

CONSTRUINDO ÁRVORES FILOGENÉTICAS COM O USO DE CAMINHADAS ALEATÓRIAS E GEOMETRIA FRACTAL

por

Luciana Renata de Oliveira

Monografia apresentada ao Curso de Bacharelado em Física do Centro de Ciência Naturais e Exatas, da Universidade Federal de Santa Maria (UFSM, RS), com requisito para a obtenção do grau de **Bacharel em Física.**

Orientador: **Prof. José Carlos Merino Mombach**

**Santa Maria, RS, Brasil
2008**

**UNIVERSIDADE FEDERAL DE SANTA MARIA
DEPARTAMENTO DE FÍSICA
CENTRO DE CIÊNCIAS NATURAIS E EXATAS**

A Comissão Examinadora, abaixo assinada,
a prova a Monografia de Graduação em Física

**CONSTRUINDO ÁRVORES FILOGENÉTICAS COM O USO DE
CAMINHADAS ALEATÓRIAS E GEOMETRIA FRACTAL**

elaborada por
Luciana Renata de Oliveira

como requisito parcial para obtenção do grau de
Bacharel em Física

COMISSÃO EXAMINADORA:

José Carlos Merino Mombach, Dr.
(Presidente/Orientador)

José Antônio Trindade Borges da Costa, Dr. (UFSM)

Celso Arami Marques da Silva, Dr. (UFSM)

Santa Maria, 28 de Novembro de 2008

A minha querida Mãe que tanto amo, ao meu Pai e ao
Marcelo.

AGRADECIMENTOS

A minha Mãe que é minha maior incentivadora e meu exemplo. Obrigada pelo apoio, dedicação, carinho, confiança, educação e principalmente pelo amor incondicional. Ao meu Pai, que além do apoio moral foi meu patrocinador durante toda a graduação. E a vocês dois por me darem a vida. Só por vocês eu cheguei onde estou e ainda vou muito mais longe. Amo vocês acima de tudo!

Ao meu orientador José Carlos Merino Mombach, pelos ensinamentos e confiança.

Ao meu melhor amigo Marcelo que é a minha família em Santa Maria. Obrigada pela paciência e por acreditar em mim mesmo quando nem eu mesma acreditava. Pelo companheirismo e amor de sempre.

Aos professores Juca, Celso, Orimar e Copetti por fazer a minha graduação valer a pena.

Aos meus colegas Carol, Jader e Caitano. E em especial ao Lucas que com certeza foi o que mais me ajudou durante esse período.

Aos meus queridos Mari e Fer por tudo que já fizeram por mim.

Ao Google que me salvou em todos os momentos de dificuldade.

RESUMO

Monografia

Curso de Bacharelado em Física
Centro de Ciências Naturais e Exatas
Universidade Federal de Santa Maria

Construindo Árvores Filogenéticas com o Uso de Caminhadas Aleatórias e Geometria Fractal

AUTOR (A): LUCIANA RENATA DE OLIVEIRA

ORIENTADOR: JOSÉ CARLOS MERINO MOMBACH

Data e Local da Defesa: Santa Maria, 28 de Novembro de 2008.

O objetivo deste trabalho é propor um método alternativo para construir árvores filogenéticas usando a dimensão fractal calculada a partir de caminhadas (pseudo) aleatórias realizadas sobre DNA mitocondrial e RNA ribossomal de algumas espécies de seres vivos. As caminhadas são realizadas atribuindo-se coordenadas espaciais a cada uma das bases nitrogenadas do DNA ou do RNA. As figuras geradas apresentam morfologias complexas e por isso utilizamos parâmetros como a dimensão fractal para caracterizarmos as diferentes morfologias encontradas. Os valores obtidos de dimensão fractal destas caminhadas são agrupados para gerarmos árvores de distância (dendogramas) aplicando o método de agrupamento hierárquico. Usamos o software MATHEMATICA 6.0 para programar as rotinas que constroem as caminhadas aleatórias e calcula os valores de dimensão fractal e também utilizamos a rotina deste software que faz o agrupamento e gera os dendogramas. Escolhemos dez espécies eucariotas para o estudo: *Anopheles gambiae* (mosquito da malária), *Arabidopsis thaliana* (uma planta), *Danio rerio* (peixe zebra), *Drosophila melanogaster* (mosca das frutas), *Homo sapiens* (humano), *Mus musculus* (camundongo), *Pan troglodytes* (chimpanzé), *Rattus norvegicus* (rato cobaia de laboratório), *Strongylocentrotus purpuratus* (ouriço do mar), *Xenopus laevis* (uma espécie de sapo). Obtivemos que para esses organismos o nosso método consegue fazer separação entre reinos, separando a planta dos animais. Entre os animais ele separa de vertebrados e invertebrados. Entre os invertebrados faz separação entre os insetos (mosca e mosquito) e o ouriço (equinoderma). Já entre os vertebrados, separa os mamíferos do anfíbio e do peixe. Concluimos que o método gera resultados consistentes com os preditos pela filogenia tradicional.

Palavras-chave: Evolução, Árvores Filogenéticas, Dimensão Fractal, Caminhadas Aleatórias.

ABSTRACT

Monograph

Curso de Bacharelado em Física
Centro de Ciências Naturais e Exatas
Universidade Federal de Santa Maria

**CONSTRUCTING PHYLOGENETIC TREES USING RANDOM
WALKS AND FRACTAL GEOMETRY**

AUTHOR: LUCIANA RENATA DE OLIVEIRA

ADVISER: JOSÉ CARLOS MERINO MOMBACH

Santa Maria, November 28th 2008

The purpose of this study is to propose an alternative method to build phylogenetic trees using the fractal dimension calculated from (pseudo) random walks on mitochondrial DNA and ribosomal RNA of some species of living beings. The walks are performed attributing coordinates to each of the nitrogenous bases of DNA or RNA. The generated figures show complex morphologies and therefore we use parameters such as fractal dimension to characterize the different morphologies found. The obtained fractal dimensions of these figures are grouped to generate distance trees (dendograms) using the method of hierarchical clustering. We use the software Mathematica 6.0 to program the routines that build the random walks and calculates the values of fractal dimension and we also use the routine of the software that does the clustering and generates the dendograms. We have chosen ten eukaryotic species for the study: *Anopheles gambiae* (malaria mosquito), *Arabidopsis thaliana* (a plant), *Danio rerio* (zebra fish), *Drosophila melanogaster* (the fruit fly), *Homo sapiens* (human), *Mus musculus* (mouse), *Pan troglodytes* (chimpanzee), *Rattus norvegicus* (laboratory rat), *Strongylocentrotus purpuratus* (sea hedgehog), *Xenopus laevis* (a species of frog). We found that for these organisms our method can distinguish kingdoms, e.g. the plant from the animals. Among the animals, it distinguishes vertebrates from invertebrates. Among invertebrates, it distinguishes insects (fly and mosquito) from echinoderms (hedgehog). Among the vertebrates, it distinguishes the amphibian from mammals and fish. We conclude that the method generates results consistent with those predicted by the traditional phylogeny.

Key words: Evolution, Phylogenetic trees, Fractal Dimension, Random Walks.

LISTA DE TABELAS

TABELA 3.1 – Representação do alinhamento de três seqüências de RNA.....	24
TABELA 5.1– Valores das dimensões fractais para linhas com diferentes comprimentos.....	46
TABELA 5.2 – Valores das dimensões fractais para caminhadas feitas sobre mtDNA.....	48
TABELA 5.3 – Valores das dimensões fractais para caminhadas feitas sobre rRNA.....	51

LISTA DE FIGURAS

FIGURA 3.1.1 – Célula animal.....	14
Fonte: http://www.obrasill.com/celulas	
FIGURA 3.1.2– (a) Exemplo da estrutura de um nucleotídeo; (b) Dupla fita do DNA.....	16
Fonte: (a) http://www.nutritotal.com.br/textos/files/44--nucleotideo.jpg ; (b) http://www.ncbi.nlm.nih.gov:80/books/bookres.fcgi/mga/ch2f2.gif	
FIGURA 3.1.3 – DNA mitocondrial.....	17
Fonte: www.fairfaxforensicslab.com	
FIGURA 3.1.4– Estrutura do RNA.....	19
Fonte: www.heroesbrasil.org	
FIGURA 3.1.5– Representação de uma árvore filogenética gerada por cladística.....	23
FIGURA 3.2.1– Representação do conjunto de Cantor.....	26
Fonte: www.unama.br	
FIGURA 3.2.2– Construção Básica do tapete de Sierpinski.....	27
Fonte: www.unama.br	
FIGURA 3.2.3– Construção da curva de Koch.....	28
Fonte: http://www.ceticismoaberto.com/imagens4/curvakoch.jpg .	
FIGURA 3.2.4– Esponja de Menger.....	29
Fonte: www.unama.br	
FIGURA 3.2.5– Conjunto de Cantor com número de réguas e seus respectivos tamanos para cada um dos processos.....	32
Fonte: http://www.if.ufg.br/projetocaos/minicurso/aula5.htm	
FIGURA 3.3.1– Representação de um dendograma gerado por análise de agrupamento hierárquica.....	35
FIGURA 3.3.2– Exemplo de agrupamento hierárquico considerando a distância entre cidades do interior paulista.....	36
Fonte: www.uspleste.usp.br/rvicente/BioInfo_jefferson_PIBIC05.pdf	
FIGURA 3.3.3– Dendograma gerado com análise aglomerativa hierárquica para a distância entre as cidades.....	37
Fonte: www.resample.com	

FIGURA 4.1.1-(a) Representação do eixo de coordenadas atribuído as bases do mtDNA; (b) Representação do eixo de coordenadas atribuído as bases do rRNA.....	39
FIGURA 4.1.2– Gráfico das coordenadas geradas pela rotina de caminhada aleatória sobre seqüenciamento de bases arbitrário.....	40
FIGURA 4.2.1– Método de Box Counting aplicado a Curva de Koch, para dois diferentes tamanhos de	42
Fonte: http://www.cbpf.br/~maysagm/	
FIGURA 4.3.1– Dendograma gerado com análise aglomerativa hierárquica.....	43
FIGURA 5.1 – (a) Caminhada sobre mtDNA de <i>Drosophila melanogaster</i> por C.L.Berthelsen em seu trabalho (b) Caminhada sobre mtDNA de <i>Drosophila melanogaster</i> gerada pela nossa rotina.....	45
FIGURA 5.2 – Gráficos das caminhadas feitas sobre mtDNA.....	48
FIGURA 5.3 – Dendograma gerado a partir do valor da dimensão fractal do mtDNA.....	49
FIGURA 5.4 – Gráficos das caminhadas feitas sobre rRNA.....	50
FIGURA 5.5 – Dendograma gerado a partir do valor da dimensão fractal do rRNA.....	51
FIGURA 5.6 – Árvore filogenética construída a partir do valor da D_F do mtDNA e rRNA dos organismos.....	52

LISTA DE APÊNDICES

APÊNDICE A – Algoritmo de Caminhada Aleatória.....	12
APÊNDICE B– Algoritmo de Cálculo de Dimensão Fractal.....	18

SUMÁRIO

1. INTRODUÇÃO.....	12
2. OBJETIVOS.....	13
3. CONSIDERAÇÕES GERAIS.....	14
3.1 Biologia Molecular e Análise Filogenética.....	14
3.1.1 Células	14
3.1.2 DNA.....	15
3.1.3 DNA Mitocondrial	17
3.1.4 RNA.....	18
3.1.5 A Evolução da Vida.....	20
3.1.6 Sistemática Filogenética.....	21
3.1.7 Cladística.....	22
3.1.8 Análise filogenética molecular.....	23
3.1.9 Importância da construção de Árvores Filogenéticas.....	24
3.2 Análise de Agrupamento.....	24
3.2.1 Agrupamento Hierárquico.....	25
3.3 Fractais.....	28
3.3.1 O conjunto de Cantor.....	28
3.3.2 Tapete de Sierpinski.....	29
3.3.3 Curva de Koch.....	31
3.3.4 Esponja de Menger.....	32
3.3.5 Benoit Mandelbrot e a Geometria Fractal.....	32
4. METODOLOGIA	37
4.1 Caminhada Aleatória.....	39
.....	41
4.2 Cálculo da Dimensão Fractal (Método do Box Counting).....	41
4.3 Agrupamento.....	42
6. CONCLUSÃO.....	52
7. REFERÊNCIAS.....	53
APÊNDICES.....	56
APÊNDICE A – Algoritmo de Caminhada Aleatória.....	57
APÊNDICE B– Algoritmo de Cálculo de Dimensão Fractal.....	58

1. INTRODUÇÃO

O desenvolvimento do projeto Genoma humano proporcionou um grande avanço científico, trazendo uma vasta quantidade de dados a serem analisados. Dando vez a contribuições de várias áreas do conhecimento, pois para uma descrição completa dos organismos, não basta determinar a sua seqüência de nucleotídeos (DNA), precisa-se também saber quais são as leis dinâmicas que as regem e o modo como estas seqüências sofreram modificações ao longo do tempo.

Uma ferramenta útil para tratamento de dados é o uso da geometria fractal, pois ela traz uma descrição alternativa de dimensão frente à definição dada pela geometria euclidiana, assumindo que a medida da dimensão possa resultar em valores não inteiros como 1.5, 2.7, $\pi/6$, $\log 2$, etc. Essa característica é importante devido ao fato que algumas figuras irregulares não possuem medidas satisfatórias de sua dimensão em termos da geometria euclidiana. Dessa maneira, um objeto que não é uma reta e nem um plano, terá uma dimensão que será um valor entre um e dois. É fácil encontrar exemplos naturais de figuras que têm a dimensão determinada de forma mais eficiente pela geometria fractal, entre elas podemos citar a superfície de uma árvore, o limite entre a costa de uma ilha e o mar, as nuvens ou o gráfico gerado pela atribuição de coordenadas à seqüência de nucleotídeos constituintes do DNA.

Glazier e colaboradores propuseram o uso da geometria fractal (e do espectro multifractal) para construção de árvores filogenéticas que estabelecem distancia evolutiva entre espécies diferentes a partir do DNA mitocondrial (mtDNA). Os autores representam a seqüência de mtDNA como uma caminhada aleatória em espaços com dimensões de 2 a 4 e calculam a dimensão fractal (e espectro multifractal) do objeto gerado o que fornece informação detalhada sobre ele. Baseando-se nessa idéia, pretende-se construir árvores filogenéticas a partir apenas do valor da dimensão fractal de caminhadas aleatórias feitas sobre DNA mitocondrial e RNA ribossomal.

