolados de *Acanthamoeba* pode ser completado utilizando-se, no máximo, dezessete oligonucleotídeos (cuja posição na seqüência estão demonstradas na figura 11). Aqueles que resultaram em reação de seqüenciamento bem sucedidas nos primeiros experimentos, sendo portanto utilizados em todas as reações subseqüentes, estão listados abaixo:

- JAP 1U (5' CGG GGA TGT ATT TAT TAG 3')
- JAP 2U (5' AGC CTG AGA AAT GGC TAC 3')
- JAP 2D (5' GTA GCC ATT TCT CAG GCT 3') * rev
- JAP 3U (5' TAA TTC CAG CTC CAA TAG 3')
- JAP 3D (5' TTG GAG CTG GAA TTA CCG 3') * rev
- JAP 4U (5' AAA TTA GAG TGT TCA 3')
- JAP 5U (5' GCC AAG GAT GTT TTC 3')

- JAP 5D (5' GAA AAC ATC CTT GGC 3') * rev
- JAP 6U (5' GGT GGT GCA TGG CCG TTC 3')
- JAP 6D (5' CCA ACT AAG AAC GGC CAT 3') * rev
- JAP 7U (5' TTA GAT GTT CTG GGC CGC 3')
- JAP 7D (5' GCG GCC CAG AAC ATC TAA 3') * rev
- JAP 8U (5' GCT CCT ACC GAT TGA ATG 3')
- JAP 8D (5' CAT TCA ATC GGT AGG AGC 3') * rev
- Ac 18.3 (5' CAG CCT TGC GAC CAT ACT 3') * rev
- Ac 18.4 (5' CCT GCC AGT AGT CAT ATG 3')
- Ac 18.5 (5' AGA CCT TGT TAC GAC TTC 3') * rev

Os oligonucleotídeos marcados com * rev são aqueles que amplificam a fita anti-sense. O oligonucleotídeo Ac 18.3, selecionado da mesma maneira que os oligonucleotídeos Ac 18.4 e 18.5, mas localizado no interior da seqüência (Fig. 11), foi utilizado em experimentos prévios de amplificação por PCR de um fragmento menor do SSU rDNA de *Acanthamoeba*.

```

Ac 18.4 ⇨
TCATATGCTT GTCTCAAAGA TTAAGCCATG CATGTCTAAG TATAAGCTTG TTTATACGGC      60
GAGACTGCGG ATGGCTCATT AAATCAGTTA TAGTTTATTT GATGGTCTCT TTTTCTTTT      120
TTACCTACTT GGATAACCGT GGTAATTCTA GAGCTAATAC ATGCGCAAGG TCCCAAGCGC      180
GAGGTAAGGG GTGTCAAAGC CTCTTATCTG CATGCGCAGCG GGGATGTATT TATTAGGTTA      240
AAAACCAGCG CATTCAAAAC CCCTGGTGAT TCATAATAAC TCTTTCGGAT CGCATTTCATA      300
TCCTCCTTGT GGGGATGGCG ACGATTCATT CAAATTTCTG CCCTATCAAC TTTCGATGGT      360
AGGATAGAGG CCTACCATGG TCGTAACGGG TAACGGAGAA TTAGGGTTCG ATTCCGGAGA      420
GGGAGCCTGA GAAATGGCTA CCACTTCTAA GGAAGGCAGC AGGCGCGCAA ATTACCCAAT      480
CCCACACGGG GGAGGTAGTG ACAATAAATA ACAATACAGG CGCTCGACAA GAGTCTTGTA      540
ATTGGAATGA GTACAATTTA AACCCCTTAA CGAGTAACAA TTGGAGGGCA AGTCTGGTGC      600
CAGCAGCCGC GGTAATTCCTA GCTCCAATAG CGTATATTAA AGTTGTTGCA GTTAAAAAGC      660
TCGTAGTTGG ATTTAGGGAC GCGCATTCAA GCGTCCGCAT CATCGGGTCA AACCGGTGGT      720
GCGTGGCGTT GCGGGCTCGG TCCGTGGAGG ACCAGCGTGT CAACCGGCCC GCCCGTCCCC      780
TCCTTCTGGA TTCCCGTTCC TGCTATTGAG TTAGTGGGGA CGTCACAGGG GATCGCTCAT      840
CGGGGTGGCA AACTCCGGT GGGTGGTCTC TGGGGCCCAG ATCGTTTACC GTGAAAATAAT      900
TAGAGTGTTC AAAGCAGGCA GATTTTACCA TTCTGCCACC GAATACATTA GCATGGGATA      960
ATGGAATAGG ACCCTGTCCT CCTATTTTCC GCGTTGGTTT TTGAGGACCA GGGTAATGAT      1020
TAATAGGGAT AGTTGGGGGC ATTAATATTT AATTGTCAGA GGTGAAATTC TTGGATTAT      1080

```

GAAAGATTAA	CTTCTGCGAA	AGCATCT	JAP 5U ⇨ GCC AAGGATGTTT	TCATTAATCA	AGAACGAAAG	1140
TTAGGGGATC	GAAGACGATC	AGATACCGTC	GTAGTCTTAA	CCATAAACGA	TGCCGACCAG	1200
CGATTAGGAG	ACGTTGAATA	CAAAACACCA	CCATCAGGCA	GTGGGGTCGT	GCTTCGCTTT	1260
TCCGGCAACG	GGGAAGTGGA	GGCGGTCTCA	TTCCCTGAT	GGCCCGGTGA	ATGACTCCCC	1320
TAGCAGCTTG	TGAGAAATCA	TAAGTCTTTG	GGTTCCGGGG	GGAGTATGGT	CGCAAGGCTG	1380
AAACTTAAAG	GAATTGACGG	AAGGGCACCA	CCAGGAGTGG	AGCCTGCGGC	TTAATTTGAC	1440
TCAACACGGG	GAAACTTACC	AGGTCCGGAC	ATAGTAAGGA	TTGACAGATT	GATAGCTCTT	1500
TCTTGATTCT	ATGGGT	JAP 6U ⇨ GGTG GTGCATGGCC	GTTCTTAGTT	GGTGGAGTGA	TTTGTCTGGT	1560
TAATTCGGTT	AACGAACGAG	ACCTTAACCT	GCTAAATATG	CCGCGCTAAC	CCGTCCATCA	1620
AAACCATGC	GTGGCTCACG	AGGTCCGCTG	CGTGGTGGTC	ATCGCTGGCA	ACGGCGGTGG	1680
TCATCACGCC	GGCAGGGCCC	GGGTCCGTGT	GGGCGGTAGG	GTTGCGCGTC	TGTGCTTCTT	1740
AGAGGGACTG	CTGCGCGCCT	AGTCAGCGGA	AGTTTGAGGC	AATAACAGGT	CTGTGATGCC	1800
JAP 7U ⇨ CTTAGATGTT	CTGGGCCGCA	CGCGCGCTAC	ACTGATTAAT	CCAACGAGTC	CGCTTCAATC	1860
GAGGCGCGGT	GCGGTCATCG	GCAACGGTGG	CGTGTCGTCTG	CTGTCCCTCGA	TCGCGCCTGG	1920
GCCGATAGGT	CCGGGTAATC	TTTGCAAATT	TAATCGTGCT	GGGGATAGAT	CATTGTAATT	1980
ATTGATCTTC	AACGAGGAAT	TCCTAGTAAG	CGCGAGTCAT	CAGCTCGCGT	TGATTACGTC	2040
CCTGCCCTTT	GTACACACCG	CCCGTC	JAP 8U ⇨ GCTC CTACCGATTG	AATG	GTCCGG	2100
GGAGCCGTGG	CCTCTACGCA	ATCCGGGCAA	CCGGGTGTG	AGGTCTCTCG	TGTCAGCGAT	2160
GGCACGGGAT	CCTGCGGCGA	AGTCGATTGA	ACCTTACCAT	TTAGAGGAAG	GAGAAGTCGT	2220
AACAAGGTTT	CCGTAG					2236
						⇨ Ac 18.5

Fig. 11 – Posicionamento e orientação dos oligonucleotídeos utilizados para amplificação e seqüenciamento do SSU rDNA de *Acanthamoeba*. A seqüência representada é proveniente de *A. castellanii* (CDC:0981:V006) sob o número de acesso do GenBank U07400.

2.4. Reações de amplificação por PCR e purificação dos produtos

As reações foram realizadas utilizando-se cerca de 100 ng de DNA total de cada um dos isolados, em um volume final de 50 µL. Foram utilizadas 2,5 U de *Taq* DNA polimerase por reação, e seu respectivo tampão foi adicionado na quantidade especificada pelo fabricante. Após experimentos de padronização, foram adicionados 1,5 mM de MgCl₂, 100 pM de cada um dos oligonucleotídeos e 0,25 mM de cada um dos deoxinucleotídeos (dATP; dCTP; dGTP; dTTP). As reações foram feitas em 35 ciclos, sendo que os tempos e temperaturas utilizados foram os seguintes: desnaturação a 94^o C por 30 s, associação a 60^o C por 1 min e extensão a 72^o C por 1,5 min. No primeiro ciclo, onde a

desnaturação completa do DNA total é crucial para o bom aproveitamento da reação nos ciclos subseqüentes, o tempo a 94^o C foi aumentado para 3 min. Já no último ciclo, a temperatura de 72^o C (extensão pela *Taq* DNA polimerase) foi mantida por 10 min, de modo a garantir que todas as moléculas presentes fossem completamente sintetizadas.

Ao final dos ciclos, o conteúdo total de cada reação foi submetido a eletroforese em gel de agarose a 2% contendo 0,4 µg/mL de brometo de etídio. A única banda presente, correspondente ao SSU rDNA, foi cortada do gel e o DNA nela presente foi purificado por eletroeluição seguida de extração por fenol/clorofórmio (1:1) e clorofórmio e precipitação em 0,3M de acetato de sódio e 2 volumes de etanol. Alternativamente, os fragmentos também foram purificados em colunas de acetato de celulose (Spin-X, Costar). A purificação dos produtos das amplificações por PCR se mostrou necessária devido à presença de restos de DNA genômico (o molde inicial da PCR) e de pequenas quantidades de produtos inespecíficos na reação, uma vez que a alta sensibilidade dos géis de poliacrilamida e do equipamento utilizados nos seqüenciamentos tornariam visíveis quaisquer outros DNAs presentes, impossibilitando a correta leitura das seqüências.

2.5. Clivagem por enzimas de restrição dos produtos de PCR

Foram utilizadas várias enzimas de restrição que reconhecem sítios de 4 e de 6 nucleotídeos. A triagem inicial, para determinar quais enzimas seriam utilizadas, foi feita simulando-se digestões do SSU rDNA em computador (programa Clone Manager v.3.11, Scientific & Educational Software), utilizando-se para tanto uma seqüência do SSU rDNA de *A. castellanii* (número de acesso no GenBank M13435). O critério de seleção das endonucleases foi o número de sítios de restrição presentes na simulação e o tamanho das bandas resultantes,

sendo que a enzima deveria apresentar pelo menos cinco sítios para ser considerada útil para os experimentos subseqüentes, além de resultar em bandas de tamanhos suficientemente diferentes para que a separação e visualização nos géis fosse otimizada. As enzimas escolhidas para a investigação da variabilidade no SSU rDNA de *Acanthamoeba* foram as seguintes: *AluI*, *MspI*, *HaeIII*, *HinfI*, *HhaI*, *AvaII* e *Sau3AI*.

As condições utilizadas para as reações de clivagem foram aquelas indicadas pelos fabricantes (New England Biolabs e Pharmacia). O DNA digerido foi então fracionado por eletroforese em gel de poliacrilamida a 8%, corado por nitrato de prata segundo o protocolo previamente descrito⁽¹⁰⁹⁾ e digitalizado para documentação, medida das bandas resultantes e introdução dos dados nos programas de computador adequados (ver mais adiante).

2.6. Cálculo das estimativas de distâncias genéticas entre os isolados

Os comprimentos aproximados dos fragmentos visualizados nos géis foram determinados com o auxílio do programa RFLPscan Plus v.3.0 (Scanalytics CSP Inc., USA). A partir dos dados obtidos, foram construídas matrizes de caracteres discretos, isto é, com algarismos 0 e 1 representando, respectivamente, a ausência ou a presença de determinado fragmento. A partir dessas matrizes, foi feito o cálculo (algoritmo baseado em Jaccard, 1901) de uma matriz de distância genética entre os isolados, utilizando-se o programa RAPDistance v.1.04⁽²⁾. Com o programa Neighbor, do PHYLIP v.3.573c⁽²⁸⁾, foram construídos diagramas representando o agrupamento dos isolados em estudo, pelo método UPGMA ("Unweighted Pair Group Method with Arithmetic Mean"). Esses diagramas são um modo de visualização mais rápido e prático dos agrupamentos do que a análise das matrizes de distância.

2.7. Reações de seqüenciamento do SSU rDNA em termociclador

Excetuando-se *A. poly.#1*, *A. poly#3*, *A. cast.#1* e *A. royreba*, cujas seqüências do SSU rDNA já estão disponíveis no GenBank, todos os outros isolados utilizados neste estudo tiveram essa seqüência determinada.

As reações de seqüenciamento foram feitas utilizando-se 200 ng de SSU rDNA purificado, 10 pmoles de cada oligonucleotídeo e o kit Big-Dye Terminator Mix (marcação do DNA por fluorescência, com uma cor diferente para cada uma das bases), em um termociclador Progene (Techne, UK). Os ciclos utilizados foram: 96°C por 10 segundos; 50°C por 5 segundos; e 60°C por 4 minutos. A taxa de mudança de temperatura entre um ciclo e outro foi de 50°C por minuto. As reações de seqüenciamento foram concentradas por precipitação com isopropanol a 75% (15 minutos à temperatura ambiente), lavadas em etanol a 70% e ressuspensas em 3 µL de tampão desnaturante (formamida + azul de dextran) e 1,5 µL foram aplicados em um seqüenciador automático ABI 377 (sistema gel, PE Applied Biosystems). Algumas das seqüências foram determinadas em um seqüenciador automático ABI Prism 310 (sistema capilar, PE Applied Biosystems). Nesse caso, a ressuspensão foi feita em 20 µL do tampão TSR ("Template Suppression Reagent", PE Applied Biosystems), e foi feita injeção de DNA por 30s. A construção dos oligonucleotídeos para seqüenciamento e os experimentos realizados em ABI 377 foram feitos em colaboração com os Profs. Drs. Luiz R. Nunes e Regina L.C. Oliveira e da técnica Márcia de Oliveira Rocha, do Núcleo Integrado de Biotecnologia da Universidade de Mogi das Cruzes (NIB-UMC).

2.8. Alinhamento das seqüências do SSU rDNA de ***Acanthamoeba***

As seqüências resultantes foram introduzidas no programa de alinhamento manual GeneDoc v. 2.6.01⁽⁹³⁾ para comparação com os doze tipos de seqüências

do SSU rDNA de *Acanthamoeba* já disponíveis no GenBank. O alinhamento original de Stothard e cols., 1998 foi utilizado como guia para o alinhamento das seqüências determinadas neste trabalho, para que fossem levadas em consideração as informações de estrutura secundária da molécula. Para isso, foram preservados os "gaps" presentes no alinhamento original, exceto nos poucos casos de inserção que exigiram inserção de novos "gaps", sendo ajustadas apenas as novas seqüências. Depois de alinhadas, as seqüências foram exportadas para o formato PHYLIP, exigido por todos programas de inferência filogenética, exceto PAUP* v.4.0b6⁽¹²⁰⁾, que utiliza o formato NEXUS, mas que tem a capacidade de importar alinhamentos PHYLIP e fazer as conversões necessárias. O alinhamento resultante, convertido para o formato de diferenças, está apresentado na figura 17.

2.9 Análise filogenética do SSU rDNA de ***Acanthamoeba***

A formatação e a apresentação das árvores foram realizadas com o auxílio dos programas TreeExplorer v. 2.12 (Tamura, K., distribuído pelo autor na página http://evolgen.biol.metro-u.ac.jp/TE/TE_man.html) e TreeView v. 1.6.1⁽⁹⁸⁾. Modificações na apresentação gráfica, quando necessárias e não disponíveis nos programas citados acima, foram feitas utilizando-se os programas Corel Draw v. 9.0 e Adobe Photoshop v. 5.5.

Todas as análises foram realizadas em computadores AMD K6-II de 450 MHz e 56 Mb de memória RAM, sob MS Windows 98, ou em um Intel Pentium Pro de 200 MHz com 64 Mb de memória RAM, sob MS Windows 95.

2.9.1 Análises de distância

As diferentes análises de distância entre as seqüências foram realizadas utilizando-se os programas Point Replacer v.2.0, escrito por J.M.P.A. em

linguagem C, utilizando MS Visual C++ 6.0, para a conversão de alinhamentos entre os formatos PHYLIP e de diferenças (utilizado na figura 17) e para o cálculo das porcentagens de diferença entre as seqüências (disponível na página <http://www.geocities.com/alvesjmp/software.html>). Foi utilizado também o programa PAUP* v. 4.0b6⁽¹²⁰⁾, para o cálculo de distâncias com as correções referentes ao modelo GTR (com heterogeneidade de taxa de substituição e proporção de sítios invariantes) e cálculo de fenogramas por “neighbor-joining” (NJ)⁽¹⁰⁷⁾, com mil réplicas de “bootstrap”. As fórmulas utilizadas para o cálculo das matrizes de distâncias foram aquelas correspondentes ao modelo de evolução disponível que melhor explica o alinhamento utilizado. Tal escolha foi feita por análise de máxima verossimilhança, descrita a seguir.

Tendo sido calculadas as distâncias par a par, procedeu-se à construção de uma árvore por NJ, à qual foram adicionados graficamente os valores de “bootstrap” anteriormente obtidos. Dos métodos algorítmicos, este é o que mostra melhores resultados em estudos de simulação⁽⁵²⁾, sendo muito útil para a definição de árvores que sirvam como ponto de partida para análises mais profundas e melhor fundamentadas. Por não exigir que os dados sejam ultramétricos, ou seja, com todas as linhagens evoluindo na mesma velocidade, o método de NJ é muito menos suscetível a erros sistemáticos do que o UPGMA, por exemplo, conforme discutido na Introdução.

Além da árvore de NJ, outro método de distância empregado foi o de mínimos quadrados, que tenta encontrar a árvore que minimize a soma dos quadrados das diferenças entre as distâncias observadas e aquelas computadas na árvore analisada, como na seguinte fórmula geral⁽¹²¹⁾:

$$\sum_{i,j} \frac{(D_{ij} - d_{ij})^2}{w_{ij}} \quad (3)$$

onde D_{ij} é a distância observada entre os *taxa* i e j , d_{ij} é a distância esperada computada a partir da árvore analisada, α é igual a 1 ou 2 e w_{ij} é um peso que depende da incerteza que se assume para a medida das distâncias. A árvore escolhida, entre todas as analisadas, foi aquela que apresentou o menor valor para essa fórmula, ou seja, a menor discrepância entre a matriz e a árvore.

2.9.2 Análises de parcimônia e métodos de procura da árvore ótima

As análises de parcimônia foram realizadas com a utilização dos programas DNAPARS do PHYLIP e PAUP* v. 4.0b6. Conforme detalhado na Introdução, os métodos de parcimônia são de critério de optimalidade e, portanto, dependem muito dos algoritmos de exploração do espaço de árvores possíveis, que cresce de forma extremamente rápida conforme se analisam mais *taxa*⁽²³⁾.

Desse modo, são necessários meios de se encontrar a melhor árvore sem, entretanto, analisar todas as possibilidades. Alguns dos métodos mais comuns, utilizados neste trabalho, estão detalhados na seção 3.5.4.2 de Resultados e Discussão.

As árvores de máxima parcimônia foram procuradas utilizando diversas combinações desses métodos, no programa PAUP*. Cerca de 150 repetições de adição aleatória dos organismos foram feitas na tentativa de encontrar alguma árvore com o valor mais baixo de número de passos. As melhores árvores de cada tentativa foram utilizadas então como ponto de partida para novas recombinações. Foram empregados os três métodos descritos na seção 3.5.4.2, sobre construção das árvores filogenéticas, do item Resultados e Discussão, utilizando várias ordens e com diversos valores para alguns dos parâmetros envolvidos (opções KEEP e HOLD do comando HSEARCH), de modo a tentar evitar a estagnação da procura em um ponto ótimo local.

Tendo sido obtidas muitas árvores com o número mínimo de passos encontrado, foi feito um consenso por regra de maioria de todas essas árvores (comando CONTREE MAJRULE=YES do PAUP*), de modo a encontrar os pontos mais críticos da árvore (grupos ou ramos apoiados por baixo número de árvores). O objetivo foi tentar analisá-los mais detalhadamente de modo separado, o que facilita análises mais sofisticadas, impossíveis de serem feitas com todos as 60 seqüências.

2.9.3 Análises de máxima verossimilhança

As análises de máxima verossimilhança foram realizadas utilizando-se os programas PAUP* v. 4.0b6 e MODELTEST v. 3.0⁽¹⁰⁰⁾, de modo a escolher, dentre 56 diferentes modelos de evolução de seqüência disponíveis, qual melhor explicaria o alinhamento do SSU rDNA de *Acanthamoeba*. O programa DNAML do PHYLIP fez a procura da árvore de máxima verossimilhança utilizando o método mais simples possível, a adição seqüencial com ordem aleatória de adição (ver seção 3.5.4.2 dos Resultados e Discussão). Foi utilizado também o programa Tree-Puzzle v. 5.0 para a procura de árvores de máxima verossimilhança pelo método de "quartet puzzling"⁽¹¹⁹⁾, quando o número de seqüências analisado torna essa análise possível.

Para fazer essa escolha, utiliza-se como base uma árvore de NJ, calculada a partir de uma matriz de distâncias construída com a utilização da fórmula de Jukes & Cantor, 1969. O programa PAUP* calcula então a verossimilhança dessa árvore para cada um dos modelos de evolução disponíveis. Tais modelos são diferentes variantes do GTR, além do próprio GTR, incorporando fatores como heterogeneidade de taxa de substituição (correção gama) e proporção de posições invariáveis.

Dentre os modelos mais utilizados em filogenia molecular, citados anteriormente na Introdução, o mais simples é o JC, que considera só haver um tipo de substituição, ou seja, qualquer base muda para qualquer outra com a mesma probabilidade, e que todas as bases estão na mesma frequência (25%). Assim, o único parâmetro estimado é a taxa de substituição, e as probabilidades de ocorrer, ou não, mudança de nucleotídeo são dadas por:

$$P_{ij}(t) = \frac{1}{4} + \frac{3}{4}e^{-\mu t} \text{ para } (i = j) \text{ e} \quad (4)$$

$$P_{ij}(t) = \frac{1}{4} - \frac{1}{4}e^{-\mu t} \text{ para } (i \neq j) \quad (5)$$

onde $P_{ij}(t)$ é a probabilidade do nucleotídeo mudar do estado i para o estado j ($i, j = A, C, G, T$) e μt é a taxa de substituição.

Uma generalização do modelo JC é o F81, que também considera somente um tipo de substituição de bases, mas assume que cada nucleotídeo pode ter uma frequência diferente:

$$P_{ij}(t) = \pi_j + (1 - \pi_j)e^{-\mu t} \text{ para } (i = j) \text{ e} \quad (6)$$

$$P_{ij}(t) = \pi_j(1 - e^{-\mu t}) \text{ para } (i \neq j) \quad (7)$$

onde π_j é a frequência da base j ($j = A, C, G, T$). Se π_j for igual a $1/4$, este modelo se torna igual ao modelo JC, ou seja, JC é uma situação especial, uma restrição, de F81, assim como dos próximos modelos apresentados.

Uma outra forma de generalização do modelo JC é o K2P, que considera que todas as bases ocorrem na mesma frequência, mas que existem dois tipos de substituição: transições - ts, mudanças de uma purina (A ou G) para outra ou de uma pirimidina (C ou T) para outra e transversões - tv, mudanças de purina para pirimidina ou vice-versa:

$$P_{ij}(t) = \frac{1}{4} + \frac{1}{4}e^{-\mu t} + \frac{1}{2}e^{-\mu t \left(\frac{\kappa+1}{2}\right)} \text{ para } (i = j), \quad (8)$$

$$P_{ij}(t) = \frac{1}{4} + \frac{1}{4}e^{-\mu t} - \frac{1}{2}e^{-\mu t \left(\frac{\kappa+1}{2}\right)} \text{ para } (i \neq j, ts) \text{ e} \quad (9)$$

$$P_{ij}(t) = \frac{1}{4} - \frac{1}{4}e^{-\mu t} \text{ para } (i \neq j, tv), \quad (10)$$

onde $\kappa = ts/tv$, a razão entre as taxas de transição e transversão. Pode ser facilmente percebido nas fórmulas acima que se κ for igual a 1 esse modelo passa a ser idêntico ao JC.

Já o modelo HKY reúne as características de K2P e F81, ou seja, considera tanto que há dois tipos de substituição de bases (ts e tv) como que cada base pode estar em uma frequência diferente:

$$P_{ij}(t) = \pi_j + \pi_j \left(\frac{1}{\Pi_j} - 1 \right) e^{-\mu t} + \left(\frac{\Pi_j - \pi_j}{\Pi_j} \right) e^{-\mu t (1 + \Pi_j (\kappa - 1))} \text{ para } (i = j), \quad (11)$$

$$P_{ij}(t) = \pi_j + \pi_j \left(\frac{1}{\Pi_j} - 1 \right) e^{-\mu t} - \left(\frac{\pi_j}{\Pi_j} \right) e^{-\mu t (1 + \Pi_j (\kappa - 1))} \text{ para } (i \neq j, ts) \text{ e} \quad (12)$$

$$P_{ij}(t) = \pi_j (1 - e^{-\mu t}) \text{ para } (i \neq j, tv), \quad (13)$$

onde $\Pi_j = \pi_A + \pi_G$ se a base j é uma purina ou $\pi_C + \pi_T$ se j é uma pirimidina. Novamente, a simples substituição das frequências de base (π_j) por $1/4$ e da taxa de transição/transversão (κ) por 1 reduz este modelo ao JC.

Por fim, generalizando ainda mais o modelo de evolução das seqüências, chega-se ao modelo TrN, que além de considerar a ocorrência de diferentes frequências para cada base, supõe que haja três tipos de substituições - transversões, transições entre purinas e transições entre pirimidinas:

$$P_{ij}(t) = \frac{2\pi_A\pi_G}{\pi_R} \left(\pi_R + \pi_Y e^{-\beta t} - e^{-(\pi_R\alpha_R + \pi_Y\beta)t} \right) \text{ para } (i \neq j, ts_R), \quad (14)$$

$$P_{ij}(t) = \frac{2\pi_T\pi_C}{\pi_Y} \left(\pi_Y + \pi_R e^{-\beta t} - e^{-(\pi_Y\alpha_Y + \pi_R\beta)t} \right) \text{ para } (i \neq j, ts_Y) \text{ e} \quad (15)$$

$$P_{ij}(t) = 2\pi_R\pi_Y (1 - e^{-\beta t}) \text{ para } (i \neq j, tv) \quad (16)$$

sendo π_R a proporção de purinas, π_Y a proporção de pirimidinas, α_R a taxa de transições entre purinas, α_Y a taxa de transições entre pirimidinas e β a taxa de transversões.

Conforme anteriormente observado, todos esses modelos são casos especiais do modelo GTR, que considera tanto frequências de bases diferentes quanto todos os seis tipos de substituição possíveis em uma matriz de substituição simétrica⁽¹⁴⁰⁾. Outros modelos utilizados neste trabalho, também simplificações do GTR, estão enumerados com suas principais características na tabela 1.

Tabela 1 – Modelos analisados pelo programa MODELTEST e suas principais características.

Modelo	Frequências de bases	p.l.	Tipos de substituição
JC	1	0	1
F81	4	3	1
K2P	1	1	2 (ts e tv)
HKY	4	4	2 (ts e tv)
TrNef	1	2	3 (ts A↔G, ts C↔T e tv)
TrN	4	5	3 (ts A↔G, ts C↔T e tv)
K3P	1	2	3 (ts, tv A↔C, G↔T e tv A↔T, C↔G)
K3Puf	4	5	3 (ts, tv A↔C, G↔T e tv A↔T, C↔G)
TIMEf	1	3	4 (2 ts, tv A↔C, G↔T e tv A↔T, C↔G)
TIM	4	6	4 (2 ts, tv A↔C, G↔T e tv A↔T, C↔G)
TVMef	1	4	5 (ts, 4 tv)
TVM	4	7	5 (ts, 4 tv)
SYM	1	5	6 (2 ts, 4 tv)
GTR	4	8	6 (2 ts, 4 tv)

ts: transições; tv: transversões; p.l.: parâmetros livres.

Os modelos básicos apresentados acima que não foram descritos anteriormente estão detalhados em Posada & Crandall, 1998, na documentação que acompanha o programa MODELTEST e nas referências neles contidas. Cada

um dos modelos enumerados anteriormente foi utilizado levando em conta também correção gama, proporção de invariantes e ambos (estimadas dos dados), resultando portanto em quatro variantes de cada um dos 14 modelos (total de 56 modelos). O número de parâmetros livres nada mais é do que o número de parâmetros sendo estimado para cada modelo. Quando se utiliza correção gama ou proporção de invariante, deve-se adicionar mais um parâmetro livre para cada um desses fatores adicionais.

O resultado do cálculo de verossimilhança para cada modelo foi colocado no programa MODELTEST, que fez o teste da razão das verossimilhanças (LRT), de modo a escolher o modelo que resulta no maior valor de $\ln L$, sem entretanto utilizar parâmetros excessivos. Outro teste feito por esse programa é o "Akaike Information Content" ($AIC = -2\ln L + 2n$, onde n é o número de parâmetros livres de cada modelo sendo testado, e o escolhido é aquele com o menor valor de AIC), que difere do LRT por penalizar mais fortemente os modelos com maior número de parâmetros, e é muito menos comum na literatura filogenética⁽⁷⁵⁾.

Após a escolha do modelo que melhor descreveu a evolução das seqüências utilizadas, procedeu-se às análises de distância (comando DSET distance=gtr rates=gamma shape=x pinv=y do PAUP*, onde x e y são os valores estimados pelo MODELTEST para α e proporção de sítios invariantes) e "bootstrap" (comando bootstrap search=nj nreps=1000) utilizando este modelo. Infelizmente estes valores de "bootstrap" não são a estimativa mais exata possível para este método, uma vez que são obtidos fixando-se os valores do parâmetro α da distribuição gama e a proporção de sítios invariantes, e não utilizando distância de ML e fazendo a estimativa destes valores e das matrizes de substituição para cada uma das subréplicas analisadas. Essa estimativa seria muito demorada,

impossibilitando as análises, uma vez que o cálculo para apenas um alinhamento não pode ser completado.

A procura da árvore de máxima verossimilhança não pode ser feita utilizando os métodos de ML devido ao número muito grande de árvores possíveis, como detalhado na seção 3.5.4.2 dos Resultados. Considerando que a máxima verossimilhança é o método mais complexo e cujos cálculos tomam mais tempo para cada árvore, a velocidade de processamento teria de ser extremamente alta mesmo para poucas seqüências e utilizando algoritmos heurísticos de busca. Além disso, independentemente do método empregado, a quantidade de bits de memória necessários para se armazenar todas possibilidades se tornaria rapidamente maior do que o próprio número de partículas do universo observável (cerca de 10^{80}), conforme fossem adicionados mais *taxa*.

Resultados e Discussão

3. RESULTADOS E DISCUSSÃO

3.1 – Amplificação aleatória de DNA polimórfico (RAPD)

O método de RAPD, assim como o de RFLP ("Restriction Fragment Length Polymorphism") do DNA total, é um método multi-locus muito útil na caracterização molecular de organismos^(135, 136). Esse método se baseia na amplificação de seqüências polimórficas de DNA utilizando uma variante de PCR onde somente um oligonucleotídeo pequeno (cerca de 10 pb) é utilizado com baixa temperatura de associação (37°C). Os resultados da aplicação dessa técnica são padrões de bandas, semelhantes àqueles obtidos por RFLP. Assim como o método de RFLP, o RAPD indiretamente indica diferenças em seqüências, uma vez que mudanças nessas seqüências podem criar novos pontos de associação para o oligonucleotídeo, assim como alterar outros anteriormente existentes, conseqüentemente mudando o padrão de bandas apresentado.

Para a caracterização dos isolados de *Acanthamoeba* por RAPD foram inicialmente investigados os padrões resultantes da utilização de diversos oligonucleotídeos, a fim de selecionar aqueles que resultassem em padrões de bandas para todos os isolados estudados. Os três oligonucleotídeos escolhidos e outros detalhes dos experimentos estão descritos nos itens Material e Métodos e Anexo 1.

Apesar da utilidade desses padrões como indicadores da variabilidade genética, ainda há limitações inerentes à análise de dados de RAPD que precisam ser levadas em consideração. A principal limitação diz respeito à homologia entre as bandas comparadas: sem a certeza de que as bandas são homólogas não é possível fazer análises filogenéticas. Como não se tem indicação da natureza dessas seqüências, que podem ser tanto regiões de grande importância biológica (reguladoras, estruturais, codificadoras de proteínas) quanto regiões cuja seqüência não tem função, é difícil estabelecer a significância dos diferentes

níveis de similaridade entre isolados. Esse problema não é tão grande quando o objetivo é simplesmente fazer um esboço da variabilidade genética, sem nenhuma pretensão filogenética. Ainda assim, neste trabalho, foram utilizados outros métodos no intuito de corroborar as conclusões resultantes dos experimentos de RAPD: RFLP do DNA total (descrito na próxima seção e Anexo 2) e RFLP do SSU rDNA.

Apesar das limitações citadas, o RAPD pode ser muito útil na caracterização dos isolados, permitindo o cálculo de distâncias genéticas aproximadas. Para tanto, utilizam-se os mesmos métodos anteriormente descritos para a análise de RFLP de DNA total e SSU rDNA (matrizes de presença/ausência de banda, matrizes de distância, cálculo de fenogramas e consenso). Assim como visto com os dois métodos de análise de fragmentos de restrição, o método de RAPD evidenciou a alta variabilidade genética do gênero *Acanthamoeba*. Apesar disso, permitiu o agrupamento de alguns isolados mais proximamente relacionados, de acordo com a similaridade dos padrões de amplificação. Os resultados de RAPD, obtidos para os três oligonucleotídeos utilizados, confirmaram os agrupamentos anteriormente definidos por RFLP de DNA total e de SSU rDNA, conforme pode ser visto no fenograma apresentado no Anexo 1 ou, mais claramente, na comparação dos fenogramas calculados pelos diferentes métodos, apresentada no Anexo 2.