2. OBJETIVOS

O objetivo deste trabalho é propor um método alternativo para construção de árvores filogenéticas, que comparem proximidade evolutiva, a partir do valor da dimensão fractal das caminhadas (pseudo) aleatórias geradas sobre o DNA mitocondrial e RNA ribossomal de dez espécies diferentes de organismos. Os valores obtidos de dimensão fractal destas caminhadas são agrupados para gerar árvores de distância (dendogramas) aplicando o método de agrupamento hierárquico.

3. CONSIDERAÇÕES GERAIS

3.1 Biologia Molecular e Análise Filogenética

3.1.1 Células

As células são consideradas as unidades fundamentais da vida. Nelas realizam-se a maior parte dos processos metabólicos dos seres vivos e é onde está contido o material genético (DNA).

A organização interna de uma célula divide o mundo dos seres vivos em dois domínios de organismos: eucariotos e procariotos. Nas células dos eucariotos são encontrados compartimentos internos bem definidos como o núcleo e organelas. Já as células de procariotos não possuem compartimentos internos e em particular não possuem núcleo. Nas células de organismos eucariotos o material genético reside no núcleo, e nas células dos procariotos o material genético permanece livre no citoplasma. Na Figura 3.1.1 está representada uma célula eucariota animal, com suas organelas indicadas.

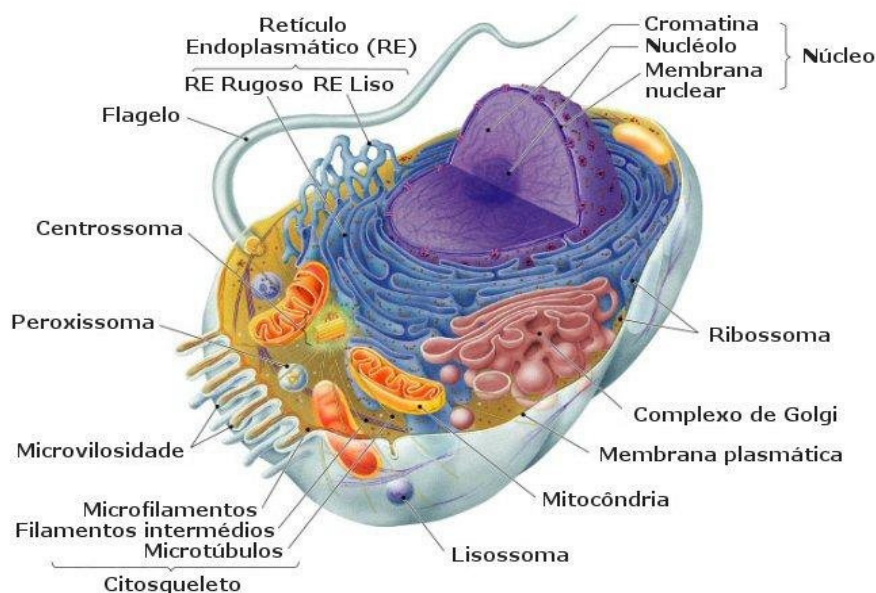


Figura 3.1.1- Célula animal

As células determinam de forma geral as características de cada organismo, sendo importante a sua estrutura e função. Diante disso, precisamos saber de que

forma os genes de cada célula determinam as características de uma espécie e de que maneira contribuem para a variação dentro das espécies. Assim, precisamos conhecer a natureza do material genético, sua localização nas células e a maneira como funcionam em ambiente celular.

3.1.2 DNA

O DNA é onde está contida toda a informação genética de um organismo, e apresenta três características que são responsáveis pela continuidade da vida, e são elas:

- (i) Função Genotípica – replicação. O material genético deve estocar a informação genética e transmitir com precisão essa informação para a prole, geração após geração.
- (ii) Função Fenotípica – expressão gênica. O material genético deve controlar o desenvolvimento do fenótipo de um organismo. Isto é, o material genético deve ditar o crescimento do organismo desde o zigoto unicelular até o adulto.
- (iii) Função Evolutiva – as mutações. O material genético deve sofrer mudanças para produzir variações que permitam aos organismos adaptar-se a modificações no ambiente de modo que possa ocorrer evolução.

Sua estrutura é composta por um grupo fosfato, um açúcar chamado desoxirribose e uma base nitrogenada. Existem quatro tipos de base nitrogenada, que se dividem em dois subgrupos devido às suas diferentes estruturas, Citosina (C) e Timina (T) possuem uma estrutura em anel chamado de pirimidina e as outras são a Adenina (A) e a Guanina (G), que possuem uma estrutura de anel duplo conhecidas como purina. As bases nitrogenadas são chamadas assim, pois são formadas por moléculas de carbono e nitrogênio.

O arranjo dos componentes químicos do DNA é chamado de nucleotídeo, cada um é formado por um grupo fosfato, o açúcar e qualquer uma das quatro bases. Por conveniência nos referimos a cada uma das bases pela primeira letra do seu nome, ou seja, A, C, T, G. Quem guarda toda a informação genética no DNA são as bases nitrogenadas, sendo que a combinação açúcar mais fosfato apresenta

apenas função estrutural, e são invariáveis nos nucleotídeos. Um exemplo de estrutura de um nucleotídeo pode ser visto na Figura 3.1.2(a).

O DNA é uma dupla hélice retorcida composta por duas cadeias de nucleotídeos. O suporte do filamento é feito pela repetição do polímero fosfato-açúcar da desoxirribose. São chamadas de ligações fosfodiéster as ligações fosfato-açúcar nesse suporte. Os filamentos permanecem ligados por fracas ligações de pontes de hidrogênio, na dupla fita do DNA os dois suportes estão em sentido inverso como podemos conferir na Figura 3.1.1(b). Essa estruturação é justificada pela complementaridade das fitas de DNA, se possuirmos apenas uma das fitas de DNA podemos determinar de forma precisa como é a sua fita complementar. Isso por que as bases do DNA seguem uma regra simples de pareamento, A paraia com T e G paraia com C, o que também pode ser conferido na Figura 3.1.2(b)

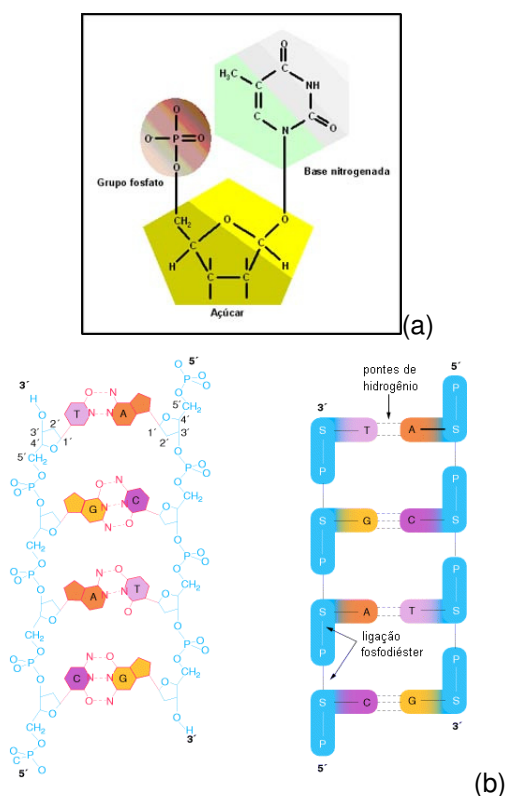


Figura 3.1.2- (a) Exemplo da estrutura de um nucleotídeo. (b) Dupla fita do DNA.

A dupla hélice é automaticamente assumida, devido ao pareamento das bases. Porém é importante salientar que uma única fita de DNA é retilínea, ficando reservada a dupla fita à estrutura de hélice.

3.1.3 DNA Mitocondrial

As células eucariontes apresentam dois tipos de materiais genéticos: o nuclear e o organelar, esse último está presente em mitocôndrias e em cloroplastos, os quais contêm um tipo específico de cromossomo da organela. Os genes dessas organelas contêm funções específicas para cada uma delas, mas mesmo assim a maioria das funções descritas por elas são comandadas pelos genes de origem nuclear.

O DNA mitocondrial (mtDNA) está contido na organela celular chamada mitocôndria, que é responsável pela respiração celular, essas organelas contêm DNA próprio e independente do DNA nuclear, os três tipos de RNA (rRNA, mRNA e tRNA) e todo o sistema molecular para a síntese de algumas proteínas. O mtDNA é relativamente pequeno, tendo tamanho na faixa de 16 mil pares de bases, e se apresenta na forma circular como o DNA bacteriano. O mtDNA é uma organela de origem exclusivamente materna. Podemos verificar a forma do DNA mitocondrial na Figura 3.1.3.

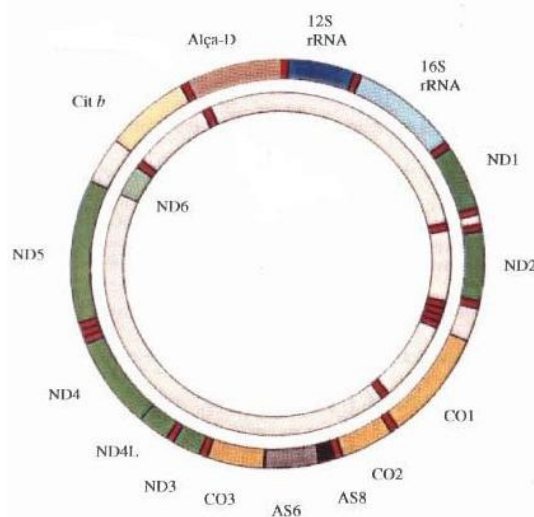


Figura 3.1.3. - DNA mitocondrial

O mtDNA está presente em grande quantidade em todas as células (entre 1.000 e 10.000 moléculas de mtDNA por célula), o que lhe confere um menor risco de degradação em relação ao DNA nuclear. Por ele estar protegido por uma membrana, o torna extremamente importante quando do estudo de espécimes biológicas antigas, cujo DNA nuclear, em cópia única, em geral, encontra-se

degradado. Por meio do conhecimento adquirido sobre a estrutura do mtDNA, o estudo da evolução humana experimentou enormes avanços, tendo em vista a possibilidade de se analisar com êxito, restos humanos antigos, como tecido cerebral de 7.000 anos de idade e ossos de até 5.500 anos.

Griffiths (2000, p.34) comenta a respeito da origem dessas organelas que “As mitocôndrias e os cloroplastos provavelmente foram originalmente células procarióticas que entraram e assumiram uma relação simbiótica em outra célula”. O genoma da mitocondrial na maioria das vezes não pode ser eliminado do organismo, por ainda existir uma interdependência entre ele e o organismo, mas em alguns casos, como por exemplo, em leveduras ele pode ser eliminado sem causar nenhum mal a célula. Já os cloroplastos só podem ser eliminados em células fotossintéticas, que conseguem sobreviver captando nutrientes pré-formados do meio ambiente.

3.1.4 RNA

Os ácidos ribonucléicos (RNAs) são os produtos iniciais de todos os genes, são produzidos através de um processo conhecido como transcrição, que faz uma cópia dos nucleotídeos de DNA. O processo de transcrição do RNA é semelhante ao processo de replicação do DNA, mas com algumas diferenças básicas.

GRIFFITHS (2000, p.50) faz uma referência entre a função do genoma e do RNA dizendo que “o DNA é o manual de instruções para a produção de todos os RNAs de que a célula precisa, enquanto o RNA é a informação impressa que pode ser apagada desta parte do manual relevante a qualquer tarefa determinada.”

Apesar de tanto DNA quanto RNA serem ácidos nucleicos eles possuem diferenças essenciais. A primeira é que o RNA é uma cadeia unifilar de nucleotídeos, e não uma dupla hélice. Como consequência o RNA tem a possibilidade de formar uma variedade muito maior de formas moleculares tridimensionais complexas. A segunda é que o RNA possui açúcar ribose em seus nucleotídeos, em vez de desoxirribose. O que difere esses dois açúcares é a presença ou ausência de apenas um oxigênio. Por fim os nucleotídeos do RNA transportam bases adenina, guanina e citosina, porém a base uracila encontra-se no lugar da base timina. Podemos ver a estrutura do RNA na Figura 3.1.4.

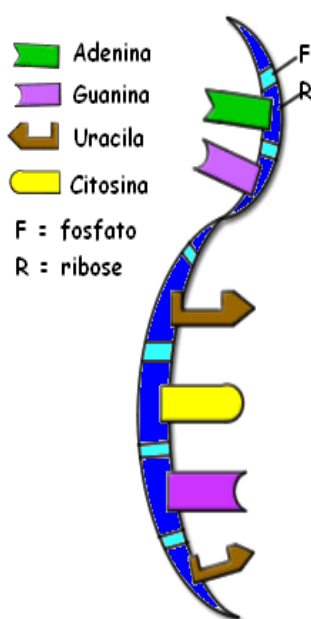


Figura 3.1.4- Estrutura do RNA

Levando em consideração apenas aspectos estruturais, os RNAs são separados em três grupos:

1. RNA de transferência (tRNA): Ele deve transferir os aminoácidos para as suas posições corretas nas cadeias polipeptídicas em formação nos complexos de ribossomos e RNA mensageiro.
2. RNA mensageiro (mRNA): sua função é de ser intermediário entre o DNA e o produto funcional, a proteína. A seqüência de nucleotídeos no mRNA é convertida na seqüência de aminoácidos nas cadeias polipeptídicas.
3. RNA ribossômico (rRNA), como esse é o RNA mais importante para esse trabalho ele será discutido mais detalhadamente.

O RNA ribossômico existe combinado com proteínas, formando partículas livres chamadas ribossomos. Os ribossomos são as máquinas de produção de proteínas de todas as células vivas. A função do rRNA é fornecer um mecanismo de decodificação de mRNA em aminoácidos e no caso dos tRNAs, os rRNAs são componentes gerais da tradução que podem ser utilizados para traduzir o mRNA de qualquer gene que codifica a proteína.

Tanto em eucariotos como em procariotos os rRNAs podem ser divididos em duas subunidades, uma maior e outra menor, com características funcionais e

estruturais diferentes. As subunidades se prendem de modo reversível no início da síntese da molécula protéica, separando-se quando a proteína está terminada. A subunidade maior dos ribossomos das células eucariontes contém três tipos de RNA, com dimensões de 28S, 58S e 5S, e a dos ribossomos dos procariontes possui dois tipos de RNA: um de 23S e outro de 5S. A subunidade menor apresenta apenas um tipo de RNA: 18S nas células eucariontes e 16S nas células procariontes. A letra S representa o coeficiente de sedimentação determinado na ultracentrífuga, expressa a unidade Svedberg.

Os ribossomos das mitocôndrias e dos cloroplastos são iguais aos das células procariontes, dado que apóia a interpretação de que essas duas organelas produtoras de energia se originam de bactérias que se tornaram simbiontes nas células eucariontes.

RNA ribossomal têm características que são importantes na medicina e na evolução. Na medicina por serem alvo de muitos antibióticos. E para a evolução, pois são as moléculas mais conservadas (menos variáveis) em todas as células. Por este motivo, os genes que codificam rRNA são seqüenciados e usados para construir a filogenia de organismos.

3.1.5 A Evolução da Vida

A descrição da evolução das espécies está toda baseada nos trabalhos dos naturalistas ingleses Charles Darwin e Alfred Russel Wallace. VIANA (p.9, 2007) menciona que “Pesquisas em diferentes áreas da Biologia têm evidenciado que o processo evolutivo proposto é o responsável pela diversidade, ou seja, as espécies se diversificam originando outras pela seleção natural”.

Atualmente, devido ao avanço e aprimoramento de técnicas de análise a biologia molecular tem acesso a diversos tipos de informação sobre os organismos, como por exemplo, a descrição completa da genética de alguns deles. O que fica evidente é que as espécies estão relacionadas através de sua estrutura molecular, isso por que todos são formados por estruturas compostas por quatro tipos de nucleotídeos e pelos mesmos vinte tipos de aminoácidos. A manifestação dessas características nos diferentes organismos, segundo o evolucionismo, podem unir ou separar uma classificação biológica. Uma forma de mostrar isto é através das árvores filogenéticas, discutidas na próxima seção. Segundo JUNQUEIRA (2005,

p.16) “Os avanços tecnológicos da biologia molecular possibilitaram o estudo mais aprimorado das relações evolutivas entre as espécies”.

Conseguir construir um histograma que descreva de que maneira as espécies evoluíram desde os seus ancestrais mais remotos não é uma tarefa simples. A origem de uma nova espécie, a partir de um ancestral, ocorre a partir do processo conhecido como especiação. Esse não é um processo que possa ser observado ao longo do tempo, e a análise das características dos fósseis, em alguns casos, não é possível devido a seu estado de conservação. Sendo assim, as relações filogenéticas devem ser inferidas, ou seja, o estudo das relações entre as espécies leva a uma estrutura provável de suas origens, pois existem inúmeras possibilidades de ramificações que representam os relacionamentos. Segundo FUTUYAMA (1996, p. 308) “Assim como outras hipóteses científicas, as árvores filogenéticas não podem ser provadas com certeza absoluta.” Apesar disso, os resultados obtidos com análise filogenética são valiosos e importantes, sendo aceitos por grande parte dos cientistas que acreditam.

3.1.6 Sistemática Filogenética

O estudo da sistemática evolutiva está baseado na compreensão das relações entre espécies para inferir a história da vida desde a sua origem, fazendo sua representação através de um diagrama chamado de *Árvore Filogenética* ou *Filogenia*.

As árvores filogenéticas eram construídas utilizando-se métodos intuitivos com base no fenótipo das espécies, o que levava a conclusões errôneas, pois nada impede que dois organismos que se originam de ancestrais distintos possam ter desenvolvido características semelhantes. Filogenia pode ser definida como o estudo das relações ancestrais entre espécies conhecidas. A origem da palavra filogenia está baseada nas palavras gregas “phylon” referente à tribo e “genetikos” relativo à origem. Viana (p.21, 2007) “Uma árvore pode retratar a evolução de toda a vida, de todas as grandes linhagens evolutivas ou de somente um pequeno grupo de organismos.”. Uma maneira de se fazer análise filogenética é chamada cladística e é descrita na seção 3.1.7.

3.1.7 Cladística

Segundo VIEIRA (p.15, 2007) “Willi Hennig, entomólogo alemão, desenvolveu as idéias que levariam à criação da Cladística”. A Cladística está baseada no princípio fundamental de que os organismos devem ser classificados de acordo com as suas relações evolutivas e que a forma de descobrir essas relações é analisar aquilo que se designa como características ancestrais (“primitivas”) e características derivadas (“evoluídas”).

As características ancestrais são compartilhadas por todos os seres vivos (plantas ou animais), herdadas com pouca ou nenhuma alteração desde ancestrais remotos. Quanto às características derivadas são as que sofreram mudanças recentes. O cérebro volumoso dos humanos, por exemplo, é uma característica derivada comparada com o de outros primatas, enquanto nosso membro pentálico (com cinco dedos) é um estado primitivo no sentido de que o número de dedos não mudou desde nossos ancestrais anfíbios do Devoniano.

O resultado obtido por análise cladística pode ser representado por uma árvore filogenética, que representa as relações evolutivas entre várias espécies que tem um antepassado comum, como a representada na Figura 3.1.5.



Figura 3.1.5- Representação de uma árvore filogenética gerada por cladística

A partir da Figura 3.1.5 podemos definir alguns conceitos importantes sobre a representação em árvores da análise filogenética.

Clado é um grupo de organismos originados de um único ancestral comum. O clado faz parte de uma hipótese científica de modelo relacional evolucionário entre os organismos incluídos na análise filogenética. Um clado particular pode ser

sustentado ou não diante de uma análise subsequente usando um conjunto diferente de dados ou de um modelo distinto de evolução.

Táxon é um grupo de organismos reais que preenchem uma categoria, no caso de uma árvore filogenética podendo ser um gênero, família, ordem, classe ou filo. Um táxon é compreendido como retratando o fato de que seus membros são derivados de uma ancestral comum.

Nodo é o ponto onde o galho se bifurca, representando a separação evolutiva entre as espécies.

Na análise cladística o comprimento dos galhos representa as divergências, isto é, os organismos que são mais distantes têm galhos mais longos, e os mais próximos mais curtos. Podemos verificar na figura 3.1.5 que o rato e o homem são mais próximos do que o homem e a mosca devido à análise dos comprimentos dos galhos da árvore.

3.1.8 Análise filogenética molecular

Para construção de árvores filogenéticas a partir da informação molecular, usa-se alinhamento de seqüências. VIANA (p.30, 2007) diz que “Um alinhamento procura determinar o grau de dissimilaridade entre as seqüências comparadas, na sua totalidade ou entre fragmentos das mesmas.” Essa técnica está baseada em se sobrepor duas ou mais seqüências de nucleotídeos, ou aminoácidos, e comparar as diferenças entre elas. A partir dessas diferenças se faz uma medida da dissimilaridade ou proximidade entre as espécies analisadas. O tamanho das seqüências normalmente é diferente entre as espécies, sendo necessário introduzir outros tipos de caracteres para que elas fiquem alinhadas. Esses caracteres são chamados de *gaps*, usualmente representados por “-”. Segundo BRINKMAN (p.333, 2001) “No alinhamento estão inclusas as variações no comprimento, o conjunto de dados filogenéticos normalmente não é igual aos dados alinhados.” Um exemplo de alinhamento feito sobre três seqüências hipotéticas de RNA é representado na Tabela 3.1.

Seqüência	Caracteres	Tamanho
1	AUGUCA AUGGACACAUAUCACGACU	25
2	AU - - CAAAGGACACGUAUC - CGAUU	22
3	GUGUCA - GGCCACA - - - ACGACU	20

Tabela 3.1- Representação do alinhamento de três seqüências de RNA

Podemos verificar na tabela que existem diferenças nas seqüências tanto de tamanho quanto em substituições de base. Na análise filogenética molecular essas são as informações levadas em conta para se fazer a comparação e a construção da árvore filogenética. Essa área da biologia é bastante extensa e aqui não teremos espaço para abordar o assunto com muita profundidade recomenda-se a bibliografia Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins. BAXEVANIS, A.D.; OUELLETTE, B.F.F. 2ª Ed. New York: Wiley-Liss, 2001.

3.1.9 Importância da construção de Árvores Filogenéticas

As árvores filogenéticas contêm informações úteis na investigação científica de uma variedade de questões biológicas. Na biologia molecular, por exemplo, técnicas filogenéticas são usadas para compreender a diferenciação de macromoléculas em grandes grupos, através do uso de técnicas de seqüenciamento de proteínas e ácidos nucleicos. Auxiliam na solução de problemas práticos como o controle e combate a parasitas (protozoários, vírus, metazoários e fungos). Pois, sabendo as relações entre esses organismos a farmacologia tem a possibilidade de procurar fármacos de alcance imunológico mais amplo. Outra utilidade muito importante é que através da ordenação dos caracteres pode-se determinar o nível evolutivo do surgimento de enzimas e mecanismos de controle encontrados no ser humano, sendo possível assim testar a reação imunológica ou toxicidade em outros sistemas biológicos similares antes de sua aplicação no homem.

3.2 Análise de Agrupamento

A técnica de agrupamento (do inglês *Clustering*) consiste basicamente em se formar grupos (*clusters*) visando maximizar suas similaridades. O pressuposto básico é que os elementos de um grupo devem ser mais semelhantes entre si do

são com os membros dos outros grupos. “Intuitivamente, padrões pertencentes a um dado cluster devem ser mais similares entre si do que em relação a padrões pertencentes a outros clusters.”¹

Existem diferentes métodos de análise de agrupamento, segundo Pires (p.4, 2008) “A análise de agrupamento de dados é realizada por diversos métodos e cada um com uma diferente abordagem.” Dentre esses métodos os mais conhecidos são os hierárquicos, por particionamento e incrementais. Nossa ênfase será nos métodos hierárquicos, pois o usamos no desenvolvimento desse trabalho.

3.2.1 Agrupamento Hierárquico

Esse tipo de agrupamento é classicamente representado por uma árvore conhecida como dendograma, com elementos individuais em uma extremidade e um único grupo contendo todos os elementos em outra.

O agrupamento hierárquico pode ser aglomerativo ou divisório, o que diferencia esses dois métodos é a maneira como acontece o procedimento. No agrupamento aglomerativo, para um certo conjunto de dados, cada membro do conjunto representa inicialmente um grupo, entre todos esses se procura os mais semelhantes e agrupasse dependendo da sua similaridade, isso é repetido em etapas até que no final todos os elementos fazem parte de um único grupo. No divisivo, acontece o contrário, o processo inicia-se com apenas um grupo contendo todos os membros e vai se dividindo até que cada elemento represente um único grupo.

A distância entre os grupos pode ser calculada de maneiras diferentes, entre as mais conhecidas estão² a Distância Euclidiana, Distância Manhattan, Distância *City Block*, Distância de Chebychev, Distância Potencial e Percentagem de Discordância. Cada uma calculada de maneira diferente. Usamos em nosso trabalho a Distância Manhattan, e seu cálculo aparece na seção 4.3 da Metodologia.

Também existem diferentes maneiras de se medir a dissimilaridade, entre os grupos. Entre os mais populares estão o vizinho próximo (*single-link*) e o vizinho distante (*complete-link*). No vizinho próximo o que se considera é a menor distância entre dois grupos, se os grupos contêm mais de um elemento considera-se os

¹ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/ia368_02/topico5_02.pdf

² <http://www.dmat.furg.br/~taba/posanaagr.htm>

elementos mais próximos de cada um dos grupos para se fazer a comparação. No caso do vizinho distante se considera os elementos mais distantes. Segundo PIRES (p.4, 2008) “a diferença no resultado entre os dois métodos é o formato dos grupos formados.” Na Figura 3.2.1 pode-se conferir a representação de um dendograma gerado pelo método de agrupamento hierárquico, as flechas indicam qual o sentido de uma análise aglomerativa e divisível.

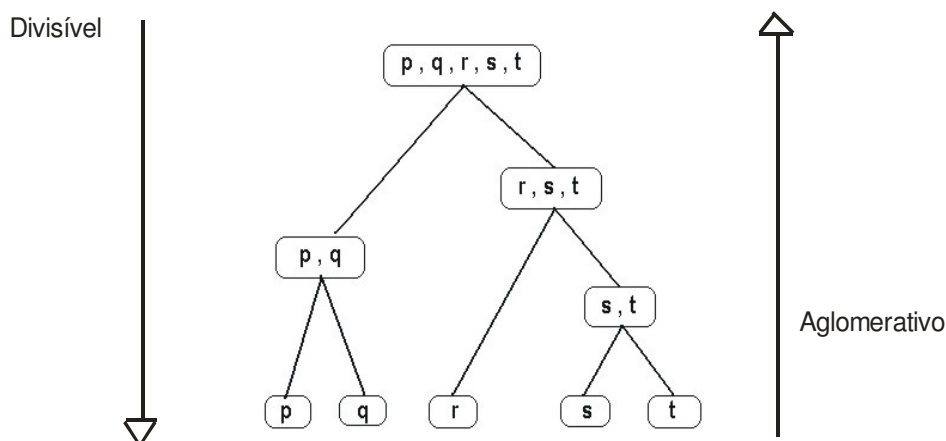


Figura 3.2.1- Representação de um dendograma gerado por análise de agrupamento hierárquica.

Consideremos o exemplo dado por VICENTE (2006 apud FREITAS, 2006, p.11) para esclarecer como o acontece o método de agrupamento hierárquico aglomerativo. Vamos analisar a distância entre algumas cidades do interior paulista. Observemos a Figura 3.2.2, nela podemos ver o mapa com as cidades de Ribeirão Preto, São Carlos, Piraçununga, Rio Claro, Araras, Piracicaba, Americana, Botucatu e Campinas. O primeiro passo é a união das cidades que tem a menor distância de separação, no caso, Figura 3.2.2 (b), são as cidades Rio Claro e Araras. Em seguida devem-se unir as cidades mais próximas depois de Rio Claro e Araras, que podemos ver em 3.2.1(c) que são Piracicaba e Americana, esse processo deve ser repetido sucessivamente até que todas as cidades já estejam ligadas, como em 3.2.2(i).