3.2 – RFLP de DNA genômico total

A análise do polimorfismo no tamanho de fragmentos de restrição provenientes da clivagem de DNA genômico total (RFLP do DNA total) pode ser utilizada como ferramenta de investigação preliminar da organização genômica, geralmente empregando sondas radioativas para a identificação de seqüências. Para isso, é necessário que o DNA digerido seja transferido do gel para um

suporte (membrana de náilon) e, em seguida, hibridado a uma ou mais sondas. Os fragmentos resultantes da digestão aparecerão, então, de forma indireta, apenas na autoradiografia. Em alguns organismos, como *Acanthamoeba*, *Naegleria* e alguns outros^(86, 21), podem ser visualizados padrões de bandas complexos, porém bastante nítidos, já no gel corado com brometo de etídio (figura 1 do Anexo 2 e Fig. 12), dispensando, portanto, o processo de transferência do DNA e hibridação com sondas. A presença dessas bandas indica a existência de um grande número de repetições de algumas poucas seqüências, o que pode ser devido tanto à baixa complexidade do genoma quanto a um grande número de cópias de DNA mitocondrial (ou uma combinação de ambas as possibilidades).

Fig. 12 – Exemplos de padrões de restrição do DNA genômico total de alguns isolados de *Acanthamoeba*. **A** – enzima *HhaI*; **B** – enzima *EcoRI*; **C** – enzima *HincII*. Marcador: λ /HindIII.

Assim, o DNA genômico total de diversos isolados foi submetido à clivagem por diversas enzimas de restrição, a fim de selecionar aquelas que resultassem

Fig. 13 – Exemplo de matriz de presença/ ausência de bandas no DNA genômico total clivado de cada isolado. Esta matriz corresponde aos dados obtidos com a enzima *EcoRI*.

Apesar de compartilhar algumas das mesmas limitações do RAPD e outros métodos de análise de padrões de fragmentos (detalhadas na seção anterior), em especial quanto à homologia entre bandas, o RFLP também pode ser muito útil na caracterização dos isolados. A comparação dos padrões dos diferentes isolados permitiu a estimativa preliminar da distância genética entre eles, além de evidenciar o alto nível de variabilidade presente no gênero *Acanthamoeba*, conforme esperado pelos dados da literatura. Puderam, também, ser evidenciados alguns agrupamentos de organismos mais geneticamente relacionados, especialmente entre os isolados brasileiros de úlcera de córnea. Um resultado interessante foi a completa (ou quase completa) similaridade entre isolados obtidos da ATCC classificados, por critérios morfológicos, como espécies de grupos morfológicos diferentes (*A. castellanii*, *A. polyphaga*, pertencentes ao grupo II, e *A. royreba*, pertencente ao grupo III), enquanto quatro isolados, classificados como uma mesma espécie (*A. polyphaga*), apresentaram sempre padrões bastante distintos. Outros isolados consistentemente tiveram seus padrões de restrição bastante diferentes daqueles apresentados pelos demais isolados, como foi o caso de *A. astronyxis*, *A. poly.* #4, U/E2 e U/HC1. Assim, esses isolados não puderam ser agrupados, com segurança, com nenhum dos outros isolados analisados.

3.3 RFLP do SSU rDNA

Devido ao forte significado biológico do RNA da subunidade menor do ribossomo, essa sequência tem alto nível de conservação entre os diversos organismos, além de várias outras características vantajosas, conforme anteriormente detalhado na Introdução. Desse modo, a clivagem do gene para o SSU rRNA por enzimas de restrição gerará bandas que podem ser consideradas homólogas com uma margem bastante grande de segurança, permitindo

conclusões muito mais robustas acerca do relacionamento genético entre os isolados.

Com essa finalidade, foram construídos dois oligonucleotídeos (ver Material e Métodos) com seqüências complementares às regiões que flanqueiam o SSU rDNA de *Acanthamoeba*. Para isso, nos baseamos no alinhamento de cerca de 20 seqüências desse gene, provenientes de várias espécies do gênero, disponíveis no GenBank. Essa região foi amplificada por PCR e os produtos de amplificação foram então purificados (por eletroeluição ou eluição em coluna de acetato de celulose) e digeridos por diversas nucleases de restrição. Os fragmentos resultantes foram fracionados em géis de poliacrilamida e corados com nitrato de prata. As análises de similaridade foram realizadas conforme descrito para os RFLPs de DNA total, e estão detalhadas no item Material e Métodos e no Anexo 2.

Os resultados das análises por RFLP do SSU rDNA confirmaram aqueles obtidos por RAPD e RFLP do DNA total, em especial no que se refere aos agrupamentos encontrados. Embora os padrões tenham se mostrado muito menos polimórficos do que aqueles obtidos a partir do DNA total, o que era esperado dada a conservação do SSU rDNA, ainda assim podemos perceber a grande variabilidade existente no gênero *Acanthamoeba*, conforme pode ver visto nas figuras 14 e 15, além da figura 2 do Anexo 2. Dados da literatura mostram que a divergência de seqüência do SSU rDNA em *Acanthamoeba* pode chegar a mais de 37%⁽¹¹⁷⁾, um nível semelhante àquele encontrado na comparação entre vertebrados e invertebrados.

Novamente, foram obtidos padrões bastante diferentes para os quatro isolados de *A. polyphaga*, enquanto os isolados de três espécies diferentes (*A. polyphaga*, *A. castellanii* e *A. royreba*) apresentaram padrões geralmente idênticos. Além disso, os grupos de isolados de úlcera de córnea, anteriormente obtidos pelas

outras análises, se repetiram com o RFLP do SSU rDNA. Os isolados com os padrões mais divergentes observados nas análises anteriores continuaram sendo também os mais distintos nas análises por RFLP do SSU rDNA.

A concordância entre os resultados obtidos por RAPD, RFLP de DNA total e RFLP do SSU rDNA nos permite concluir que as informações resultantes do RFLP de DNA total não são aleatórias, tendo realmente uma base genética que permite o agrupamento e caracterização prévios dos isolados de *Acanthamoeba*, assim como aquelas derivadas do RAPD.

Fig. 14 - Eletroforese em gel de poliacrilamida a 6% dos produtos de clivagem do SSU rDNA com a enzima *Ava*I.

Fig. 15 - Eletroforese em gel de poliacrilamida a 6% dos produtos de clivagem do SSU rDNA com a enzima *Hha*I.

3.4 – Determinação da seqüência do SSU rDNA

A análise de seqüências de macromoléculas biológicas tem se revelado o método mais refinado de inferência filogenética, devido às várias características intrínsecas a esse tipo de abordagem. Entre elas, podemos citar a abundância de características disponíveis, a precisão na determinação dessas características, a facilidade de comparação entre estudos diferentes e a possibilidade da aplicação de modelos explícitos de evolução, o que possibilita testes de hipóteses muito mais significativos quanto aos processos de evolução. Entre as diversas seqüências utilizadas, aquelas que codificam os RNAs que participam da estrutura do ribossomo (rDNA) são, até o momento, as de maior potencial para inferências filogenéticas. Estão presentes em todos os seres vivos, apresentam alta conservação na seqüência (devido à forte seleção que é exercida sobre essa molécula) mas, ainda assim, contêm regiões com diferentes taxas de substituição, permitindo tanto comparações entre organismos distantes quanto entre outros mais proximamente relacionados filogeneticamente. Dentre os diferentes componentes do ribossomo, o RNA da subunidade menor (SSU RNA) é o mais extensivamente estudado e utilizado para fins filogenéticos, uma vez que, além de apresentar todas as características acima, possui também um tamanho estatisticamente favorável às análises sem, no entanto, ser tão grande a ponto de dificultar a obtenção da seqüência por PCR.

Desse modo, foram amplificadas e purificadas as seqüências do SSU rDNA de todos os isolados estudados, utilizando os mesmos oligonucleotídeos e métodos anteriormente descritos para a determinação do RFLP do SSU rDNA. Diversos novos oligonucleotídeos foram utilizados para o seqüenciamento automático da molécula inteira, resultando, conforme esperado, seqüências de 2.204 (U/Oft1) a 2.538 (*A. astronyxis*) pb.

Das 17 seqüências determinadas, 13 (aquelas sem nenhuma ambigüidade e com mais de três leituras de boa qualidade em cada trecho) foram depositadas no GenBank, estando acessíveis pelos seguintes números de acesso:

<i>A. astronyxis</i> (ATCC 30901)	AY026242	<i>A. poly.</i> #2 (ATCC 30461)	AY026243
<i>A. poly.</i> #4 (ATCC 30872)	AY026244	<i>A. cast.</i> #2 (ATCC 30868)	ND
U/E 1	NT	U/HC 1	AY026245
U/E 3	AY026747	U/E 2	ND
U/E 5	AY026748	U/E 4	ND
U/E 7-1	AY026247	U/E 6	AY026246
U/E 8R	AY026251	U/E 7-2	AY026249
U/E 8X	AY026250	U/E 8L	ND
U/Oft1	AY026248	U/E 10	AY026749

NT: não determinada; ND: ainda não depositada no GenBank

As seqüências de U/E7-1 e U/E7-2 se mostraram idênticas, assim como aquelas de U/E8R e U/E8X. O isolado U/E8L, cuja seqüência esperávamos ser praticamente idêntica às dos outros isolados U/E8, apresentou 1,47% de diferença (o correspondente a 33 diferenças de nucleotídeos) em relação a esses. Entretanto, o refinamento de algumas regiões de baixa qualidade, apresentando diversas ambigüidades, consideradas como diferenças pelo programa Point Replacer, e deleções, poderia fazer com que esse valor de diferença diminua bastante, mas talvez não chegue a zero, uma vez que alguns dos RFLPs do SSU rDNA apresentaram pequenas diferenças no padrão de bandas entre U/E8R e U/E8L. Somente o término da seqüência de U/E8L poderia esclarecer o significado dessas diferenças.

Outros dois isolados que apresentaram seqüências do SSU rDNA idênticas foram *A. poly.*#1, anteriormente disponível no GenBank e presente em nosso

alinhamento identificada como A21, e *A. cast.*#2. Esse resultado já era esperado pelas conclusões prévias dos métodos de análise de fragmentos de DNA anteriormente utilizados, que sempre mostraram padrões de bandas idênticos, ou muito semelhantes, para esses dois isolados. Além desses, *A. royreba*, cuja sequência do SSU rDNA também está disponível no GenBank, apresentou sequência idêntica àquela identificada como A23 (*A. rhysodes* Singh ATCC 30973).

Apenas U/E1 ainda não teve a sequência completada, mas já foram obtidos três fragmentos, de 1.300, 700 e 300 pb de comprimento. Embora a soma desses fragmentos seja do tamanho esperado para a molécula completa, a qualidade das sequências das extremidades do gene ainda não é suficiente para juntá-los, provavelmente devido a problemas na pureza do DNA utilizado nas últimas reações (que tiveram os cromatogramas curtos e de baixa qualidade). Espera-se que a sequência completa seja muito semelhante, se não idêntica, àquela apresentada por U/E3, uma vez que esses dois isolados sempre mostraram padrões de RAPD, RFLP de DNA total e de rDNA idênticos. A comparação entre os fragmentos de U/E1 e a sequência completa de U/E3 (eliminando tanto ambigüidades quanto deleções maiores) resultou em 4,16% de diferença entre as sequências, mas somente o sequenciamento completo e com boa qualidade do SSU rDNA de U/E1 poderá confirmar se esses isolados são realmente idênticos, conforme sugerido pelos resultados das abordagens anteriormente empregadas.

3.5 – Análises filogenéticas

3.5.1 – Alinhamento e determinação do tipo de sequência do SSU rDNA

O alinhamento manual das sequências completas determinadas neste trabalho foi feito levando-se em consideração, de maneira indireta, as informações de

estrutura secundária, utilizando-se como guia o alinhamento de Stothard e cols., 1998. O alinhamento completo está apresentado na figura 17.

A análise preliminar do alinhamento revelou que 1.804 posições (63,85%) são constantes, ou seja, apresentam o mesmo nucleotídeo em todas as seqüências alinhadas. Todas essas posições são descartadas em análises de parcimônia, uma vez que, assim como vários outros tipos de padrão de diferença entre seqüências, são considerados não informativos para este método^(32, 121). Isso leva a uma perda de grande quantidade dos dados analisados, uma vez que o número de sítios informativos (definidos como aqueles cuja distribuição das diferenças ocorre em pares de taxa e possibilita a escolha de uma topologia entre outras concorrentes) tende a ser pequeno em relação ao total seqüenciado. Em nosso caso, o programa PAUP* indicou que foram obtidos 769 sítios informativos em 2.819 posições de alinhamento, o que dá uma proporção de 27,3%, tendo sido eliminadas da análise, portanto, 246 posições variáveis não informativas para a parcimônia e as 1.804 posições constantes, anteriormente mencionadas. Se forem eliminadas as posições em que alguma das seqüências analisadas contenha um indel, o que é aconselhável fazer ao se considerar que tais regiões costumam apresentar uma maior incerteza na determinação do alinhamento, além de levar a superestimativas das divergências, teremos somente 282 sítios informativos, em um total de 1.818, resultando, portanto, em apenas 15,5% de utilização da seqüência determinada. Desse modo, conforme sugerido por diversos estudos de simulação^(50,52), a eficiência do método de parcimônia tende a ser menor do que no caso de métodos como os de distância ou os de máxima verossimilhança, uma vez que é necessária a coleta de uma maior quantidade de dados para se obter a mesma probabilidade de inferir a árvore filogenética correta.

Os métodos de máxima verossimilhança e de distâncias, por outro lado, utilizam muito mais os dados disponíveis. Em nosso caso, a máxima verossimilhança utilizou todos os 1.818 sítios que não possuem indel em nenhuma das seqüências, sendo que há 948 diferentes padrões presentes (dos 4^{60} possíveis para 60 seqüências). Ou seja, cerca de 64,5% do alinhamento foi considerado por esse método de análise. Os métodos de distância, por outro lado, utilizaram o alinhamento inteiro, mas não é possível mencionar uma única porcentagem de utilização, uma vez que os métodos de distância fazem o cálculo par a par, desconsiderando as posições onde há indel nas duas seqüências do par. Desse modo, cada distância vai ter o seu valor de posições comparadas em relação ao total do alinhamento global mas, dentro de cada comparação de um par, sempre são utilizados 100% dos dados. Por exemplo, entre U/E2 e U/E3 foram comparados 2.269 sítios, significando que 550 posições do alinhamento geral continham indel tanto em U/E2 quanto em U/E3, mas todos esses 2.269 sítios entraram no cálculo das distâncias.

O resultado do cálculo das distâncias não corrigidas, calculadas pelo programa Point Replacer, foi uma matriz com os valores entre cada um dos isolados e todos os outros, e está apresentado na figura 16, com destaque para os isolados representando novos tipos de seqüência do SSU rDNA. Conforme brevemente mencionado na Introdução, cada tipo de seqüência está separado dos outros por, pelo menos, 5% de divergência não corrigida (porcentagem de diferença entre as seqüências), sugerindo que cada tipo possa representar uma espécie no gênero *Acanthamoeba*⁽¹¹⁷⁾.

A matriz apresenta valores de divergência não corrigida variando de 0,09%, entre A39 e A44 (isolados tipo T5), a 37,98%, entre alguns isolados tipo T5 e o isolado A49 (tipo T9), mostrando mais uma vez a alta variabilidade genética presente nos organismos do gênero *Acanthamoeba* como concebido atualmente.

Conforme se pode perceber pela análise da matriz obtida, os SSU rDNAs seqüenciados puderam ser classificados em quatro tipos, um deles previamente conhecido (T4) e outros três novos tipos (provisoriamente chamados T13 a T15). Dos três novos tipos de seqüência encontrados, um deles provém de um isolado de córnea brasileiro, U/HC1 (tipo T13). Os outros dois são de isolados provenientes do ambiente, obtidos da ATCC: *A. poly.*#4 (tipo T14) e *A. astronyxis* (tipo T15). Os isolados classificados como T4 foram os outros 13 isolados de córnea brasileiros, além dos isolados ATCC *A. poly.*#2 e *A. cast.*#2, ambos também provenientes de casos de úlcera de córnea. O isolado U/E1, embora ainda não completamente seqüenciado, certamente deverá ser classificado também como T4, uma vez que a comparação entre as seqüências parciais e a molécula completa de U/E3 mostrou alta similaridade entre os dois isolados. Eliminando-se as ambigüidades e imprecisões na seqüência de U/E1, a diferença entre as seqüências desses isolados deverá diminuir, conforme sugerido pelos experimentos prévios de RAPD e RFLP.

Desse modo, fortalece-se a associação entre o tipo T4 de SSU rDNA e a ocorrência de úlcera de córnea, uma vez que 16, entre 17 novas seqüências provenientes de casos clínicos, apresentaram esse tipo. Entretanto, é conveniente salientar que são necessários estudos mais abrangentes de isolados provenientes do ambiente, a fim verificar se essa associação não seria decorrente de uma maior prevalência ecológica desse tipo de *Acanthamoeba*. Stothard e cols., 1998 descreveram 53 seqüências completas e sete incompletas (não apresentadas) do SSU rDNA de isolados de *Acanthamoeba* de diversas partes do mundo, embora nenhum deles proveniente do hemisfério sul. Dessas, 32 pertencem ao tipo T4, sendo 24 provenientes de isolados de úlcera de córnea, oito de ambiente e 28 distribuídos por outros tipos. Nos outros 11 tipos de seqüência descritos, é encontrada a seguinte distribuição: tipos T1, T6, T7, T8,

T9, T10 e T12 - um isolado; T11 - dois isolados; T2 - três isolados; T3 - quatro isolados e T5 - 12 isolados. O que se pode perceber nesses números é que, embora a diversidade de tipos no ambiente ainda pareça ser maior do que aquela encontrada entre os isolados de córnea, ainda assim foram observados oito isolados T4 no ambiente (provenientes da Inglaterra e do México, EUA e Japão), o que só é menor do que o observado para o tipo T5, *A. lenticulata*, um tipo muito coeso geneticamente e bem diferenciado filogeneticamente em relação aos outros (referência 117 e Fig. 20). Entretanto, é conveniente notar que quatro destes isolados vieram de um estudo em uma mesma localidade (isolados de mucosa nasal na Alemanha⁽¹¹⁷⁾), o que poderia explicar uma parte dessa coesão como um efeito de amostragem.

É interessante observar a presença de um intron do grupo I ("self-splicing intron") no SSU rDNA de oito entre doze isolados de *A. lenticulata* sendo que esses introns pertencem a três tipos bem diferentes^(36, 117). Outra espécie do gênero *Acanthamoeba*, *A. griffini*, também pode apresentar esse tipo de intron no SSU rRNA, mas é bem distinto daquele encontrado em *A. lenticulata*. Tais observações mostram o potencial desses organismos no estudo de mecanismos de evolução molecular. Um exemplo é a presença de seqüências móveis horizontalmente, muito importantes por levar a problemas na inferência de homologia entre moléculas que, isoladas de diferentes organismos, podem não representar exatamente a história evolutiva justamente por terem sido transferidas lateralmente e não por relação ancestral-descendente.

1	
2	12.78 2
3	12.69 1.52 3
4	13.72 5.15 4.89 4
5	11.40 11.17 10.97 11.61 5
6	11.35 11.15 11.08 11.72 0.76 6
7	11.53 11.13 11.06 11.65 0.63 0.85 7
8	11.91 11.54 11.25 12.02 1.91 2.04 1.91 8
9	10.99 11.35 11.23 11.87 6.56 6.60 6.83 7.30 9
10	10.84 11.45 11.20 12.02 6.02 5.75 6.11 6.54 1.92 10
11	11.17 11.82 11.53 12.38 6.63 6.55 6.90 7.33 2.61 2.80 11
12	11.18 11.79 11.50 12.42 6.39 6.30 6.61 7.12 3.50 3.20 3.10 12
13	10.83 11.53 11.16 12.06 6.48 6.44 6.75 7.17 1.07 1.70 2.35 2.93 13
14	11.28 11.68 11.57 12.20 6.43 6.21 6.52 6.91 2.53 1.65 3.32 3.58 2.01 14
15	11.05 11.64 11.35 12.08 6.57 6.52 6.79 7.21 2.09 2.27 2.70 2.79 1.12 2.27
16	11.65 11.64 11.31 12.41 6.46 6.20 6.68 7.20 3.81 3.78 3.88 3.63 3.81 4.07
17	10.77 11.32 11.08 11.94 6.19 5.93 6.28 6.67 1.83 0.18 2.97 3.25 1.65 1.74
18	11.86 11.72 11.39 12.41 6.41 6.16 6.64 7.16 3.72 3.82 3.79 3.81 3.72 3.89
19	10.69 11.46 11.17 11.99 6.02 5.76 6.11 6.50 1.83 0.18 2.89 3.20 1.61 1.79
20	10.92 11.57 11.20 12.10 6.53 6.39 6.70 7.13 1.21 1.74 2.48 3.06 0.13 2.05
21	10.95 11.22 10.98 11.61 6.40 6.52 6.66 7.26 2.93 3.52 3.76 3.19 2.89 3.68
A.cast.#2	10.95 11.22 10.98 11.61 6.40 6.52 6.66 7.26 2.93 3.52 3.76 3.19 2.89 3.68
22	11.07 11.84 11.77 12.49 6.73 6.52 6.83 7.12 2.71 1.43 3.54 3.72 2.40 1.07
23	11.24 11.31 11.02 11.61 6.53 6.65 6.79 7.39 2.93 3.51 3.45 3.33 2.80 3.37
24	10.70 11.53 11.20 11.89 6.43 6.43 6.70 7.13 1.47 2.28 2.92 3.10 1.03 2.09
25	10.60 11.48 11.27 11.95 6.60 6.51 6.78 7.12 0.80 1.92 2.88 3.10 1.43 2.62
26	11.31 11.70 11.41 12.42 6.34 6.03 6.51 6.99 3.32 2.85 2.87 0.80 2.75 3.23
27	10.88 11.58 11.20 12.10 6.66 6.53 6.79 7.22 1.25 1.70 2.53 2.84 0.18 2.01
A.poly.#2	11.06 11.72 11.43 12.25 6.95 6.91 7.26 7.64 2.58 2.89 3.32 3.54 1.61 2.89
U/E2	12.05 12.79 12.55 13.15 7.94 7.82 8.21 8.54 3.60 3.91 4.46 5.25 3.56 4.21
U/E3	11.35 11.54 11.29 12.14 6.90 6.81 7.11 7.32 3.41 3.37 3.35 1.55 3.06 3.80
U/E4	12.06 12.16 11.95 12.60 7.64 7.43 7.78 8.12 4.22 3.56 4.86 4.64 3.83 2.89
U/E5	11.67 11.85 11.69 12.34 7.34 7.12 7.47 7.77 3.73 3.03 4.55 4.24 3.38 2.45
U/E6	11.67 11.81 11.61 12.33 7.25 7.04 7.38 7.68 3.51 3.07 4.42 4.07 3.07 2.13
U/E7-1	11.54 12.01 11.90 12.50 7.46 7.25 7.60 7.89 3.55 2.89 4.37 4.46 3.33 2.44
U/E7-2	11.54 12.01 11.90 12.50 7.46 7.25 7.60 7.89 3.55 2.89 4.37 4.46 3.33 2.44
U/E8L	12.04 13.12 12.91 13.58 8.25 8.17 8.52 8.90 2.85 4.09 4.68 4.82 3.43 4.61
U/E8R	10.82 11.94 11.73 12.45 7.02 6.94 7.29 7.67 1.65 2.76 3.49 3.62 2.18 3.28
U/E8X	10.82 11.94 11.73 12.45 7.02 6.94 7.29 7.67 1.65 2.76 3.49 3.62 2.18 3.28
U/E10	11.29 12.02 11.86 12.50 7.29 7.33 7.55 8.10 2.62 3.42 3.92 4.50 2.98 3.63
U/Oft1	10.99 12.16 11.88 12.65 7.14 6.92 7.27 7.56 2.89 2.06 3.72 3.95 2.59 2.93
29	10.77 11.56 11.40 12.04 6.77 6.68 7.00 7.38 0.71 2.01 3.05 3.36 1.60 2.62
30	10.85 11.51 11.35 12.20 6.73 6.64 6.95 7.37 1.16 2.23 3.14 3.18 1.65 2.89
32	11.03 11.72 11.61 12.37 6.61 6.39 6.70 7.00 2.66 1.43 3.54 3.72 2.32 0.89
33	11.73 11.77 11.43 12.37 6.41 6.16 6.64 7.16 3.54 3.64 3.62 3.68 3.63 3.84
U/HC1	13.26 14.46 14.26 15.37 12.37 12.17 12.54 12.21 11.16 11.37 11.56 11.30 11.05 11.17
34	19.82 17.82 17.64 18.47 17.96 18.01 18.05 18.72 17.47 17.47 17.94 17.93 17.22 17.35
36	19.75 17.80 17.66 18.49 17.86 17.91 17.95 18.62 17.37 17.36 17.87 17.91 17.19 17.24
38	19.69 17.91 17.77 18.43 17.97 18.02 18.06 18.73 17.48 17.48 17.98 17.94 17.31 17.35
39	19.58 17.80 17.61 18.40 17.89 17.95 17.98 18.65 17.40 17.40 17.87 17.86 17.23 17.28
42	19.77 17.78 17.60 18.43 17.92 17.97 18.01 18.68 17.43 17.43 17.90 17.89 17.17 17.31
44	19.54 17.75 17.61 18.40 17.82 17.87 17.91 18.58 17.33 17.32 17.83 17.79 17.15 17.20
45	19.71 17.92 17.78 18.44 17.98 18.04 18.07 18.74 17.50 17.49 17.99 17.96 17.32 17.37
A.poly.#4	16.83 12.15 12.08 11.30 14.85 14.96 14.82 15.48 14.97 15.08 15.51 15.36 15.19 15.00
46	14.24 5.76 5.59 5.90 12.67 12.70 12.64 13.21 13.18 13.20 13.51 13.65 13.15 13.29
A.astronyxis	36.53 36.10 36.09 35.54 35.62 35.44 35.70 35.93 35.14 35.28 35.00 35.36 35.18 35.18
47	36.03 35.81 35.56 35.73 35.22 35.22 35.22 35.38 34.51 34.53 34.71 34.90 34.49 34.24
48	35.34 35.31 35.23 35.07 34.98 34.90 34.92 35.18 34.17 34.35 34.43 34.75 34.24 34.05
49	37.49 37.02 37.07 36.62 36.49 36.37 36.61 36.83 36.31 36.41 36.17 36.59 36.39 36.38
50	17.70 17.63 17.61 18.02 17.57 17.61 17.66 17.74 18.63 17.90 18.51 18.60 18.38 18.09
51	11.93 11.70 11.46 12.17 5.62 5.61 5.84 6.40 5.85 4.99 5.84 5.59 5.69 5.68
52	11.95 11.94 11.78 12.07 5.74 5.60 5.96 6.69 6.19 5.25 6.13 5.58 5.86 5.67
53	18.40 18.26 18.31 18.80 19.26 19.13 19.22 19.40 20.11 19.87 20.13 20.27 20.13 20.07

	47						
48	12.21	48					
49	20.19	19.25	49				
50	35.65	35.15	36.99	50			
51	34.63	34.34	36.42	17.82	51		
52	34.87	34.42	36.04	18.64	2.68	52	
53	36.38	35.57	37.20	12.36	19.05	20.07	

Fig. 16 – Matriz de distâncias não corrigidas (% de diferença) entre os isolados de *Acanthamoeba* utilizados neste estudo e aqueles presentes na literatura. Os valores marcados em cores indicam os novos tipos de seqüência encontrados neste estudo. Os cálculos foram feitos pelo programa Point Replacer (ver texto).

		*	20	*	40	*	60	*	80
A9	:	TCATATGCTTGTCTCAAAGATTAAGCCATGCATGTCTAAGTATAAGCTTG	-TTT	-ATACGGCGAGACTGCGGATGGCTCAT					
A1	:								
A2	:								
A3	:								
A4	:								
A5	:								
A6	:								
A7	:								
A8	:								
A10	:								
A11	:								
A12	:								
A13	:								
A14	:								
A15	:								
A16	:								
A17	:								
A18	:								
A19	:								
A20	:								
A21	:								
A.cast.#2	:								
A22	:								
A23	:								
A24	:								
A25	:								
A26	:								
A27	:								
A.poly.#2	:								
U/E2	:								
U/E3	:								
U/E4	:	A							
U/E5	:								
U/E6	:								
U/E7-1	:								
U/E7-2	:								
U/E8L	:								
U/E8R	:								
U/E8X	:								
U/E10	:	G							
U/Oft1	:								
A29	:								
A30	:								
A32	:								
A33	:								
U/HC1	:								
A34	:	T							
A36	:	T							
A38	:	T							
A39	:	T							
A42	:	T							
A44	:	T							
A45	:	T							
A.poly.#4	:								
A46	:								
A.astronyx	:	A . GA . A . T T							
A47	:	A . GATA . T T							
A48	:	A . GAA . . T T							
A49	:	T A . GA . A . T T T							
A50	:								
A51	:								
A52	:								
A53	:								

```

          *      100      *      120      *      140      *      160
A9       : TAAATCAGTTATAGTTTATTGATGGTCTCTTTGTCTT----TTTTT---ACC-TACTTGGATAACCGTGGTAATTCTA
A1       : .....T.....-.....
A2       : .....GT.....C.....
A3       : .....GT.....C.....
A4       : .....GT.....C.....
A5       : .....G.....GATT.....
A6       : .....G.....GATT.....
A7       : .....G.....GATT.....
A8       : .....G.....GATT.....
A10      : .....
A11      : .....
A12      : .....
A13      : .....
A14      : .....TT.....
A15      : .....TT.....
A16      : .....
A17      : .....
A18      : .....T.....
A19      : .....
A20      : .....
A21      : .....
A.cast.#2 : .....
A22      : .....T.....
A23      : .....
A24      : .....
A25      : .....
A26      : .....
A27      : .....
A.poly.#2 : .....TT.....
U/E2     : .....T.....
U/E3     : .....T.....
U/E4     : .....T.....
U/E5     : .....T.....
U/E6     : .....T.....
U/E7-1   : .....T.....
U/E7-2   : .....T.....
U/E8L    : .....
U/E8R    : .....
U/E8X    : .....
U/E10    : .....T.....
U/Oft1   : .....
A29      : .....
A30      : .....
A32      : .....T.....
A33      : .....
U/HC1    : .....TT.....
A34      : .....GACA.GGCGCAAGCCG.ATC.TT.....T.....
A36      : .....GACA.GGCGCAAGCCG.ATC.TT.....T.....
A38      : .....GACA.GGCGCAAGCCG.ATC.TT.....T.....
A39      : .....GACA.GGCGCAAGCCG.ATC.TT.....T.....
A42      : .....GACA.GGCGCAAGCCG.ATC.TT.....T.....
A44      : .....GACA.GGCGCAAGCCG.ATC.TT.....T.....
A45      : .....GACA.GGCGCAAGCCG.ATC.TT.....T.....
A.poly.#4 : .....GT.....C.....-C.....
A46      : .....A.AA.....C.....-.....
A.astronx : .....T-T.....C.....T.....-.....
A47      : .....T-T.....-.....-.....
A48      : .....T-T.....-.....-.....A.....
A49      : .....T-T.....C.....T.....-.....T.....
A50      : .....AC.....TGT..CAGAGTT....C.....
A51      : .....C.A.....-.....
A52      : .....C.A.A.....-.....
A53      : .....AC.....TGT..A.GAGTT.--.-.....

```

```

      *      180      *      200      *      220      *      240
A9      : GAGCTAATACATGCGCAAGGTCCCGAGCGCGGGGGGC-----GGGGCT-----TCA-CGGCCT-----CG
A1      : .....A.....-A.-.TAA.....TG.....A-A.....T
A2      : .....A.....T.....A.....TTC.....TA
A3      : .....A.....T.....A.....TC.....TA
A4      : .....A.....T.....A.....TC.....TA
A5      : .....A.....A.....TG.....A-A.....
A6      : .....A.....A.....TG.....A-A.....
A7      : .....A.....A.....TG.....A-A.....
A8      : .....A.....A.....A.....TG.....A-A.....
A10     : .....
A11     : .....A.....C.....T
A12     : .....A.....TC.....T
A13     : .....C.....
A14     : .....T.....
A15     : .....C.....
A16     : .....T.....TC.....A
A17     : .....
A18     : .....T.....TC.....A
A19     : .....
A20     : .....C.....
A21     : .....AT.....TC.....T
A.cast.#2 : .....AT.....TC.....T
A22     : .....A.....T.....
A23     : .....AT.....TC.....T
A24     : .....
A25     : .....
A26     : .....A.....TC.....T
A27     : .....C.....
A.poly.#2 : .....C.....
U/E2     : .....TC
U/E3     : .....A.....TC.....T
U/E4     : .....A.....-TC.....TC
U/E5     : .....A.....-TC.....TC
U/E6     : .....A.....-TC.....TC
U/E7-1   : .....A.....-TC.....TC
U/E7-2   : .....A.....-TC.....TC
U/E8L    : .....T.....
U/E8R    : .....TC
U/E8X    : .....TC
U/E10    : .....TA
U/Oft1   : .....TC
A29      : .....
A30      : .....
A32      : .....A.....T.....
A33      : .....T.....TC.....A
U/HCl    : .....A.....TC.....CTCG...CGGGC..
A34      : .....T..AA..T.....TGGTCTT...TGCAGGTGG..A.ACTTGTGCTCCGTTTCGGGC.A
A36      : .....T..AA..T.....TGGTCTT...ATGCAGGTGG..A.ACTTGTGTTCCGTTTCGGGC.A
A38      : .....T..AA..T.....TGGTCTT...TGCAGGTGG..A.ACTTGTGCTCCGTTTCGGGC.A
A39      : .....T..AA..T.....TGGTCTT...TGCAGGTGG..A.ACTTGTGCTCCGTTTCGGGC.A
A42      : .....T..AA..T.....TGGTCTT...TGCAGGTGG..A.ACTTGTGCTCCGTTTCGGGC.A
A44      : .....T..AA..T.....TGGTCTT...TGCAGGTGG..A.ACTTGTGCTCCGTTTCGGGC.A
A45      : .....T..AA..T.....TGGTCTT...TGCAGGTGG..A.ACTTGTGTTCCGTTTCGGGC.A
A.poly.#4 : .....A.....A.A.TAA...--.T.....A.....T.....TT
A46      : .....A.....A.A.T.....A.....TA
A.astronx : .....T.....T..T.TGCCT..TTTTTCGTTTCAT.GTG...GCCA.TGT.CGCGCACTGGCGCA.A
A47      : .....T.....T..C.TGCC.CC.TG...T...TG.....CA.....TGGCG..
A48      : .....T.....T..T.TGCC.CT.TG...T...TG.....CA.....GCA..
A49      : .....T.T.....T..T.TGCCT..TTTTTCGTTTCAT.GTG...ACCA.TGT.TGTGCACTGGTGCA.A
A50      : .....CAAA...G.A.TG.AAAAA.G.TT.TTTG.....T
A51      : .....A.A.....TC.....A.....T..
A52      : .....A.....AAGGATC.....T..
A53      : .....A.....ATAAGGAAGT...G.AAAAAAC...TTTT.....--

```

```

*          260          *          280          *          300          *          320
A9      : TCCTCGC-ATGCG----CAGAGGGATGTATTATTAGGTTAAA--AACCAGC--A--GCCGGCAAC-GGC----TTCA
A1      : .ATCT.....CG.....-...-...-C...TTCAAAACCC----
A2      : ...T.....C.....-...-...TTTAAACAATCAAT
A3      : ...T.....C.....-...-...TTTTCATATCA-T
A4      : ...T.....T.....C.....-...-...T.TTTTCATT.---
A5      : .T.CT.....A.....GC..T.A....-T...G.....
A6      : .T.CT.....A.....GC..T.A....-T.A.G.....
A7      : .T.?T.....A.....GC..T.A....-T.A.G.....
A8      : .T.CT.....A.....GC..T.A....-T...G.....C
A10     :
A11     : .....GGC.GGG.T.A....-T...CCCTGC..
A12     : .T.....GGC.G....A....-T...C..TG...
A13     :
A14     : .T.....
A15     : .....G.....T...C.....
A16     : .T.....G.T....T.A....-T.....A..
A17     :
A18     : .T.....G.T....T.A....-T.....A..
A19     :
A20     :
A21     : .....A....-T.....A..
A.cast.#2 : .....A....-T.....A..
A22     : .T.....
A23     : .....G.T....A....-T....TAC...
A24     :
A25     :
A26     : .T.....GGC.G....A....-T...C..TG...
A27     :
A.poly.#2 : .....G.....C.....
U/E2     : GT.CTCGC.....GGC.G....A....-T...C..TG...
U/E3     :
U/E4     : GTTCTCGC.....
U/E5     : GTTCTCGC.....
U/E6     : GTTCTCGC.....
U/E7-1   : GTTCTCGC.....
U/E7-2   : GTTCTCGC.....
U/E8L    : GT.CTCGC.....
U/E8R    : GT.CTCGC.....
U/E8X    : GT.CTCGC.....
U/E10    : GT.CTCGC.....
U/Oft1   : GT.CTCGC.....
A29      :
A30      :
A32      : .T.....
A33      : .T.....G.T....T.A....-T.....A..
U/HC1    : --...T.....C.....-...-...AC...
A34      : ..T.T.....GA.....ACAC.C..A.....-...-...TTCTTTT.A.----
A36      : ..T.T.....GA.....ACAC.C..A.....-...-...TTCTTTT.A.----
A38      : ..T.T.....GA.....ACAC.CACA.....-...-...TTCTTTT.A.----
A39      : ..T.T.....GA.....ACAC.C..A.....-...-...TTCTTTT.A.----
A42      : ..T.T.....GA.....ACAC.C..A.....-...-...TTCTTTT.A.----
A44      : ..T.T.....GA.....ACAC.C..A.....-...-...TTCTTTT.A.----
A45      : ..T.T.....GA.....ACAC.C..A.....-...-...TTCTTTT.A.----
A.poly.#4 : ATTCTTGC.....T..C.....C.....-...-...A.TTCTTT.AAAACAG
A46      : ..T..T.....T.....C.....-...-...TTTCAATATCAT.
A.astronyx : GGTG..GGCCTTTGGTAG.A.....A.....A.....G.....-G....T-..CTCCCT.CG-
A47      : GTT.T.G.-...-G.A.....A.....A.....G.....-G....C-..TTCTCTC.-
A48      : GTT.G.G.-...-G.A.....A.....A.....G.....-G....T-..CTCTCTC.-
A49      : GGTGG...C.TT.GTA.G.A.....A.....A.....T.....-GT....T-..TTCCCT.CGT
A50      : GT.ATT...-T.....-...-...ACACTTTAT.AAA.----
A51      : ..TCT.....GC..T.A....-T...G.....
A52      : ...CT.....GC..T.A....-T...G..T...
A53      : GT.-T.....T.....-...-...A-..TTTAAAAATCAATT

```



```

          *       340           *       360           *       380           *       400
A9       : A---CTC-CTGGTGATTCATAGTAACTCT-TTCGGATCGCA-TT-CATGCCCTCCTT-GT-----
A1       : -...---.....A.....AT.....
A2       : -...---.....
A3       : -...---.....?.....
A4       : -...---.....T.....
A5       : .ACA.A-.....AT.....
A6       : .ACA.A.A.....AT.....
A7       : .ACA.A-.....AT.....
A8       : .ATA.A.A.....AT.....
A10      : ..CA.....T.....
A11      : .ACA.....T.....
A12      : CAC...A.....T.....C.....
A13      : ..CA.....T.....
A14      : ..CA.....T.....
A15      : .A.....T.....
A16      : .AC.TC.C.....T.....C.....
A17      : ..CA.....T.....
A18      : .AC.TC.C.....T.....C.....
A19      : ..CA.....T.....
A20      : ..CA.....T.....
A21      : .T.....T.....T.....
A.cast.#2 : .T.....T.....T.....
A22      : ..CA.....T.....
A23      : .T.....T.....
A24      : .A.....T.....
A25      : ..CA.....T.....
A26      : CAC...A.....T.....C.....
A27      : ..CA.....T.....
A.poly.#2 : .A.....T.....
U/E2     : ...TCTC.....T.....
U/E3     : -CAC...A.....T.....C.....
U/E4     : ..CA.-TC.....T.....
U/E5     : ..CA.-TC.....T.....
U/E6     : ..CA.-TC.....T.....
U/E7-1   : ..CA.-TC.....T.....
U/E7-2   : ..CA.-TC.....T.....
U/E8L    : ..CA.....T.....
U/E8R    : ..CA.....T.....
U/E8X    : ..CA.....T.....
U/E10    : ...TCTC.....T.....
U/Oft1   : ..CA.....T.....
A29      : ..CA.....T.....
A30      : ..CA.....T.....
A32      : ..CA.....T.....
A33      : .AC.TC.C.....T.....C.....
U/HC1    : .AAC...A.....G.....T.....C.....-C.....
A34      : -...---.....T.AT.T.....
A36      : -...---.....T.AT.T.....
A38      : -...---.....T.AT.T.....
A39      : -...---.....T.AT.T.....
A42      : -...---.....T.AT.T.....
A44      : -...---.....T.AT.T.....
A45      : -...---.....T.AT.T.....
A.poly.#4 : .ATT.ATT.....A.....
A46      : T...---.....T.C...-C.....
A.astronyx : CGGGGA-GG..TCT.CAA..G-----CGG.CC.-----TT-.-.---.G.GGG--..CCGATCTCTGGTGATTCATAGT
A47      : GGGAGGG.G..TCT.CAA..GCAGT--T.....--TTCT..CTTG.GGGGGGAGA.CTGATCTCTGGTGATTCATAGT
A48      : GGGGGAG.G..TCT.CAA..GCAGT--TG.CAA---GGTT-.....C-.G.CTG.G..CTGATCTCTGGTGATTCATAGT
A49      : GGGGAA-.G..TCT.CAA..G--T--GG.CC.---.TG-.-.---.GGG-....CCG?TCTCTGGTGATTCATAGT
A50      : -...---.....A.....TA.TA.ATAT.CT.C.T-GT.....
A51      : CACAA--.....T.....
A52      : CACAA--.....G..T.....
A53      : -...---.....A.....T.T...AT...A..TT.CGAT.....

```

		*	420	*	440	*	460	*	480
A9	:	-----					GGGG-GCGGCGACGATTCAATTC		AAATTTCTGC
A1	:						AT.		
A2	:								
A3	:						?		
A4	:						A.		
A5	:						AT.		
A6	:						AT.		
A7	:						AT.		
A8	:						AT.		
A10	:						A.		
A11	:						A.		
A12	:						A.		
A13	:						A.		
A14	:						A.		
A15	:						A.		
A16	:						A.		
A17	:						A.		
A18	:						A.		
A19	:						A.		
A20	:						A.		
A21	:						A.		
A. cast. #2	:						A.		
A22	:						A.		
A23	:						A.		
A24	:						A.		
A25	:						A.		
A26	:						A.		
A27	:						A.		
A. poly. #2	:						A.		
U/E2	:						A.		
U/E3	:						A.		
U/E4	:						A.		
U/E5	:						A.		
U/E6	:						A.		
U/E7-1	:						A.		
U/E7-2	:						A.		
U/E8L	:						A.		
U/E8R	:						A.		
U/E8X	:						A.		
U/E10	:						A.		
U/Oft1	:						A.		
A29	:						GA.		
A30	:						A.		
A32	:						A.		
A33	:						A.		
U/HC1	:						A.		
A34	:						A. AT.		
A36	:						A. AT.		
A38	:						A. AT.		
A39	:						A. AT.		
A42	:						A. AT.		
A44	:						A. AT.		
A45	:						A. AT.		
A. poly. #4	:						G-T.		
A46	:						A.		
A. astronyx	:	AACTCTTTCGGATCGCAATTAC.		AAACAGTCGCCTCGTCGCAAGGCG.		CG.			
A47	:	AACTCTTTCGGATCGCAATTACAC.		AGCCGCCTCGTCGCAAGGCG.		CG.			
A48	:	AACTCTTTCGGATCGCAAAATTG.		AGCCGCCTTGTGCAAGGCA.		CG.			
A49	:	AACTCTTTCGGATCGCAATTAAACAATATAGTCGCCTCGTTGCAAGGCG.		CG.T.		T.T.			
A50	:					A.		AT.	
A51	:					A.			
A52	:					A.			
A53	:					A.		AT.	

	*	500	*	520	*	540	*	560
A9	:	CCTATCAACTTTTCGATGGTAGGATAGAGGCCTACCATGGTCGTAACGGGTAACGGAGAATTAGGGTTCGATTCCGGAGAGG						
A1	:							
A2	:							
A3	:							
A4	:							
A5	:							
A6	:							
A7	:							
A8	:							
A10	:							
A11	:							
A12	:							
A13	:							
A14	:							
A15	:							
A16	:							
A17	:							
A18	:							
A19	:							
A20	:							
A21	:							
A.cast.#2	:							
A22	:							
A23	:							
A24	:							
A25	:							
A26	:							
A27	:							
A.poly.#2	:							
U/E2	:							
U/E3	:							
U/E4	:							
U/E5	:							
U/E6	:							
U/E7-1	:							
U/E7-2	:							
U/E8L	:							
U/E8R	:							
U/E8X	:							
U/E10	:							
U/Oft1	:							
A29	:							
A30	:							
A32	:							
A33	:							
U/HC1	:							
A34	:							
A36	:							
A38	:							
A39	:							
A42	:							
A44	:							
A45	:							
A.poly.#4	:							
A46	:							
A.astronyx	:							
A47	:							
A48	:	T.....						
A49	:	A.T...T.A...T.....T....T....T.						
A50	:							
A51	:							
A52	:							
A53	:							

	*	580	*	600	*	620	*	640
A9	:	GAGCCTGAGAAATGGCTACCACTTCTAAGGAAGGCAGCAGGCGCGCAAATTACCCAATCCCGACACGGGGAGGTAGTGACA						
A1	:							
A2	:							
A3	:							
A4	:							
A5	:							
A6	:							
A7	:							
A8	:							
A10	:							
A11	:							
A12	:							
A13	:							
A14	:							
A15	:							
A16	:							
A17	:							
A18	:							
A19	:							
A20	:							
A21	:							
A.cast.#2	:							
A22	:							
A23	:							
A24	:							
A25	:							
A26	:							
A27	:							
A.poly.#2	:							
U/E2	:							
U/E3	:							
U/E4	:							
U/E5	:							
U/E6	:							
U/E7-1	:							
U/E7-2	:							
U/E8L	:							
U/E8R	:							
U/E8X	:							
U/E10	:							
U/Oft1	:							
A29	:							
A30	:							
A32	:							
A33	:							
U/HC1	:							
A34	:							
A36	:							
A38	:							
A39	:							
A42	:							
A44	:							
A45	:							
A.poly.#4	:							
A46	:							
A.astronyx	:							
A47	:							
A48	:							
A49	:					T.T		
A50	:							
A51	:							
A52	:							
A53	:			A				

	*	660	*	680	*	700	*	720
A9	:	ATAAA-TAACAATACAGGCGCTCGATAAGAGTCTTGTAAATTGGAATGAGTACAATTAAACCCCTTAACGAGTAACAATTG						
A1	:	C.....						
A2	:							
A3	:							
A4	:							
A5	:							
A6	:							
A7	:							
A8	:							
A10	:							
A11	:							
A12	:							
A13	:							
A14	:							
A15	:							
A16	:							
A17	:							
A18	:							
A19	:							
A20	:							
A21	:							
A.cast.#2	:							
A22	:							
A23	:							
A24	:							
A25	:							
A26	:							
A27	:							
A.poly.#2	:							
U/E2	:							
U/E3	:							
U/E4	:							
U/E5	:							
U/E6	:							
U/E7-1	:							
U/E7-2	:							
U/E8L	:							
U/E8R	:							
U/E8X	:							
U/E10	:							
U/Oft1	:							
A29	:							
A30	:							
A32	:							
A33	:							
U/HC1	:	G.....C.....						
A34	:	T.....C.....G.....						
A36	:	T.....C.....G.....						
A38	:	T.....C.....G.....						
A39	:	T.....C.....G.....						
A42	:	T.....C.....G.....						
A44	:	T.....C.....G.....						
A45	:	T.....C.....G.....						
A.poly.#4	:							
A46	:							
A.astronyx	:	T.....C.....						
A47	:	T.....C.....						
A48	:	T.....C.....						
A49	:	T.....T.....C.....T.....A.....						
A50	:	A.....C.....						
A51	:							
A52	:							
A53	:	C.....						

	*	740	*	760	*	780	*	800	*
A9	:	GAGGGCAAGTCTG	-	GTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGCGTATATTAAAGTTGTTGCAGTTAAAAAGCTCG					
A1	:								
A2	:								
A3	:								
A4	:								
A5	:								
A6	:								
A7	:								
A8	:								
A10	:								
A11	:								
A12	:								
A13	:								
A14	:								
A15	:								
A16	:								
A17	:								
A18	:								
A19	:								
A20	:								
A21	:								
A.cast.#2	:								
A22	:								
A23	:								
A24	:								
A25	:								
A26	:								
A27	:								
A.poly.#2	:								
U/E2	:								
U/E3	:								
U/E4	:								
U/E5	:								
U/E6	:								
U/E7-1	:								
U/E7-2	:								
U/E8L	:								
U/E8R	:								
U/E8X	:								
U/E10	:								
U/Oft1	:								
A29	:								
A30	:								
A32	:								
A33	:								
U/HC1	:								
A34	:								
A36	:								
A38	:								
A39	:								
A42	:								
A44	:								
A45	:								
A.poly.#4	:								
A46	:								
A.astronyx	:								
A47	:								
A48	:								
A49	:								
A50	:								
A51	:								
A52	:								
A53	:								

```

      820          *      840          *      860          *      880          *
A9      : TAGTTGGATCTAGGGACGCGCATTTC-----AAGCGCCCGTGTCTCGTCCGGGTCAAACCGGCGACTGCG--
A1      : .....T.....-.....T...CA..A.....T.G-.....
A2      : .....-ATTT.....AT.....A.T.....T.T.....
A3      : .....-ATTT.....AT.....A.....-.....AT.....
A4      : .....-ATTTTT.....?..ATTT..A.TA.....T.AT..T..A..
A5      : .....T.....A.....T.....
A6      : .....T.....A.....T.....
A7      : .....T.....A.....T.....
A8      : .....A.....T...C.....
A10     : .....T.....A.....
A11     : .....
A12     : .....C.A.....T.G.....
A13     : .....
A14     : .....T.....A.....
A15     : .....C.A.....T.G.....
A16     : .....C.A.....T.G.....
A17     : .....T.....A.....
A18     : .....
A19     : .....T.....A.....
A20     : .....
A21     : .....C.A.....T.G.....
A.cast.#2 : .....C.A.....T.G.....
A22     : .....T.....A.....
A23     : .....
A24     : .....C.A.....T.G.....
A25     : .....C.A.....T.G.....
A26     : .....
A27     : .....C.....G.....
A.poly.#2 : .....C.A.....T.G.....
U/E2    : .....
U/E3    : .....
U/E4    : .....Y..T.....A.....
U/E5    : .....T.....A.....
U/E6    : .....T.....A.....
U/E7-1  : .....T.....A.....
U/E7-2  : .....T.....A.....
U/E8L   : .....C.A.....T.G.....
U/E8R   : .....C.A.....T.G.....
U/E8X   : .....C.A.....T.G.....
U/E10   : .....
U/Oft1  : .....T.....A.....
A29     : .....C.A.....T.G.....
A30     : .....C.A.....T.G.....
A32     : .....T.....A.....
A33     : .....
U/HC1   : .....-ACA.....GCGT.G..C.....G.-.....
A34     : .....T.T.....T.....-ACATTCATAAAGGTGTA..AGC.TGTG.AA.CGTCT.ACTG.TCTATTGGCGCTA
A36     : .....T.T.....T.....-ACATTCATAAAGGTGTA..AGC.TGTG.AA.CGTCT.ACTG.TCTATTGGCGCTA
A38     : .....T.T.....T.....-ACATTCATAAAGGTGTA..AGC.TGTG.AA.CGTCT.ACTG.TCTATTGGCGCTA
A39     : .....T.T.....T.....-ACATTCATAAAGGTGTA..AGC.TGTG.AA.CGTCT.ACTG.TCTATTGGCGCTA
A42     : .....T.T.....T.....-ACATTCATAAAGGTGTA..AGC.TGTG.AA.CGTCT.ACTG.TCTATTGGCGCTA
A44     : .....T.T.....T.....-ACATTCATAAAGGTGTA..AGC.TGTG.AA.CGTCT.ACTG.TCTATTGGCGCTA
A45     : .....T.T.....T.....-ACATTCATAAAGGTGTA..AGC.TGTG.AA.CGTCT.ACTG.TCTATTGGCGCTA
A.poly.#4 : .....CA..TAATTAT.....ATTT..A..ACT.....T..T..T...-..
A46     : .....T.....-ATT..ATTATAT.....A.ATTT..A..A.T.....T.AT..T..A..
A.astronyx : .....T....G..ACACCG..TTCTGTGTGTGCCGTGGTG...TT-----CTC.GG----
A47     : .....T....G..ACACCG..TTTCGTGTGCTGCACGTCC..CA..CG...TCTCTC.TTTT.CCTC.GGCG.CG
A48     : .....T....G..ACACCG..TTTCGTGTGCGTGCAATGTC...TT-----CTC.GG----
A49     : .....T....T.ACACCG..TTCTGTGTGTGCCGTGGTG...TT-----CTC.GG----
A50     : .....T.....T.....A..ATTTTAAATGACG??TCT??T...T.TAT.....GTAAA.G.AAAA..
A51     : .....?..CTA.T.....AT..G..T..
A52     : .....T...CACTA.....T..G.....
A53     : .....T..A.A.....T.C...ATCTTATTGTATG.....TTT..T.TCGAA.TGT.TTTATA.ATT.TT.A.

```

	900	*	920	*	940	*	960	*
A9	:	-----						
A1	:							
A2	:							
A3	:							
A4	:							
A5	:							
A6	:							
A7	:							
A8	:							
A10	:							
A11	:							
A12	:							
A13	:							
A14	:							
A15	:							
A16	:							
A17	:							
A18	:							
A19	:							
A20	:							
A21	:							
A.cast.#2	:							
A22	:							
A23	:							
A24	:							
A25	:							
A26	:							
A27	:							
A.poly.#2	:							
U/E2	:							
U/E3	:							
U/E4	:							
U/E5	:							
U/E6	:							
U/E7-1	:							
U/E7-2	:							
U/E8L	:							
U/E8R	:							
U/E8X	:							
U/E10	:							
U/Oft1	:							
A29	:							
A30	:							
A32	:							
A33	:							
U/HC1	:							
A34	:	GGTCACCTTCACGGGTGGTCTGGTGTCTCAAGGTAGT.....						
A36	:	AGTCACCTTCACGGGTGGTTTGGTGTC..AAGGTAGT.....						
A38	:	AGTCACCTTCACGGGTGGTTTGGTGTC..AAGGTAGT.....						
A39	:	AGTCACCTTCACGGGTGGTTTGGTGTC..AAGGTAGT.....						
A42	:	GGTCACCTTCACGGGTGGTTTGGTGTCTCAAGGTAGT.....						
A44	:	AGTCACCTTCACGGGTGGTTTGGTGTC..AAGGTAGT.....						
A45	:	AGTCACCTTCACGGGTGGTTTGGTGTC..AAGGTAGT.....						
A.poly.#4	:							
A46	:							
A.astronyx	:	CTTCTGC...CT....GGGCAGGCGTG CATG....CGGT.....GCATGCAAGGGTGCGTGCTTTTCAGTG						
A47	:	GCGGGGTCATACCCGTTTTTCTCGGGGGGGGGGGAGTCGTGTGGCGCGTGCGCATGCGAGGGTGCGTGCTTTTCAGTG						
A48	:	CTTTCGGGT..CT.....GGGGAGT..TGGGCAGTGTCGTATGCGAGGGTGCGTGCTTTTCAGTG						
A49	:	CTTCTGC...CT....GGGCAGGCGTG CATG....TGGT.....GCATGCAAGGGTGCGTGCTTTTCAGTG						
A50	:							
A51	:							
A52	:							
A53	:							


```

          980          *          1000          *          1020          *          1040          *
A9       : -----TTGGCGTTGCGGGCTCGGTCCGTCCGTG-CCCCAC-AA
A1       : .....-..G.AG..-----..
A2       : .....T.....A.TA..CAT.....A.-
A3       : .....T.....A.TAA.CAT.....A.-
A4       : .....-?..T.....T.A.CAT.....A.-
A5       : .....T.....T.-..T..
A6       : .....-..T.-..T..
A7       : .....-A.....T.-..T..
A8       : .....-..T.-..-
A10      : .....-
A11      : .....GA...T...-
A12      : .....-
A13      : .....
A14      : .....G..A.....
A15      : .....G.....TC..
A16      : .....T.-.....
A17      : .....-.....
A18      : .....T.-.....
A19      : .....-.....
A20      : .....
A21      : .....
A.cast.#2 : .....
A22      : .....GTG..A.....
A23      : .....
A24      : .....
A25      : .....
A26      : .....T-.....
A27      : .....
A.poly.#2 : .....G.....TC..
U/E2     : .....TTC.
U/E3     : .....T-.....
U/E4     : .....GTG..A.....
U/E5     : .....GTG..A.....
U/E6     : .....G..A.....
U/E7-1   : .....GTG..A.....
U/E7-2   : .....GTG..A.....
U/E8L    : .....
U/E8R    : .....
U/E8X    : .....
U/E10    : .....C...ATTG..
U/Oft1   : .....--.....
A29      : .....
A30      : .....
A32      : .....GTG..A.....
A33      : .....?..-.....
U/HC1    : .....C.....CT..A.G-----..
A34      : .....A.....A..A..TG..TCGC..
A36      : .....A.....A..A..TG..TCGC..
A38      : .....A.....A..A..TG..TCGC..
A39      : .....A.....A..A..TG..TCGC..
A42      : .....A.....A..A..TG..TCGC..
A44      : .....A.....A..A..TG..TCGC..
A45      : .....A.....A..A..TG..TCGC..
A.poly.#4 : .....-AA.T.....A.TAA.CAT.....A.-
A46      : .....-AA.T.....TAA.CAT.....A.-
A.astronyx : TGCGGGG.CACCTGCACATGGCACGAGGAGGGCTCGGTCCGC.ACC.TACCGTT..G-CACG.A..AAAGGTG.GTTGC
A47      : TGTGGGGG.CACCTGCACATGGCACGAGGAGGGCTCGGTCCGC..AC.TGCCGTT..G----.AC..AAAGGTGGGTGCGT
A48      : TGTGGGGCCGACCTGCACATGGCACGAGGAGGGCTCGGTCCGC.-AC.CGCCGTT..A----TA..AAAGGTGTGT..-
A49      : TGCAAGGT.GACTTGCACATGGCATGAGGAGAGCTTGGTCCGC.ACT.TACCATT.TG-CA.G.ACT.AAAGGTG..TTGT
A50      : .....AATGA.AT...T.....ACT..A.G-----..
A51      : .....T.....-.....
A52      : .....T.....-.....T..
A53      : .....AAATGA-AT...T..A.....ACT..A.G-----..

```

```

      1060          *      1080          *      1100          *      1120          *
A9      : A-GGGCT---ACTGGCGTGTCAACCGGCCCCGCCCCGTCCCCTCCTTCTGGATTCCCGTTCCT-----GCTATTGAGTT
A1      : -.-----..CA.....
A2      : -.-----TG.T.A.T....T.....
A3      : -.-----TG.T.A.T....T.....
A4      : -.-----TGGT.A....T.....
A5      : .A---..TC.....
A6      : .A---..TC.....
A7      : .A---..TC.....
A8      : .A---..TC.....
A10     : .....C.....
A11     : GT..T..TA..TC.....
A12     : .....TC.....
A13     : .....TC.....
A14     : .T..-C.....
A15     : .C...T.TT..TC.....
A16     : .A---..TC.....
A17     : .....C.....
A18     : .A---..TC.....
A19     : .....C.....
A20     : .....TC.....
A21     : .....T....TC.....
A.cast.#2 : .....T....TC.....
A22     : .T..-C.....
A23     : .....T....TC.....
A24     : .....TC.....
A25     : .....TC.....
A26     : .....TC.....
A27     : .....TC.....
A.poly.#2 : .C...T.TT..TC.....
U/E2    : .A.....TC.....
U/E3    : .....GC.TC.....
U/E4    : .T..-C.....G...A.....T..T.....
U/E5    : .T..-C.....
U/E6    : .T..-C.....
U/E7-1  : .T..-C.....
U/E7-2  : .T..-C.....
U/E8L   : .....TC.....
U/E8R   : .....TC.....
U/E8X   : .....TC.....
U/E10   : .....TC.....
U/Oft1  : .....TC.....
A29     : .....
A30     : .....TC.....
A32     : .T..-C.....
A33     : .A---..TC.....
U/HCl   : .....CA.....C.....
A34     : G....CTCC.T...T...T.....T...
A36     : G....CTCC.T...T...T.....T...
A38     : G....CTCC.T...T...T.....T...
A39     : G....CTCC.T...T...T.....T...
A42     : G....CTCC.T...T...T.....T...
A44     : G....CTCC.T...T...T.....T...
A45     : G....CTCC.T...T...T.....T...
A.poly.#4 : -.-----TGGT.A.T....T.....
A46     : -.-----TGGT.A....T.....A.....
A.astronyx : GT.C.-GT.T-G..G.C..GC..T.....CT.....T....-..CAG..TTTG.....A.C..
A47     : GGC..-TGT--G..G.C..GC..T.....CT.....T....-..CAG..CCG.....A.C..
A48     : GGC..-T.T-GG..G.C..GC..T.....CT.....T....-..CCAGTTTCG.....A.C..
A49     : GT.CA-GT.T-G..G-..GC..T...T..T.CT.....T.T..-..CAGTATTG.....A.C..
A50     : -.-----..CA.T.....-G.....A...A.....A.....
A51     : .....TCA.....
A52     : .....TCA.....
A53     : -.-----..CA.T.....-G....TT.A....T.T.....A..T.....

```

```

1140      *      1160      *      1180      *      1200      *
A9       : AG---TGGGGACGTCACAGGGGGCTCATCGTCGT-GCGGCGTCAAAAC--CGTGGCG---GCGGTG-----
A1       : .....ATCGC..A...GG.-T.G--...--AC.-C.....T...
A2       : .....A...GTATCATTG...C.T...--G.G--C.A...T...A...
A3       : .....A...-TATCATTG...C-T...--G.G--C.A...T...A...
A4       : .....AAT.-TATCATTG...-C.T...--G.G--C.A...T...A...
A5       : .....A.....CA.....--TG.A-C.....
A6       : .....A.....CA.....--T.....
A7       : .....A.....CA.....--TG.A-C.....
A8       : .....A.....CA.....G.--TG.A-C.....
A10      : .....A.....CA.....--C.....
A11      : .....CAT.....--CG.....
A12      : .....C.....C.CAT.....--
A13      : .....CGT.....--
A14      : .....ATCGT.....-..G--
A15      : .....TC.....CGT.....--
A16      : .....C.....TC-ATG.....--ACG
A17      : .....A.....CA.....--C.....
A18      : .....C.....TCCATG.....--ACG
A19      : .....A.....CA.....--C.....
A20      : .....CGT.....--
A21      : .....T.....G.CGT--T.....G.....
A. cast.#2 : .....T.....G.CGT--T.....G.....
A22      : .....CCA.....--TG.C.....
A23      : .....TC.....A.CGT--T.....G.....
A24      : .....TC.....CG.....--GC--
A25      : .....TC.....
A26      : .....C.....CAT.....--
A27      : .....CGT.....--
A. poly.#2 : .....TC.....CGT.....--
U/E2     : .....TC.....-GT.....--C.TGCG
U/E3     : .....C.....--CATG.....--
U/E4     : .....A.....--CATG..C--G.....
U/E5     : .....C.AG..T--G.C.....
U/E6     : .....A.--TCGTG.....-G.....
U/E7-1   : .....C.AG.....--CG
U/E7-2   : .....C.AG.....--CG
U/E8L    : .....TC.....
U/E8R    : .....TC.....
U/E8X    : .....TC.....
U/E10    : .....TC.....
U/Oft1   : .....C.....GC.A.--CG
A29      : .....TC.....
A30      : .....TC.....
A32      : .....CCA.....--T.G--C.....
A33      : .....C.....TCCATG.....--CACG
U/HC1    : G.....CGCG.....--A.....CC.GC
A34      : .....A.....TAAAA.....CTG
A36      : .....A.....TAAAA.....CTG
A38      : .....A.....TC.....TAAAA.....CTG
A39      : .....A.....TC.....TAAAA.....CTG
A42      : .....A.....TAAAA.....CTG
A44      : .....A.....TC.....TAAAA.....CTG
A45      : .....A.....TC.....TAAAA.....CTG
A. poly.#4 : .....AAGCA.A...TTG.CT.AC...A...T.....
A46      : .....AAT.-TATCATCG...-T.T...G.GA--...T...A...
A. astronyx : .CCC...G..C.TAG..C..TGTGC-.C.-.AAACAGCGTGC.TTTT-----C...G.CATCTGGCGCTTTG
A47      : .CTC...G..C.T.G..T.TGCTG.T.C.C.TAGCCG.G..GTGGTTTT-.A...CTCAG.GCACTGTGCTTTT
A48      : .CTC...G..C.T.G..T.TGCTG.T.C.C.TAGCCG.GT.GCGG..TT-.A...CTCAG.GCATCTGTGCTTC
A49      : .CCC.A...G.AC.TAG..CA.TGTG-.C.-.AAACAGTGTGC.TTTT-----CT..G.CATCTGGTGTCTTTG
A50      : .....T.....AAT.G..A...A.....CAA.GT
A51      : .....T.....CA.....--TG.A--
A52      : .....TC.....GTCA.....--TG....C.....
A53      : .....A.T.....A.AT.G-.AA..A.....-T....T..TT-----TTAAA.....T

```

```

1220          *          1240          *          1260          *          1280          *
A9      : -----GGTCCCTGGGGCCCAGATCGTTTACCGTGAAAAAATTAGAGTGTTCAAAGCAGGCAGATC-CAATTTT-C-TGC
A1      : .....T...T.....T.TT.CCA.T....
A2      : .....CA..TT.....T.....
A3      : .....A.CTT.....T.....
A4      : .....A..TT.....T.....
A5      : .....A..T.....
A6      : .....A..T.....
A7      : .....A..T.....
A8      : .....A..T.....
A10     : .....T.....
A11     : .....
A12     : .....T.....
A13     : .....
A14     : .....
A15     : .....
A16     : .....
A17     : .....T.....
A18     : .....A.....
A19     : .....T.....
A20     : .....
A21     : .....A..T.....T.....
A.cast.#2 : .....A..T.....T.....
A22     : .....
A23     : .....A..T.....T.....
A24     : .....
A25     : .....
A26     : .....T.....
A27     : .....
A.poly.#2 : .....
U/E2    : .....
U/E3    : .....T.....
U/E4    : .....
U/E5    : .....
U/E6    : .....
U/E7-1  : .....
U/E7-2  : .....
U/E8L   : .....
U/E8R   : .....
U/E8X   : .....
U/E10   : .....
U/Oft1  : .....
A29     : .....
A30     : .....
A32     : .....
A33     : .....
U/HCl   : .....C.....CAT...CCA.....
A34     : .....-...T.....G.....-...T.C...
A36     : .....-...T.....G.....-...T.C...
A38     : .....-...T.....G.....-...T.C...
A39     : .....-...T.....G.....-...T.C...
A42     : .....-...T.....G.....-...T.C...
A44     : .....-...T.....G.....-...T.C...
A45     : .....-...T.....G.....-...T.C...
A.poly.#4 : .....-TGTTTCT.....T.....
A46     : .....A..T.....T.....
A.astronyx : CGGTGCCA.GTGT...A.....A.TTTCA.-A....
A47     : GCGGCACA.GTGT..A.....A.TTT.A..A....
A48     : GCGGTACA.GTGT..A.....T.TTTCAA.A....
A49     : TGATGCCA.GTGT...A.....T.....T.....A.TTTCA.-A....
A50     : GATTATGAA.T.....T.T..CCA.T....
A51     : .....A.....
A52     : .....A.....
A53     : GGTATGA-.TT.....T.T..CCA.T....