Figura 3.2.2- Exemplo de agrupamento hierárquico considerando a distância entre cidades do interior paulista. (a) Mapa das cidades; (b) Entre as nove cidades as mais próximas são Rio Claro e Araras por isso elas são as primeiras a serem agrupadas; (c) Depois de Rio Claro e Araras as mais próximas são Piracicaba e Americana; (d) Nesse passo Rio Claro e Piracicaba são as mais próximas (e), (f), (g) e (h) repetindo o processo. (i) Finalmente todas as cidades estão ligadas, representando um único grupo.

Depois de medidas essas distâncias e aplicado o método de agrupamento hierárquico aglomerativo obtém-se como resultado o dendograma da Figura 3.2.3.

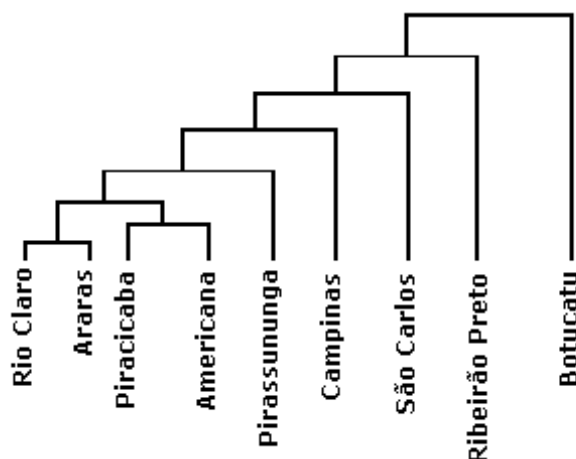


Figura 3.2.3- Dendrograma gerado com análise aglomerativa hierárquica para as distâncias entre as cidades.

3.3 Fractais

A idéia dos fractais teve a sua origem no trabalho de alguns cientistas entre 1857 e 1913. Esses trabalhos introduziram alguns objetos que foram catalogados como "demônios" ou "monstros" e que se supunha não conterem grande valor científico. Nessa época os fractais eram chamados de curva sem derivada e os principais fractais são:

3.3.1 O conjunto de Cantor

Esse fractal foi construído pelo matemático alemão Georg Cantor (1845-1918), a ele é atribuído o mérito de apresentar a teoria dos conjuntos como conhecemos modernamente. Esse conjunto consiste de um conjunto infinito de pontos contidos em um intervalo unitário $[0,1]$. Isto é, podemos interpretar como um conjunto de certos números, como 0, 1, $1/3$, $2/3$, $1/9$, $2/9$, $7/9$, $8/9$, $1/27$...

Partimos do segmento $[0,1]$ e retiramos o terço central, designado por $]1/3,2/3[$. Resultando em dois intervalos fechados $[0,1/3]$ e $[2/3, 1]$, com comprimento $1/3$ cada. Repetindo o processo com os intervalos $[0,1/3]$ e $[2/3, 1]$ removemos os terços centrais, o que resulta em intervalos de comprimento $1/9$. Então, em cada um dos terços restantes se retira o terço do meio, e assim por diante até o infinito. Observando que no estágio inicial temos um intervalo, depois dois, depois quatro e assim por diante, podemos inferir que a regra que rege o crescimento do número de intervalos e a diminuição do comprimento é dada por 2^n intervalos de comprimento

$1/3n$ depois de n passos. Na Figura 3.3.1 podemos ver a representação do conjunto de Cantor.

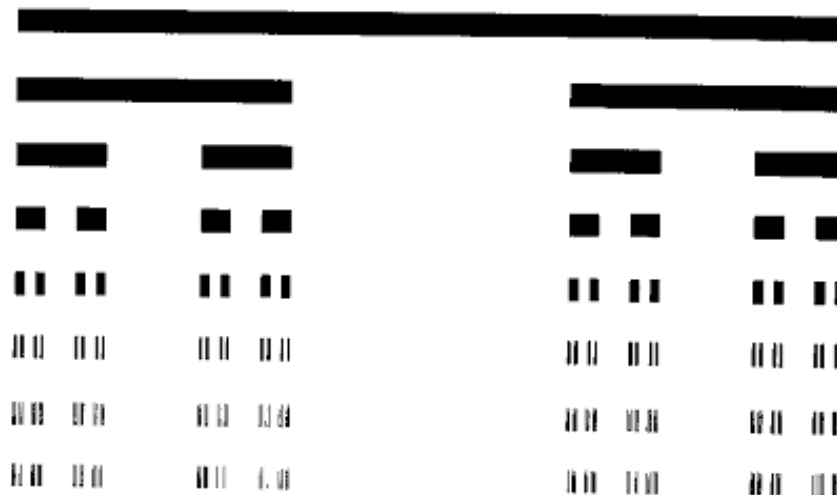


Figura 3.3.1 -Representação do conjunto de Cantor

Esse conjunto possui a característica especial ter comprimento nulo, devido ao fato de a cada etapa do processo seu comprimento ser reduzido por um fator $2/3$. Para o limite do número de passos tendendo para o infinito temos que o comprimento será dado por $L = \lim_{n \rightarrow \infty} (2/3)^n = 0$. Outra característica importante é que a dimensão desse conjunto é fracionária e aproximadamente $0,6$ caracterizando um objeto fractal.

3.3.2 Tapete de Sierpinski

Esse fractal foi contruído pelo matemático plones Waclaw Sierpinski (1882-1969) no ano de 1916. A construção do tapete de Sierpinski começa considerando-se um quadrado preenchido no plano, dividimos então esse quadrado em nove pedaços, em seguida removemos o quadrado central. Ficamos então com oito quadrados idênticos, então repetimos o processo para cada um deles indefinidamente, depois de infinitas interações vamos ter construído o tapete de Sierpinski, o qual podemos verificar na Figura 3.3.2

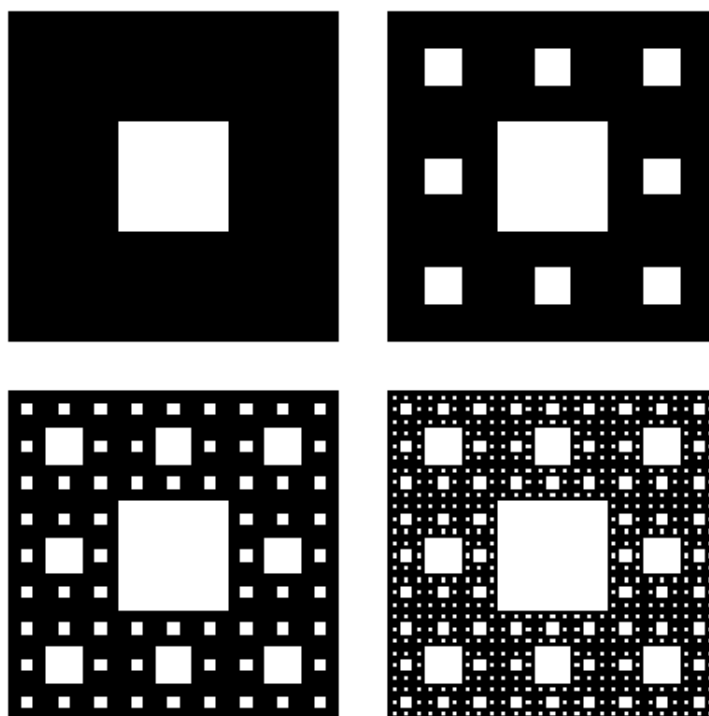


Figura 3.3.2 – Construção Básica do tapete de Sierpinski

Como em cada interação a área é diminuída por um fator $8/9$ a área total desse tapete, após infinitas interações, é zero. Isso fica claro se pensarmos que o tapete de Sierpinsky não é formado por pedacinhos que não possuem área, mas sim que ele é formado por pequeninos tapetes de Sierpinsky que não têm área. Outra característica curiosa é que se você somar o perímetro de todos os buracos, ele é infinito. Sierpinsky construiu também outro fractal famoso, que é conhecido como triângulo de Sierpinsky.

3.3.3 Curva de Koch

A autoria desse fractal é do matemático sueco Niels Fabian Helge von Koch . Para construir a curva de Koch partimos de uma linha de comprimento qualquer L . Para simplificar definimos esse comprimento como sendo por exemplo, $L=3\text{km}$. Dividimos o comprimento em três partes iguais e substituímos a parte do meio por dois pedaços de tamanho igual, formando uma entrada ou cavidade, que podemos pensar como uma baía ou península. Qual é o comprimento dessa nova linha? É quatro quilômetros, pois cada linha tem um quilômetro. Repetimos o processo, a

nova linha tem 16 pedacinhos, e cada um possui $1/3$ de quilômetro, logo o comprimento dela deve ser 16 vezes $1/3$. Se repetirmos mais uma vez, verificaremos que teremos 64 pedaços, e que cada um possui $1/9$ de quilômetro.

Devemos continuar esse processo, sempre dividindo e acrescentando o novo segmento na parte central. Repetindo esse procedimento indefinidamente, chega-se ao limite que se dá o nome de “curva de Koch”, podemos acompanhar o processo até chegar na curva na Figura 3.3.3:

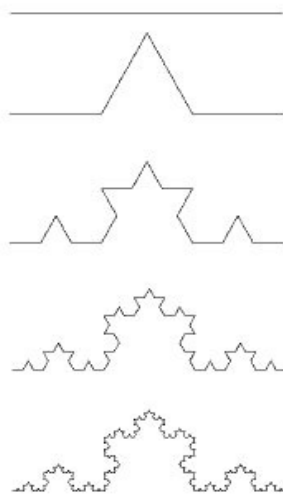


Figura 3.3.3- Construção da curva de von Koch

Ela possui a fascinante propriedade de ser semelhante entre todas as suas partes. Em particular a curva de von Koch tem a característica de possuir área nula. Percebemos que em cada etapa o seu comprimento aumenta $4/3$. Quando extrapolamos pensando em repetir o processo um número de vezes infinito, temos que o comprimento da curva de von Koch é infinito. Além disso, ela é contínua, mas não possui tangente na maioria de seus pontos. Mandelbrot (1991, p.36) comenta sobre esse fato “é um ser geométrico próximo de uma função contínua, mas sem derivada”.

3.3.4 Esponja de Menger

A esponja de Menger é um bom exemplo que podemos considerar para visualização um fractal com dimensão entre dois ou três. Ela é um fractal obtido a partir de um cubo, onde são retirados sistematicamente outros cubos de modo semelhante ao

que é feito na construção do tapete de Sierpinski. O resultado desse procedimento é uma figura com muitos furos, como podemos ver na Figura 3.3.4.

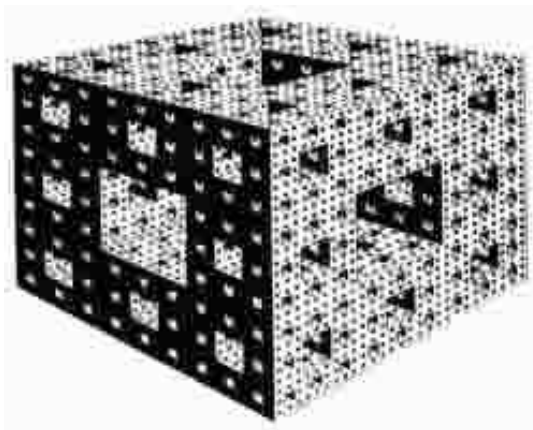


Figura 3.3.4- Esponja de Menger

Apesar dos primeiros fractais terem surgido no começo do século XX foi só na década de setenta que Benoit Mandelbrot formalizou a teoria dos fractais. O que faremos a seguir é introduzir a idéia de Mandelbrot sobre geometria fractal.

3.3.5 Benoit Mandelbrot e a Geometria Fractal

Benoît Mandelbrot nasceu Varsóvia, Polônia no ano de 1924. Seu principal trabalho foi a proposta de um novo conceito de geometria, que hoje conhecemos como geometria fractal. Moreira³ relata que Mandelbrot escreveu logo no primeiro parágrafo de seu famoso livro "*The fractal geometry of nature*", a seguinte frase "Nuvens não são esferas, montanhas não são cones, linhas costeiras não são círculos, cascas de árvores não são suaves nem o raio se propaga em linha reta". Entre os méritos que ele carrega estão o fato dele ser a primeira pessoa a se referir a estas figuras geométricas como fractais e a descoberta do conjunto de Mandelbrot. Segundo as palavras do próprio Mandelbrot (1991, p.13) "fractal ou objeto fractal são termos que formei a partir do adjetivo *fractus*, que significa irregular ou quebrado."

Vamos partir do exercício mental proposto por Mandelbrot em seu livro

³ http://br.geocities.com/projeto_caos_ufg/fractais/fractais.htm

“Objetos Fractais” para tentar compreender do que se tratam os fractais. Faremos isso usando o exemplo da medida do comprimento da linha costeira considerando diferentes tamanhos de “régua”.

Consideremos um pedaço de linha costeira, como a costa de um continente ou de uma ilha, como a fronteira que separa a terra do mar. Nosso objetivo é medir o comprimento dela. Em uma primeira aproximação, podemos dizer que a costa mede a distância, em linha reta, entre seu ponto inicial e final. Se observarmos à costa com algum detalhe, podemos deduzir que essa aproximação está longe de descrever corretamente o verdadeiro comprimento da costa pois, a geografia dela é bastante irregular. Então, podemos ter certeza que o comprimento real vai ser maior do que o comprimento da linha que une os pontos inicial e final.

Uma maneira prática de medir o comprimento da costa seria enfileirar barcos ao longo dela. Começamos com barcos de tamanho relativamente grande. Para saber o comprimento da costa temos que contar o número de barcos e multiplicar pelo tamanho de cada barco, designado pela letra N , e encontramos assim $L(N)$ que é o comprimento da costa. Mesmo com o tamanho do barco grande já percebemos que o comprimento da costa é maior em relação ao comprimento que encontramos quando usamos a linha. Pois, com os barcos estamos detalhando mais a nossa medida. Para conseguir valores cada vez mais próximos do valor real podemos diminuir cada vez mais o tamanho do barco, verificando que cada vez mais o valor de $L(N)$ aumenta, sem um limite definido.

Quando se tenta medir a costa de Brest, as medidas sempre dão um resultado crescente, até ao ponto em que a unidade N desce abaixo de 20m. Encontra-se então uma zona em que $L(N)$ varia muito pouco, só recomeçando a aumentar quando N atinge valores da ordem dos 20 cm ou inferiores, ou seja, valores tão pequenos que o comprimento começa a levar em conta as irregularidades das pedras. (MANDELBROT, 1991, p.30)

Agora já somos capazes de entender do que se tratam os fractais, e se analisamos tais figuras não conseguimos dizer de forma precisa qual é o valor da sua dimensão. Nesse contexto que Mandelbrot formaliza uma nova teoria, que dá conta de fazer a medida da dimensão fractal dos objetos fractais, a geometria fractal.

Podemos encontrar diferentes maneiras de se determinar a dimensão de um objeto, segundo Moreira⁴ “neste século surgiram várias noções de dimensão, (...) a primeira delas é a denominada capacidade, definida por Kolmogorov, que mede o

⁴ http://br.geocities.com/projeto_caos_ufg/fractais/fractais.htm

quanto o conjunto ou objeto considerado preenche o espaço em que está imerso.” A dimensão de Hausdorff é uma definição muito próxima da de capacidade, por isso nos referiremos tanto à dimensão de capacidade, dimensão de Hausdorff e dimensão fractal como sendo sinônimos. “Dimensão de Hausdorff é aquela em que valores não inteiros são permitidos”.⁵ Para entender como se calcula essa dimensão, segundo o que foi descrito por Backes (2005), consideremos uma linha de comprimento L e uma régua de comprimento ϵ , onde $\epsilon < L$. Se cobrirmos a reta L completamente por N régua de tamanho ϵ , encontramos que $N = L/\epsilon$, que nada mais é do que a medida da linha. Estendendo esse raciocínio para um quadrado de lado L , e cobrindo toda sua área com quadrados de lado ϵ , obteremos a relação $N = (L/\epsilon)^2$.

Este processo leva a uma relação do tipo $N = (L/\epsilon)^D$ ou, aplicando logaritmo em ambos os lados,

$$D = \frac{\log N}{\log\left(\frac{L}{\epsilon}\right)} \quad (3.1)$$

$$D = \frac{\log N}{\log L + \log(1/\epsilon)} \quad (3.2)$$

Quando ϵ tende a zero podemos desconsiderar o termo que depende de L , sendo assim:

$$D = - \lim_{\epsilon \rightarrow 0} \frac{\log N}{\log(\epsilon)} \quad (3.3)$$

Agora já podemos calcular o valor da dimensão dos objetos fractais, sendo assim, calcularemos como exemplo o valor da dimensão fractal do conjunto de Cantor visto na seção anterior. Observemos a Figura 3.2.5.

⁵ <http://mathworld.wolfram.com/CapacityDimension.html>

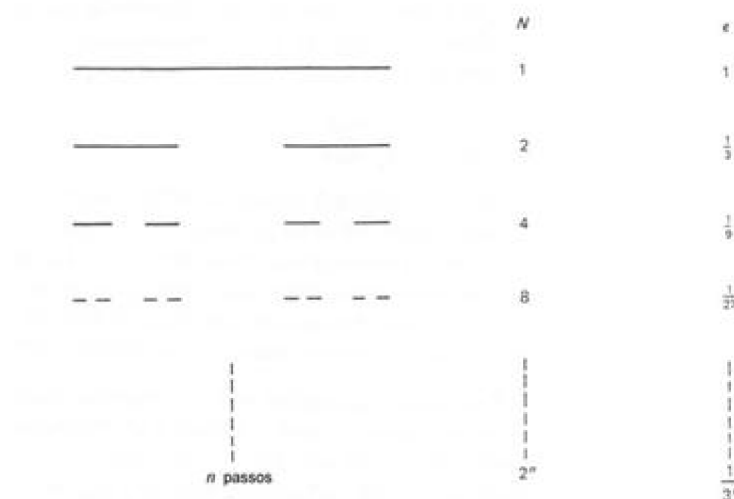


Figura 3.2.5- Conjunto de Cantor com o número de réguas e seus respectivos tamanhos para cada um dos passos.

Se considerarmos a definição de dimensão dada por (1) vemos que a regra que rege N e ϵ é $N = 2^n$ e $\epsilon = 1/3^n$, sendo assim:

$$D = \frac{\log 2}{\log 3} \cong 0,63 \quad (3.4)$$

Agora definiremos dimensão fractal segundo as próprias palavras de Mandelbrot (1991, p.14), “Ao contrario do que é observado na Geometria Euclidiana, onde o valor da dimensão representa a dimensionalidade do espaço em que dado objeto está inserido, a Dimensão Fractal representa o nível de irregularidade de um fractal”.

É necessário esclarecer que formas da natureza não são fractais “perfeitos”, pois são muito irregulares e não possuem auto-similaridade para todas as escalas. A esses fractais se dá o nome de fractais estatísticos.

Embora nesse tipo de fractal natural o resultado imediato de cada ampliação não possa ser mapeado linearmente sobre o todo, se forem feitas ampliações em diferentes pontos, em média, as características estruturais dessas ampliações tenderão para um valor próximo das características médias do todo SOARES (1994, p.52).

4. METODOLOGIA

Partindo da proposta feita por Glazier e colaboradores, descrita a seguir, usei caminhadas aleatórias, a partir daqui vou me referir a caminhadas (pseudo) aleatórias simplesmente como caminhadas aleatórias, sobre a seqüência completa de DNA mitocondrial (mtDNA) e sobre o segmento de 18s de RNA ribossomal (rRNA) de dez diferentes espécies de organismos eucariotos e do cálculo da dimensão fractal (D_F) das figuras geradas por essas caminhadas, para determinar se a D_F da morfologia das caminhadas é capaz de fazer diferenciação entre espécies. Usando como parâmetros, para o método de agrupamento hierárquico, os valores das dimensões fractais calculadas em cima das caminhadas geradas construímos as árvores filogenéticas.

A proposta de Glazier é o uso da geometria fractal (em particular, do espectro multifractal) para construção de árvores filogenéticas que relacionam distância evolutiva entre espécies diferentes a partir do DNA mitocondrial (mtDNA). Os autores fazem a caminhada aleatória sobre mtDNA atribuindo coordenadas espaciais para pares específicos de bases do mtDNA, que são mapeadas em um espaço com dimensão, entre 2 a 4, a ser escolhida. Estes conjuntos de coordenadas quando aplicados a seqüências de mtDNA geram trajetórias aleatórias com características fractais, típicas da região do DNA analisado. A partir do objeto gerado eles obtêm o espectro multifractal e com essa informação constroem as árvores. GLASIER (p. 2665, 1995) “Nós mapeamos mtDNA animal com caminhadas aleatórias quadridimensionais e estimamos as correlações hereditárias de longo alcance utilizando espectro multifractal.”

No ano de 1992 Cheryl L. Bertelsen e colaboradores publicaram o artigo “*Global fractal dimension of human DNA sequences treated as pseudorandom walks*” no qual fazem um estudo das maneiras como se pode calcular a dimensão fractal de caminhadas aleatórias feitas sobre seqüências de DNA. A idéia apresentada nesse artigo é a de se representar a seqüência de DNA como uma caminhada pseudo aleatória bidimensional, isso devido à complementaridade das bases, ou seja, o fato de A ser complementar a T e de G ser complementar a C.

Por esse motivo escolhemos fazer as caminhadas em duas dimensões, como o descrito na seção 4.1. Enquanto Glazier usa apenas um parâmetro para

comparação entre espécies, nós usamos dois, que são os valores das dimensões fractais das caminhadas geradas tanto para o mtDNA quanto para o segmento 18s do rRNA. A escolha dessas moléculas, que são componentes celulares de quase todos os organismos (organismos procariotos não apresentam DNA mitocondrial em sua composição celular), é devido ao fato de que elas são muito conservadas ao longo do tempo, por resistirem a mutações, pois pequenas alterações em suas ordens de nucleotídeos pode fazer com que elas percam suas funções. Devido a essa característica elas se mostram apropriadas para o uso como parâmetros de diferenciação e correlação entre espécies, ou seja, com elas podemos saber se espécies têm um ancestral comum e em que momento aconteceu a sua separação. Onde e como ela aconteceu.

Precisava de bancos de dados públicos. Três sítios atenderam a essa exigência, o *download* das seqüências de mtDNA foi feito no banco de dados do *National Center for Biotechnology Information* (NCBI), cujo endereço é <http://www.ncbi.nlm.nih.gov>. Nele encontra-se toda a informação genética de quase todos os organismos seqüenciados até hoje. Existem vários tipos de arquivos disponíveis para download, mas os que me interessavam eram os que tivessem toda a seqüência de nucleotídeos dos organismos. Os arquivos do tipo *.fna* são os que estão nesse formato. Algumas seqüências do segmento 18s de rRNA foram encontradas no sítio *The European ribosomal RNA database*, <http://bioinformatics.psb.ugent.be/webtools/rRNA>, e outras no sítio *Silva comprehensive ribosomal RNA database*, <http://www.arb-silva.de>. Nas seqüências de rRNA foram precisas alterações, pois elas não continham apenas a seqüência real de rRNA, continham também caracteres para alinhamento de seqüência, como foi descrito na seção 3.1.8.

Usei apenas espécies eucariotas pela característica de todas possuírem mtDNA e rRNA. Todas são espécies que bem estudadas, e das quais já foram inferidas as relações evolutivas. E são elas: *Anopheles gambiae* (mosquito da malária), *Arabidopsis thaliana* (uma espécie de planta), *Danio rerio* (peixe zebra), *Drosophila melanogaster* (mosca das frutas), *Homo sapiens* (humano), *Mus musculus* (camundongo), *Pan troglodytes* (chimpanzé), *Rattus norvegicus* (cobaia de laboratório), *Strongylocentrotus purpuratus* (ouriço do mar), *Xenopus laevis* (uma espécie de sapo).

O *software* usado para todos os cálculos foi desenvolvido com o MATHEMATICA 6.0, as rotinas que fazem a caminhada aleatória e calculam o valor de D_F foram implementadas por mim, mas a que faz o agrupamento e gera o dendograma é um pacote que está pronto dentro do MATHEMATICA 6.0.

4.1 Caminhada Aleatória

A caminhada aleatória, segundo o modelo de Glazier, é feita atribuindo coordenadas para pares específicos de bases do mtDNA e rRNA que são mapeadas no plano. O “mapeamento” do mtDNA, atribuindo-se a base A o par ordenado (1,0), para a base G o par é (0,1), para T é (-1,0) e para C é (0,-1). No caso do rRNA no lugar da base Timina temos a base Uracila, e a caminhada é equivalente, apenas leva-se em conta essa substituição. Na figura 4.1.1 é possível ver de que forma acontece a caminhada aleatória.

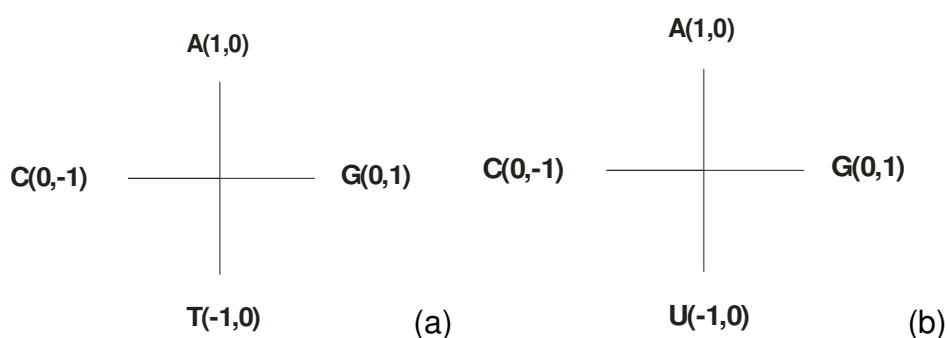


Figura 4.1.1-(a) Representação do eixo de coordenadas atribuído as bases do mtDNA;
(b) Representação do eixo de coordenadas atribuído as bases do rRNA.

A construção da caminhada aleatória, dentro do MATHEMATICA 6.0, se inicia com a importação do arquivo onde está salva a seqüência de mtDNA, ou rRNA, com extensão *.txt*⁶. A partir desse arquivo a rotina lê o seqüenciamento de bases, atribuindo a cada base a sua respectiva coordenada. Por conveniência escolhi que a caminhada fosse centrada na origem, ou seja, na coordenada (0,0). A coordenada da primeira base da seqüência é atribuída e somada à coordenada inicial. Como um processo recursivo, a coordenada da segunda base é somada a da primeira, passa para a terceira base e a essa é atribuída uma coordenada que deve ser somada a

⁶ Dentro do MATHEMATICA é possível importar arquivos com essa extensão.

coordenada da base anterior e assim sucessivamente até o fim da seqüência. O valor de cada coordenada é dada pela seguinte fórmula:

$$\text{Caminhada}[i+1]=\text{Caminhada}[i]+\text{Coordenada da base}[i+1]$$

Onde: Caminhada [i+1] representa a coordenada atual, Caminhada [i] é o valor da coordenada anterior e Coordenada [i+1] é o valor de par ordenado atribuído para base atual, (0,1), (1,0), (0,-1) ou (-1,0).

Depois de todas as bases já possuírem um valor de coordenada, a rotina gera uma lista com os valores dessas coordenadas e em seguida constrói o gráfico com os pares ordenados gerados, visto que essa é uma caminhada em duas dimensões.

Para visualizar a maneira como o método funciona consideremos uma seqüência hipotética de DNA dada por:

ATGCAAATCATAGAAGGGAAATTGCAATTACAAGGGAATGAAAGAGTCGCTATTT
TAACATCGCGCTTCAATCATATCATCACAGACAGAT

Depois de ler essa seqüência, as coordenadas geradas pela rotina são:

{1,0}, {0,0}, {0,1}, {0,0}, {1,0}, {2,0}, {3,0}, {2,0}, {2,-1}, {3,-1}, {3,0}, {4,0}, {5,0}, {5,1},
{5,2}, {5,3}, {6,3}, {7,3}, {8,3}, {7,3}, {6,3}, {6,4}, {6,3}, {7,3}, {6,3}, {7,3}, {7,2}, {8,2},
{9,2}, {9,3}, {9,4}, {9,5}, {10,5}, {11,5}, {10,5}, {10,6}, {11,6}, {12,6}, {13,6}, {13,7},
{14,7}, {14,8}, {13,8}, {13,7}, {13,8}, {13,7}, {12,7}, {11,7}, {10,7}, {9,7}, {10,7}, {11,7},
{11,6}, {12,6}, {11,6}, {11,5}, {11,6}, {11,5}, {10,5}, {9,5}, {9,4}, {10,4}, {11,4}, {10,4},
{10,3}

E para as mesmas bases, a Figura 4.1.2 ilustra o gráfico gerado.