```

	1300	*	1320	*	1340	*	1360	*	1
A9	:	CACCGAATACATTAGCATGGGATAATGGAATAGGACCCTGTCTCCTATTTTCAG--TTGGTTTGGCAGCGC-----							
A1	:	C.CG.....--TT-----							
A2	:	C.....AACTT-.TA-							
A3	:	C.....AACTT-.TA-							
A4	:	C.C.....AATTT.CGTGT							
A5	:	--..--							
A6	:	--..--							
A7	:	--..--							
A8	:	--..--							
A10	:								
A11	:								
A12	:								
A13	:								
A14	:								
A15	:								
A16	:								
A17	:								
A18	:								
A19	:								
A20	:								
A21	:								
A.cast.#2	:								
A22	:								
A23	:								
A24	:								
A25	:								
A26	:								
A27	:								
A.poly.#2	:								
U/E2	:								
U/E3	:								
U/E4	:								
U/E5	:								
U/E6	:								
U/E7-1	:								
U/E7-2	:								
U/E8L	:								
U/E8R	:								
U/E8X	:								
U/E10	:								
U/Oft1	:								
A29	:								
A30	:								
A32	:								
A33	:								
U/HC1	:	C-----							
A34	:	A.....TTTA.AGC							
A36	:	A.....TTTA.AGC							
A38	:	A.....TTTA.AGC							
A39	:	A.....TTTA.AGC							
A42	:	A.....TTTA.AGC							
A44	:	A.....TTTA.AGC							
A45	:	A.....TTTA.AGC							
A.poly.#4	:	C.....AATTT.TGTGC							
A46	:	T.....AATT..CAAGA							
A.astronyx	:	T.ATG...-.CG....G..C..CGTGCGTGCCT							
A47	:	T.ATG...-.CG....G..CACC AAAAGGACAAG							
A48	:	T.ATG...-.CG....G..CA-----TTTT							
A49	:	...T.....A...T....TTATG..T-.TG....G..C..TGTGCATGCACT							
A50	:	A.....-AC.TG.....T-----							
A51	:	-.-.-?							
A52	:	-.-.-							
A53	:	-AC.TG.....T-----							

	380	*	1400	*	1420	*	1440	*	14
A9	:	-----							-GAGG-
A1	:								
A2	:								
A3	:								
A4	:								
A5	:								
A6	:								
A7	:								
A8	:								
A10	:								
A11	:								
A12	:								
A13	:								
A14	:								
A15	:								
A16	:								
A17	:								
A18	:								
A19	:								
A20	:								
A21	:								
A. cast. #2	:								
A22	:								
A23	:								
A24	:								
A25	:								
A26	:								
A27	:								
A. poly. #2	:								
U/E2	:								
U/E3	:								
U/E4	:	TCGCGG.....							
U/E5	:								
U/E6	:								
U/E7-1	:								
U/E7-2	:								
U/E8L	:								
U/E8R	:								
U/E8X	:								
U/E10	:								
U/Oft1	:								
A29	:								
A30	:								
A32	:								
A33	:								
U/HC1	:								
A34	:								T
A36	:								T
A38	:								T
A39	:								T
A42	:								T
A44	:								T
A45	:								T
A. poly. #4	:								
A46	:								
A. astronox	:	TCGCGG.....							TGGCATGGTTTCGTCTAAAG.....
A47	:	AATGGTTGTGTGTGCGCGTGCAAACGCGTACCATAGCTATTTTTTGT							TTTGGTGATGGTTTCGTCTAAAG.....
A48	:	TCTAGC.GCGTGTT.....							TTGGGGGGTGATGGTTTCGTCTAAAG.....
A49	:	TCGCGGTGTGTG.....							CATG.....TGGCATGGTTTGTCTAAAG.....
A50	:								
A51	:								
A52	:								
A53	:								

	60	*	1480	*	1500	*	1520	*	154
A9	:	---	ACTAGGGTAATGATTAATAGGGATAGTTGGGGGCATTAATATTTAATTGTCAGAGGTGAAATTCTTGGATTATGAAA						
A1	:		C.....						
A2	:		TC.....						
A3	:		TC.....						
A4	:		TC.....						
A5	:		C.....						
A6	:		C.....						
A7	:		C.....						
A8	:		C.....						
A10	:								
A11	:								
A12	:								
A13	:								
A14	:								
A15	:								
A16	:								
A17	:								
A18	:								
A19	:								
A20	:								
A21	:								
A.cast.#2	:								
A22	:								
A23	:								
A24	:								
A25	:								
A26	:								
A27	:								
A.poly.#2	:								
U/E2	:								
U/E3	:								
U/E4	:		C.....						
U/E5	:								
U/E6	:								
U/E7-1	:		C.....						
U/E7-2	:								
U/E8L	:								
U/E8R	:		C.....						
U/E8X	:								
U/E10	:								
U/Oft1	:								
A29	:								
A30	:				C.....				
A32	:								
A33	:								
U/HC1	:		C.....						
A34	:	TAT.	TC.....						
A36	:	TAT.	TC.....						
A38	:	TAT.	TC.....						
A39	:	TAT.	TC.....						
A42	:	TAT.	TC.....						
A44	:	TAT.	TC.....						
A45	:	TAT.	TC.....						
A.poly.#4	:		TC.....						
A46	:		AC.....						
A.astronyx	:		C.....						
A47	:		C.....						
A48	:		C.....						
A49	:		C.....			A.....			
A50	:		C.....						
A51	:		C.....						
A52	:		C.....						
A53	:		C.....						


```

*      1640      *      1660      *      1680      *      1700
A9      : CCGTCGTAGTCTTAACCATAAACGATGCCGACCAGCGATTAGGAGACGTTGAATACAAAACACCACCATCGGC----GCGG
A1      : .....A.GCAGT.G..
A2      : .....ACATAATTCA.T
A3      : .....ATATAATTCA.T
A4      : .....TT.ATATAATTCAAT
A5      : .....GC...ATT...
A6      : .....GC...ATT...
A7      : .....GC.-.ATT...
A8      : .....GC.-.ATT...
A10     : .....T.....
A11     : .....
A12     : .....T.....
A13     : .....
A14     : .....T...T...
A15     : .....T.....
A16     : .....
A17     : .....T.....
A18     : .....
A19     : .....T.....
A20     : .....
A21     : .....
A.cast.#2 : .....
A22     : .....T...T...
A23     : .....
A24     : .....
A25     : .....T.....
A26     : .....T.....
A27     : .....T.....
A.poly.#2 : .....T.....
U/E2     : ...C...YTA.....-.....
U/E3     : .....T.....
U/E4     : .....T...T...
U/E5     : .....T...T...
U/E6     : .....T...T...
U/E7-1   : .....T...T...
U/E7-2   : .....T...T...
U/E8L    : GA-----
U/E8R    : .....
U/E8X    : .....
U/E10    : .....
U/Oft1   : .....T.....
A29      : .....
A30      : .....
A32      : .....T...T...
A33      : .....
U/HC1    : .....GCGGGT...T
A34      : .....G.TG.TAATCCTTT...
A36      : .....G..G.TAATCCTTT...
A38      : .....G.TG.TAATCCTTT...
A39      : .....G..G.TAATCCTTT...
A42      : .....G.TG.TAATCCTTT...
A44      : .....G..G.TAATCCTTT...
A45      : .....G.TG.TAATCCTTT...
A.poly.#4 : .....G.T.AA.ATAATTTCATT
A46      : .....T.T..ATA.AATTCA.T
A.astronyx : .....C.....G.....---...C.CGGGAGTGG..CCGGG...---
A47      : .....C.....G.....---...C.CGGAGGTGGC.TCGG...-G..
A48      : .....C.....G.....---...C.CGGAGGTGGC.TCGGG...---
A49      : .....C.....G.....---...C.CGGGGGTGG..CC.AT...G..
A50      : .....T.CATTTAGCA?.GTC
A51      : .....T.....
A52      : .....T.....
A53      : .....TAA.ACGAT...T

```

```

          *      1720      *      1740      *      1760      *      1780
A9       : TCGTCCTTGGCGTC--GGTCCTTCAC-----GGGGCC--GGCGCGAGGGCGGCTTAGCCCCG-G--TGGCACCGGTGAATGA
A1       : ...G...C...TTT--C.G-G--A...AC...GAA..TG.A.GC..TCT.A.TC...TGA.....-.....
A2       : A.A.GGCCA-----GTCAA...AT..TTGTGTATTGTTT.TG-----
A3       : A.A.GGCCA-----GTCAA...AT..TTGTGTATTGTTT.TG-----
A4       : ATA.GGC.T-----A...C..CTGTAT.TTGTTT.TA--TAA.....-C.A.....
A5       : ....T...T...C....A...A..C.G.A....CG.....T..CA.....
A6       : ....T...T...GTC....A...A..C.G.ATCAT--G..A.....T..CA.....
A7       : ....T...T...C....A...A..C.G.AT----G..A.....T..CA.....
A8       : ....T...T...C....C...AGCGAT.T.TGGT...ATCG..AT.....T..CA.....
A10      : ....TC...T....GG..T...G..A.....
A11      : ....TGTCC.T...AC.....G.A.....T.....
A12      : ....T....AA...A---A...G...T.....
A13      : ....T....G.....
A14      : ....TC...T....GG.....G..A...T.....
A15      : ....T....G.....
A16      : ....GTCTCGG.....CTGGGGC.....T.....
A17      : ....TC...T....GG..T...G..A.....
A18      : ....GTCTCGG.....CCGGGGC.....T.....
A19      : ....TC...T....GG..T...G..A.....
A20      : ....T....G..A.....
A21      : ....TC...G.....CACG-A-.....A.....
A.cast.#2 : ....TC...G.....CACG-A-.....A.....
A22      : ....TC...T....GG.....G..AT.....
A23      : ....TC...G.....CACG-A-.....T.....A.....
A24      : ....T....G...T.....
A25      : ....TC.....GG.....G...T.....
A26      : ....CGTT....AA...A---AGC...G.....
A27      : ....T....G..A.....
A.poly.#2 : ....TT-----C.....G.....
U/E2     : ....GGTCCTT-----C..GGCC.....
U/E3     : ....GT....AA...A---G...T.....
U/E4     : ....TC.T.TT-----C.CG.GG.....G..AT.....
U/E5     : ....TC...TT-----C...GG.....G..AT.....
U/E6     : ....TC...TT-----C...GG.....G..AT.....
U/E7-1   : ....TC...TT-----C...GG.....G..AT.....
U/E7-2   : ....TC...TT-----C...GG.....G..AT.....
U/E8L    : ....TCT.....C.A...CGA...R.....
U/E8R    : ....TC.....GG.....G.....
U/E8X    : ....TC.....GG.....G.....
U/E10    : ....GGTCCTT-----C...GG-----
U/Oft1   : ....TC...TT-----C...GG.....G..AT.....
A29      : ....TC.....GG.....G.....
A30      : ....TC.....GG.....G.....
A32      : ....TC...T....GG.....G..AT.....
A33      : ....GTCTCGG.....CCGGGGC.....T.....
U/HC1    : C.C.GGCC..T...GTCCGT.GCGG.AACGC..C.GGCG.C..GTG...GC.G.CC.TTTCCCCGAT.GC.....
A34      : GG-----A-----A--TT.....
A36      : GG-----A-----A--TT.....
A38      : GG-----A-----A--TT.....
A39      : GG-----A-----A--TT.....
A42      : GG-----A-----A--TT.....
A44      : GG-----A-----A--TT.....
A45      : GG-----A-----A--TT.....
A.poly.#4 : ATA.GGC.TCACGGCT.TA-----TA.TG.TTGT.TT-TA.TTCAT.....
A46      : ATA.GG.CA-----G--A...AT..TTGTATATTGTTT.TA--TA.....A-----
A.astronx : AGAGTGCGT.G.GGTG.CCG...GTG...C..TAG.CTCCTTG.CTTT.CCT..G.GG--CTGC.CC...TAAA.G...
A47      : G.A.TT...CT..TCC.TCG.GCA-....C...GGGC..AATGT.T.CCT..TTGGG.CTGC.TC...TAAA.G...
A48      : -...G..C--C.CT.TCG.GCA-....C...AGGC...A-----CT..TTGGG-CTGC.TC..GTAAA.G...
A49      : AGAGTGCGT.G.GGTG.CCG...TG...C..T.G.CTCCTCGCCTTT.CCT..G.GGG.CTGC.CC...TAAA.G...
A50      : GTT.T-CAAATA.TCCTT.TTGCG.AG...GTT.T--TT----GAACGA-T.CGT...-T.A...AT.T...
A51      : ....ACGTC.TGG.-G..A....CCATGGC.G.....AT.....T..A.....T.....
A52      : ....CGTC.TGG....CTG.GGC.T.....T.....T.....T.....
A53      : .TT.TGCAAATA.GCCACATGCG..AG...T.T.TGGTT.TGTTTGAA.GAA.GATTGTCTTAA...-C.....

```

		*	1800	*	1820	*	1840	*	1860
A9	:	CTCC	CCTAGCAGCTTGTGAGAAATCAT	-AAGT	CTTTGGGTTCCGGGGGGAGTATGGT	CGCAAGGCTGAAACTTAAAGGAAT			
A1	:								
A2	:								
A3	:								
A4	:								
A5	:								
A6	:								
A7	:								
A8	:								
A10	:								
A11	:								
A12	:		T						
A13	:								
A14	:								
A15	:								
A16	:								
A17	:								
A18	:								C
A19	:								
A20	:								
A21	:								
A.cast.#2	:								
A22	:								
A23	:								
A24	:								
A25	:								
A26	:								
A27	:								
A.poly.#2	:								
U/E2	:								
U/E3	:								
U/E4	:								
U/E5	:								
U/E6	:								
U/E7-1	:								
U/E7-2	:								
U/E8L	:								
U/E8R	:								
U/E8X	:								
U/E10	:								
U/Oft1	:								
A29	:								
A30	:				?				
A32	:								
A33	:								
U/HC1	:								
A34	:								
A36	:								
A38	:								
A39	:								
A42	:								
A44	:								
A45	:								
A.poly.#4	:								
A46	:								
A.astronyx	:		C		C				
A47	:		C		C				
A48	:		C		C				
A49	:		C		C				
A50	:								
A51	:								
A52	:								
A53	:								

	*	1880	*	1900	*	1920	*	1940
A9	:	TGACGGAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGGAAACTTACCAGGTCCGGACATAGT						
A1	:							
A2	:							
A3	:							
A4	:							
A5	:							
A6	:							
A7	:							
A8	:							C.....
A10	:							
A11	:							
A12	:		A.....					
A13	:							
A14	:							
A15	:							
A16	:							
A17	:							
A18	:							
A19	:							
A20	:							
A21	:							
A.cast.#2	:							
A22	:							
A23	:							
A24	:							
A25	:							
A26	:							
A27	:							
A.poly.#2	:							
U/E2	:							
U/E3	:							
U/E4	:							
U/E5	:							
U/E6	:							
U/E7-1	:							
U/E7-2	:							
U/E8L	:							
U/E8R	:							
U/E8X	:							
U/E10	:							
U/Oft1	:							
A29	:							
A30	:							
A32	:							
A33	:							
U/HC1	:							
A34	:							
A36	:							
A38	:							
A39	:							
A42	:							
A44	:							
A45	:							
A.poly.#4	:							
A46	:							
A.astronyx	:							
A47	:							
A48	:							
A49	:							
A50	:							
A51	:							
A52	:							
A53	:							

	*	1960	*	1980	*	2000	*	2020
A9	:	AAGGATTGACAGATTGATAGCTCTTTCTTGATTCTATGGGTGGTGGTGCATGGCCGTTCTTAGTTGGTGGAGTGATTGTC						
A1	:							
A2	:							
A3	:							
A4	:							
A5	:							
A6	:							
A7	:							
A8	:							
A10	:							
A11	:							
A12	:							
A13	:							
A14	:							
A15	:							
A16	:							
A17	:							
A18	:							
A19	:							
A20	:							
A21	:							
A.cast.#2	:							
A22	:							
A23	:							
A24	:							
A25	:							
A26	:							
A27	:							
A.poly.#2	:							
U/E2	:							
U/E3	:							
U/E4	:							
U/E5	:							
U/E6	:							
U/E7-1	:							
U/E7-2	:							
U/E8L	:							
U/E8R	:							
U/E8X	:							
U/E10	:							
U/Oft1	:							
A29	:							
A30	:							
A32	:							
A33	:							
U/HC1	:							
A34	:	G						C
A36	:	G						C
A38	:	G						C
A39	:	G						C
A42	:	G						C
A44	:	G						C
A45	:	G						C
A.poly.#4	:							
A46	:							
A.astronyx	:							
A47	:							
A48	:							
A49	:							
A50	:							
A51	:							
A52	:							
A53	:							

	*	2040	*	2060	*	2080	*	2100
A9	:	TGGTTAATTCCGTTAACGAACGAGACCTTAACCTGCTAAATATGCCGCGCTAACCC	-----	GTCC	-----			
A1	:							
A2	:					T		
A3	:							
A4	:							
A5	:							
A6	:							
A7	:							
A8	:							
A10	:							
A11	:							
A12	:							
A13	:							
A14	:							
A15	:							
A16	:							
A17	:							
A18	:							
A19	:							
A20	:							
A21	:							
A.cast.#2	:							
A22	:							
A23	:							
A24	:							
A25	:							
A26	:							
A27	:							
A.poly.#2	:					Y		
U/E2	:							
U/E3	:							
U/E4	:							
U/E5	:							
U/E6	:							
U/E7-1	:							
U/E7-2	:							
U/E8L	:							
U/E8R	:							
U/E8X	:							
U/E10	:							
U/Oft1	:							
A29	:							
A30	:							
A32	:							
A33	:							
U/HC1	:							
A34	:	T		G		A		A
A36	:	T		G		A		A
A38	:	T		G		A		A
A39	:	T		G		A		A
A42	:	T		G		A		A
A44	:	T		G		A		A
A45	:	T		G		A		A
A.poly.#4	:							
A46	:					T		
A.astronyx	:					C.T..TTGGCCA..AAAGCCCGCGTATGT		
A47	:					C.T..TTGGCCA..AAAGCCCGCATGTGT		
A48	:					C.T..TTGGCCA..AAAGCCCGTGTGTGT		
A49	:					C.T..TTGGCCA..AAAGCCCGCGTATGT		
A50	:							
A51	:							
A52	:							
A53	:					T		

```

*      2120      *      2140      *      2160      *      2180
A9      : ----ATCAAAACCCATGCGCGGCTCACGCGGTCCGCTGCGGGGTGGTG-TC-----GCTT--CGCGGCGACGTC-ATCCC
A1      : .....T.....T.....A.....T.....CA.....GCT.G--C.AA....GT.GTC...A.
A2      : .....T...AT.T.A.....T.....ATAT.A...C.G...CC.AG--C.AA-T.GTGGTGTT.ATAT
A3      : .....AT.T.A.....T.....ACA..A...C.G...TC.AG.-C.AAAT.ATGGTGTT.ATGT
A4      : .....T.T.A.....T.....ACAC.AA..C.G...CCC.T.AAAA.GT.GTGTTA-.GGG-T
A5      : .....T.....T.....ATACCAA..C.G...CCGCG..CG.....T.GT..T.G.TGT
A6      : .....T.....T.....AT.CCAA..C.G...CCGCG..CG.....T.GT..T.G.TAT
A7      : .....T.....T.....ATACCAA..C.G...CC.-G..CG.....T.GT..T.G.TAT
A8      : .....T.....T.....ACAACAA..C.G...CCGCG..CA.....T.GT..T.G.TGT
A10     : .....T.....T.....T.....T.....T.....
A11     : .....T.....T.....T.....T.....T.....
A12     : .....T.....T.....T.....T.....T.....
A13     : .....C.....T.....T.....T.....T.....
A14     : .....T.....T.....T.....T.....T.....
A15     : .....T.....T.....T.....T.....T.....
A16     : .....T.....T.....T.....T.....T.....
A17     : .....T.....T.....T.....T.....T.....
A18     : .....T.....T.....T.....T.....T.....
A19     : .....T.....T.....T.....T.....T.....
A20     : .....C.....T.....T.....T.....T.....
A21     : .....T.....T.....T.....T.....T.....
A.cast.#2 : .....T.....T.....T.....T.....T.....
A22     : .....T.....T.....T.....T.....T.....
A23     : .....T.....T.....T.....T.....T.....
A24     : .....T.....T.....T.....T.....T.....
A25     : .....T.....T.....T.....T.....T.....
A26     : .....T.....T.....T.....T.....T.....
A27     : .....C.....T.....T.....T.....T.....
A.poly.#2 : .....G.....AGACGTC.....
U/E2    : .....T.....T.....T.....T.....T.....
U/E3    : .....T.....T.....T.....T.....T.....
U/E4    : .....T.....T.....T.....T.....T.....
U/E5    : .....T.....T.....T.....T.....T.....
U/E6    : .....T.....T.....T.....T.....T.....
U/E7-1  : .....T.....T.....T.....T.....T.....
U/E7-2  : .....T.....T.....T.....T.....T.....
U/E8L   : .....T.....T.....T.....T.....T.....
U/E8R   : .....T.....T.....T.....T.....T.....
U/E8X   : .....T.....T.....T.....T.....T.....
U/E10   : .....T.....T.....T.....T.....T.....
U/Oft1  : .....T.....T.....T.....T.....T.....
A29     : .....T.....T.....T.....T.....T.....
A30     : .....T.....T.....T.....T.....T.....
A32     : .....T.....T.....T.....T.....T.....
A33     : .....T.....T.....T.....T.....T.....
U/HC1   : .....A.....C.AC-.C.....GTGCCT..AA.CG..T.CG..GCG..G.
A34     : .....TC..A.....AA...C.....G.TGTC..TTT
A36     : .....TC..A.....AA...C.....A.....G.TGTC..TTT
A38     : .....TC..A.....AA...C.....A.....G.TGTC..TTT
A39     : .....TC..A.....AA...C.....A.....G.TGTC..TTT
A42     : .....TC..A.....AA...C.....A.....G.TGTC..TTT
A44     : .....TC..A.....AA...C.....A.....G.TGTC..TTT
A45     : .....TC..A.....AA...C.....A.....G.TGTC..TTT
A.poly.#4 : .....T...T.T.A.....T.....ACAC.AA..T.-.....--GCC..TTAAACGGTG
A46     : .....TT...T.T.A..T...T...T.A...ATAT.AA..T.G...TC.AG.-C.AAAT.GTGATATT.AGAT
A.astronyx : TTTT--..CGGT.TGCGCGT.GC.G.CTCTC.GCTGCGTCC-AAAA-..GCA.GTGC-----GGG.
A47     : CTTTT.A..CGG..TGCGCGT.G.TCTTT.TC..TG.TT.CCCA.CAAT--GGCGC..AAGCAA--ATT..TTGG.GG
A48     : CTTTT--..CGG..TGCGCGT.G.TC-----CATT.--TGCAT..GCAAGC--ATTAGGT----
A49     : TTTTAAAT.CGGT.TGC.-GT.GC.G.CTCTC.GCTGCGTCC-AAAA-..GCA.GC-----GGG.
A50     : .....A..T...CTTT.A.A.....ATTAGT.ATGAGGCGCGTT.GCGCGCGC--TTTA..AT.AGT
A51     : .....T.....A..ACA.C.....A.....T..T.G.T.T
A52     : .....A..ACA...C.....C.....T.GT.CT.G.T.T
A53     : .....T..AT.T...CTTT..A.....ATTAAT.ACA..GTGTGT.T..ACACACATACGAGGGT.GTT

```

```

      *      2200      *      2220      *      2240      *      2260
A9      : GCCGCGCAGGGCCCGGGTCCGTGTGGGCGGTA--GGGT--TCGG----CGT-CCGTGCTTCTTAGAGGGACTGCTGCGC-
A1      : .....T.....
A2      : .....T.A...A.....A.....T.....
A3      : .....T.A.....T.....
A4      : .....T.A.....
A5      : .....CT..A.....
A6      : .....CT..A.....
A7      : .....CT..A.....
A8      : .....CT..A.....
A10     : .....
A11     : .....
A12     : .....
A13     : .....
A14     : .....
A15     : .....
A16     : .....
A17     : .....
A18     : .....
A19     : .....
A20     : .....
A21     : .....
A.cast.#2 : .....
A22     : .....
A23     : .....
A24     : .....
A25     : .....
A26     : .....
A27     : .....
A.poly.#2 : .....
U/E2    : .....
U/E3    : .....
U/E4    : .....
U/E5    : .....
U/E6    : .....
U/E7-1  : .....
U/E7-2  : .....
U/E8L   : .....
U/E8R   : .....
U/E8X   : .....
U/E10   : .....
U/Oft1  : .....
A29     : .....
A30     : .....
A32     : .....
A33     : .....
U/HC1   : .....C.....T.....-...
A34     : ....TG..AT...A.CT.....C.....ATT.
A36     : ....TG..AT...A.CT.....C.....ATT.
A38     : ....TG..AT...A.CT.....C.....ATT.
A39     : ....TG..AT...A.CT.....C.....ATT.
A42     : ....TG..AT...A.CT.....C.....ATT.
A44     : ....TG..AT...A.CT.....C.....ATT.
A45     : ....TG..AT...A.CT.....C.....ATT.
A.poly.#4 : ATGTTAG..TG...CAGG.CCC...TTCATATA..CGGTAG..TTCGG...T.....
A46     : ....T...A...A...T.A.A.A...A.....T.....
A.astronyx : .G-C---...-ATGG..C.C.T.-CA.GCGC...GG...ATGGG...T.C.....TCTGCGC
A47     : .AG.....-ATGG..C.C.T.-CA.GTGC...GG...ATGGG...T.T.C.....TCTGCGC
A48     : -----..G...-ATGG..C.C.T.-CAC.TGC...GG...ATGGG...T.C.....TCTGCGC
A49     : .GGC---...-ATGG..C.C.T.-CA.GCGC...GG...ATGG...T.C.....TCTGCGC
A50     : .....T...T..A...G...A.....TT.....CTG....C
A51     : .....CT...?.....
A52     : .....CT.....
A53     : AGT.C.G.CTGGGTCCAGGTATATATA..CGGTA.....TT.....CTG....C

```



```

      *      2280      *      2300      *      2320      *      2340
A9      : -----GCCTAGCCAGCGGAAGTTTGAGGCAATAACAGGTCTGTGATGCCCTTAGATGTTCTGGGCCGCACGCGCGCT
A1      : .....T.....
A2      : .....
A3      : .....
A4      : .....
A5      : .....
A6      : .....
A7      : .....
A8      : .....
A10     : .....
A11     : .....
A12     : .....
A13     : .....
A14     : .....
A15     : .....
A16     : .....
A17     : .....
A18     : .....
A19     : .....
A20     : .....
A21     : .....
A.cast.#2 : .....
A22     : .....
A23     : .....
A24     : .....
A25     : .....
A26     : .....
A27     : .....
A.poly.#2 : .....
U/E2    : .....
U/E3    : .....
U/E4    : .....
U/E5    : .....
U/E6    : .....
U/E7-1  : .....
U/E7-2  : .....
U/E8L   : .....A..hC.....A.K.T.t.....
U/E8R   : .....
U/E8X   : .....
U/E10   : .....
U/Oft1  : .....
A29     : .....
A30     : .....
A32     : .....
A33     : .....
U/HC1   : .....T.....
A34     : ATTTAAAGC.TT.....
A36     : ATTTAAAGC.TT.....
A38     : ATTTAAAGC.TT.....
A39     : ATTTAAAGC.TT.....
A42     : ATTTAAAGC.TT.....
A44     : ATTTAAAGC.TT.....
A45     : ATTTAAAGC.TT.....
A.poly.#4 : .....
A46     : .....
A.astronyx : T.....TAGC..A.G.---.....
A47     : T.....TAGC..A.G.---.....
A48     : T.....TAGC..A.G.---.....
A49     : T.....TAGC..A.G.---.....
A50     : .....TAG.CAG.G.---.....
A51     : .....
A52     : .....
A53     : .....TAG.CAG.G.---.....

```

```

*      2360      *      2380      *      2400      *      2420      *
A9      : ACACTGATTAATCCAACGAGTCCGCTTCAATCGAGGC-GCGATGCCGTT---GGG-----GTCAAA
A1      : .....G...G..CATC---.....G...C
A2      : .....TA.-TAT.CGT.....
A3      : .....TA.-TATA.GT.....
A4      : .....TAG.AAATG.....
A5      : .....TAGC.TATG.....
A6      : .....TAGC.TATG.....
A7      : .....TAGC.TATG.....
A8      : .....TAGC.TATG.....
A10     : .....
A11     : .....
A12     : .....T.G...-TGGTGC.....
A13     : .....GC.....
A14     : .....T.GC..T.G.C.....
A15     : .....GC.....
A16     : .....T..C-TT-A.CAT.....
A17     : .....C.....
A18     : .....T.GCTTT-A.CAT.....
A19     : .....
A20     : .....GC.....
A21     : .....T.GC..GTC.GTC.....
A.cast.#2 : .....T.GC..GTC.GTC.....
A22     : .....T.G.GTC.....
A23     : .....T.GC..CG.C.....
A24     : .....T.GC..T..C.....
A25     : .....
A26     : .....T.G..G.TGGTGC.....
A27     : .....GC.....
A.poly.#2 : .....GC.....
U/E2     : .....GC...CG.T.....
U/E3     : .....T.G..G.TGGTGC.....
U/E4     : .....T.G-CG.TGGT.C.....
U/E5     : .....T.G-CG.TGGT.C.....
U/E6     : .....T.G-CG.TGGT.C.....
U/E7-1   : .....T.G.GTC.....
U/E7-2   : .....T.G.GTC.....
U/E8L    : .....GGG---.....
U/E8R    : .....GGG---.....
U/E8X    : .....GGG---.....
U/E10    : .....GC...CG.TGG---.....
U/Oft1   : .....GGG---.....
A29      : .....
A30      : .....TG.CCGTT.....
A32      : .....T.G.GTC.....
A33      : .....T.GCCTTCA..AT.....
U/HC1    : .....T.GC..G.CGGGGC.....
A34      : .....GCA..TTCATGT.CGGT.....
A36      : .....GCA..TTCATGT.CAGT.....
A38      : .....GCA..TTCATGT.CAGT.....
A39      : .....GCA..TTCATGT.CAGT.....
A42      : .....GCA..TTCATGT.CGGT.....
A44      : .....GCA..TTCATGT.CAGT.....
A45      : .....GCA..TTCATGT.CAGT.....
A.poly.#4 : .....T.....TAG-TTT...-GT.....
A46      : .....A.....TA.-TTTA.AT.....
A.astronx : .....T.....T.-.-.-.-.CT.C.CTT-GTCGTGGCGGTTG.CCGTCGTGCTGCTGTGCTCGGTCG...G
A47      : .....T.....T.-.-.-.-.CT.C.CTT.TCG.TTTT.GGGT...?CGTCGTGC...TTCG..GTCG...G
A48      : .....T.....T.-.-.-.-.CT.C.CTT-T-G.CGT.-GG.T...GCGTCGTGC...TCTG..GTCG...G
A49      : .....T.....T.-.-.-.-.CT.C.CTTTGTGTCGTGGCGGTTG.CCGTCGTGCTG...TGCTCGGTCG...G
A50      : .....T.....TAGATTA...TT...TTTG.....A.C.
A51      : .....TA?...TATA.GT.....
A52      : .....T....C.TAGCC-GTG.GT.....
A53      : .....T.....T.....TT.GATATA.GAA.T.ACC.....-...C-

```

```

      2440          *      2460          *      2480          *      2500          *
A9      : -----CCC---A--ACTG-CGTC-GCTGTCCTCGATCGCGCCTGGGCCGATAGGTCCGGGTAATCTT-TGCAAATT
A1      : .....GGTGG.-...G..TG.-.....
A2      : .....T.ATG..TA.-GTT--.....T.....
A3      : .....T.ATG..TA.-ATT--.....T.....
A4      : .....T.ATG..TT.-ATT--.....
A5      : .....ATG..TA...TC.-.....
A6      : .....ATG..TA...TC.-.....
A7      : .....ATG..TA...TC.-.....
A8      : .....T..ACG..TA...TC.-.....
A10     : .....G.....
A11     : .....T.....
A12     : .....GTGC..CG...A.....
A13     : .....G..C.....
A14     : .....G..-...TCCG.....
A15     : .....G..C.....
A16     : .....T.ATG..T-GA.-.....T.....
A17     : .....G.....
A18     : .....ATG..T-GA.-.....T.....
A19     : .....
A20     : .....G..C.....
A21     : .....GAT...GT.....
A.cast.#2 : .....GAT...GT.....
A22     : .....G..-...TCCG.....
A23     : .....GATG..GTC.....
A24     : .....G..-...C.....
A25     : .....
A26     : .....GTGC..CA..C.A.....
A27     : .....G..C.....
A.poly.#2 : .....G..C.....
U/E2    : .....G..GTC.....
U/E3    : .....GTGC..CG..C.A.....
U/E4    : .....G..-...TCC.GCGTC.....
U/E5    : .....G..-...TCC.GCGTC.....
U/E6    : .....G..-...TCC.GCGTC.....
U/E7-1  : .....G..-...TCC.GCGTC.....
U/E7-2  : .....G..-...TCC.GCGTC.....
U/E8L   : .....G...GTC.....
U/E8R   : .....G...GTC.....
U/E8X   : .....G...GTC.....
U/E10   : .....G...GTC.....
U/Oft1  : .....T.....
A29     : .....
A30     : .....
A32     : .....G..-...TCCG.....
A33     : .....ATG..T?CA.G.....T.....
U/HC1   : .....GTTCC..GTC.T.....
A34     : .....-ATC..GTA-TT..GA...GTC-.....
A36     : .....-ATT..GTA-TT..GA...GTC-.....
A38     : .....-ATT..GTA-TT..GA...GTC-.....
A39     : .....-ATT..GTA-TT..GA...GTC-.....
A42     : .....-ATC..GTA-TT..GA...GTC-.....
A44     : .....-ATT..GTA-TT..GA...GTC-.....
A45     : .....-ATT..GTA-TT..GA...GTC-.....
A.poly.#4 : .....T.AT...TAA-ATT--.....T.....
A46     : .....T.GTG..TAA-ATT--...C...AT.....
A.astronyx : GCCGGGTGCGGTGCGCGCACGGTTG.CGCTGC.G.GG..C...TC...T.....TGTG....
A47     : GCTGA.GGTG.GGCGTC..-TGCCT.GGC.GC.G.GCTCG.--...T.....TGTG....
A48     : GCTGGTGGTG.GGCGTC..GTGCCT--GC.GC.G.GCTCG.--...T.....TGTG....
A49     : ACCGGGTGCGGTGCGCG..CGGTTG.CGCTGC.G.GG..C...TC...T.....TGTG....
A50     : .....CAAA.AAGAC..TA.-.TTCTG-.....T.....
A51     : .....ATG..CA..GTC.-.....T.....
A52     : .....ATG..CGC.GTC.-.....
A53     : .....GGTTATTTTGT...TA.-ATTCTGA.....TA.....T.....

```

	2520	*	2540	*	2560	*	2580	*
A9	:	TAATCGTGCTGGGGATAGATCATTGTAATTATTGATCTTCAACGAGGAATTCCTAGTAAGCGCGAGTCATCAGCTCGCGTT						
A1	:							
A2	:							
A3	:							
A4	:							
A5	:							
A6	:							
A7	:							
A8	:							
A10	:							
A11	:							
A12	:							
A13	:							
A14	:							
A15	:							
A16	:							
A17	:							
A18	:							
A19	:							
A20	:							
A21	:							
A.cast.#2	:							
A22	:							
A23	:							
A24	:							
A25	:							
A26	:							
A27	:							
A.poly.#2	:							
U/E2	:							
U/E3	:							
U/E4	:							
U/E5	:							
U/E6	:							
U/E7-1	:							
U/E7-2	:							
U/E8L	:							
U/E8R	:							
U/E8X	:							
U/E10	:							
U/Oft1	:							
A29	:							
A30	:							
A32	:							
A33	:							
U/HC1	:							
A34	:							
A36	:							
A38	:							
A39	:							
A42	:							
A44	:							
A45	:							
A.poly.#4	:							
A46	:							
A.astronyx	:							
A47	:							
A48	:							
A49	:							
A50	:							
A51	:							
A52	:							
A53	:							

	2600	*	2620	*	2640	*	2660	*
A9	:	GATTACGTC	CCTGCCCTTTGTACACACCGCCCGTCGCTCCTACCGATTGAATGGTCCGGTGAAATCCTCGGAGCCGTGGCC					
A1	:							
A2	:							
A3	:							
A4	:							
A5	:							
A6	:							
A7	:							
A8	:							
A10	:							
A11	:							
A12	:							
A13	:							
A14	:							
A15	:							
A16	:							
A17	:							
A18	:							
A19	:							
A20	:							
A21	:							
A.cast.#2	:							
A22	:							
A23	:							
A24	:							
A25	:							
A26	:							
A27	:							
A.poly.#2	:							
U/E2	:							
U/E3	:							
U/E4	:							
U/E5	:							
U/E6	:							
U/E7-1	:							
U/E7-2	:							
U/E8L	:							
U/E8R	:							
U/E8X	:							
U/E10	:							
U/Oft1	:							
A29	:							
A30	:							
A32	:							
A33	:							
U/HCl	:						A	A
A34	:						A	A
A36	:						A	A
A38	:						A	A
A39	:						A	A
A42	:						A	A
A44	:						A	A
A45	:						A	A
A.poly.#4	:							
A46	:							
A.astronyx	:						C	-
A47	:						C	?
A48	:						C	-
A49	:						C	-
A50	:							
A51	:							
A52	:							
A53	:							

```

      2680          *          2700          *          2720          *          2740          *
A9      : TCTACGCAATCC-GGG-----CAACCGGG-TGTGAGGTCTCCCC-----
A1      : .....T.....T.GTGTC.A.GCGA....T
A2      : .....T.....AA.GGTGATGTGC.AAA...C
A3      : .....T.....AA.GGTGGTGTGC.AAG...C
A4      : .....T.....AA.GGTGG.TGCGCAAG...C
A5      : .....---.....T
A6      : .....---.....T
A7      : .....---.....T
A8      : .....T---.....T
A10     : .....
A11     : .....
A12     : .....
A13     : .....
A14     : .....
A15     : .....
A16     : .....
A17     : .....
A18     : .....
A19     : .....
A20     : .....C.....-.....
A21     : .....
A.cast.#2 : .....
A22     : .....C.....
A23     : .....
A24     : .....
A25     : .....
A26     : .....
A27     : .....
A.poly.#2 : .....
U/E2    : .....
U/E3    : .....
U/E4    : .....
U/E5    : .....
U/E6    : .....
U/E7-1  : .....
U/E7-2  : .....
U/E8L   : .....
U/E8R   : .....
U/E8X   : .....
U/E10   : .....
U/Oft1  : .....
A29     : .....-.....
A30     : .....
A32     : .....
A33     : .....
U/HC1   : .....--ACGCGTTGTAAAA...
A34     : .....G.....
A36     : .....G.....
A38     : .....G.....
A39     : .....G.....
A42     : .....G.....
A44     : .....G.....
A45     : .....?.....G.....
A.poly.#4 : .....T.....AA.GGT.....GGT
A46     : .....T.....AA.GGTGCATGCGCAAGTG.T
A.astronox : ...CT.TC.C..A.T.....TCCC.TTTCTTTCCTTC..-C-G..TGG--A---GG-----AGGG.....
A47     : ...CT.TC.C..A...GTCTTTCCCC.TTTTGGCTTCCCTTT.CGG...GGCCATTT.GGAG.ATGGGGG.....
A48     : ...CT.TC.C..A...CCTCCCCACTCTGCCTTCTC.T..CGG...GGGCA---GGAG--TGGGG.....
A49     : ...CT.TC.C..A.T.....TCCCCTTTCTCTCTTCCCTTCG.G..TGG--A---GG-----GGAGGG.....
A50     : ...?.....GTGGGTGGCTTGAGAGTGG..
A51     : .....
A52     : .....
A53     : .....GTAG-TGTTTGA.AGTGGGT

```

```

      2760          *      2780          *      2800          *
A9      : -----TTTGGCGGC-GAAGTCGATTGAACCTTACCATTAGAGGAAG
A1      : GGCAC.....GGGA.CC.-.....
A2      : ATCACC.....TA.....
A3      : ATCACC.....TA.....
A4      : ATCACC.....T?.....
A5      : AAC.....CAC.....
A6      : AAC.....CAC.....
A7      : AAC.....CAC.....
A8      : AAC.....CAC.....
A10     : .....
A11     : .....
A12     : .....T.....
A13     : .....
A14     : .....
A15     : .....
A16     : .....CA.....
A17     : .....
A18     : .....CA.....
A19     : .....
A20     : .....
A21     : .....
A.cast.#2 : .....
A22     : .....GGCGA.....
A23     : .....
A24     : .....
A25     : .....
A26     : .....
A27     : .....
A.poly.#2 : .....
U/E2    : .....--GC.....W.N.....
U/E3    : .....
U/E4    : .....
U/E5    : .....
U/E6    : .....
U/E7-1  : .....
U/E7-2  : .....
U/E8L   : .....
U/E8R   : .....
U/E8X   : .....
U/E10   : .....
U/Oft1  : .....
A29     : .....
A30     : .....
A32     : .....
A33     : .....CA.....
U/HC1   : .....GGCGCGTG.CCT.T.....-
A34     : ..GCAA.....GGCGA....T.....
A36     : ..GCAA.....GGCGA....T.....
A38     : ..GCAA.....GGCGA....T.....
A39     : ..GCAA.....GGCGA....T.....
A42     : ..GCAA.....GGCGA....T.....
A44     : ..GCAA.....GGCGA....T.....
A45     : ..GCAA.....GGCGA....T.....
A.poly.#4 : GTGCAAACAT.....CACCTA.....A.....
A46     : GT.ACC.....TA.....A.....
A.astronx : ...CCAC...TGTTGAGAGGCC.C.TT.A.GC...C.....
A47     : .....TTTCTGTGAGAGGCC.C.AT.A.GC...C.....
A48     : ...CC...TGTTGAGAGGCC.C..T.A.-C...C.....
A49     : ...CCAC...TGTTGAGAGGCC.C.TT.A.GC...C.....
A50     : ..CAACACTC...TTAGGTGCCCC..CGC.GTA.....
A51     : .....CA.....
A52     : .....CA.....
A53     : GTCACCACCCACTTTTAAATAC.CC..CGC.GTA.....

```

Fig. 17 – Seqüências do SSU rDNA dos 46 isolados utilizados para a definição dos 12 tipos de SSU rDNA de *Acanthamoeba*, alinhadas manualmente com as seqüências determinadas neste trabalho. Para facilitar a visualização das diferenças, posições conservadas entre a primeira seqüência e as restantes foram substituídas por pontos (utilizando-se o programa Point Replacer, ver texto).

Um resultado interessante, obtido a partir da análise do alinhamento das seqüências, foi a observação de que o isolado *A. astronyxis* apresentou maior similaridade de seqüência em relação ao tipo T9 (*A. comandoni* ATCC 30135 - 6,45% de diferença) do que em relação ao tipo T7 (*A. astronyxis* ATCC 30137 - 18,61% de diferença). Esse fato indica que, mantidos os binômios dos isolados mais antigos, o isolado ATCC 30901 - isolado em 1964 e, até o momento, considerado como *A. astronyxis* - deveria ser reclassificado como *A. comandoni*, isolada em 1935, caso não tivesse se qualificado como um novo tipo de seqüência do SSU rDNA. A grande divergência observada entre o isolado *A. astronyxis* e os outros já era esperada devido às diferenças encontradas entre eles nos resultados preliminares de RAPD, RFLP de DNA genômico total e de SSU rDNA. Como não dispúnhamos de nenhum dos outros isolados do grupo morfológico I (*A. comandoni*, *A. tubiashi* e outros isolados de *A. astronyxis*) em nosso laboratório, não foi possível estimar a relação entre *A. astronyxis* ATCC 30901 e os outros, pelos métodos de análise de fragmentos de DNA e, portanto, não tínhamos nenhuma indicação prévia quanto à sua classificação específica. Conforme mencionado na Introdução, não ser necessária a disponibilidade dos organismos em cada laboratório é uma das maiores vantagens da análise de seqüências de macromoléculas biológicas, em relação aos outros métodos de estimativa da variabilidade genética. Isso porque essas seqüências, se já determinadas por outros pesquisadores, ficam acessíveis em bancos de dados, onde todos os detalhes sobre sua origem podem ser obtidos, bem como as referências bibliográficas dos trabalhos originais, quando publicados.

3.5.2 – Conceitos de espécie

Para se discutir a validade e as bases da proposta de Stothard e cols., 1998 de considerar cada tipo de sequência do SSU rDNA como uma espécie do gênero *Acanthamoeba*, seria necessário, pelo menos, se ter uma noção dos conceitos de espécie mais comuns descritas na literatura e suas características principais.

3.5.2.1 Conceitos reprodutivos

Dos diversos conceitos, os reprodutivos (o biológico^(148, 150) e o de reconhecimento⁽¹⁵¹⁾) são, historicamente, aqueles mais aceitos, uma vez que se baseiam em características intrinsecamente biológicas (o isolamento reprodutivo ou mecanismos de reconhecimento) e por serem de natureza bastante intuitiva e de simples compreensão. Entretanto, além de apresentar muitas dificuldades na aplicação, até mesmo em diversos organismos de reprodução sexuada^(84, 125), esses conceitos não podem ser aplicados àqueles cuja reprodução seja assexuada, como vários animais e plantas partenogénéticos e à grande maioria dos organismos unicelulares, eucarióticos ou procarióticos. Cada indivíduo está, por definição, isolado reprodutivamente de todos os outros, o que levaria à absurda conclusão de que cada indivíduo é uma espécie. Se tais conceitos fossem rigorosamente aplicados, mesmo duas células irmãs, derivadas de uma fissão binária, teriam de ser consideradas espécies diferentes,.

3.5.2.2 Conceito ecológico⁽¹⁵³⁾

Outros conceitos foram desenvolvidos, na tentativa de se estudar tanto os organismos de reprodução assexuada quanto aqueles com modos de reprodução diferente do padrão básico dos Tetrapoda. Entre eles, o conceito ecológico poderia ser o mais adequado, uma vez que independe totalmente do modo de reprodução. Por este conceito, pertenceriam à mesma espécie aqueles indivíduos

que ocupassem o mesmo nicho ecológico fundamental ou que tivessem a mesma função em um ecossistema. A participação de um dado organismo em um nicho seria inferida a partir de suas características fisiológicas, morfológicas ou bioquímicas, uma vez que as pressões às quais o organismo está exposto certamente selecionaram o estado em que tais características se encontram, levando a correlações que possibilitam sua utilização para identificação. Embora seja muito comum, não se deve, entretanto, inverter o conceito, ou seja, considerar que as características diagnósticas é que definem a espécie.

Embora seja simples na definição, o conceito ecológico de espécie é geralmente de aplicação difícil, uma vez que existem muitas sutilezas, dificilmente perceptíveis, quando se estudam os organismos e seu meio, especialmente quando se trata de organismos microscópicos. Esse conceito deixa em aberto quais seriam os critérios para definir o que constituiria um nicho para diferentes tipos de organismos (parasitas, patógenos oportunistas, apenas de vida livre, etc.) e a partir de que ponto o indivíduo já seria considerado como pertencente a outro nicho. Nesse caso, o principal problema é considerar que os nichos sejam entidades discretas e bem distintas entre si, o que, na realidade, muitas vezes não ocorre. Além disso, outro fator que enfraquece o conceito ecológico é o fato de que diferentes fases do ciclo de vida de um organismo podem ter formas de vida e hábitos muito diversos, fazendo com que pertençam a nichos ecológicos completamente diferentes. Tem-se de considerar, também, o fato de que a ocupação de todo o nicho fundamental (conjunto de todo o ambiente onde é possível a existência do organismo) dificilmente é alcançada, levando apenas à ocupação do chamado nicho efetivo, ou seja, aquele subconjunto do nicho fundamental que é realmente ocupado pelo organismo. Considerando que o nicho efetivo é influenciado por fatores extrínsecos ao

organismo, ambientais, pode ser difícil o reconhecimento do nicho fundamental ao qual uma população realmente pertence.

Por fim, um outro problema em relação ao conceito ecológico, pelo menos do ponto de vista do evolucionista, é a falta da utilização de dados referentes ao relacionamento filogenético entre as populações comparadas, o que aumenta ainda mais o risco de se considerar grupos de hábitos convergentes como agrupamentos naturais.

3.5.2.3 Conceito evolutivo⁽¹⁵²⁾

O conceito evolutivo de espécie define como pertencentes à mesma espécie aquelas populações que compartilham um destino evolutivo comum ao longo do tempo, ou seja, aquelas que pertencem à mesma linhagem ancestral-descendente, levando a uma coesão fenotípica, geralmente utilizada como diagnóstico. Tal definição é aplicável tanto a organismos vivos quanto fósseis, de reprodução sexuada ou não. Como o “destino evolutivo” de uma linhagem depende de uma grande variedade de fatores, a unidade de uma espécie pode estar relacionada a características genéticas, ecológicas ou relativas ao desenvolvimento de um organismo, tornando o conceito bastante versátil, em teoria. Na verdade, é muito semelhante ao conceito ecológico de espécie, adicionado de um componente temporal. Por outro lado, fica indefinido o que seria um “destino evolutivo comum”, uma vez que as populações sempre apresentam um certo nível de polimorfismo. Assim, é necessário escolher, arbitrariamente, o nível de variabilidade que se aceitará dentro de um “destino comum”. Além disso, no caso de microrganismos, como *Acanthamoeba*, as limitações nas possibilidades de investigação de vários dos fatores envolvidos na definição do conceito evolutivo (como os fósseis, a maior parte da ecologia ou a coesão fenotípica em linhagens onde a morfologia não é confiável) reduzem esse

conceito a uma espécie de conceito filogenético menos rigoroso, uma vez que a filogenia pode ser o único daqueles fatores passível de ser determinado com certa segurança.

3.5.2.4 Conceito filogenético ou cladístico⁽¹⁴⁷⁾

Apesar de partir do pressuposto completamente válido e correto de que as classificações devem refletir a história evolutiva, o conceito filogenético ou cladístico é de difícil implementação, uma vez que considera que uma espécie seria o menor grupo de organismos pertencentes a uma rede de ancestrais comuns, diagnosticado por uma ou mais sinapomorfias. Ou seja, o conjunto de populações situadas entre duas ramificações da filogenia estudada (ou entre uma ramificação e o tempo presente). Desse modo, a aplicação estrita desse conceito acabaria criando um número enorme de grupos, tornando as classificações confusas e de difícil compreensão, além de instáveis, uma vez que novas características estudadas podem mudar a classificação freqüentemente⁽¹⁰⁴⁾. Além disso, existe a arbitrariedade de se escolher quantas sinapomorfias serão suficientes como critério de diferenciação entre categorias, escolha essa que não pode ser facilmente justificada.

Um problema ainda maior em relação ao conceito filogenético de espécie, apontado pela teoria da coalescência na genética de populações⁽⁴⁾, é o fato de que as relações genealógicas de muitos alelos, logo após os eventos de especiação, tendem a indicar grupos parafiléticos. Na verdade, são apenas reflexo da estrutura das populações que deram origem às espécies atuais e da amostragem dos alelos quando da separação física entre as populações em questão. Desse modo, diferentes genes podem indicar diferentes filogenias, levando a confusões freqüentes entre "árvores de gene" e "árvores de espécies", se não se levar em consideração este tipo de efeito de amostragem.

3.5.2.5 Conceito de coesão

Um conceito interessante e bastante geral, em relação aos organismos aos quais pode ser aplicado, é o conceito de coesão⁽¹²⁵⁾, na verdade uma tentativa de combinar os melhores aspectos dos conceitos evolutivo, ecológico e reprodutivos. Nesse caso, espécie é definida como a população mais inclusiva de indivíduos que apresentam potencial para coesão fenotípica, através de mecanismos intrínsecos de coesão. Tais mecanismos são aqueles que promovem o parentesco genético e determinam as fronteiras populacionais para a ação de forças microevolutivas: seleção, deriva genética e fluxo gênico. Os mecanismos de coesão podem ser divididos em dois tipos: os que delimitam o intercâmbio genético (sistemas de fertilização e de desenvolvimento e mecanismos de isolamento) e aqueles que delimitam o intercâmbio demográfico (fatores que delimitam o nicho fundamental e os limites de dispersão das novas variantes genéticas, como fatores ecológicos, históricos, genéticos, etc.).

Por ser uma combinação dos pontos mais bem sucedidos de outros conceitos, o conceito de coesão é teoricamente bastante flexível ao poder ser aplicado a uma ampla gama de organismos. Entretanto, algumas das dificuldades de implementação foram transferidas junto com as vantagens, como a determinação de quais organismos separados espacial ou temporalmente apresentariam “potencial para coesão fenotípica” ou quais seriam seus mecanismos de coesão, analogamente ao que ocorre com o reconhecimento de mecanismos de isolamento ou reconhecimento. Ou, então, se os mecanismos de intercâmbio demográfico estariam suficientemente bem determinados, analogamente ao que ocorre com a determinação do nicho fundamental do conceito ecológico.

3.5.2.6 A situação das espécies em *Acanthamoeba*

Classicamente, o conceito de espécie aplicado a *Acanthamoeba* pode ser identificado como morfológico com aspectos de um “ecológico brando”, uma vez que as espécies eram definidas tanto com base na forma e tamanho dos cistos como, por vezes, em algumas características do ambiente, como tolerância a diferentes temperaturas ou capacidade de causar patologias. Entretanto, conforme citado anteriormente, tais características são muito variáveis conforme o ambiente, com diferentes condições de cultura fazendo um mesmo isolado ter diversas formas e, conseqüentemente, de determinação muito subjetiva⁽¹¹⁸⁾.

Entretanto, apesar dos grandes avanços no conhecimento da organização genética e filogenia no gênero *Acanthamoeba*, surgidos com as técnicas de análise molecular, a determinação das espécies continuou baseada nos critérios antigos, embora os novos dados indicassem total inadequação da nomenclatura vigente.

Foi proposta, recentemente, uma mudança objetiva nos grupos incluídos como espécie no gênero⁽¹¹⁷⁾, de modo a refletir as relações filogenéticas entre os isolados, determinadas pela análise da sequência do SSU rDNA. Embora não seja explicitado no artigo, os autores parecem adotar um critério de espécie situado entre o evolutivo e o filogenético. Essa conclusão pode ser feita uma vez que consideram, como pertencentes a uma mesma espécie, os isolados que estejam incluídos em grupos monofiléticos e dentro de uma certa margem de distância evolutiva em relação a outros grupos, que seriam considerados outras espécies. Entretanto, pode-se notar como desvantagem a arbitrariedade envolvida no nível (distância ou ponto da árvore filogenética) escolhido como limite para se considerar dois grupos como espécies diferentes, e não como populações de uma mesma espécie, por exemplo.

Uma vez que, até o momento, não existe outro tipo de abordagem objetiva, bem desenvolvida e de aplicação prática conhecida para *Acanthamoeba*, esse foi o critério adotado no presente trabalho, apesar das limitações anteriormente citadas. Além disso, por envolver somente grupos naturais de *Acanthamoeba*, consideramos que essa estratégia representa um bom ponto de partida para outras abordagens futuras, possibilitando conclusões melhor fundamentadas e mais amplas acerca de fenômenos biológicos, ecológicos e epidemiológicos, que podem ser importantes em definições de espécie mais refinadas no gênero.

Conforme sugerido por diversos autores, provavelmente a enorme diversidade presente no mundo vivo não permita a escolha de um único conceito de espécie que possa ser aplicável na prática desde os vírus até os eucariotos mais complexos⁽⁸⁴⁾. De qualquer modo, seja qual for o conceito adotado, tem-se sempre de diferenciar a existência das características que definem a espécie (e são, portanto, determinadas e selecionadas pela aplicação do conceito) daquelas características meramente diagnósticas, que servem apenas para identificação, mas que podem causar confusão com o conceito morfológico ou fenético⁽¹⁰⁴⁾. Esse é mais um motivo para se manter sempre explícitos os critérios utilizados na definição das espécies, destacando-os bem das ferramentas de simples diagnóstico.

3.5.3 – Determinação do modelo evolutivo disponível mais adequado para o alinhamento do SSU rDNA

O cálculo da verossimilhança de uma árvore, considerados os 56 diferentes modelos de evolução de seqüência e nosso alinhamento, levou cerca de uma semana em um computador AMD K6-II 450 MHz, 56 Mb de RAM.

Foram feitas análises para todos os modelos de evolução molecular descritos no item 2.9.3 de Material e Métodos, assim como para outros casos especiais do modelo GTR (total de 14 modelos básicos), cada um deles também levando em consideração heterogeneidade de taxa entre sítios (correção gama) e/ou proporção de sítios invariáveis. A listagem de todos os modelos analisados e suas principais características encontram-se na tabela 1 do item Material e Métodos. As referências bibliográficas para os modelos que não foram anteriormente descritos na Introdução podem ser encontradas em Posada & Crandall, 1998. A análise de LRT hierárquico dos valores de verossimilhança obtidos foi feita pelo programa MODELTEST, resultando nas seguintes conclusões para os diversos testes de hipóteses feitos (com $P=0,05$ como limite e a hipótese aceita em cada teste em negrito):

Iguais frequências de base

Hipótese nula = JC $-\ln L_0 = 18645,6953$

Hipótese alternativa = F81 $-\ln L_1 = 18609,7715$

$2(\ln L_1 - \ln L_0) = 71,8477$ (graus de liberdade = 3)

$P < 0,000001$ (probabilidade de o valor 71,8477 ser obtido na comparação dos modelos caso a hipótese nula fosse a realmente correta)

Transições = Transversões

Hipótese nula = F81 $-\ln L_0 = 18609,7715$

Hipótese alternativa = HKY $-\ln L_1 = 18459,2246$

$2(\ln L_1 - \ln L_0) = 301,0938$ (gl = 1)

$P < 0,000001$

Iguais transições

Hipótese nula = HKY

 $-\ln L_0 = 18459,2246$ **Hipótese alternativa = TrN** $-\ln L_1 = 18416,1621$ $2(\ln L_1 - \ln L_0) = 86,1250$ $(g) = 1)$ $P < 0,000001$ Iguais transversões

Hipótese nula = TrN

 $-\ln L_0 = 18416,1621$ **Hipótese alternativa = TIM** $-\ln L_1 = 18411,7051$ $2(\ln L_1 - \ln L_0) = 8,9141$ $(g) = 1)$ $P = 0,002830$ Apenas duas transversões

Hipótese nula = TIM

 $-\ln L_0 = 18411,7051$ **Hipótese alternativa = GTR** $-\ln L_1 = 18375,9883$ $2(\ln L_1 - \ln L_0) = 71,4336$ $(g) = 2)$ $P < 0,000001$ Taxas iguais entre sítios

Hipótese nula = GTR

 $-\ln L_0 = 18375,9883$ **Hipótese alternativa = GTR+G** $-\ln L_1 = 16474,3516$ $2(\ln L_1 - \ln L_0) = 3803,2734$ $(g) = 1)$ $P < 0,000001$ Ausência de sítios invariáveis

Hipótese nula = GTR+G

 $-\ln L_0 = 16474,3516$ **Hipótese alternativa = GTR+I+G** $-\ln L_1 = 16296,1572$

$$2(\ln L_1 - \ln L_0) = 356,3887 \quad (gI = 1)$$

$$P < 0,000001$$

Modelo selecionado: GTR+I+G

Taxas de substituição estimadas (relativas e normalizadas em relação a $G \leftrightarrow T$):

$$A \leftrightarrow C = 0,7858 \quad A \leftrightarrow G = 1,8926$$

$$A \leftrightarrow T = 1,1687 \quad C \leftrightarrow G = 1,0008$$

$$C \leftrightarrow T = 3,6119 \quad G \leftrightarrow T = 1,0000$$

Frequências de base estimadas:

$$A = 0,2102 \quad C = 0,2135 \quad G = 0,2971 \quad T = 0,2792$$

Proporção estimada de sítios invariáveis (I) = 0,4769

Estimativa do parâmetro de forma da distribuição gama (α) = 0,4917

Conforme se pode observar pelos números acima, uma grande proporção dos sítios do SSU rDNA é invariável: observado 63,85%, estimado por máxima verossimilhança 47,69%. Levar esse fato em consideração eleva muito a razão da verossimilhança da árvore analisada ($\Delta \ln L = 356,39$), ou seja, torna significativamente mais eficiente a explicação da evolução das seqüências obtidas e alinhadas. Pode-se notar também, pelos valores de P sempre bem baixos (o maior foi $P = 0,002830$, para a comparação entre iguais e diferentes taxas para transversões), que cada complexidade adicional introduzida nos modelos de evolução eleva muito a eficiência da explicação dos padrões presentes em nosso alinhamento. Deste modo, nada nos garante que um modelo mais complexo do que o GTR+G+I, o mais detalhado à nossa disposição, não fosse também ser ainda mais eficiente. Convém salientar que tais testes de modelos nunca podem provar qual deles

teria sido o “verdadeiro” na evolução das seqüências, mas apenas comparam diferentes hipóteses de explicação, mostrando qual delas melhor se encaixa aos dados. Desenvolvimentos nos modelos de evolução molecular devem portanto levar a hipóteses ainda mais realistas, especialmente no estudo de moléculas complexas como os DNAs ribossômicos.

O valor de α , o parâmetro de forma da distribuição gama, indica que a maioria dos sítios apresenta baixas taxas de substituição, enquanto alguns poucos apresentam altas taxas. Esse parâmetro é inversamente proporcional a essa variabilidade: quando α é infinito, a variabilidade de taxa entre sítios é zero, e quanto menor o valor de α , maior a heterogeneidade. O ponto de mudança mais importante do α é o valor 1,0: quando $\alpha \leq 1$, a distribuição das taxas de substituição tem a forma de um L (tanto mais pronunciada quanto menor o α), ou seja, muitos sítios com baixas taxas e alguns poucos com altas taxas. Quando $\alpha > 1$, a distribuição assume a forma de sino, mais simétrica e mais fina, conforme o valor de α aumenta. Um gráfico ilustrativo desses comportamentos da distribuição gama, para diferentes valores de α , e maiores detalhes sobre a variação de taxas entre sítios podem ser encontrados em Yang, 1996a e Swofford e cols., 1996. A figura 18 mostra a distribuição das taxas de substituição para o α obtido nas análises do MODELTEST, sendo que o eixo horizontal representa a quantidade de sítios e o eixo vertical representa a taxa de substituição. A tabela com os limites para cada categoria da aproximação discreta da distribuição gama, realçada no gráfico por diferentes cores, mostra as taxas médias em cada uma e os valores mínimo e máximo, que efetivamente delimitam cada categoria.

Pode-se perceber que levar em consideração uma proporção de sítios invariáveis (ou seja, com taxa zero de substituição) deve aumentar o valor

estimado de α , uma vez que uma parte dos sítios com baixa taxa foi considerada como parte de um outro processo. Isso pôde ser claramente observado em nossas análises, uma vez que o valor de α obtido nas análises do MODELTEST para o modelo GTR com correção gama e com proporção de sítios invariantes (GTR+G+I) foi de 0,4917, enquanto o α para GTR+G, sem levar em consideração sítios invariantes, foi de 0,1714, indicando muito maior heterogeneidade de taxa do que quando se considerou uma proporção de sítios invariantes.

Limites e categorias de taxas para a distribuição gama discreta

----- limites -----			
categoria	inferior	superior	taxa (média)
1	0.00000000	0.09770002	0.03176852
2	0.09770002	0.44775425	0.24613220
3	0.44775425	1.31932616	0.81375822
4	1.31932616	infinito	2.90834106

Fig. 18 - Distribuição gama para o alinhamento utilizado, com $\alpha=0,4912$, obtida no programa PAUP*.

3.5.4 – Construção das árvores filogenéticas

Após os diversos passos citados anteriormente, como o alinhamento das seqüências, a escolha do modelo evolutivo mais bem sucedido em explicar os dados, o cálculo de distâncias entre as seqüências de acordo com esse modelo e diversas análises de procura da melhor árvore para os vários critérios, chegou-se a uma árvore consenso (Fig. 19) que pode ser considerada a melhor inferência disponível para os recursos utilizados. Pode-se notar, no que concerne à constituição dos tipos de seqüência do SSU rDNA, a alta concordância entre essa árvore e aquela obtida por Stothard e cols., 1998, exceto pela situação dos isolados 2, 3, 4 (SSU rDNA tipo T2) e 46 (T6): nesse caso, os isolados 4 e 46 estão reunidos no mesmo ramo, excluindo 2 e 3, o que tornaria o tipo T2 parafilético. Nosso isolado *A. poly.#4* (T14) sempre se agrupou com esses isolados (como grupo irmão de 46 ou de 4 e 46), embora o comprimento do ramo entre esse isolado e os outros indique que a divergência se deu há muito tempo ou que a seqüência do SSU rDNA de *A. poly.#4* evoluiu mais rapidamente do que a dos outros isolados filogeneticamente próximos (Figs. 20 a 22 e 25). Desse modo, o tipo de seqüência T2 seria inválido, uma vez que não forma um grupo monofilético, mas sim parafilético, conforme detalhado na seção 1.2.3 da Introdução (Fig. 6). Todos os outros tipos de seqüência estão com a mesma composição apresentada na literatura⁽¹¹⁷⁾, embora a relação filogenética entre tipos não esteja completamente concordante: o único grupo de tipos de seqüência que sempre se apresentou estável em todas as inferências, por diversos métodos, foi o agrupamento ((T3, T11), T4). Embora não tenham sido apresentados nas árvores presentes em Stothard e cols., 1998, os tipos T7, T8 e T9, utilizados para enraizar a árvore com todos os outros tipos, também permaneceram sempre juntos, na estrutura (T8, (T7, (T15, T9))), já incluindo o novo tipo T15 (*A. astronyxis* 30901). Conforme anteriormente discutido, o isolado *A. astronyxis*

30901 mostrou maior similaridade de seqüência com *A. comandoni* (tipo T7), e as análises filogenéticas confirmam a proximidade desses dois isolados, que sempre formaram um grupo monofilético em todas as análises.

A árvore consenso discutida acima foi feita com base em árvores inferidas por diversos métodos de naturezas muito diferentes, como métodos de distância algorítmicos (NJ e LogDet, ver abaixo), de optimalidade (mínimos quadrados), métodos de parcimônia e análises de máxima verossimilhança de todas as árvores encontradas que pareciam explicar igualmente bem os dados, seguidas do teste de Kishino-Hasegawa. Este teste estatístico permite a comparação de diferentes topologias com base no valor de verossimilhança das árvores avaliadas e suas variâncias⁽¹²¹⁾, diferentemente do LRT, anteriormente citado, que compara diferentes modelos de evolução em relação a uma topologia fixa.

Fig. 19 – Consenso estrito das árvores ótimas encontradas por diversos métodos de inferência filogenética, com os tipos de sequência representados por mais de um isolado destacados por cores. Os novos tipos caracterizados neste trabalho estão destacados em retângulos.

3.5.4.1 – Análises de distância

Além dos métodos de distância anteriormente citados, como NJ com distâncias GTR (Fig. 20) e mínimos quadrados (Fig. 21), foi feita também a análise do alinhamento pela distância LogDet ou paralinear^(68, 121). Este tipo de distância tem como diferencial em relação a outros métodos de inferência o fato de não depender da composição de bases aproximadamente igual entre os *taxa* analisados, da constância da matriz de substituição em toda a árvore e da reversibilidade do processo, ou seja, que as frequências de bases estão em equilíbrio. Além disso, as variâncias associadas às distâncias nesse método são bastante baixas, comparáveis às aquelas de métodos de distância mais simples. Entretanto, existe a desvantagem de não haver um modo direto de se incorporar a heterogeneidade de taxa de substituição ao longo da molécula, um fator muito significativo em moléculas como o SSU rDNA.