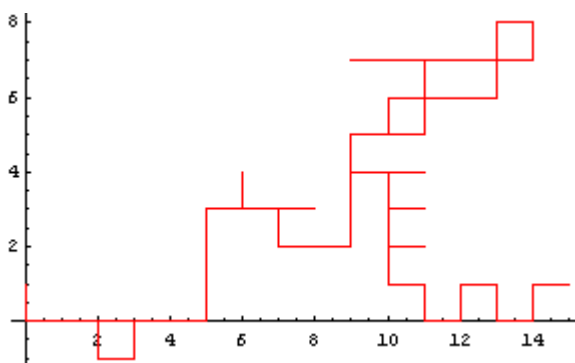


Figura 4.1.2- Gráfico das coordenadas geradas pela rotina de caminhada aleatória sobre seqüenciamento de bases arbitrário.

O algoritmo que gera a caminhada aleatória está no Apêndice A.

4.2 Cálculo da Dimensão Fractal (Método do Box Counting)

O método de Box Counting consiste em cobrir um objeto fractal com uma malha de quadrados (para um objeto em 2D) igualmente divididos de tamanho ε e contar quantos desses quadrados contêm o objeto, e número é $N(\varepsilon)$. Conforme a Figura 4.2.1

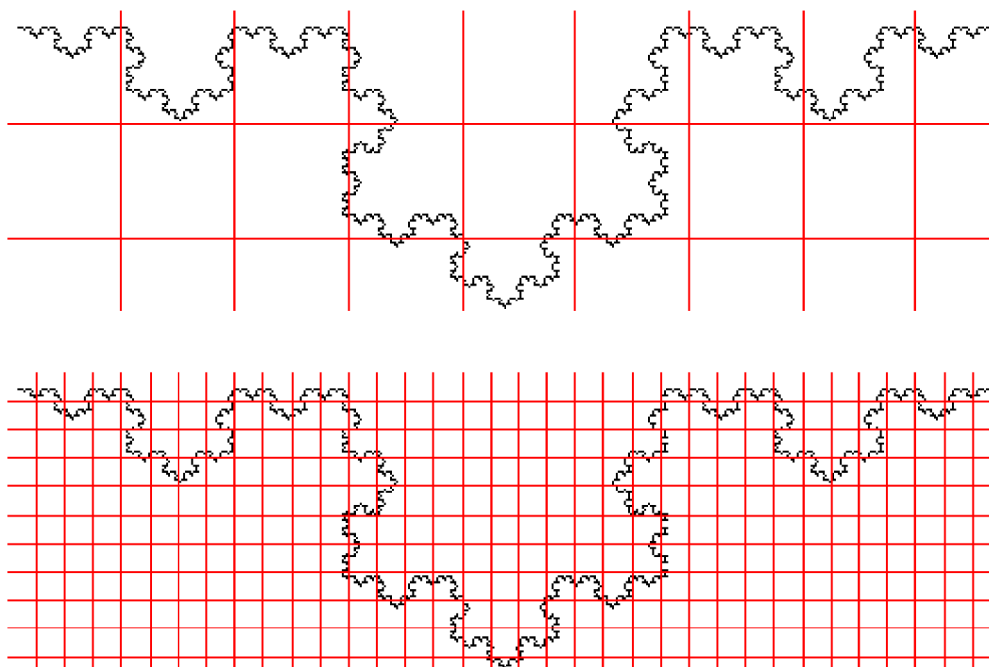


Figura 4.2.1- Método de Box Counting aplicado a Curva de Koch, para dois diferentes tamanhos de ε

A contagem é repetida para tamanhos de caixa cada vez menores, ou seja, para $\lim_{\varepsilon \rightarrow 0}$. Com essa variação construímos o gráfico de $\log N(\varepsilon)$ contra $\log (1/\varepsilon)$. A inclinação da reta que ajusta os dados fornece o expoente da relação

$$N(\varepsilon) = \left(\frac{1}{\varepsilon}\right)^D \quad (4.1)$$

dado por

$$D = \lim_{\varepsilon \rightarrow 0} \frac{\log N(\varepsilon)}{\log\left(\frac{1}{\varepsilon}\right)} \quad (4.2)$$

O expoente D medido pela equação (4.2) é a definição da dimensão de capacidade, que coincide com a dimensão fractal do objeto.

O algoritmo que calcula a dimensão fractal está no Apêndice B.

4.3 Agrupamento

Para poder analisar e construir as árvores filogenéticas que determinam a separação entre as espécies eu uso a técnica de agrupamento do pacote pronto que se encontra dentro do software MATHEMATICA 6.0, ele usa a função **HierarchicalClustering**.

O MATHEMATICA 6.0 usa o método de agrupamento aglomerativo hierárquico, e a função DendrogramPlot gera o gráfico com a representação do agrupamento. Dentre as maneiras existentes de se medir a distância entre dois grupos, o MATHEMATICA 6.0 usa a distância Manhattan, essa distância é definida da seguinte forma⁷:

Sejam dois vetores $A=\{a,b\}$ e $B=\{x,y\}$, a distância Manhattan entre eles é dada por:

$$D = \text{Abs}[a-x] + \text{Abs}[b-y] \quad (4.3)$$

Como exemplo podemos considerar o seguinte conjunto de dados. Lista= {1, 2, 10, 4, 8}. O processo começa com a comparação da distância de todos os membros da lista, onde cada um representa um grupo, usando o método de vizinho próximo agrupam-se os membros mais próximos 1 e 2. Esses dois números vão formar um grupo, chamaremos de 1;2. No próximo passo comparam-se o grupo 1;2 e os outros, na hora de se fazer a medida da distância do grupo 1;2 com os outros considera-se o membro que é mais próximo dos outros, no caso 2. Nessa

⁷ <http://www.nist.gov/dads/HTML/manhattanDistance.html>

comparação a distância entre 4 e 2 é menor do que a distância entre 4 e os outros membros, por isso ele forma um novo grupo 4;1;2, nesse mesmo passo os elementos 10 e 8 são agrupados. Por último agrupa todos os grupos em um único. O Dendograma gerado pelo MATHEMATICA 6.0 está representado na Figura 4.3.1:

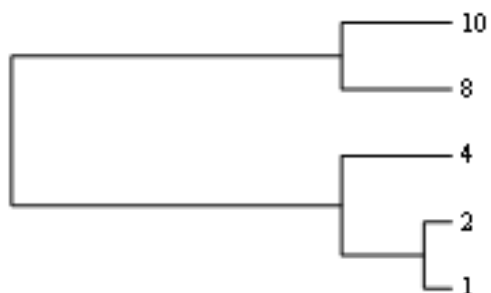


Figura 4.3.1-Dendograma Gerado pelo MATHEMATICA 6.0 para o conjunto de dados.

5.RESULTADOS

Para saber se os resultados das rotinas eram confiáveis, e qual era a porcentagem de erro que eles forneciam, comparamos nossos com os resultados já aceitos na literatura. Para a rotina que constrói a caminhada aleatória usei como referência a caminhada feita sobre mtDNA de *Drosophila melanogaster* (mosca da fruta) que aparece na tese de doutorado de C.L.Berthelsen⁸. Na Figura 5.1 (a) esta representado o gráfico presente na tese de C.L.Berthelsen e na Figura 5.1(b) esta representada a caminhada gerada pela minha rotina.

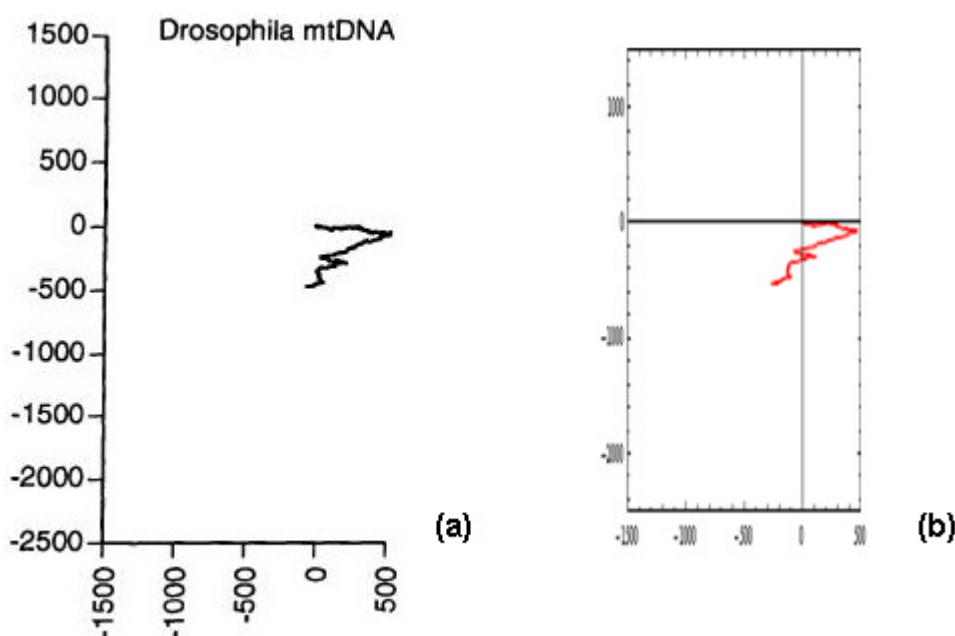


Figura 5.1-(a) Caminhada sobre mtDNA de *Drosophila melanogaster* por C.L.Berthelsen em seu trabalho (b) Caminhada sobre mtDNA de *Drosophila melanogaster* gerada pela nossa rotina.

Já para o cálculo de D_F usamos como referência a dimensão euclidiana de uma linha, como já foi discutido na seção 3.2.5, a dimensão fractal de figuras regulares devem coincidir com o valor da dimensão euclidiana. Usamos diferentes tamanhos de linha, todos gerados dentro do MATHEMATICA 6.0. Isso foi feito, pois dessa maneira foi possível estimar o erro no cálculo de D_F com o tamanho da caminhada. Na Tabela 5.1 estão os diferentes valores de dimensão fractal para os diferentes comprimentos da caminhada.

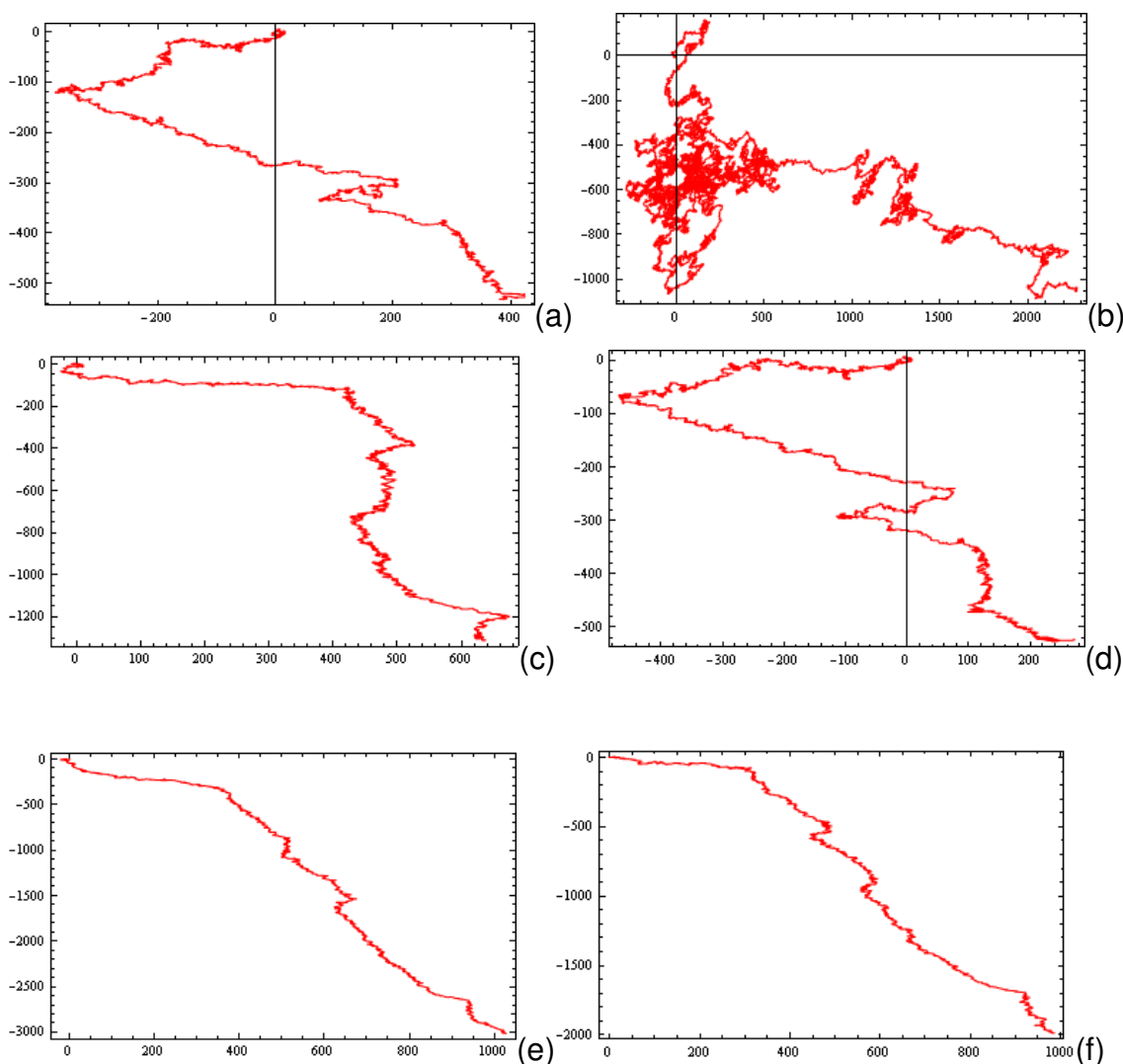
⁸Cheryl L. Berthelsen foi orientada por de James A. Glazier em seu doutorado.