A árvore de LogDet calculada para nosso alinhamento pelo programa PAUP* está apresentada na figura 22, e foi feita utilizando como método de construção de árvore o NJ, que se mostra bastante eficiente quando as distâncias utilizadas são aditivas⁽¹²¹⁾, como geralmente é o caso para distâncias paralineares. As análises estatísticas (ML) dessa árvore em relação a árvores obtidas por outros métodos mostraram que a árvore de LogDet/NJ explica os dados igualmente bem, não podendo ser descartada (ver abaixo, seção 3.5.4.3).

A árvore de NJ utilizando distâncias calculadas de acordo com o modelo escolhido por LRT, conforme descrito anteriormente, está apresentada na figura 20, apresentando nos nós os valores de "bootstrap" para 1.000 réplicas. Esta árvore também pode ser considerada estatisticamente válida em relação às outras determinadas nesse trabalho.

A árvore de mínimos quadrados, um método de critério de optimalidade, foi calculada pelo programa FITCH do PHYLIP a partir da matriz de distâncias calculada pelo PAUP* para o modelo GTR+G+I, e está apresentada na figura 21. Conforme anteriormente citado, os métodos de optimalidade dependem fortemente dos métodos de procura de árvores ótimas no espaço de possibilidades (métodos descritos a seguir). Como o método de procura utilizado pelo PHYLIP (adição seqüencial) é o mais simples e, portanto, com menor chance de ter encontrado a melhor árvore para o critério de mínimos quadrados, não é possível saber se esta realmente é a árvore ótima. Para aumentar a chance de se encontrar uma árvore melhor para o critério em questão, foi ativada a opção de rearranjos globais do programa. Assim, depois da adição, em ordem aleatória, de todas as seqüências à árvore, o programa fez rearranjos semelhantes aos descritos abaixo, conhecidos como TBR ("tree bisection-reconnection"). Isso aumenta muito o tempo de análise, mas é a única forma de refinar a procura da árvore ótima no programa utilizado. A procura levou algumas horas e analisou cerca de 17 mil topologias até encontrar a árvore ótima, que mostrou topologia muito similar às obtidas por outros métodos.

Fig. 20 – Árvore de NJ, calculada utilizando distâncias GTR+G+I. Os números nos nós indicam o suporte de “bootstrap” para 1.000 replicatas. Os tipos de seqüência representados por mais de um isolado estão destacados por cores e os novos tipos caracterizados neste trabalho, destacados em retângulos. InL=16.316,17.

Fig. 22 - Árvore de NJ calculada a partir de distâncias LogDet. Os tipos de seqüência representados por mais de um isolado estão destacados por cores e os novos tipos caracterizados neste trabalho, destacados em retângulos. InL= 16.306,06.

3.5.4.2 – Procura de árvores ótimas por parcimônia

Apesar de terem sido analisadas diversas árvores filogenéticas diferentes por vários métodos de inferência e procura de árvores ótimas (detalhado abaixo), ainda assim esse número foi ínfimo perto do número teórico de possibilidades. Conforme anteriormente citado, o número de árvores possíveis aumenta muito rapidamente conforme o número de organismos analisados, como se pode notar na tabela a seguir⁽²³⁾:

Tabela 2 – O número de *taxa* e o correspondente número de possíveis árvores bifurcadas.

Número de <i>taxa</i>	Árvores <i>bifurcadas</i> possíveis
2	1 (10^0)
3	3 (10^0)
4	15 (10^1)
5	105 (10^2)
6	945 (10^3)
7	10.395 (10^4)
8	135.135 (10^5)
9	2.027.025 (10^6)
10	34.459.425 (10^7)
11	654.729.075 (10^8)
12	13.749.310.575 (10^{10})
13	316.234.143.225 (10^{11})
14	7.905.853.580.625 (10^{12})
15	213.458.046.676.875 (10^{14})
16	6.190.283.353.629.375 (10^{15})
17	191.898.783.962.510.625 (10^{17})
18	6.332.659.870.762.850.625 (10^{18})
19	221.643.095.476.699.771.875 (10^{20})
20	8.200.794.532.637.891.559.375 (10^{22})
21	319.830.986.772.877.770.815.625 (10^{23})
22	13.113.070.457.687.988.603.440.625 (10^{25})

Pela simples observação dos números acima, referentes apenas às árvores bifurcadas e com raiz, pode-se concluir que é inexeqüível a análise de todas as possibilidades quando se tem mais de nove ou dez *taxa*, dependendo da carga

computacional do método utilizado, pois o tempo de computação necessário seria impraticável.

No nosso caso, com 60 seqüências, existem cerca de $5,86 \times 10^{96}$ diferentes árvores com raiz ou $5,01 \times 10^{94}$ árvores sem raiz, de acordo com a seguinte fórmula⁽²³⁾:

$$T_{(N)} = \prod_{i=3}^N (2i - 5) \quad (17)$$

onde N é o número de organismos colocados na árvore. Essa fórmula calcula o número de possíveis árvores sem raiz ($T_{(N)}$). Como a raiz pode ser considerada como um “organismos a mais” na árvore, pelo menos para efeito de cálculos matemáticos, para se ter o número de possíveis árvores com raiz deve-se considerar N+1 no cálculo acima o que, em nosso caso, resultou em cerca de 100 vezes mais árvores enraizadas do que sem raiz.

Assim sendo, são necessários meios de fazer a procura das árvores ótimas sem examinar todas essas possibilidades. Um meio exato, mas que não analisa todas as árvores, é o “branch-and-bound” (BandB). Esse método garante a obtenção da melhor árvore possível para determinado critério, contanto que se saiba que o valor otimizado nunca diminui com a adição de mais *taxa*, o que é o caso para os critérios de optimalidade utilizados em filogenia⁽¹²¹⁾. O BandB funciona, inicialmente, adicionando seqüencialmente os organismos a todos os pontos possíveis de uma árvore bem simples. Só serão mantidas, para continuar as adições, as melhores árvores (pelo critério em questão), abandonando, portanto, os pontos onde não há chance de se obter o melhor valor possível. A velocidade (ou seja, número de árvores analisadas até chegar à solução final) desse método depende muito dos dados utilizados, uma vez que dados com pior relação sinal filogenético/ ruído terão tempos de análise muito maiores do que dados em que houver maior quantidade de informação filogenética. Além disso,

fica claro que a velocidade também será muito influenciada pela quantidade de seqüências analisadas e, em nosso caso, não foi possível fazer esse tipo de procura.

Métodos não-exatos, conhecidos como heurísticos, são portanto necessários, embora não possam garantir a obtenção da melhor árvore global. Há diversos tipos de métodos heurísticos, variando na complexidade envolvida e, portanto, no tempo gasto na procura, diretamente proporcional, sendo possível fazer análises combinando tais abordagens. O que têm em comum é o fato de serem métodos que utilizam a estratégia de "hill climbing" (subida de colina), ou seja, essas abordagens continuam manipulando determinada região do espaço de árvores somente enquanto melhore o valor do critério procurado, e nunca levam em consideração opções que diminuam este valor. Desse modo, é possível que a procura fique presa em um ponto de optimalidade local, impedindo que chegue ao máximo global, dependendo da árvore inicial utilizada e da extensão dos rearranjos tentados nessa árvore. Em geral, rearranjos maiores irão explorar pontos mais esparsos do espaço de árvores, aumentando a chance de que algum desses novos pontos de início seja parte da "colina" que leva à árvore de valor mais alto possível.

Normalmente, obtém-se uma árvore inicial que é depois "perturbada" sistematicamente de modo a tentar encontrar variantes que satisfaçam melhor o critério empregado. Tal árvore inicial pode ser uma árvore aleatória, pode ser fornecida pelo usuário ou obtida pelo programa por métodos como adição seqüencial (na ordem presente no alinhamento, aleatoriamente ou outros modos). Pode ser, também, uma "árvore razoável" - alguma árvore que já reflita, mesmo que de maneira apenas aproximada, como deve ser o relacionamento entre os *taxa* analisados, como uma árvore de NJ, por exemplo. Foram feitas, em nosso caso, cerca de 150 repetições de adição aleatória dos

isolados. As árvores vão sendo construídas pela adição das seqüências em diferentes ordens a cada réplica, sendo possíveis 60 ou cerca de 8×10^{81} diferentes ordens de adição para 60 seqüências, e são submetidas então às perturbações na topologia (descritas abaixo). Nessa fase, foram analisadas pelo critério de parcimônia cerca de 5 bilhões de diferentes topologias, sempre obtendo como melhor resultado os 2.720 passos.

Após obter a árvore inicial, foram então feitas as modificações em sua topologia. Tais modificações podem ser de vários tipos, sendo que o programa PAUP* (opção SWAP do comando HSEARCH) possibilita três variantes: NNI ("nearest-neighbor interchange"), SPR ("subtree pruning-regrafting") e TBR ("tree bisection-reconnection"), colocados aqui em ordem de complexidade. Na verdade, NNI é um subconjunto dos rearranjos SPR que por sua vez é um subconjunto dos TBR. A escolha do método depende do número de organismos e de árvores iniciais analisados, uma vez que o tempo de computação cresce razoavelmente conforme a complexidade do algoritmo ($NNI < SPR < TBR$). O NNI faz permutações entre braços próximos da árvore e, portanto, leva a rearranjos locais. Já o SPR retira uma sub-árvore e vai procurando um ponto onde ela possa ser reinserida, melhorando o valor do critério utilizado. Por fim, o TBR funciona cortando a árvore em duas sub-árvores, que serão religadas de todas as maneiras até encontrar um modo que melhore o valor do critério de optimalidade. Em qualquer dos casos, um rearranjo que não leve à melhoria nesse valor é abandonado e o processo continua em outro ramo ou sub-árvore. Figuras ilustrativas desses métodos e alguns detalhes adicionais podem ser encontrados em Swofford e cols., 1996.

Embora muitas tentativas de encontrar alguma árvore com menos passos tenham sido feitas, combinando esses métodos, a procura do resultado mais parcimonioso chegou a quase 36 mil árvores com 2.720 passos para explicar a

configuração atual dos 769 sítios informativos para parcimônia presentes no alinhamento. O consenso de maioria entre todas as árvores obtidas com o menor número de passos está apresentado na figura 23, e indica os ramos da filogenia que se mostram com maior dificuldade de resolução. Esses pontos são aqueles que apresentam baixos valores associados a eles (porcentagem das 36 mil árvores que concordam entre si), e são os responsáveis pelo grande número de árvores encontrado, uma vez que não há informação suficiente nos dados para definir, nesses ramos, uma só configuração com o menor número de passos.

Fig. 23 – Árvore consenso das cerca de 36 mil árvores de parcimônia com 2.720 passos. Os números nos nós indicam a porcentagem dessas árvores que possuem o respectivo grupamento. Os tipos de seqüência representados por mais de um isolado estão destacados por cores e os novos tipos caracterizados neste trabalho, destacados em retângulos. InL= 16.284,64.

3.5.4.3 – Análises de verossimilhança das árvores obtidas e procura da árvore de ML para um alinhamento parcial

Essa dificuldade de resolução em algumas áreas da árvore pode ser devida, principalmente, a uma grande proximidade filogenética entre os isolados de posição pouco clara ou então a momentos de divergências muito semelhantes entre ramos maiores. Em ambos os casos, devido ao pouco tempo decorrido entre as separações de isolados, seria difícil ter divergência de sequência suficientemente significativa nesses curtos espaços de tempo para poder resolver, com segurança, as relações filogenéticas envolvidas. Em casos como esses pode ser necessária a escolha de algum outro gene com uma taxa de evolução um pouco mais elevada, mas não tanto que a relação entre sinal filogenético e ruído (devido a múltiplas substituições e convergências) seja deteriorada a ponto de também impedir uma conclusão segura. Outra possibilidade é a análise parcial do alinhamento, retirando a maior parte dos ramos bem resolvidos, de modo a diminuir consideravelmente o número de seqüências envolvidas. Desse modo, é possível fazer análises muito mais profundas dos possíveis arranjos de árvores, relacionando as seqüências que sobraram, podendo-se utilizar, também, métodos mais poderosos de análise. Uma desvantagem desse método é a possibilidade de introduzir erros de amostragem (ou seja, retirar seqüências importantes para a definição de ramos intermediários), que podem até mesmo piorar a definição dos grupamentos procurados e de outros anteriormente bem resolvidos, conforme discutido na literatura.

Ainda assim, foram feitas análises de parte dos organismos dos grupos T3, T11 e T4 apenas, sendo que os dois primeiros foram considerados como grupo externo do último, de modo a tentar resolver melhor as relações dentro do grupo T4. Foram consideradas 30 seqüências nessas análises, retirando a maioria

daquelas cuja relação filogenética já estava bem estabelecida em análises anteriores. Isso possibilitou a utilização de máxima verossimilhança com procura de árvores por "quartet puzzling"⁽¹¹⁹⁾, considerando o modelo TrN com correção gama. A árvore obtida está apresentada na figura 24, com os valores de "bootstrap" correspondentes a cada ramo obtidos por NJ, utilizando as distâncias calculadas pelas fórmulas do método de TrN+G+I (escolhido pelo MODELTEST, como descrito anteriormente). Conforme se pode notar, não foi possível resolver adequadamente a maioria das relações filogenéticas entre os isolados do grupo T4 com a utilização da seqüência do SSU rDNA, o que indica que outro gene, talvez de evolução mais rápida, possa ser de maior utilidade se houver interesse em se determinar as relações dentro do grupo. Uma indicação de que o motivo da falha, na resolução dessas relações, pode mesmo ser a falta de variabilidade entre as seqüências dos isolados do grupo T4 são as categorias de taxas relativas estimadas pelo TreePuzzle para a distribuição gama: das oito categorias, cinco têm taxa igual a zero; uma igual a 0,0003; uma igual a 0,0733; e a última igual a 7,9264. Isso significa que quase todos os sítios são invariáveis, enquanto alguns poucos (os da categoria 7,9264) apresentam muita variação. Conforme discutido anteriormente, esses dados são resumidos pelo valor de α que, no caso, foi estimado em 0,02, um valor muito baixo (ou seja, grande heterogeneidade de taxa de substituição). Quando foi considerada uma proporção de sítios invariantes (pelo método do MODELTEST no PAUP*, conforme anteriormente descrito), o valor de α aumentou muito, passando a ser 0,43, uma vez que boa parte dos sítios de evolução lenta foi considerada pela proporção de invariantes (0,79). Entretanto, apesar da pouca resolução dos dados para a distinção dos grupos internos do tipo T4, é possível perceber que o grupo, como um todo, é mesmo coeso filogeneticamente e bem apoiado por essa seqüência,

uma vez que apresentou 100% de “bootstrap” em relação aos grupos externos (T3 e T11).

Fig. 24 – Árvore de máxima verossimilhança encontrada pelo programa TreePuzzle para parte do alinhamento envolvendo isolados dos tipos T3, T4 e T11. O modelo de evolução molecular utilizado foi TrN+G+I. Os números nos nós indicam o suporte de “bootstrap” para 1.000 replicatas.

O teste de árvores obtidas por diferentes métodos pode ser feito utilizando a verossimilhança como critério. Para tanto, faz-se o cálculo da verossimilhança para cada uma das árvores obtidas, de acordo com o modelo de evolução escolhido como o que melhor explica os padrões no alinhamento (em nosso caso, GTR+G+I). Com os valores obtidos é possível fazer testes estatísticos que

indiquem qual árvore é a melhor por esse critério, ou se mais de uma árvore explica igualmente bem os dados. O teste de Kishino-Hasegawa foi feito com as verossimilhanças das seguintes árvores: 1) consenso das cerca de 36 mil árvores com 2.720 passos; 2 a 6) cinco árvores de parcimônia com 2.720 passos (cada uma delas representando uma “colina” (ou “ilha”, conforme denominado pelo PAUP*) diferente de árvores); 7) a árvore de LogDet/NJ; 8) a árvore de máxima verossimilhança encontrada pelo DNAML do PHYLIP; 9) a árvore de NJ com distâncias GTR+G+I; e 10) a árvore de mínimos quadrados encontrada pelo FITCH do PHYLIP.

Interessantemente, as cinco árvores de parcimônia foram muito piores do que a árvore consenso das 36 mil árvores mais parcimoniosas. Já as árvores DNAML, GTR+G+I/NJ e LogDet/NJ se mostraram estatisticamente iguais à árvore consenso, conforme pode ser visto nos números da tabela 3 abaixo:

Tabela 3 – Resultado do teste Kishino-Hasegawa (discriminação das árvores no texto acima)

Árvore	-ln L	Δ -ln L	s.d. (Δ)	T	P
1	16284.64100	(melhor)			
2	22249.93498	5966.22981	260.41835	22.9102	<0.0001*
3	22684.84918	6401.14401	284.40572	22.5071	<0.0001*
4	22568.78988	6285.08472	273.59117	22.9725	<0.0001*
5	22689.71235	6406.00719	284.76748	22.4956	<0.0001*
6	22706.37159	6422.66643	286.44326	22.4221	<0.0001*
7	16306.05563	22.35047	26.01152	0.8593	0.3903
8	16330.84045	47.13529	29.35753	1.6056	0.1085
9	16316.17480	32.46964	28.18497	1.1520	0.2494
10	16326.43935	24.48356	31.84879	0.7687	0.4421

* $P < 0.05$, árvores rejeitadas em comparação com a melhor

Utilizando-se um método mais simples de procura da melhor alternativa no espaço de árvores é possível se fazer uma estimativa da árvore de máxima

verossimilhança para o alinhamento completo (60 seqüências). Essa procura, em nosso caso, foi feita utilizando o programa DNAML do PHYLIP, por adição seqüencial em ordem aleatória, e o resultado está apresentado na figura 25. Nesse método, começa-se com uma árvore de três seqüências, e as restantes vão sendo adicionadas a todos os pontos possíveis, sendo mantida a árvore que melhor satisfizer o critério utilizado (no caso, ML). O ciclo continua até que todas as seqüências estejam colocadas na árvore, e são examinadas muito menos topologias do que por outros métodos. No nosso caso, o programa examinou cerca de 8.300 diferentes árvores, a maior parte delas com poucos *taxa*, tendo levado cerca de 20 horas para completar a tarefa. A desvantagem desta abordagem é sua dependência em relação à ordem de adição das seqüências, aumentando muito a chance de se escolher uma árvore localizada em um ponto ótimo apenas local. Uma maneira de se diminuir esse efeito no programa utilizado seria ativar a opção de rearranjos globais, mas isso aumenta muito o tempo de processamento, impossibilitando a análise de um alinhamento com muitas seqüências como o nosso.

Fig. 25 – Árvore de máxima verossimilhança encontrada pelo programa DNAML. Os tipos de sequência representados por mais de um isolado estão destacados por cores e os novos tipos caracterizados neste trabalho destacados em retângulos. InL= 16.330,84.

O consenso estrito entre todas as árvores que explicaram igualmente bem os dados foi calculado pelo PAUP* e está apresentado na figura 19, evidenciando os grupos que aparecem em todas as árvores. Apesar de estar no formato radial, esse consenso também poderia ter sido enraizado no ponto de inserção dos isolados T7, T8, T9 e T15 (*A. astronyxis* 30901), como as outras árvores. Além de, conforme anteriormente mencionado, sempre formarem um grupo monofilético bem apoiado pelos dados, as seqüências desses isolados são as mais divergentes em relação às dos outros presentes no alinhamento, chegando a 37,98% (entre alguns isolados T5 e T9, ver Fig. 16), o que levou à proposta de separação destes isolados, pertencentes ao grupo morfológico I, em um ou mais gêneros irmãos de *Acanthamoeba*⁽¹¹⁷⁾.

Conclusões

4 – CONCLUSÕES

A utilização de métodos moleculares como RAPD, RFLP (tanto do DNA genômico total quanto do SSU rDNA) e seqüenciamento do SSU rDNA, bem como a comparação dos resultados obtidos com dados morfológicos, permitiram a caracterização, identificação e o estabelecimento das relações filogenéticas dentro do isolados do gênero *Acanthamoeba*, provenientes da ATCC ou isolados de casos brasileiros de úlcera de córnea.

Foi possível confirmar os dados da literatura que freqüentemente indicaram, principalmente por diversos métodos moleculares, a inconsistência generalizada na taxonomia específica, classicamente feita por métodos morfológicos, desse grupo de amebas de vida livre. Assim, pôde-se concluir que a morfologia em *Acanthamoeba* não pode, de maneira alguma, ser utilizada como único, ou mesmo principal, critério na classificação ou identificação de isolados. Isso porque a grande variabilidade e a falta de correlação suficiente entre a morfologia e a posição filogenética podem, muito freqüentemente, levar a conclusões errôneas quanto à classificação. Esse fato pode ter enorme influência negativa em estudos sobre a epidemiologia ou a ecologia do gênero, uma vez que impediria conclusões corretas em qualquer dessas áreas em que a determinação da espécie é essencial para se saber quais isolados causam determinados tipos de patologias, ou quais deles são mais comuns em diferentes tipos de ambiente, etc..

Um novo critério para a classificação das espécies do gênero foi proposto recentemente⁽¹¹⁷⁾, utilizando a divergência não corrigida da seqüência do SSU rDNA como parâmetro de classificação: cada tipo de seqüência, proposto como equivalente a uma espécie em *Acanthamoeba*, teria, pelo menos, 5% de diferença em relação aos outros tipos, além de tornar possível também a análise do relacionamento filogenético entre os isolados. Desse modo, foram definidos

doze tipos de seqüência do SSU rDNA em *Acanthamoeba*, sendo que o tipo T4 foi correlacionado fortemente com a ocorrência de ceratite: apenas um isolado de ceratite, em 25, não se mostrou pertencente a esse tipo. O seqüenciamento do SSU rDNA dos 14 isolados brasileiros de ceratite e dos oito isolados provenientes da ATCC (4 deles de ceratite) fortaleceu ainda mais a associação entre tipo de seqüência e patogenicidade, uma vez que 16 das 17 novas seqüências puderam ser classificadas como T4. Entretanto, nossos resultados mostraram também que essa correlação não pode ser considerada absoluta, uma vez que um dos isolados brasileiros de úlcera de córnea, U/HC1, apresenta um tipo de seqüência (provisoriamente chamado T13) bastante diferente de todos os outros. Dos 4 isolados de ambiente analisados, provenientes da ATCC, dois foram classificados como T4 (*A. poly.*#2 e *A. cast.*#2), enquanto outros dois não se encaixaram em nenhum dos outros tipos previamente estabelecidos. Para esses propomos, portanto, dois novos tipos de seqüência: T14 (*A. poly.*#4) e T15 (*A. astronyxis*). Embora o isolado *A. astronyxis* 30901 possa não mais ser considerado como pertencente à espécie em que foi inicialmente classificado, ainda assim ele continua filogeneticamente muito ligado aos outros isolados do grupo morfológico I (*A. comandoni*, *A. tubiashi* e *A. astronyxis*), uma vez que apresentou 6,45% de divergência em relação a *A. comandoni* (tipo T7). Isso sugere que o grupo morfológico I é consistente morfológica, genética e filogeneticamente, sendo que a divergência dos isolados desse grupo, em relação a todos os outros do gênero *Acanthamoeba*, é tanta que foi sugerida a possibilidade de separar o grupo I em um ou mais gêneros irmãos a *Acanthamoeba*⁽¹¹⁷⁾. No entanto, ainda falta uma maior quantidade de informações para justificar tal mudança, embora se possa argumentar que seu nível de arbitrariedade seja exatamente o mesmo da decisão que envolve a delimitação dos tipos de seqüência.

Referências Bibliográficas

5 – REFERÊNCIAS BIBLIOGRÁFICAS

1. AMORIM, D. S. (1997) **Elementos básicos de sistemática filogenética**. 2ª ed., Holos Editora e Sociedade Brasileira de Entomologia, Ribeirão Preto.
2. ARMSTRONG, J., GIBBS, A., PEAKALL, R. & WELLER, G. (1994) **RAPDistance Programs, Version 1.03 for the analysis of patterns of RAPD fragments**. Australian National University, Canberra.
3. AVERY, O. T., McLEOD, C. M. & McCARTY, M. (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types. **J. Exp. Med**, **79(2)**: 137-158.
4. AVISE, J. C. & WOLLENBERG, K. (1997) Phylogenetics and the origin of species. **Proc. Natl. Acad. Sci. USA**, **94(15)**: 7748-7755.
5. BADENOCH, P. R., ADAMS, M. & COSTER, D. J. (1995) Corneal virulence, cytopathic effect on human keratocytes and genetic characterization of *Acanthamoeba*. **Int. J. Parasitol.**, **25**: 229- 239.
6. BOGLER, S. A., ZARLEY, C. D., BURIANEK, L. L., FUERST, P. A. & BYERS, T. J. (1983) Interstrain mitochondrial DNA polymorphism detected in *Acanthamoeba* by restriction endonucleases analysis. **Mol. Biochem. Parasitol.**, **8**: 145-163.
7. BYERS, T. J., BOGLER, S. A. & BURIANEK, L. L. (1983) Analysis of mitochondrial DNA variation as an approach to systematic relationships in the genus *Acanthamoeba*. **J. Protozool.**, **30**:198-203.
8. BYERS, T. J., HUGO, E. R. & STEWART, V. J. (1990) Genes of *Acanthamoeba*: DNA, RNA and protein sequences (a review). **J. Protozool.**, **37**:17S-25S.
9. CALORE, E. E., CAVALIERE, M. J. & CALORE, N. M. (1997) Cerebral amebiasis in the acquired immunodeficiency syndrome. **Acta Neurol. Belg.**, **97**: 248-250.
10. CASTELLANI, A. (1930) An amoeba found in cultures of yeast: preliminary note, **J. Trop. Med. Hyg.**, **33**:160.
11. CASTRO, C., HERNÁNDEZ, R. & CASTAÑEDA, M. (1981) *Trypanosoma cruzi* ribosomal RNA: internal break in the large molecular-mass species and number of genes. **Mol. Biochem. Parasitol.**, **2**:219-233.
12. CHALWATZIS, N., BAUR, A., STETZER, E., KINZELBACH, R. & ZIMMERMAN, F.K. (1995) Strongly expanded 18S rRNA genes correlated with a peculiar morphology in the insect order of Strepsiptera. **Zoology**, **98**:115-126.
13. CLARK, C. G. & DIAMOND, L. S. (1994) Pathogenicity, virulence and *Entamoeba histolytica*. **Parasitol. Today**, **10**: 46-47.
14. COSTAS, M., EDWARDS, S. W., LLOYD, D., GRIFFITHS, A. J. & TURNER, G. (1983) Restriction enzyme analysis of mitochondrial DNA of members of the genus *Acanthamoeba* as an aid in taxonomy. **FEMS Microbiol. Letters**, **17**:231-234.
15. COSTAS, M. & GRIFFITHS, A. J. (1985) Enzyme composition and the taxonomy of *Acanthamoeba*. **J. Protozool.**, **32**:604-607.
16. CULBERTSON, C. G. (1958) *Acanthamoeba*: observations on animal pathogenicity, **Science**, **127**:1506.
17. CURSONS, R. T. M., BROWN, T. J., KEYES, E. A., MORIARITY, K. M. & TILL, D. (1980) Immunity to pathogenic free-living amoebae: role of humoral antibody. **Infect. Immun.**, **29**:401.
18. DAGGETT, P., LIPSCOMB, D., SAWYER, T. K. & NERAD, T. A. (1985) A molecular approach to the phylogeny of *Acanthamoeba*. **BioSystems**, **18**:399-405.

19. D'ALESSIO, J. M., HARRIS, G. H., PERNA, P. J. & PAULE, M. R. (1981) Ribosomal nucleic acid repeat unit of *Acanthamoeba castellanii*: cloning and restriction endonuclease map. **Biochemistry**, **20**:3822-3827.
20. DARWIN, C. R. (1859) **On the origin of species by means of natural selection**. J. Murray, London.
21. DE JONCKHEERE, J. F. (1987) Characterization of Naegleria species by restriction endonuclease digestion of whole-cell DNA. **Mol. Biochem. Parasitol.**, **24**: 55-66.
22. DE JONCKHEERE, J. F. (1991) Ecology of *Acanthamoeba*. **Rev. Infect. Dis.**, **13**: S385-S387.
23. FELSENSTEIN, J. (1978a) The number of evolutionary trees. **Syst. Zool.**, **27**: 27-33.
24. FELSENSTEIN, J. (1978b) Cases in which parsimony and compatibility methods will be positively misleading. **Syst. Zool.**, **27**: 401-410.
25. FELSENSTEIN, J. (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. **J. Mol. Evol.**, **17**: 368-376.
26. FELSENSTEIN, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. **Evolution**, **39**(4): 783-791.
27. FELSENSTEIN, J. (1988) Phylogenies from molecular sequences: inference and reliability. **Annu. Rev. Gen.**, **22**: 521-565.
28. FELSENSTEIN, J. (1989) **Phylip 3.2 Manual**, University of California, Herbarium, Berkeley, CA.
29. FERRANTE, A. & BATES, E. J. (1988) Elastase in the pathogenic free-living amoebae *Naegleria* and *Acanthamoeba* spp. **Infect. Immun.**, **56**: 3320-3321.
30. FERRANTE, A. (1991) Free-living amoebae: pathogenicity and immunity. **Parasite Immunol.**, **13**:31-47.
31. FICKER, L., SEAL, D., WARHURST, D. & WRIGHT, P. (1990) *Acanthamoeba* keratitis: resistance to medical therapy. **Eye**, **4**: 835-838, 1990.
32. FOREY, P. L., HUMPHRIES, C. J., KITCHING, I. L., SCOTLAND, R. W., SIEBERT, D. J. & WILLIAMS, D. M. (1992) **Cladistics, a practical course in systematics**. Oxford University Press, Oxford.
33. FRITSCH, T. R., HORN, M., WAGNER, M. HERWIG, R. P., SCHLEIFER, K. H. & GAUTOM, R. K. (2000) Phylogenetic diversity among geographically dispersed Chlamydiales endosymbionts recovered from clinical and environmental isolates of *Acanthamoeba* spp. **Appl. Environ. Microbiol.**, **66**(6): 2613-2619.
34. GAETH, A. P., SHORT, R. V. & RENFREE, M. B. (1999) The developing renal, reproductive, and respiratory systems of the African elephant suggest an aquatic ancestry. **Proc. Natl. Acad. Sci. USA**, **96**(5): 5555-5558.
35. GARDNER, H. A., MARTINEZ, A. J., VISVESVARA, G. S. & SOTREL, A. (1991) Granulomatous amebic encephalitis in an AIDS patient. **Neurology**, **41**: 1993-1995.
36. GAST, R. J., FUERST, P. A. & BYERS, T. J. (1994) Discovery of group I introns in the nuclear small subunit ribosomal RNA genes of *Acanthamoeba*. **Nucl. Acids Res.**, **22**:592-596.
37. GAST, R. J., LEDEE, D. R., FUERST, P. A. & BYERS, T. J. (1996) Subgenus systematics of *Acanthamoeba*: four nuclear 18S rDNA sequence types. **J. Euk. Microbiol.**, **43**:498-504.
38. GAUTON, R. K., LORY, S., SEYEDIRASHTI, S., BERGERON, D. L. & FRITSCH, T. R. (1994) Mitochondrial DNA fingerprinting of *Acanthamoeba* spp. isolated from clinical and environmental sources. **J. Clin. Microbiol.**, **32**(4): 1070-1073.
39. GOLDMAN, N. (1993a) Statistical tests of models of DNA substitution. **J. Mol. Evol.**, **36**: 182-198.

40. GOLDMAN, N. (1993b) Simple diagnostic statistical tests of models for DNA substitution. **J. Mol. Evol.**, **37**: 650-661.
41. GOULD, S. J. (1977) **Ontogeny and Phylogeny**. Harvard University Press, Belknap Press, Cambridge, Massachusetts.
42. GRAY, M. W. (1979) The ribosomal RNA of trypanosomatid protozoa *Crithidia fasciculata*: physical characteristics and methylated sequences. **Can. J. Biochem.**, **57**:914-926.
43. GUNDERSON, J. H. & SOGIN, M. L. (1986) Length variation in eukaryotic rRNAs: small subunit rRNAs from the protists *Acanthamoeba castellanii* and *Euglena gracilis*. **Gene**, **44**:63-70.
44. HAECKEL, E. (1866) **Generelle Morphologie der Organismen: Allgemeine Grundzüge der organischen Formen-Wissenschaft, mechanisch begründet durch die von Charles Darwin reformirte Descendenz-Theorie**. 2 vols. Georg Reiner, Berlin.
45. HAMBURG, A. & DE JONCKHEERE, J. F. (1980) Amoebic keratitis. **Ophthalmologica**, **181**: 74.
46. HASEGAWA, M., KISHINO, H. & YANO, T. (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. **J. Mol. Evol.**, **22**: 160-174.
47. HILLIS, D. M. & DIXON, M. T. (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. **Q. Rev. Biol.**, **66**, 411-446.
48. HILLIS, D. M., BULL, J. J., WHITE, M. E. BADGETT, M. R. & MOLINEUX, I. J. (1992) Experimental phylogenetics: generation of a known phylogeny. **Science**, **255**: 589-592.
49. HILLIS, D. M. & BULL, J. J. (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. **Syst. Biol.**, **42(2)**: 182-192.
50. HILLIS, D. M. (1994) Application and accuracy of molecular phylogenies. **Science**, **264**: 671-677.
51. HILLIS, D. M. (1995) Approaches for assessing phylogenetic accuracy. **Syst. Biol.**, **44(1)**: 3-16.
52. HUELSENBECK, J. P. (1995) Performance of phylogenetic methods in simulation. **Syst. Biol.**, **44(1)**: 17-48.
53. HUELSENBECK, J. P. & CRANDALL, K. A. (1997) Phylogeny estimation and hypothesis testing using maximum likelihood. **Annu. Rev. Ecol. Syst.**, **28**: 437-466.
54. HULL, D. L. (1967) Certainty and circularity in taxonomy. **Evolution**, **21**: 174-189.
55. ILLINGWORTH, C. D. & COOK, S. D. (1998) *Acanthamoeba* keratitis. **Survey Ophthalmol.**, **42(6)**: 493-508.
56. JACCARD, P. (1901) Etude comparative de la distribution florale dans une portion des Alpes et des Jura. **Boll. Soc. Vaudoise Sci. Nat.**, **44**: 223-270.
57. JACOBSON, L. M. & BAND, R. N. (1987) Genetic heterogeneity in a natural population of *Acanthamoeba polyphaga* from soil, an isoenzyme analysis. **J. Protozool.**, **34**:83-86.
58. JOHNSON, A. M., FIELKE, R., CHRISTY, P. E. & BAVERSTOCK, P. R. (1990) Small subunit ribosomal RNA evolution in the genus *Acanthamoeba*. **J. Gen. Microbiol.**, **136**:1689-1698.
59. JONES, D. B. (1986) *Acanthamoeba*: the ultimate opportunist? **Am. J. Ophthalmol.**, **Oct.**: 527-530.
60. JONES, D. B., VISVESVARA, G. S. & ROBINSON, N. M. (1975) *Acanthamoeba polyphaga* keratitis and *Acanthamoeba* uveitis associated with a fatal meningoencephalitis. **Trans. Ophthalmol. Soc. UK**, **95**: 221-232.