Tamanho da caminhada (L)	Dimensão Fractal(DF) / Erro
100	0,975086 / 2,4914%
1000	0,98781/1,219%
10000	0,995625/0,4345%
100000	0,99941/0,059%

Tabela 5.1- Valores das dimensões fractais para linhas com diferentes comprimentos.

Da tabela é possível verificar que para caminhadas acima de 10000 já se tem um erro menor do que 1%.

Comecei a minha tentativa de construção de árvores filogenéticas com o uso de um único parâmetro, o valor da dimensão fractal do DNA mitocondrial, com o intuito de fazer da forma mais simples possível essa construção. Os gráficos gerados para esse tipo de caminhada para cada um dos organismos estão listados na Figura 5.2:



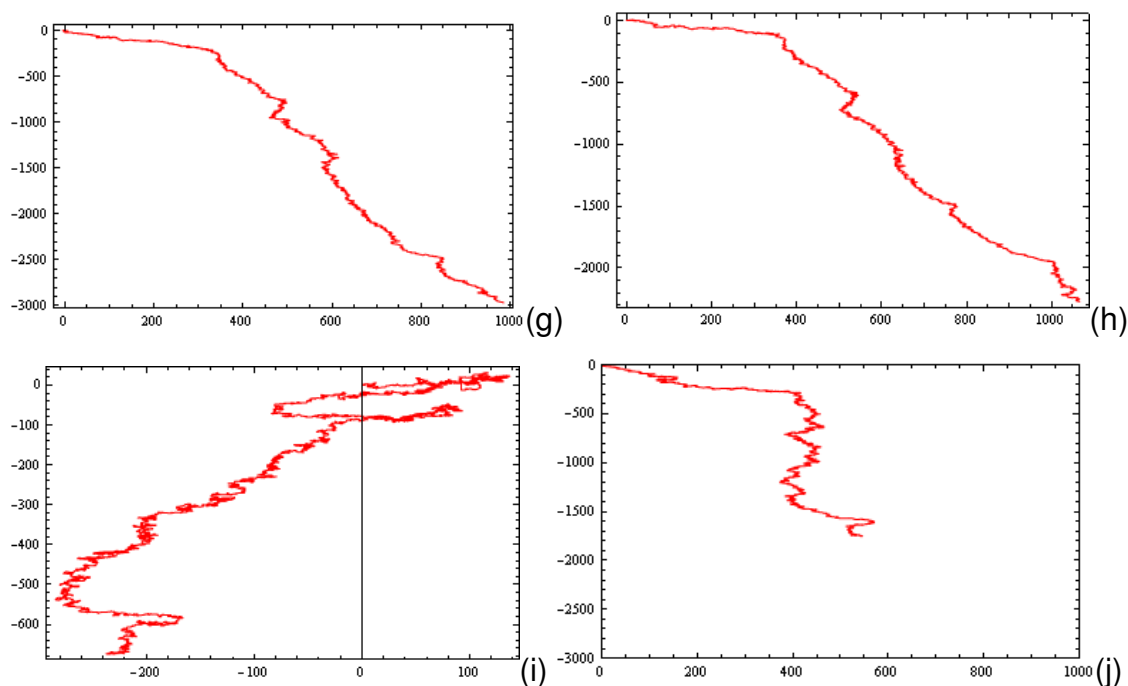


Figura 5.2- Gráficos das caminhadas feitas sobre mtDNA (a) *Anopheles gambiae* (mosquito da malária), (b) *Arabidopsis thaliana* (uma espécie de planta), (c) *Danio rerio* (peixe zebra), (d) *Drosophila melanogaster* (mosca das frutas), (e) *Homo sapiens* (humano), (f) *Mus musculus* (camundongo), (g) *Pan troglodytes* (chimpanzé), (h) *Rattus norvegicus* (cobaia de laboratório), (i) *Strongylocentrotus purpuratus* (ouriço do mar), (j) *Xenopus laevis* (uma espécie de sapo).

Fazendo uma análise qualitativa das caminhadas geradas podemos verificar que a de organismos mais próximos são mais parecidas. Por exemplo, analisando as figuras em grupos, primeiro as Figuras 5.2(a), 5.2(b), 5.2(c), 5.2(d) podemos verificar que o gráfico 5.2(b), referente ao agrião, é o mais diferente de todos, pois é uma planta, enquanto todos os outros são animais. Que 5.2(a) e 5.2(d) são os mais parecidos, sendo eles o mosquito e a mosca, ambos insetos, e ainda que 5.2(c) difere dos outros três por se tratar de um peixe. Analisando as outras seis figuras, 5.2(e), 5.2(f), 5.2(g), 5.2(h), 5.2(i), 5.2(j), verificamos que as quatro primeiras, que pertencem ao camundongo, chimpanzé, humano e rato, como eles são mamíferos as suas caminhadas são muito parecidas. Sendo bem diferentes a caminhada do ouriço e do sapo. É importante salientar que essa é uma análise apenas qualitativa, não servindo de parâmetro para a construção do dendograma. Os valores das DF s para cada um dos organismos, calculados sobre as caminhadas feitas sobre seus DNAs mitocondriais, estão listadas na tabela 5.2.

Organismo	D_F mtDNA
Anopheles gambiae (mosquito da malária)	1.3182
Arabidopsis thaliana (espécie de planta)	1.5761
Danio rerio (peixe zebra)	1.3083
Drosophila melanogaster (mosca da fruta)	1.2972
<i>Homo sapiens</i> (humano)	1.1910
<i>Mus musculus</i> (camundongo)	1.2450
<i>Pan troglodytes</i> (chimpanzé)	1.1905
<i>Rattus norvegicus</i> (cobaia de laboratório)	1.2110
<i>Strongylocentrotus purpuratus</i> (ouriço do mar)	1.3790
<i>Xenopus laevis</i> (espécie de sapo)	1.2700

Tabela 5.2- Valores das dimensões fractais para caminhadas feitas sobre mtDNA

A partir desses valores construímos a filogenia das espécies. O dendograma gerado está na Figura 5.3.

O método separou a planta dos animais, deixando o humano e o chimpanzé como espécies mais próximas entre as analisadas. Os méritos acabam por aí, pois ele não separa nem vertebrados de invertebrados unindo, por exemplo, o peixe com o mosquito da malária.

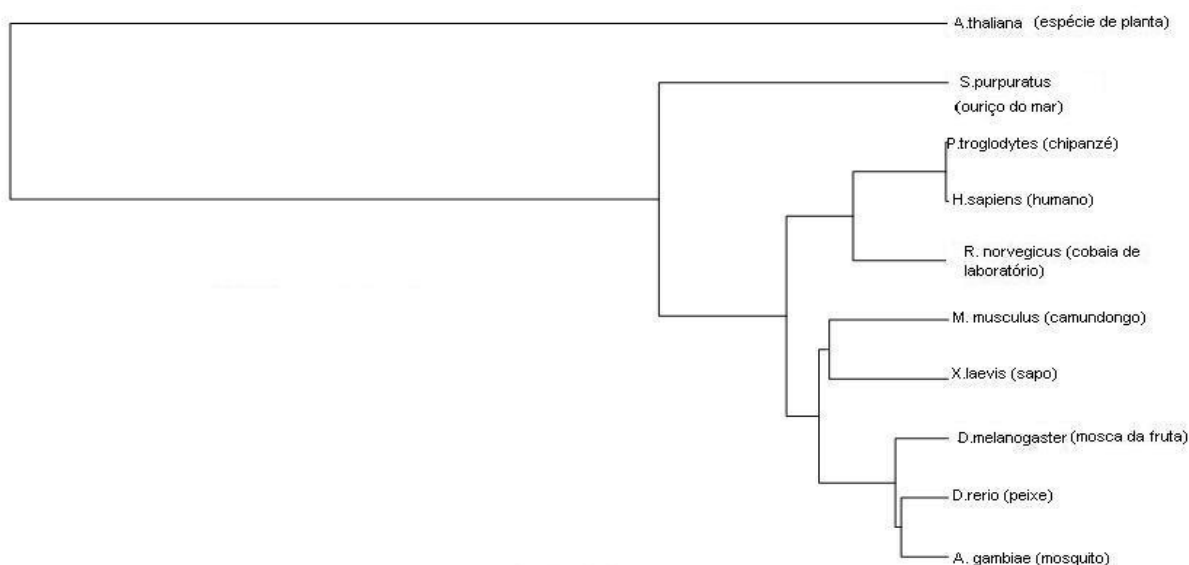
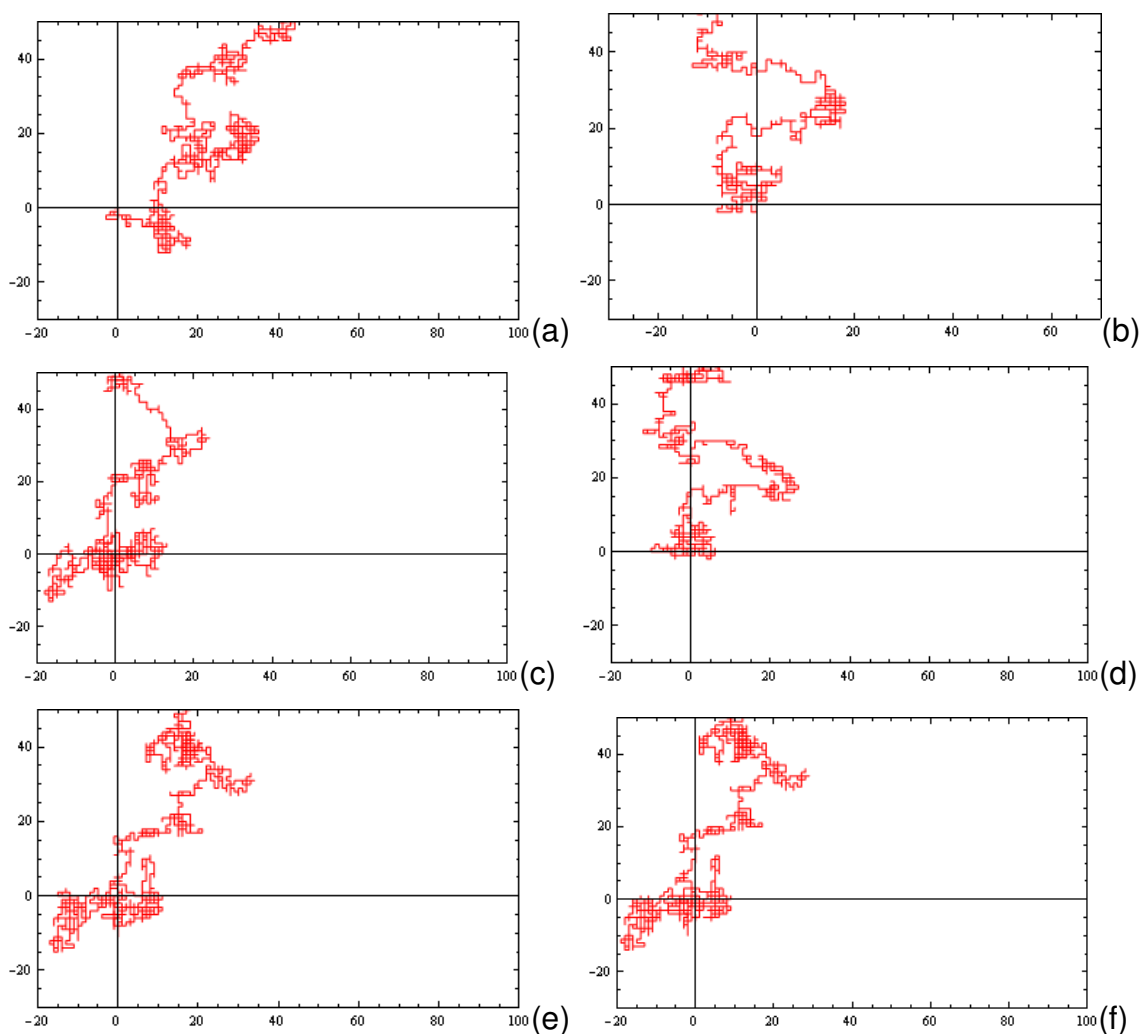


Figura 5.3- Dendograma gerado a partir do valor da dimensão fractal do mtDNA

Como pode ser verificado na Figura 5.3 o resultado obtido não foi satisfatório, queríamos encontrar um método que fizesse separação entre espécies. Precisávamos encontrar uma molécula que fosse tanto, ou mais, conservada durante a evolução, pesquisamos e verificamos que o segmento de rRNA é a molécula mais conservada ao longo do tempo, e que usada, por isso, para fazer análise filogenética molecular. Decidimos então construir caminhadas com o rRNA e novamente calcular os valores de D_F e reconstruir o dendograma. Na Figura 5.3 estão representadas as caminhadas feitas sobre rRNA.