61. JUKES, T.H. & CANTOR, C. R. (1969) Evolution of protein molecules. In H. N. Munro (ed.) **Mammalian protein metabolism**. Academic Press, New York.
- Clin. Microbiol., 29:**310-314.
62. KILVINGTON, S., LARKIN, D. F. P., WHITE, D. G. & BEECHING, J. R. (1990) Laboratory investigation of *Acanthamoeba* keratitis. **J. Clin. Microbiol., 28:**2722-2725.
63. KILVINGTON, S., BEECHING, J. R. & WHITE, D. G. (1991) Differentiation of *Acanthamoeba* strains from infected corneas and the environment by using restriction endonuclease digestion of whole-cell DNA. **J.**
64. KILVINGTON, S. & WHITE, D. G. (1994) *Acanthamoeba*: biology, ecology and human disease. **R. Med. Microbiol., 5:** 12-20.
65. KIMURA, M. (1980) A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. **J. Mol. Evol., 16:** 111-120.
66. KINGSTON, D. & WARHURST, D. C. (1969) Isolation of amoebas from the air. **J. Med. Microbiol., 2:** 27-36.
67. KONG, H. H. & CHUNG, D. I. (1996) PCR and RFLP variation of conserved region of small subunit ribosomal DNA among *Acanthamoeba* isolates assigned to either *A. castellanii* or *A. polyphaga*. **Kor. J. Parasitol., 34:** 127-134.
68. LAKE, J. A. (1994) Reconstructing evolutionary trees from DNA and protein sequences: paralinear distances. **Proc. Natl. Acad. Sci. USA, 91:** 1455-1459.
69. LEDEE, D. R., HAY, J., BYERS, T. J., SEAL, D. V. & KIRKNESS, C. M. (1996) *Acanthamoeba griffini*. Molecular characterization of a new corneal pathogen. **Invest. Ophthalmol. Vis. Sci. 37:**544-550.
70. LEE, M. S. Y. (1999) Circularity, evolution, systematics... and circularity. **J. Evol. Biol., 12(4):** 724-734.
71. LEHMANN, O. J., GREEN, S. M., MORLET, N., KILVINGTON, S., KEYS, M. F., MATHESON, M. M., DART, J. K. G., MCGILL, J. I. & WATT, P. J. (1998) Polymerase chain reaction analysis of corneal epithelial and tear samples in the diagnosis of *Acanthamoeba* keratitis. **Invest. Ophthalmol. Vis. Sci, 39(7):** 1261-1265.
72. LI, W. H. & GRAUR, D. (1991) **Fundamentals of molecular evolution**. Sinauer Associates, Snderland, Massachusetts.
73. LI, W.H. & ZHARKIKH, A., (1995) Statistical tests of DNA phylogenies. **Syst. Biol., 44(1):** 49-63.
74. LINDQUIST, T. M. (1998) Treatment of *Acanthamoeba* keratitis. **Cornea, 17(1):** 11-16.
75. LIÒ, P. & GOLDMAN, N. (1998) Models of molecular evolution and phylogeny. **Genome Res., 8:** 1233-1244.
76. LOGAR, J. & KRAUT, A. (1997) *Acanthamoeba* corneal infection in a contact lens wearer. **J. Infect., 35:** 237-240.
77. LONERGAN, K. M. & GRAY, M. W. (1994) The ribosomal RNA gene region in *Acanthamoeba castellanii* mitochondrial DNA - a case of evolutionary transfer of introns between mitochondria and plastids? **J. Mol. Biol., 239:**476-499.
78. MA, P., VISVESVARA, G. S., MARTINEZ, A. J., THEODORE, F. H., DAGGET, P. M. & SAWYER, T. K. (1990) *Naegleria* and *Acanthamoeba* infections: review. **Rev. Infect. Dis., 12:** 490-513.
79. MARCIANO-CABRAL, F., PUFFENBARGER, R. & CABRAL, G. A. (2000) The increasing importance of *Acanthamoeba* infections. **J. Eukaryot. Microbiol., 47(1):** 29-36.
80. MARTÍNEZ, A. J. (1985) **Free-living amebas: natural history, prevention, diagnosis, pathology, and treatment of disease**. CRC Press, Inc.: 1-156.

81. MARTÍNEZ-PALOMO, A. (1986) **Amebiasis**. eds.: RUITENBER, E. J. & MACINNIS, A. J., Elsevier, p.1-269.
82. MATIAS, R., SCHOTTELIUS, J., RADDATZ, C. F. & MICHEL, R. (1991) Species identification and characterization of an *Acanthamoeba* strain from human cornea. **Parasitol. Res.**, **77**:469-474.
83. MATSUNAGA, S., ENDO, T., YAGITA, K., HIRUKAWA, Y., TOMINO, S., MATSUGO, S. & TSURUHARA, T. (1998) Chromosome size polymorphisms in the genus *Acanthamoeba* electrokaryotype by pulsed-field gel electrophoresis. **Protist**, **149(4)**: 323-340.
84. MAYR, E. (1982) **The growth of biological thought: diversity, evolution and inheritance**. Harvard University Press, Cambridge, Massachusetts.
85. McKEY, R. M. & DOOLITTLE, W. F. (1981) Nucleotide sequences of *Acanthamoeba castellanii* 5S and 5.8S ribosomal ribonucleic acids: phylogenetic and comparative structural analyses. **Nucl. Acids Res.** **9**:3321-3334.
86. McLAUGHLIN, G. L., BRANDT, F. H. & VISVESVARA, G. S. (1988) Restriction fragment length polymorphisms of the DNA of selected *Naegleria* and *Acanthamoeba* amebae. **J. Clin. Microbiol.**, **26**:1655-1658.
87. MERGERYAN, H. (1991) The prevalence of *Acanthamoeba* in the human environment. **Rev. Infect. Dis.**, **13(Suppl. 5)**: S390-S391.
88. MITRO, K. (1994) Partial characterization of the proteolytic secretions of *Acanthamoeba polyphaga*. **Exp. Parasitol.**, **78**: 377-385.
89. MIYAMOTO, M. M. & FITCH, W. M. (1995) Testing phylogenies and phylogenetic methods with congruence. **Syst. Biol.**, **44(1)**: 64-76.
90. MOORE, M. B., UBELAKER, J. E., MARTIN, J. H., SILVANY, R., DOUGHERTY, J. M., MEYER, D. R., McCULLEY, J. P. (1991) In vitro penetration of human corneal epithelium by *Acanthamoeba castellanii*: a scanning and transmission electron microscopy study. **Cornea**, **10**: 291-298.
91. NAGINGTON, J., WATSON, P. G., PLAYFAIR, T. J., MCGILL, J., JONES, B. R. & STEELE, A. D. M. (1974) Amoebic infection of the eye. **Lancet**, **2**: 1537-1540.
92. NEFF, R. J., RAY, S. A., BENTON, W. F. & WILBORN, M. Induction of synchronous encystment (differentiation) in *Acanthamoeba* sp. In: PRESCOTT, D.M., **Methods in Cell Physiology**, Academic Press, New York, vol. 1, p.55-83, 1964.
93. NICHOLAS, K. B., NICHOLAS, Jr., H. B. & DEERFIELD, D. W. II. (1997) GeneDoc: Analysis and Visualization of Genetic Variation, **EMBNEW.NEWS** **4**:14
94. NIKAIDO, M., ROONEY, A. P. & OKADA, N. (1999) Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: hippopotamuses are the closest extant relatives of whales. **Proc. Natl. Acad. Sci. USA**, **96(8)**: 10261-10266.
95. OLSEN, G. J. (1988) Phylogenetic analysis using ribosomal RNA. **Meth. Enzymol.**, **164**: 793-812.
96. OLSEN, G. J. & WOESE, C. R. (1993) Ribosomal RNA: a key to phylogeny. **The FASEB Journal**, **7**:113-123.
97. PAGE, F. C. (1988) **A new key to freshwater and soil gymnamoebae with instructions for culture**. Freshwater Biological Association, Ambleside, Cumbria, UK.
98. PAGE, R. D. M. (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. **CABIOS** **12**: 357-358.

99. PELLEGRIN, J. L. J., ORTEGA-BARRIA, E., BARZA, M. BAUM, J. & PEREIRA, M .E. A. (1991) Neuraminidase Activity in *Acanthamoeba* Species Trophozoites and Cysts. **Inv. Ophthalmol. Vis. Sci.**, **32(12)**: 3061-3066.
100. POSADA, D. & CRANDALL, K. A. (1998) MODELTEST: testing the model of DNA substitution. **Bioinformatics**, **14(9)**: 817-818.
101. PROCA-CIANOBANU, M. (1975) Electron microscopic study of a pathogenic *Acanthamoeba castellanii* strain: the presence of bacterial endosymbionts. **Int. J. Parasitol.**, **5**: 49-56.
102. PUSSARD, R. & PONS, M. (1977) Morphologie de la paroi kystique et taxonomie du genre *Acanthamoeba* (Protozoa, Amoebeida), **Protistologica T.XIII, fasc.4**:557-598.
103. REY, L. (1991) **Parasitologia: parasitos e doenças parasitárias do homem nas Américas e na África**. 2^a. ed. Ed. Guanabara Koogan, Rio de Janeiro, RJ.
104. RIDLEY, M. (1996) **Evolution**. 2^a ed., Blackwell Science, Cambridge, Massachusetts.
105. RÖSEL VON ROSENHOF, A. J. (1755) Der kleine Proteus. **Der monat.-herausgeg, Insekten-Belustigungen**, **3**:622.
106. ROWBOTHAM, T. J. (1980) Preliminary report on the pathogenicity of *Legionella pneumophila* for freshwater and soil amoebae. **J. Clin. Pathol.**, **33**: 1179-1183,
107. SAITOU, N. & NEI, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Mol. Biol. Evol.**, **4**: 406-425.
108. SANDERSON, M. J. (1989) Confidence limits on phylogenies: the bootstrap revisited. **Cladistics**, **5**: 113-129.
109. SANGUINETTI, C. J., NETO, E. D. & SIMPSON, A. J. G. (1996) Recovery and PCR amplification of DNA fragments separated on PhastGel. **Pharmacia Biotech Appl. Note**, Feb: 39-40.
110. SCHAUMBERG, D. A., SNOW, K. K. & DANA, M. R. (1998) The epidemic of *Acanthamoeba* keratitis: where do we stand. **Cornea**, **17(1)**: 3-10.
111. SISON, J. P., KEMPER, C. A., LOVELESS, M., McSHANE, D., VISVESVARA, G. S. & DERESINSKI, S. C. (1995) Disseminated *Acanthamoeba* infection in patients with AIDS: case reports and review. **Clin. Infect. Dis.**, **20**:1207-1216.
112. SNEATH, P. H. A. & SOKAL, R. R. (1963) **Principles of Numerical Taxonomy**. W. H. Freeman, London.
113. SOGIN, S. J., SOGIN, M. L. & WOESE, C. R. (1972) Phylogenetic measurement in prokaryotes by primary structural characterization. **J. Mol. Evol.** **1**: 173-184.
114. SOGIN, M. L., ELWOOD, H. J. & GUNDERSON, J. H. (1986) Evolutionary diversity of eukaryotic small-subunit rRNA genes. **Proc. Natl. Acad. Sci. USA**, **83**:1383-1387.
115. SOGIN, M. L. & SILBERMAN, J. D. (1998) Evolution of the protists and protistan parasites from the perspective of molecular systematics. **Int. J. Parasitol.** **28**: 11-20.
116. STEVENS, A. R. & PACHLER, P. F. (1972) Discontinuity of 26S RNA in *Acanthamoeba castellanii*. **J. Mol. Biol.**, **66**: 225-237.
117. STOTHARD, D. R., SCHROEDER-DIEDRICH, J. M., AWWAD, M. H., GAST, R. J., LEDEE, D. R., RODRIGUEZ-ZARAGOZA, S., DEAN, C. L., FUERST, P. A. & BYERS, T. J. (1998) The evolutionary history of the genus *Acanthamoeba* and the identification of eight new 18S rRNA gene sequence types. **J. Euk. Microbiol.**, **45**:45-54.

118. STRATFORD, M. P. & GRIFFITHS, A. J. (1978) Variations in the properties and morphology of cysts of *Acanthamoeba castellanii*. **J. Gen. Microbiol.**, **108**:33-37.
119. STRIMMER, K. & VON HAESELER, A. (1996) Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. **Mol. Biol. Evol.** **13**: 964-969.
120. SWOFFORD, D. L., (1998) **PAUP*. Phylogenetic Analysis Using Parsimony (*And Other Methods)**. Version 4. Sinauer Associates, Sunderland, Massachusetts
121. SWOFFORD, D. L., OLSEN, G. J., WADDELL, P. J. & HILLIS, D. M. (1996) in **Molecular Systematics**. HILLIS, D. M., MORITZ, C. & MABLE, B. K., eds. Sinauer Associates, Sunderland, MA: 407-514.
122. SZÉNÁSI, Z., ENDO, T., YAGITA, K. & NAGY, E. (1998) Isolation, identification and increasing importance of "free-living" amoebae causing human disease. **J. Med. Microbiol.**, **47**: 5-16.
123. TAMURA, K. & NEI, M. (1993) Estimation of the number of nucleotide substitution in the control region of mitochondrial DNA in humans and chimpanzees. **Mol. Biol. Evol.**, **10**: 512-526.
124. TANAKA, Y., SUGURI, S., HARADA, M., HAYABARA, T., SUZUMORI, K. & OHTA, N. (1994) *Acanthamoeba*-specific human T-cell clones isolated from healthy individuals. **Parasitol. Res.**, **80**: 549-553.
125. TEMPLETON, A. R. (1989) The meaning of species and speciation: a genetic perspective. In **Speciation and its consequences**. Eds. Endler, J. A. & Otte, D., Sinauer Ass., Sunderland Massachusetts.
126. TYNDALL, R. L. (1982) Cocultivation of *Legionella pneumophila* and free-living amoebae. **App. Environ. Microbiol.**, **44**(4): 954-959.
127. VAN DE PEER, Y. & DE WACHTER, R. (1993) TREECON: a software package for the construction and drawing of evolutionary trees. **CABIOS**, **9**(2): 177-182.
128. VAN DE PEER, Y. & DE WACHTER, R. (1997) Evolutionary relationships among the eukaryotic crown taxa taking into account site-to-site variation in 18S rRNA. **J. Mol. Evol.**, **45**: 619-630.
129. VISVESVARA, G. S. & STEHR-GREEN, J. K. (1990) Epidemiology of free-living ameba infection. **J. Protozool.**, **37**:255-335.
130. VISVESVARA, G. S. (1991) Classification of *Acanthamoeba*. **Rev. Infect. Dis.**, **13**: S369-S372.
131. VISVESVARA, G. S., JONES, D. B. & ROBINSON, N. M. (1975) Isolation, identification and biological characterization of *Acanthamoeba polyphaga* from a human eye. **Am. J. Trop. Med. Hyg.** **24**:784-790.
132. VODKIN, M. H., HOWE, D. K., VISVESVARA, G. S. & McLAUGHLIN, G. L. (1992) Identification of *Acanthamoeba* at the generic and specific levels using the polymerase chain reaction. **J. Protozool.**, **39**:378-385.
133. WALKER, C. W. B. (1996) *Acanthamoeba*: ecology, pathogenicity and laboratory detection. **Br. J. Biomed. Sci.**, **53**: 146-151.
134. WATSON, J. D. & CRICK, F. H. C. (1953) Molecular structure of nucleic acids – a structure for deoxyribose nucleic acid. **Nature**, **171**: 737-738.
135. WELSH, J. & McCLELLAND, M. (1990) Fingerprint genomes using PCR with arbitrary primers. **Nucl. Acids Res.**, **18**:7213-7218.
136. WILLIAMS, J. G. K., KUBELIK, A. R., LIVAK, K. J., RAFALSKI, J. A. & TINGEY, S. V. (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. **Nucl. Acids Res.**, **18**:6531-6535.
137. WOESE, C. R., GUTELL, R., GUPTA, R. & NOLLER, H. F. (1983) Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acid. **Microbiol. Rev.**, **47**: 621-669.

138. YAGITA, K. & ENDO, T. (1990) Restriction enzyme analysis of mitochondrial DNA of *Acanthamoeba* strains in Japan. **J. Protozool.**, **37**: 570-575.
139. YAGITA, K., ENDO, E., DE JONCKHEERE, J. F. (1999) Clustering of *Acanthamoeba* isolates from human eye infections by means of mitochondrial DNA digestion patterns. **Parasitol. Res.**, **85**: 284-289.
140. YANG, Z. (1994a) Estimating the pattern of nucleotide substitution. **J. Mol. Evol.**, **39**: 105-111.
141. YANG, Z. (1994b) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. **J. Mol. Evol.**, **39**: 306-314.
142. YANG, Z. (1996a) Among-site rate variation and its impact on phylogenetic analyses. **Tr. Ecol. Evol.**, **11(9)**: 367-372.
143. YANG, Z. (1996b) Maximum-likelihood models for combined analyses of multiple sequence data. **J. Mol. Evol.**, **42**: 587-596.
144. YANG, Z., CAO, Z. & PANJWANI, N. (1997) Pathogenesis of *Acanthamoeba* keratitis: carbohydrate-mediated host-parasite interactions. **Infect. Immun.**, **65(2)**: 439-445.
145. YANG, Z., GOLDMAN, N. & FRIDAY, A. (1995) Maximum likelihood trees from DNA sequences: a peculiar statistical estimation problem. **Syst. Biol.**, **44(3)**: 384-399.
146. ZUCKERKANDL, E. & PAULING, L. (1965) Molecules as documents of evolutionary history. **J. Theor. Biol.**, **8**: 357-366.
147. CRACRAFT, J. (1989) Speciation and its ontology: the empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. In **Speciation and its consequences**. Eds. Endler, J. A. & Otte, D., Sinauer Ass., Sunderland Massachusetts.
148. DOBZHANSKY, T. (1935) A critique of the species concept. **Phil. Sci.**, **2**: 344-355.
149. LANAVE, C., PREPARATA, G., SACCONI, C. & SERIO, G. (1984) A new method for calculating evolutionary substitution rates. **J. Mol. Evol.**, **20**: 86-93.
150. MAYR, E. (1942) **Systematic and the origin of species**. Columbia University Press, New York.
151. PATERSON, H. E. H. (1985) The recognition concept of species. In **Species and speciation**. Ed. Vrba, E. S., Transvaal Museum Monograph n°4, Pretoria, South Africa.
152. SIMPSON, G. G. (1961) **Principles of animal taxonomy**. Columbia University Press, New York.
153. VAN VALEN, L. (1976) Ecological species, multispecies, and oaks. **Taxon**, **25**: 233-239.

Anexo 1 – Trabalho publicado

Random amplified polymorphic DNA profiles as a tool for the characterization of Brazilian keratitis isolates of the genus *Acanthamoeba*

J.M.P. Alves¹, C.X. Gusmão¹,
M.M.G. Teixeira¹,
D. Freitas², A.S. Foronda¹
and H.T. Affonso¹

¹Departamento de Parasitologia, Instituto de Ciências Biomédicas,
Universidade de São Paulo, São Paulo, SP, Brasil

²Departamento de Oftalmologia, Escola Paulista de Medicina,
Universidade Federal de São Paulo, São Paulo, SP, Brasil

Abstract

Correspondence

H.T. Affonso
Av. Prof. Lineu Prestes, 1374
05508-900 São Paulo, SP
Brasil
Fax: +55-11-818-7417
E-mail: mhtaffon@biomed.icb2.usp.br

Research supported by FAPESP and
CNPq. J.M.P. Alves and C.X. Gusmão
are recipients of FAPESP fellowships.

Received December 17, 1998
Accepted November 4, 1999

The genus *Acanthamoeba* comprises free-living amebae identified as opportunistic pathogens of humans and other animal species. Morphological, biochemical and molecular approaches have shown wide genetic diversity within the genus. In an attempt to determine the genetic relatedness among isolates of *Acanthamoeba* we analyzed randomly amplified polymorphic DNA (RAPD) profiles of 11 Brazilian isolates from cases of human keratitis and 8 American type culture collection (ATCC) reference strains. We found that ATCC strains belonging to the same species present polymorphic RAPD profiles whereas strains of different species show very similar profiles. Although most Brazilian isolates could not be assigned with certainty to any of the reference species, they could be clustered according to pattern similarities. The results show that RAPD analysis is a useful tool for the rapid characterization of new isolates and the assessment of genetic relatedness of *Acanthamoeba* spp. A comparison between RAPD analyses and morphological characteristics of cyst stages is also discussed.

Key words

- *Acanthamoeba*
- Keratitis
- RAPD typing
- Genetic variation

Introduction

Free-living amebae of the genus *Acanthamoeba* are opportunistic pathogens of humans and other animal species and have been isolated from a wide variety of natural and man-made environments (1,2). In humans, some species have been implicated in different pathologies, including granulomatous amebic encephalitis (GAE), which occurs in immunologically depressed individu-

als, and keratitis, a severe and painful corneal infection mainly associated with the use of contact lenses (3,4).

The identification of *Acanthamoeba* spp at the genus level is relatively easy since the trophozoite and cyst stages show some features, especially the double-walled cyst shape, that are unique to the genus. Since a wide variability of cyst sizes and shapes can be detected among different isolates, 18 different species were described and assigned to

three distinct morphological groups (I, II and III) based primarily on endo- and ectocyst features (5). However, since the shape of cyst walls can be altered by growth conditions (6), various studies have shown that division of *Acanthamoeba* sp isolates into different groups and/or species is often inconsistent with the classical group designations mentioned above (7-9). Morphological parameters alone are thus not suitable for an accurate identification at the species level.

Among the molecular approaches that have been employed to discriminate species within the genus are isoenzyme patterns (8,10), restriction analysis of mitochondrial DNA (7,11), whole-cell DNA (12) or ribosomal DNA small subunit (SSU rDNA) (13). Based on SSU rDNA sequences, at least 12 different sequence types (T1-T12) were detected (14). Despite the demonstration of a high genetic variability within the genus, and nonetheless contributing to the classification of *Acanthamoeba* sp, most of these methods require large amounts of DNA, relatively complex procedures and even sequencing of the target DNA. On the other hand, the randomly amplified polymorphic DNA (RAPD) method does not require large amounts of DNA, time-consuming procedures or any *a priori* knowledge about the target sequences (15,16). This method has been extensively used to evaluate genetic differences between related organisms (17-20).

In the present study we investigated by RAPD analysis the genetic relatedness among 19 *Acanthamoeba* spp, including 8 reference species from the American type culture collection (ATCC) and 11 not yet characterized Brazilian keratitis isolates. The results were used to construct data matrices and phenograms. RAPD profiles showed highly polymorphic patterns among most of the isolates studied but revealed some groups of more related organisms. The results of RAPD analysis were compared with those of morphological and morphometric analysis of cysts from the different isolates.

Material and Methods

Organisms, growth conditions and DNA extraction

The origin and characteristics of the organisms used are listed in Table 1. For cloning purposes, isolates were grown on 2%-soy extract agar plates previously seeded with heat-killed, plasmidless *Escherichia coli*. Two to 5 trophozoites were picked from each isolate and grown on fresh soy agar plates. After growing and encystment, one cyst of each isolate was cultured, unless otherwise specified. For large axenic cultures, cysts or trophozoites were inoculated in Neff medium (21) containing ampicillin (100 µg/ml) and/or gentamicin (40 µg/ml) and incubated at 28°C with shaking. DNA was purified from late-log-phase cultured cells by SDS lysis, phenol-chloroform extraction, and isopropanol precipitation, basically by the method of Kilvington et al. (22).

Determination of cyst features

To determine their size and structural features, a total of 50 cysts of each corneal scraping isolate and ATCC reference strains were photographed under the phase microscope and the main diameters were measured. The images were scanned and digitally processed to enhance quality. The average measurements of each isolate are presented in Table 1.

RAPD reactions and product analyses

A total of 35 decameric oligonucleotides of arbitrary sequence were used in a previous screening of some *Acanthamoeba* spp isolates. According to the number and intensity of the resulting bands, the reproducibility and discriminating potential of the amplified products, three oligonucleotides were selected for further analyses: 606 (5'-CGGTCCGCCA-3'), 688 (5'-GCAGGAGC

GT-3') and 694 (5'-GGTTTGGAGG-3'). Several of the remaining oligonucleotide-derived patterns, while still useful to confirm the results of genetic similarity estimates, were not used in the computer analyses.

The DNA amplification reactions were performed in 50 µl volumes containing 100 ng of template DNA, 250 µM of dNTPs, 100 pmol of a single primer, 2 units of *Taq* DNA polymerase (Gibco BRL, Gaithersburg, MD, USA), and 1.5 mM MgCl₂, in a Progene thermocycler (Techne, Cambridge, UK). The cycling conditions used were: initial denaturing phase of 94°C for 3 min and 35 repetitions at 94°C for 1 min, 37°C for 2 min and 72°C for 2 min. The primer extension phase was prolonged for 10 min at 72°C in the last cycle. Amplification products were fractionated by 2% agarose gel electrophoresis, stained with ethidium bromide and photographed under ultraviolet light with an Eagle Eye II (Stratagene, La Jolla, CA, USA) image processor. The molecular marker used for fragment sizing was a 100-bp DNA ladder (Pharmacia Biotech, Uppsala, Sweden). For data analysis, amplification products obtained with each primer from each *Acanthamoeba* spp isolate were fractionated in the same gel.

Computer-assisted analysis

The digitized gel images were analyzed with the use of the RFLPscan Plus software (version 3.0, Scanalytics CSP Inc., Billerica, MA, USA) in order to size and compare the DNA fragments amplified from the various strains. The discrete character matrix (absence or presence of bands) was analyzed by the RAPDistance software version 1.03 (23) for the calculation of genetic distances between all compared isolates, using the Jaccard index ($A/A + B + C$, where A is the number of shared bands among two organisms x and y , and B and C are the number of bands present only in organism x or y , respectively). To graphically represent the inferred groups, a

phenogram was constructed by the UPGMA method using the PHYLIP 3.5 Neighbor program (24). In order to assess the robustness of the branching, a hundred bootstrap replicates were constructed and only branches presenting bootstrap values of at least 50 were considered as supported clusters.

Results and Discussion

A total of 11 Brazilian keratitis isolates and 8 ATCC reference strains, the latter belonging to 4 different species representative of the three known morphological groups of *Acanthamoeba* (Table 1), were characterized based on cyst morphology and RAPD analysis.

The classification of the isolates within the genus *Acanthamoeba* was based on morphological criteria (5). Figure 1 shows the highly characteristic double-walled morphology of cyst stages of Brazilian keratitis isolates and some representative ATCC refer-

Table 1 - Characteristics of *Acanthamoeba* isolates.

ATCC: American type culture collection; A. poly.: *Acanthamoeba polyphaga*; A. cast.: *A. castellanii*; HAK: human *Acanthamoeba* keratitis; CLW: contact lenses wearer; NCLW: noncontact lenses wearer; HCC: human choriocarcinoma cells; U/E: USP/EPM; U/HCC: USP/HCC; ^aaverage measurements of 50 cysts; ^btwo morphologically distinct clones isolated from the same corneal sample; ^cisolated from the right (R) and left (L) eye of the same patient.

Isolate	ATCC	Source	Cyst diameter (µm ± SD) ^a
A. astronyxis	30901	water	19.7 ± 1.6
A. poly. #1	30873	HAK	16.4 ± 1.5
A. poly. #2	30461	HAK	16.6 ± 2.2
A. poly. #3	30871	water	16.3 ± 2.7
A. poly. #4	30872	water	16.4 ± 1.5
A. cast. #1	30011	yeast culture	17.4 ± 1.6
A. cast. #2	30868	HAK	15.4 ± 1.6
A. royreba	30884	HCC	17.2 ± 2.4
U/E 1		HAK-CLW	17.0 ± 2.2
U/E 2		HAK-CLW	14.9 ± 1.4
U/E 3		HAK-CLW	16.7 ± 1.8
U/E 4		HAK-NCLW	18.3 ± 1.6
U/E 5		HAK-NCLW	14.2 ± 1.8
U/E 6		HAK-CLW	14.8 ± 1.9
U/E 7.1 ^b		HAK-CLW	14.0 ± 1.5
U/E 7.2 ^b			18.8 ± 2.0
U/E 8R ^c		HAK-CLW	16.5 ± 2.3
U/E 8L ^c			18.8 ± 1.4
U/HCC1		HAK-CLW	10.8 ± 1.4

Figure 1 - Representative cysts of *Acanthamoeba* isolates photographed by phase microscopy and digitally processed. A, ATCC reference isolates - 1: *A. astronyxis*; 2: *A. cast.* #1; 3: *A. poly.* #3; 4: *A. cast.* #2; 5: *A. royreba*, and 6: *A. poly.* #4. B, Brazilian keratitis isolates - 1: U/HC 1; 2: U/E 1; 3: U/E 3; 4: U/E 4; 5: U/E 5; 6: U/E 6; 7: U/7.1; 8: U/E 7.2; 9: U/E 8R; 10: U/E 8L, and 11: U/E 2. Bar = 12.0 μ m. For abbreviations, see legend to Table 1.

ence strains. Except for *A. astronyxis*, which presents large cysts with a smooth ectocyst and a stellate endocyst typical of group I, and *A. royreba*, typical of group III, the shape features of cysts were closely similar among the different isolates (Figure 1). On the other hand, average cyst sizes obtained from our keratitis isolates showed wide variability, ranging from 10.8 to 18.8 μ m (Table 1). In spite of this high variability, the morphological criteria allowed us to classify keratitis isolates in the genus *Acanthamoeba*, although they proved insufficient to distinguish among most of them. Based on these criteria, they could be included in morphological group II which comprises the majority of species (25).

To determine the genetic relatedness of the isolates, DNA fingerprints were determined by RAPD analysis. Each reaction was repeated at least three times, and 100 stable bands were considered and computed in the matrices. In most cases, RAPD profiles al-

lowed detection of extensive polymorphism among isolates. Similar results were obtained with the three primers and an illustrative profile is shown in Figure 2.

Results from the RAPD analyses with the three primers were combined and used to calculate a matrix of pairwise similarities (Figure 3). This matrix was used to construct a phenogram, after determination of bootstrap values for 100 replicates (Figure 4). As expected from the RAPD profiles, some genetically more related Brazilian keratitis isolates could be detected. However, as a result of the high genetic variability, no major groups could be assigned. Moreover, none of our keratitis isolates could be surely grouped with any ATCC reference strains. Although minor differences in RAPD profiles were detected, four clusters of Brazilian keratitis isolates could be identified: a) U/E 1 and 3, b) U/E 4, 5 and 6, c) U/E 7.1 and 7.2, and d) U/E 8R and 8L. Isolates U/E 2 and U/HC 1 could not be assigned to any cluster

Figure 2 - RAPD profiles of 19 *Acanthamoeba* isolates obtained with primer 688 (5'-GCAGGA GCGT-3'). The PCR products were resolved by electrophoresis on 2% agarose gel. Designations of isolates are indicated on the top: A. astr.: *A. astronyxis*, A. poly.: *A. polyphaga*, A. cast.: *A. castellanii*, A. royr.: *A. royreba*, U/E: USP/EPM, U/H: USP/HC. Molecular size markers (100-bp DNA ladder - Gibco BRL) are shown on the left.

	A. astr.	A. cast. #1	A. cast. #2	A. poly. #1	A. poly. #2	A. poly. #3	A. poly. #4	A. royr.	USP/EPM 1	USP/EPM 2	USP/EPM 3	USP/EPM 4	USP/EPM 5	USP/EPM 6	USP/EPM 7-1	USP/EPM 7-2	USP/EPM 8R	USP/EPM 8L
A. cast. #1	10.0																	
A. cast. #2	12.9	30.4																
A. poly. #1	26.7	26.9	45.8															
A. poly. #2	11.4	25.9	20.0	18.1														
A. poly. #3	6.4	16.7	11.1	13.8	13.3													
A. poly. #4	15.4	19.0	22.7	25.0	14.8	8.7												
A. royr.	26.7	22.2	40.0	65.2	18.2	13.8	25.0											
USP/EPM 1	18.2	13.3	20.0	21.9	21.2	6.2	19.2	18.2										
USP/EPM 2	17.6	20.7	19.4	17.6	17.1	12.9	10.3	14.3	17.1									
USP/EPM 3	16.7	15.4	14.3	20.7	16.1	7.1	12.5	16.7	50.0	15.6								
USP/EPM 4	13.3	26.1	19.2	21.4	25.0	16.0	30.0	25.9	20.7	12.5	19.2							
USP/EPM 5	20.0	24.0	26.9	33.3	23.3	14.8	27.3	38.5	19.4	11.8	17.9	52.4						
USP/EPM 6	19.4	33.3	17.2	19.4	26.7	14.3	20.8	23.3	22.6	18.2	17.2	43.5	40.0					
USP/EPM 7-1	8.8	23.1	17.2	15.6	26.7	23.1	20.8	15.6	18.7	18.2	9.7	26.9	40.0	28.6				
USP/EPM 7-2	8.6	22.2	16.7	15.2	25.8	22.2	20.0	15.2	18.2	17.6	9.4	25.9	38.5	27.6	85.0			
USP/EPM 8R	10.0	16.7	15.4	22.2	13.3	12.0	19.0	22.2	25.9	9.4	11.1	16.0	14.8	14.3	18.5	17.9		
USP/EPM 8L	9.4	20.0	18.5	25.0	12.5	15.4	22.7	25.0	20.0	8.8	6.7	19.2	17.9	13.3	17.2	16.7	66.7	
USP/HC 1	15.2	13.8	9.4	15.2	11.4	17.9	7.1	18.8	11.4	11.1	9.4	13.3	20.0	19.4	8.8	8.6	10.0	12.9

Figure 3 - Matrix showing pairwise similarities calculated for RAPD profiles obtained with primers 606, 688 and 694. For abbreviations, see legend to Figure 2.

Figure 4 - Unrooted UPGMA phenogram based on RAPD profiles of 19 *Acanthamoeba* isolates obtained with primers 606, 688 and 694. For abbreviations, see legend to Figure 2. The numbers at the nodes represent bootstrap values for a hundred replicates.

since they showed high genetic variability when compared to the others. Based on the analysis of the resulting clusters, we found both agreements and disagreements between RAPD profiles and morphological cyst features. In the clusters harboring U/E 1 and 3, and U/E 5 and 6, the isolates showed very similar cyst sizes (Table 1, Figure 1) and RAPD patterns (Figure 2), despite the fact that they were isolated from different patients. In contrast, the clusters to which U/E 4 and 5, U/E 7.1 and 7.2, and U/E 8R and 8L were assigned showed discordance between cyst sizes and RAPD patterns. The cyst sizes of isolates U/E 4 and 5 were 18.33 μm and 14.17 μm , respectively, whereas their RAPD profiles were very similar (Figure 2).

It is noteworthy that U/E 7.1 and 7.2, and U/E 8 L and R also showed discrepancies between cyst sizes and RAPD profiles. The two clones of U/E 7, which were simultaneously isolated from the same corneal sample, showed quite different cyst sizes (Table 1, Figure 1). These observations may suggest that the patient had been infected with genetically different amebae. However, despite some detected differences, especially with primer 688 (similarity index of 0.87), very similar RAPD patterns were obtained (Figure 2), indicating that the two clones probably originated from the same organism. In fact, it has been reported that cyst morphology might vary depending on culture conditions, even within cloned strains (6,26), possibly explaining the polymorphism

of cyst sizes. Isolates U/E 8R and 8L, which were respectively obtained from the right and left eye of another patient, showed very similar cyst features (Table 1). However, the RAPD patterns with primer 688 (Figure 2) showed a lower similarity index (0.57) when compared with that of the two clones of U/E 7. Analyses with other primers showed similar results (data not shown), indicating that this patient could have been infected by genetically different organisms.

A remarkable feature of the RAPD-derived phenogram is the presence of strains assigned to the same species (*A. castellanii* and *A. polyphaga*) distributed through different branches. On the other hand, different species - *A. poly. #1*, *A. cast. #2* and *A. royreba* - were assigned to nearby branches. These observations do not agree with the morphological data, especially with cyst diameter. For example, the four strains designated as *A. polyphaga* (morphological group II) showed a narrow range of cyst sizes and *A. royreba* showed cyst shapes and measurements typical of morphological group III (Table 1). In contrast, *A. astronyxis*, a representative species of morphological group I, could not be clustered with any other reference species or our keratitis isolates, showing concordance between RAPD profiles and cyst features. Similar findings were also observed by other approaches (8,14,27). Therefore, it is clear that species definition in the genus *Acanthamoeba* - especially those assigned to either *A. polyphaga* or *A. castellanii* - must be reviewed, as recently proposed (14).

Our data demonstrate the ability of RAPD analysis to assign a specific genetic fingerprint to otherwise morphologically indistinguishable *Acanthamoeba* isolates. Furthermore, this approach proved to be useful in the characterization of new isolates and for the assessment of genetic relatedness and proved to be a fast and informative strategy to determine relationships among different strains. A general limitation of our technique

is its inability to properly infer phylogenetic relationships, because genetic distances calculated from RAPD could be affected by paralogy, which would make RAPD patterns inconsistent with speciation events. Even identification could be hampered by such problems, since some patterns could eventually mislead species assignment. However, our RAPD analyses are in agreement with preliminary SSU rDNA fingerprinting profile determination, which is in progress in our laboratory, indicating that this problem probably does not affect our analyses. This is nonetheless very serious and should be addressed with care in future studies.

Other approaches, such as sequencing of ribosomal DNA subunits, can be used for a more precise determination of the relationships among our strains. However, such relationships cannot be assessed if the approach used is unable to detect subtle differences between very closely related organisms, as recently found in the analysis of SSU rDNA sequences (14). Based on this last approach, at least 12 different sequence types (T1-T12) were detected, presenting dissimilarities ranging from 5.0 to 37.6% between sequence types. The detection of such divergent SSU rDNA sequences is in accordance with our data, since it also reflects the high polymorphism of these organisms. However, as stated by those authors, this strategy could not identify branching patterns within T4 sequence type strains, a

group to which 24 of 25 worldwide-distributed keratitis isolates were assigned. Although our keratitis isolates were not characterized by that approach, they showed a high degree of polymorphism and hence were distributed through different branches by our RAPD analyses. It would thus be interesting to investigate whether these isolates present the same SSU sequence types or, conversely, if RAPD patterns of the isolates designated as T4 would be so polymorphic as to allow the identification of their branching patterns. It should be noted that three ATCC reference strains - *A. poly.* #1 (30873), *A. poly.* #3 (30871), and *A. cast.* #1 (30011) - which were assigned to the same sequence type (T4), were also spread across different branches of the phenogram (Figure 4).

In conclusion, the RAPD technique can be useful to distinguish between closely related *Acanthamoeba* sp isolates and to possibly characterize additional isolates of this genus. On the other hand, more conserved sequences would still be required to determine major groups involving less related strains.

Acknowledgments

The authors thank Annamaria S. Stolf, Regina V. Milder and Silvia C. Alfieri for a critical reading of the manuscript, and Manoel A. Peres for technical assistance.

References

1. Ma P, Visvesvara GS, Martinez AJ, Theodore FH, Daggett PM & Sawyer TK (1990). Naegleria and Acanthamoeba infections: review. *Reviews of Infectious Diseases*, 12: 490-513.
2. Kilvington S & White DG (1994). Acanthamoeba: biology, ecology and human disease. *Reviews in Medical Microbiology*, 5: 12-20.
3. Logar J & Kraut A (1997). Acanthamoeba corneal infection in a contact lens wearer. *Journal of Infection*, 35: 237-240.
4. Szénási Z, Endo T, Yagita K & Nagy E (1998). Isolation, identification and increasing importance of "free-living" amoebae causing human disease. *Journal of Medical Microbiology*, 47: 5-16.
5. Pussard R & Pons M (1977). Morphologie de la paroi kystique et taxonomie du genre Acanthamoeba (Protozoa, Amoebozoa). *Protistologica*, 13: 557-598.
6. Stratford MP & Griffiths AJ (1978). Variations in the properties and morphology of cysts of Acanthamoeba castellanii. *Journal of General Microbiology*, 108: 33-37.
7. Costas M, Edwards SW, Lloyd D, Griffiths AJ & Turner G (1983). Restriction enzyme analysis of mitochondrial DNA of members of the genus Acanthamoeba as an aid in taxonomy. *FEMS Microbiology Letters*, 17: 231-234.
8. Daggett P, Lipscomb D, Sawyer TK & Nerad TA (1985). A molecular approach to the phylogeny of Acanthamoeba. *BioSystems*, 18: 399-405.
9. Visvesvara GS (1991). Classification of Acanthamoeba. *Reviews of Infectious Diseases*, 13: S369-S372.
10. Costas M & Griffiths AJ (1985). Enzyme

- composition and the taxonomy of *Acanthamoeba*. *Journal of Protozoology*, 32: 604-607.
11. Yagita K & Endo T (1990). Restriction enzyme analysis of mitochondrial DNA of *Acanthamoeba* strains in Japan. *Journal of Protozoology*, 37: 570-575.
 12. McLaughlin GL, Brandt FH & Visvesvara GS (1988). Restriction fragment length polymorphisms of the DNA of selected *Naegleria* and *Acanthamoeba* amoebae. *Journal of Clinical Microbiology*, 26: 1655-1658.
 13. Kong HH & Chung DI (1996). PCR and RFLP variation of conserved region of small subunit ribosomal DNA among *Acanthamoeba* isolates assigned to either *A. castellanii* or *A. polyphaga*. *Korean Journal of Parasitology*, 34: 127-134.
 14. Stothard DR, Schroeder-Diedrich JM, Awwad MH, Gast RJ, Ledee DR, Rodriguez-Zaragoza S, Dean CL, Fuerst PA & Byers TJ (1998). The evolutionary history of the genus *Acanthamoeba* and the identification of eight new 18S rRNA gene sequence types. *Journal of Eukaryotic Microbiology*, 45: 45-54.
 15. Welsh J & McClelland M (1990). Fingerprint genomes using PCR with arbitrary primers. *Nucleic Acids Research*, 18: 7213-7218.
 16. Williams JGK, Kubelik AR, Livak KJ, Rafalski JA & Tingey SV (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, 18: 6531-6535.
 17. Steindel M, Dias Neto E, Carvalho Pinto CJ, Grisard EC, Menezes CLP, Murta SMF, Simpson AJG & Romanha AJ (1994). Randomly amplified polymorphic DNA (RAPD) and isoenzyme analysis of *Trypanosoma rangeli* strains. *Journal of Eukaryotic Microbiology*, 41: 261-267.
 18. Pogue GP, Koul S, Lee NS, Dwier DM & Nakhasi HL (1995). Identification of intra- and interspecific *Leishmania* genetic polymorphisms by arbitrary primed polymerase chain reactions and use of polymorphic DNA to identify differentially regulated genes. *Parasitology Research*, 81: 282-290.
 19. Pélandakis M, Kaundun SS, De Jonckheere JF & Pernin P (1997). DNA diversity among the free-living amoeba, *Naegleria fowleri*, detected by the random amplified polymorphic DNA method. *FEMS Microbiology Letters*, 151: 31-39.
 20. Tibayrenc M (1998). Genetic epidemiology of parasitic protozoa and other infectious agents: the need for an integrated approach. *International Journal of Parasitology*, 28: 85-104.
 21. Neff RJ, Ray SA, Benton WF & Wilborn M (1964). Induction of synchronous encystment (differentiation) in *Acanthamoeba* sp. In: Prescott DM (Editor), *Methods in Cell Physiology*. Academic Press, New York.
 22. Kilvington S, Beeching J & White DG (1991). Differentiation of *Acanthamoeba* strains from infected corneas and the environment by using restriction endonuclease digestion of whole-cell DNA. *Journal of Clinical Microbiology*, 29: 310-314.
 23. Armstrong J, Gibbs A, Peakall R & Weller G (1994). RAPDistance Programs, Version 1.03 for the Analysis of Patterns of RAPD Fragments. Australian National University, Canberra.
 24. Felsenstein J (1989). *Phylip 3.2 Manual*. University of California, Herbarium, Berkeley, CA.
 25. Walker CWB (1996). *Acanthamoeba*: ecology, pathogenicity and laboratory detection. *British Journal of Biomedical Science*, 53: 146-151.
 26. Jacobson LM & Band RN (1987). Genetic heterogeneity in a natural population of *Acanthamoeba polyphaga* from soil, an isoenzyme analysis. *Journal of Protozoology*, 34: 83-86.
 27. Bogler SA, Zarley CD, Burianek LL, Fuerst PA & Byers TJ (1983). Interstrain mitochondrial DNA polymorphism detected in *Acanthamoeba* by restriction endonuclease analysis. *Molecular and Biochemical Parasitology*, 8: 145-163.

*Anexo 2 – Manuscrito em
preparação*

Genus *Acanthamoeba*: parity among random amplified polymorphic DNA profiles, total DNA and SSU rDNA restriction fragment length polymorphism patterns for the grouping and characterization of Brazilian keratitis isolates

João M. P. Alves, Maisa M. F. Saba, Marta M. G. Teixeira and Heloiza T. Affonso*

Departamento de Parasitologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, SP, Brazil

Research supported by FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico). J.M.P.A. is recipient of a FAPESP fellowship.

* Corresponding author. Address: Av. Prof. Lineu Prestes, 1374 - 05508-900 São Paulo, SP, Brazil; Fax: +55 (11) 3818 7417; e-mail: mhtaffon@icb.usp.br

Running title: Total DNA RFLP for the characterization of *Acanthamoeba*

Keywords: *Acanthamoeba*; keratitis; total DNA RFLP typing; polymorphism; SSU rDNA
RFLP

Abstract

The free-living amoebae of the genus *Acanthamoeba* have been proved to be opportunistic pathogens of humans and other animal species, causing severe diseases. Although different approaches have shown great genetic diversity within the genus, there are few reliable methods to the rapid identification and characterization of *Acanthamoeba* isolates. In this report we show agreement between use of total DNA and SSU rDNA RFLP patterns and previously reported RAPD typing for the characterization and grouping of *Acanthamoeba*. We analyzed RFLP patterns of 11 Brazilian isolates from cases of human keratitis and 8 American Type Culture Collection (ATCC) reference strains. As previously found for RAPD typing of these same isolates, ATCC strains belonging to the same species presented distinct patterns whereas strains of different species showed very similar ones. Besides that, most Brazilian keratitis isolates could be clustered according to RFLP pattern similarities, in spite of the great polymorphism presented. Results showed concordance between RAPD and both total DNA and SSU rDNA RFLP groupings, confirming them as useful tools for the rapid characterization and identification of new *Acanthamoeba* spp. isolates.

Introduction

Genus *Acanthamoeba* comprises free-living amoebae that are opportunistic pathogens of humans and other animal species and are ubiquitous in a large variety of natural and man-made environments (1,2). Some species have been observed causing different pathologies, especially the fatal granulomatous amoebic encephalitis (GAE), which occurs in immunologically depressed individuals, and keratitis, a severe and painful corneal infection mainly associated with use of contact lenses (3,4). There is only one published case of GAE in Brazil (5), while keratitis has been much more frequently diagnosed by some clinical laboratories.

Generic identification of *Acanthamoeba* spp. is relatively easy since trophozoite and cyst stages show some features, especially the trophozoite's acanthopodia and the cyst's double-wall and shape, that are characteristic of the genus. However, since shape of cyst walls can be altered by growth conditions and a wide range of morphological variability can be found in cultures derived from single clones (6), various studies have shown that division of *Acanthamoeba* spp. isolates into different groups and/or species is often inconsistent with the classical designations (7,8,9,10,11). Morphological parameters alone are thus not suitable for an accurate identification at the species level. Due to the many inconsistencies in classification of *Acanthamoeba* species until the moment, recent studies use only strain designation, omitting traditional specific names until a better and more objective consensus is reached (12).

On the other hand, molecular approaches that have been employed to discriminate species within the genus, such as isoenzyme patterns (9,13), and restriction analysis of mitochondrial DNA (8,12), whole-cell DNA (14, 15) or ribosomal DNA small subunit (SSU rDNA) (16) have also demonstrated high genetic variability within the genus. Twelve different SSU rDNA sequence types (T1-T12) were detected until now (11), and were suggested to define one species each, or even new genera, in the case of the most diverging isolates. Although contributing to *Acanthamoeba* spp. classification, most of these methods require large amounts of DNA, relatively complex procedures and even sequencing of large target genes. Besides that, there are little species in common between different studies using different approaches, thus hampering generalization about usefulness of identification methods. Very few studies have used the randomly amplified polymorphic DNA (RAPD) method, SSU rDNA RFLP and the total DNA RFLP analysis for identification and characterization of *Acanthamoeba* (14,15,17). These multiloci methods are advantageous because they do not require time-consuming procedures, large quantities of sample or any

prior knowledge about the target sequences, even in the case of SSU rDNA RFLP, since universal primers can be successfully used for PCR amplification.

Total DNA RFLP is generally used as a first insight into genome organization, and usually demands use of radioactive probes to identify fragments hidden in the resulting smear. For some organisms, such as *Acanthamoeba*, *Naegleria* and other related organisms (14,18), total DNA RFLP patterns with abundant and well-defined bands can be visualized directly on ethidium bromide-stained agarose gels under ultraviolet light. This suggests either a high number of mtDNA copies or a low-complexity genome for these organisms, as well as a combination of these two characteristics, as discussed below.

In this report we determined total DNA restriction fragment length polymorphism patterns for 19 *Acanthamoeba* spp. isolates (including 8 reference species from ATCC and 11 Brazilian keratitis isolates) to investigate the parity between conclusions drawn from this study and from the previously done RAPD analysis, which used the same isolates (17). As shown for RAPD profiles and in accordance with previous literature (14,15,18), total DNA RFLP showed highly polymorphic banding patterns among most of the isolates studied but could nonetheless uncover some groups of more related organisms. The resulting patterns were used to construct phenograms, which were compared to visualize which groups were supported by both methods. Groups unveiled by total DNA RFLP have been shown to support the previously found RAPD groupings. On the other hand, SSU rDNA RFLP patterns showed much less variation between different isolates, as expected, due to the more conserved nature of this much more biologically significant molecule. However, there was still enough differences and polymorphism to establish groups of isolates, which agreed with those determined by the two former methods, thus considerably strengthening their conclusions.

Material and methods

Organisms, growth conditions and DNA extraction

The origin and characteristics of the organisms used are listed in Table 1. Cloning and culturing were done as previously reported (17). DNA was purified from late-log-phase cultured cells by SDS lysis, phenol-chloroform extraction, and isopropanol precipitation, basically by the method of Kilvington et al. (15).

Restriction enzyme reactions and product analyses

Several restriction endonucleases were used in a previous screening of *Acanthamoeba* spp. total DNA RFLP. According to complexity of the resulting patterns, seven restriction enzymes were selected for further experiments: *Bgl*II, *Eco*RI, *Hae*III, *Hha*I, *Hinc*II, *Hind*III and *Msp*I. Restriction enzymes for SSU rDNA RFLP were previously selected by computer simulation using a known *Acanthamoeba* SSU gene (GenBank accession number M13435) and the Clone Manager 3.11 (Scientific & Educational Software) computer program, in order to find enzymes that would yield a significant number of adequate fragments for analysis. The chosen enzymes were *Ava*II, *Hha*I, *Hinf*I, *Msp*I and *Taq*I.

Restriction endonuclease reactions were performed in 30 µl volumes containing approximately 0.5 to 1.0 µg of DNA in the case of total DNA RFLP or approximately 200 ng in 20 µL reaction volumes for SSU rDNA RFLP. Smaller DNA quantities were more adequate for SSU rDNA RFLP due to the high sensibility of silver-stained polyacrylamide gels. Buffers and reaction conditions were those determined by enzyme manufacturers. Total DNA restriction products were fractionated by 0.8% agarose gel electrophoresis, stained with ethidium bromide and photographed under ultraviolet light, and the molecular marker used for fragment sizing was λ /*Hind*III DNA. SSU rDNA restriction products were fractionated by 27 cm, 6% polyacrylamide gel electrophoresis. Gels were stained with AgNO₃, following the method of Sanguinetti et al., 1996 (19), with modifications: solution I (10% ethanol + 0.5% acetic acid) for 5 min at 40°C; solution I + 0.2% AgNO₃ for 20 min at 40°C; two 1 min

washes with distilled water at room temperature; developer (100 mL NaOH 3% + 100 µL formaldehyde 37%) for 10 min, twice (in the second incubation time, it may be needed to wait until bands are clearly visible), at room temperature; solution I for 5 min at 40°C.

Stained polyacrylamide gels were digitized using an ordinary flatbed scanner and immediately dried before further examination.

Computer-assisted analysis

The digitized gel images were analyzed with aid of the RFLPscan Plus software (version 3.0, Scanalytics CSP Inc., USA) in order to size and compare the DNA fragments obtained from the various isolates with each enzyme, constructing discrete character matrices (absence or presence of bands) with data from each enzyme. These matrices were analyzed by the RAPDistance software version 1.04 (20) for the calculation of genetic distances between all compared isolates, using the Jaccard index (proportion of shared bands). To graphically represent the inferred groups, phenograms were constructed by the UPGMA method using the Neighbor program of PHYLIP 3.5 (21). In order to assess statistical support for the branches, a hundred bootstrap replicates were constructed to each data matrix and only branches presenting bootstrap values of at least 50 were considered as supported clusters.

Results and discussion

A total of 11 Brazilian keratitis isolates and 8 ATCC reference strains of *Acanthamoeba* (Table 1) were characterized based on total DNA and SSU rDNA analysis, and the resulting groups were compared to those previously obtained by RAPD analysis of these same isolates (17).

As expected from previously published data (14,15), clearly visible patterns with abundant and well-defined bands were observed directly on ethidium bromide stained agarose gels when total DNA was digested (Fig. 1). Although other organisms such as the

amoeboflagellate *Naegleria* spp. also present banding patterns directly on ethidium bromide-stained agarose gels (18), these patterns are less clear and abundant than those obtained by digesting *Acanthamoeba* whole-cell DNA. On the other hand, DNA must be gently extracted in order to achieve the best results in banding pattern resolution and visibility against the background smear that would otherwise be much more intense (14). SSU rDNA RFLP patterns were determined in polyacrylamide gels, since frequent presence of several smaller bands (< 100 bp) and bands with similar sizes could not be reliably separated and identified in agarose gels. Another advantage of using silver-stained polyacrylamide gels over ethidium bromide-stained gels is the increased sensibility in detection, allowing use of much smaller quantities of material in reactions and gels, thus rendering the method faster

Detection of many prominent bands directly on ethidium bromide-stained agarose gels is probably due to the low-complexity or repetitive-rich nature of *Acanthamoeba* genome, which was estimated to be formed by at least 22% of repetitive sequences (22). Using hybridization, previous reports showed that the majority of these bands are not mitochondrial or ribosomal DNA derived, although a few of them possibly are (14). On the other hand, another study (15) concluded that all bands from total DNA RFLP could have been originated from mtDNA, since it was observed that total DNA of one of the studied isolates yielded exactly the same restriction pattern when digested with *EcoRI* as that obtained in another report involving purified mitochondrial DNA from the same strain (23). However, we can not be sure of the purity and identity of this mtDNA, since the mtDNA isolation method was not specified by the authors and they did not present any experiment confirming that all observed bands were actually mtDNA derived.

This same question appears clearly in another report (24) where it was found that digestion of purified mtDNA with *EcoRI* yielded a pattern that lacked two bands when compared to previous similar findings (15). Interestingly, the lacking bands were present when whole-cell DNA rather than the mtDNA fraction alone was digested, indicating that the

conclusion by Kilvington et al., 1991 that all bands in total DNA RFLP were mtDNA derived was precipitate. Since we intended to use a rapid and simple method for identification and characterization of our isolates, the precise origin of the bands was not deemed as important for our analyses, as long as they indicated genetic similarity between organisms.

To determine the genetic relatedness between the isolates, DNA fingerprints were determined by RFLP analysis using seven restriction enzymes for total DNA and 5 enzymes for SSU rDNA. Each reaction was repeated at least twice, to assure the accuracy of the resulting patterns. In most cases, RFLP profiles allowed detection of extensive polymorphism among isolates, although SSU rDNA RFLP banding patterns were, as expected, much less variable than those obtained for total DNA digestion (Fig. 2).

As can be seen in Fig. 3, groups best supported by RAPD data were also obtained by RFLP analysis. The RFLP phenogram presented is the consensus for the seven enzymes utilized, and bootstrap values for 100 replicates are shown at the nodes. The same groupings of Brazilian keratitis isolates were obtained, namely: a) U/E 1 and 3; b) U/E 4, 5 and 6; c) U/E 7.1 and 7.2; d) U/E 8R and 8L. As reported by previous studies using several different approaches (9,12,13), strains assigned to the same species (*A. castellanii* and *A. polyphaga*) were distributed through different branches, while strains considered as belonging to different species – *A. polyphaga* and *A. royreba* – were assigned to nearby branches. In contrast, *A. astronyxis*, a morphologically very distinct strain, could not be clustered with any other reference species or our keratitis isolates, showing concordance between RAPD, morphological features and total DNA RFLP, as previously reported (9,11,17,25). Therefore, it is clear that species definition in the genus *Acanthamoeba* - especially those assigned to either *A. polyphaga* or *A. castellanii* in particular and morphological group II in general - must be reviewed, as recently proposed (11).

As previously shown for RAPD analysis, RFLP could be able to assign specific genetic fingerprints to otherwise morphologically indistinguishable *Acanthamoeba* isolates.

Furthermore, this approach also proved to be useful in the characterization of new isolates and for assessment of genetic relatedness and to be a fast and informative strategy to determine relationships among different strains. The fact that these two different approaches corroborate the same conclusions suggests that these groupings could be natural and possibly represent phylogenetically closer isolates. In spite of that, RFLP patterns alone are generally limited in its ability to properly infer phylogenetic relationships, because genetic distances calculated from RFLP patterns could be affected by paralogy and non-independent character changes, which would make conclusions inconsistent with speciation events, as is the case for RAPD-derived patterns (17, 26). Even identification could be hampered by such problems, since some patterns could eventually mislead species assignment. However, both RAPD and total DNA RFLP analyses are in good agreement with SSU rDNA fingerprinting profile determination, indicating that this problem probably does not affect our analyses, considering the conserved and biologically significant nature of this sequence.

More precise determination of relationships among our strains and possible identification of species could be reached by other approaches, such as sequencing of the ribosomal DNA small subunit. Considering that three reference strains present in this study (ATCC 30873, ATCC 30871, and ATCC 30011), which were assigned to sequence type T4 (11), presented very different total DNA RFLP and RAPD patterns and were thus spread across different branches of the phenogram (Fig. 3), it would be interesting to investigate whether our isolates present the same SSU sequence types or, conversely, if total DNA RFLP and RAPD patterns of the isolates designated as T4 would be so polymorphic as to allow the identification of their branching patterns.

In conclusion, total DNA RFLP, SSU rDNA RFLP and RAPD techniques can be useful to distinguish between closely related *Acanthamoeba* spp. isolates and to possibly characterize additional isolates belonging to this genus, although they are not well suited to determining phylogenetic relationships between species within the genus.

Acknowledgments

The authors thank Manoel A. Peres for technical assistance.

References

1. Ma P, Visvesvara GS, Martinez AJ, Theodore FH, Dagget PM & Sawyer TK (1990) *Naegleria* and *Acanthamoeba* infections: review. *Reviews of Infectious Diseases*, 12: 490-513.
2. Kilvington S & White DG (1994) *Acanthamoeba*: biology, ecology and human disease. *Reviews of Medical Microbiology*, 5: 12-20.
3. Logar J & Kraut A (1997) *Acanthamoeba* corneal infection in a contact lens wearer. *Journal of Infection*, 35: 237-240.
4. Szénási Z, Endo T, Yagita K & Nagy E (1998) Isolation, identification and increasing importance of "free-living" amoebae causing human disease. *Journal of Medical Microbiology*, 47: 5-16.
5. Calore EE, Cavaliere MJ & Calore NMP (1997) Cerebral amebiasis in the acquired immunodeficiency syndrome. *Acta Neurol. Belg.*, 97:248-250.
6. Stratford MP & Griffiths AJ (1978) Variations in the properties and morphology of cysts of *Acanthamoeba castellanii*. *Journal of General Microbiology*, 108: 33-37.
7. Pussard R & Pons M (1977) Morphologie de la paroi kystique et taxonomie du genre *Acanthamoeba* (Protozoa, Amoebida). *Protistologica*, 13: 557-598.
8. Costas M, Edwards SW, Lloyd D, Griffiths AJ & Turner G (1983) Restriction enzyme analysis of mitochondrial DNA of members of the genus *Acanthamoeba* as an aid in taxonomy. *FEMS Microbiology Letters*, 17: 231-234.
9. Daggett P, Lipscomb D, Sawyer TK & Nerad TA (1985) A molecular approach to the phylogeny of *Acanthamoeba*. *BioSystems*, 18: 399-405.

10. Visvesvara GS (1991) Classification of *Acanthamoeba*. *Reviews of Infectious Diseases*, 13: S369-S372.
11. Stothard DR, Schroeder-Diedrich JM, Awwad MH, Gast RJ, Ledee DR, Rodriguez-Zaragoza S, Dean CL, Fuerst PA & Byers TJ (1998) The evolutionary history of the genus *Acanthamoeba* and the identification of eight new 18S rRNA gene sequence types. *Journal of Eukaryotic Microbiology*, 45: 45-54.
12. Yagita K, Endo T & De Jonckheere JF (1999) Clustering of *Acanthamoeba* isolates from human eye infections by means of mitochondrial DNA digestion patterns. *Parasitology Research*, 85: 284-289.
13. Costas M & Griffiths AJ (1985) Enzyme composition and the taxonomy of *Acanthamoeba*. *Journal of Protozoology*, 32: 604-607.
14. McLaughlin GL, Brandt FH & Visvesvara GS (1988) Restriction fragment length polymorphisms of the DNA of selected *Naegleria* and *Acanthamoeba* amoebae. *Journal of Clinical Microbiology*, 26: 1655-1658.
15. Kilvington S, Beeching J & White DG (1991) Differentiation of *Acanthamoeba* strains from infected corneas and the environment by using restriction endonuclease digestion of whole-cell DNA. *Journal of Clinical Microbiology*, 29: 310-314.
16. Kong HH & Chung DI (1996) PCR and RFLP variation of conserved region of small subunit ribosomal DNA among *Acanthamoeba* isolates assigned to either *A. castellanii* or *A. polyphaga*. *Korean Journal of Parasitology*, 34: 127-134.
17. Alves JMP, de Gusmão CX, Teixeira MMG, Freitas D, Foronda AS & Affonso HT (2000) Random amplified polymorphic DNA profiles as a tool for the characterization of Brazilian keratitis isolates of the genus *Acanthamoeba*. *Brazilian Journal of Medical and Biological Research*, 33 (1): 19-26.

18. De Jonckheere JF (1987) Characterization of *Naegleria* species by restriction endonuclease digestion of whole-cell DNA. *Molecular and Biochemical Parasitology*, 24: 55-66.
19. Sanguinetti, C.J., Neto, E.D. & Simpson, A.J.G. (1996) Recovery and PCR amplification of DNA fragments separated on PhastGel. *Pharmacia Biotech Application Note*, Feb: 39-40.
20. Armstrong J, Gibbs A, Peakall R & Weller G (1994) RAPDistance Programs, Version 1.03 for the analysis of patterns of RAPD fragments. Australian National University, Canberra.
21. Felsenstein J (1989) Phylip 3.2 Manual, University of California, Herbarium, Berkeley, CA.
22. Bohnert HJ, Herrmann RG (1974) The genomic complexity of *Acanthamoeba castellanii* mitochondrial DNA. *European Journal of Biochemistry*, 50: 83-90.
23. Byers TJ, Bogler SA, Burianek LL (1983) Analysis of mitochondrial DNA variation as an approach to systematic relationships in the genus *Acanthamoeba*. *Journal of Protozoology*, 30(2): 198-203.
24. Gauton RK, Lory S, Seyedirashti S, Bergeron DL, Fritsche TR (1994) Mitochondrial DNA fingerprinting of *Acanthamoeba* spp. isolated from clinical and environmental sources. *Journal of Clinical Microbiology*, 32(4): 1070-1073.
25. Bogler SA, Zarley CD, Burianek LL, Fuerst PA & Byers TJ (1983) Interstrain mitochondrial DNA polymorphism detected in *Acanthamoeba* by restriction endonucleases analysis. *Molecular and Biochemical Parasitology*, 8: 145-163.
26. Swofford, DL, Olsen, GJ, Waddell, PJ & Hillis, DM (1996). Phylogenetic Inference. In: Hillis, DM, Moritz, C. & Mable, BK (Editors), *Molecular Systematics*. 2nd edn. Sinauer Associates, Sunderland, MA.

Table 1

Characteristics of *Acanthamoeba* isolates

Isolate	Source	SSU rDNA type ^a
Page Res 20 (<i>A. astronyxis</i> ATCC 30901)	Water	nd
Nagington (<i>A. polyphaga</i> #1 ATCC 30873)	HAK	T4
Jones (<i>A. poly.</i> #2 ATCC 30461)	HAK	nd
Page 23 (<i>A. poly.</i> #3 ATCC 30871)	Water	T4
Page 45 (<i>A. poly.</i> #4 ATCC 30872)	Water	nd
Castellani (<i>A. castellanii</i> #1 ATCC 50374)	yeast culture	T4
Nagington (<i>A. cast.</i> #2 ATCC 30868)	HAK	nd
Oak Ridge (<i>A. royreba</i> ATCC 30884)	HCC	nd
U/E 1	HAK-CLW	nd
U/E 2	HAK-CLW	nd
U/E 3	HAK-CLW	nd
U/E 4	HAK-NCLW	nd
U/E 5	HAK-NCLW	nd
U/E 6	HAK-CLW	nd
U/E 7.1 ^b	HAK-CLW	nd
U/E 7.2 ^b		nd
U/E 8R ^c	HAK-CLW	nd
U/E 8L ^c		nd
U/HC1	HAK-CLW	nd

ATCC: American Type Culture Collection; nd: not determined; HAK: Human *Acanthamoeba* keratitis; CLW: contact lenses wearer; NCLW: non contact lenses wearer; HCC: human choriocarcinoma cells; U/E: USP/EPM; U/HC; USP/HC; ^aas defined by Stothard et al., 1998; ^btwo morphologically distinct clones isolated from the same corneal sample; ^cisolated from right (R) and left (L) eyes of the same patient.

Fig 1

Fig. 2

Fig. 3

Figure legends

Fig. 1 – Agarose gel (0.8%) electrophoresis of *Acanthamoeba* total DNA cleaved by the following restriction enzymes: **A** – *Bgl*II – 1: U/HC1; 2: U/E1; 3: U/E2; 4: U/E3; 5: U/E4; 6: U/E5; 7: U/E6. **B** – *Msp*I – 1: A. poly.#1; 2: A. poly.#2; 3: *A. royreba*; 4: U/HC1; 5: U/E4; 6: U/E5; 7: U/E6. **C** – *Msp*I – 1: *A. astronyxis*; 2: A. cast.#1; 3: A. cast.#2; 4: U/E7-1; 5: U/E7-2.

Fig. 2 – Polyacrylamide gel (6%) electrophoresis of *Acanthamoeba* SSU rDNA cleaved by restriction enzyme *Hinf*I. Molecular marker used is a 50 bp ladder.

Fig. 3 – Consensus phenogram constructed with data from seven restriction enzymes used (left) and RAPD phenogram for three primers, from Alves et al., 2000 (right). Numbers at the nodes indicate how many enzyme patterns supported each group (for the RAPD phenogram) and the bootstrap support for each group (for the RFLP phenogram).