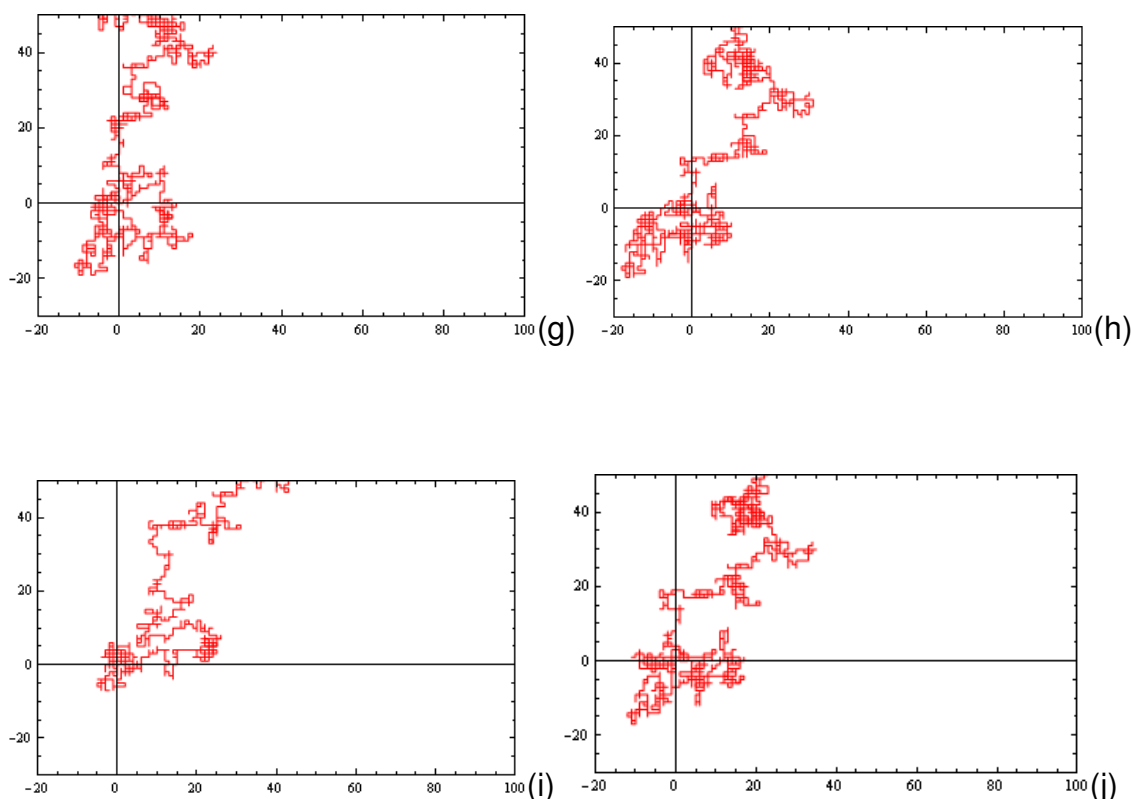


Figura 5.4- Gráficos das caminhadas feitas sobre rRNA (a) *Anopheles gambiae* (mosquito da malária), (b) *Arabidopsis thaliana* (uma espécie de planta), (c) *Danio rerio* (peixe zebra), (d) *Drosophila melanogaster* (mosca das frutas), (e) *Homo sapiens* (humano), (f) *Mus musculus* (camundongo), (g) *Pan troglodytes* (chimpanzé), (h) *Rattus norvegicus* (cobaia de laboratório), (i) *Strongylocentrotus purpuratus* (ouriço do mar), (j) *Xenopus laevis* (uma espécie de sapo).

Fica mais difícil de fazer inferências sobre as relações entre as espécies analisando apenas os gráficos das caminhadas geradas sobre o seu rRNA, como ela é a molécula mais conservada suas caminhadas são muito parecidas, ela também é uma molécula muito menor do que a mtRNA, com tamanho na faixa de 2000 pares de base. Os valores da D_F calculados sobre o rRNA estão na tabela 5.3.

Organismo	D_F rRNA
<i>Anopheles gambiae</i> (mosquito da malária)	1.39
<i>Arabidopsis thaliana</i> (espécie agrião)	1.58
<i>Danio rerio</i> (peixe zebra)	1.45
<i>Drosophila melanogaster</i> (mosca das frutas)	1.38
<i>Homo sapiens</i> (humano)	1.48
<i>Mus musculus</i> (camundongo)	1.47
<i>Pan troglodytes</i> (chimpanzé)	1.47
<i>Rattus norvegicus</i> (cobaia de laboratório)	1.53
<i>Strongylocentrotus purpuratus</i> (ouriço do mar)	1.39
<i>Xenopus laevis</i> (espécie sapo)	1.44

Tabela 5.3- Valores das dimensões fractais para caminhadas feitas sobre rRNA

Aqui consideramos apenas duas casas depois da vírgula devido à limitação de precisão do método estar relacionada com o tamanho da caminhada. Construímos um dendograma usando como único parâmetro o valor da dimensão fractal da caminhada feita sobre rRNA. E o resultado está representado na Figura 5.5.

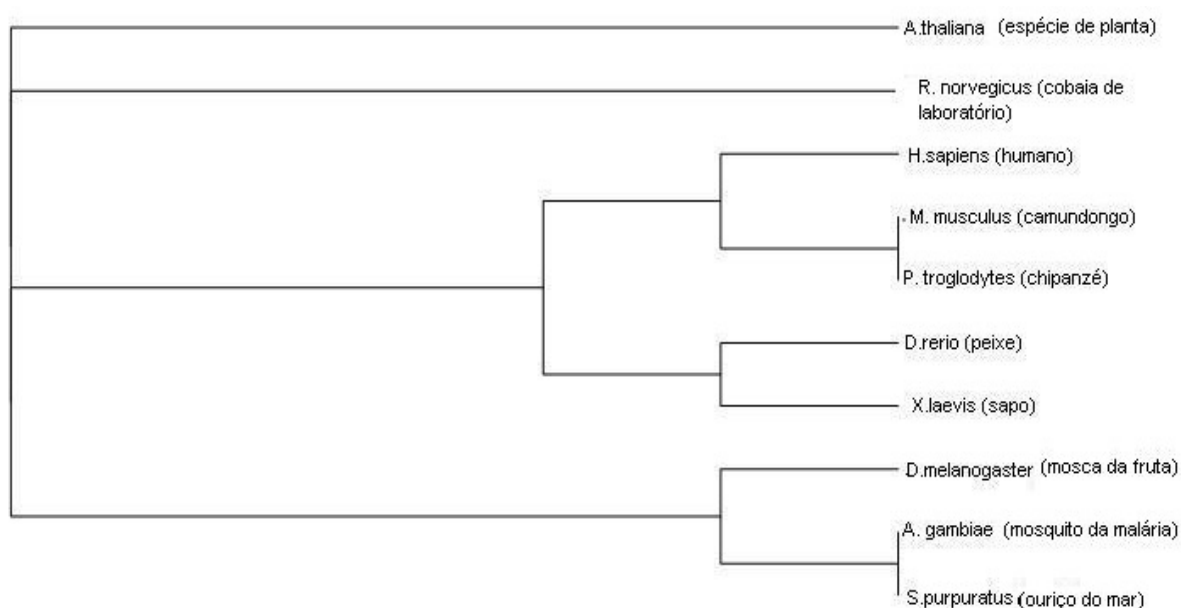


Figura 5.5- Dendograma gerado com o valor da dimensão fractal do rRNA.

Nessa tentativa o método diz que a cobaia de laboratório (rato) é tão distante evolutivamente dos animais quanto as plantas, ainda que camundongo e homem são mais próximos entre eles do que do chimpanzé. Consegue fazer a separação entre vertebrados e invertebrados, desconsiderando a cobaia, mas erra novamente em considerar que mosquito da malária e ouriço do mar são mais próximos do que o mosquito da mosca da fruta.

Os resultados obtidos levando-se em consideração apenas um parâmetro foram ruins. Então combinei os dois parâmetros para fazer a construção dos dendogramas, esse foi o que se mostrando mais eficiente, como era de se esperar, pois consideraremos mais características para se fazer a diferenciação. As diferenças de D_F aparecem na segunda ou terceira casa depois da vírgula, se

pensarmos que os valores de D_F representam “vetores de similaridade” quanto mais componentes tiverem esses vetores mais poderemos especificar mais as características dos organismos. Também se sabe que o mtDNA conserva algumas características e o rRNA outras durante a evolução. O resultando foi a seguinte árvore filogenética, representada na Figura 5.6.

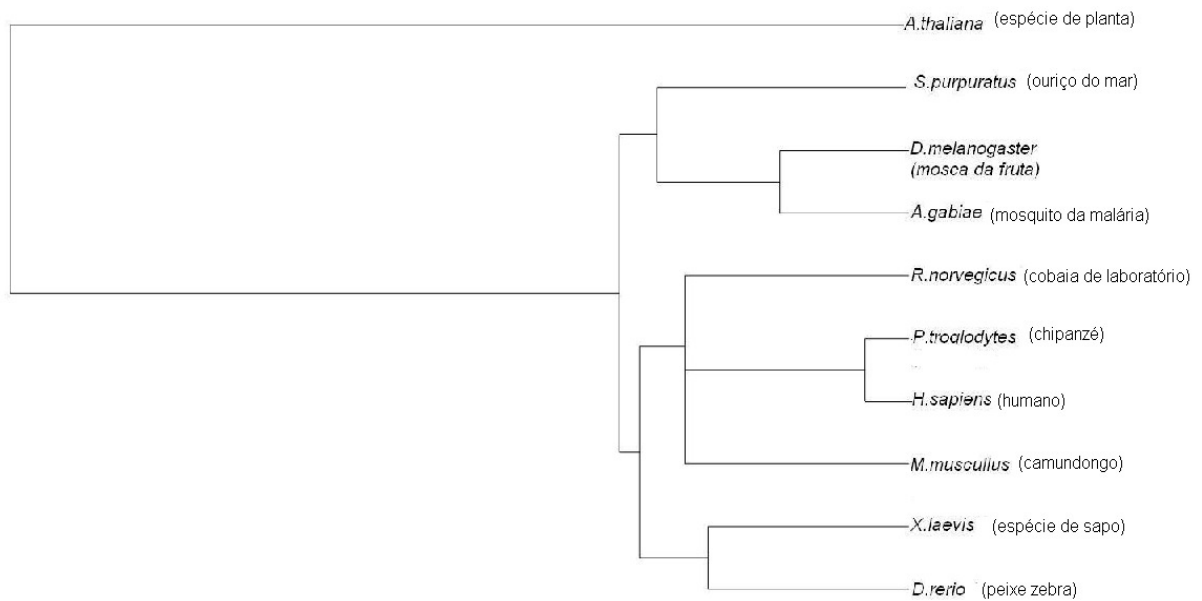


Figura 5.6- Árvore filogenética construída a partir dos valores da D_F do mtDNA e rRNA dos organismos.

Podemos verificar o maior galho está associado a planta, como já comentamos na seção 4.6 o tamanho do galho diz o quanto os grupos são separados, sendo assim, o método faz separação de reinos, deixando separados os animais da planta. Entre os animais ele faz uma separação em dois galhos distintos, separando vertebrados de invertebrados. Entre os invertebrados ficaram o ouriço (equinoderma), e o mosquito e a mosca (insetos), então ele separa insetos dos equinodermas. No outro ramo, o dos vertebrados, se subdividem dois ramos, em um deles estão o chimpanzé, humano, rato de laboratório e camundongo, todos mamíferos, e em outro coloca o sapo e o peixe. Sabe-se que evolutivamente anfíbios e peixes são mais próximos entre si do que são dos mamíferos, então essa separação também é correta. No galho onde se encontram os mamíferos ele separa em um único galho o chimpanzé e o humano, isso significa que dessa lista de organismos eles seriam os mais próximos evolutivamente.

6. CONCLUSÃO

O método separa a planta dos animais, ou seja, faz separação entre reinos. Dentre dos animais faz separação entre vertebrados e invertebrados. Entre os invertebrados faz separação entre os insetos (mosca e mosquito) e o equinoderma (ouriço do mar). Já entre os vertebrados separa os mamíferos do anfíbio e do peixe, o que está correto, pois se sabe que peixes e anfíbios são mais próximos entre si do que dos mamíferos. Segundo ele os organismos mais próximos evolutivamente são o chimpanzé e o humano. Ele só falha quando não coloca o *Rattus norvegicus* e o *Mus musculus* em um único galho. Analisando os dados obtidos percebemos que a diferença nas dimensões fractais desses organismos é para o rRNA, a justificativa é que a dimensão fractal calculada para o rRNA ainda apresenta um erro considerável devido ao tamanho da caminhada. Como estava propondo um método alternativo, frente ao método de análise filogenética molecular, concluí que ele é eficiente na separação entre espécies, apesar de ainda cometer erros. Pois gera resultados consistentes com os preditos pela filogenia tradicional. Minhas perspectivas futuras são o melhoramento no método, com a consideração de segmentos de mtDNA ao invés de fazer uma descrição global da molécula.

7. REFERÊNCIAS

Referências Bibliográficas:

BAXEVANIS, A.D.; OUELLETTE, B.F.F. **Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins**. 2ª Ed. New York: Wiley-Liss, 2001.

BERTHELSEN, C.L.; GLAZIER, J. A.; SKOLNICK, M.H.
Global Fractal Dimension of Human DNA Sequences Treated as Pseudorandom Walks. In: **Physical Review A** v.45, 8902, 1992.

BORGES-OSORÓRIO, M.R.; ROBINSON, W.M. **Genética Humana**. Porto Alegre: Editora Artes Médicas Sul Ltda, 1993.

COSTA, L. F.; BIANCHI, A. G. C. A Outra Dimensão da Dimensão Fractal. In: **Ciência Hoje**, v.31, nº 183, p.40. Rio de Janeiro: 2002

FARAH, S.B. **DNA: Segredos e Mistérios**. 5ª ed. São Paulo: Sarvier, 2000

FUTUYMA, D.J. **Biologia Evolutiva**. 2ª Ed. Ribeirão Preto: Sociedade Brasileira de Genética/CNPQ, 1992.

GLAZIER, J. A. et al. Reconstructing Phylogeny from the Multifractal Spectrum of Mitochondrial DNA. In: **Physical Review E** v. 51, 2665, 1995.

GRIFFITHS, A.J.F. et al. **Genética Moderna**. 2ª Ed. Rio de Janeiro: Guanabara-Koogan, 2000.

JUNQUEIRA, L.C.; CARNEIRO, J. **Biologia Celular e Molecular**. 8ª Ed. Rio de Janeiro: Guanabara-Koogan, 2005.

MANDELBROT, B. **Objetos Fractais: Forma, Acaso e Dimensão**. 3ª ed. Lisboa: Gradiva, 1991

MOTTA P. A. **Genética humana aplicada à psicologia e toda área biomédica**. 2^o ed. Rio de Janeiro:Guanabara-Koogan,2005

OLIVEIRA, L. H.A Matemática do Delírio. In: **Super Interessante**, v.85, p. 22, São Paulo:1994.

PEITGEN, H.; JÜRGENS, H.; SAUPE, D. **Fractals for the Classroom**: Part one Introduction to Fractals and Chaos. New York: Springer-Verlag, 1993.

PIRES, M.M. **Agrupamento Incremental e Hierárquico de Documentos**. Dissertação (Mestrado em Ciências em Engenharia Civil) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2008.

SOARES, T.T. **Sobre a Introdução da Geometria Fractal na Metalografia Qualitativa via Imagens Digitais**. Dissertação (Mestrado em Engenharia Metalúrgica e dos Materiais) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 1994

VIANA, G. V. R. **Técnicas para Construção de Árvores Filogenéticas**. Tese (Doutorado em Ciências da Computação) – Universidade Federal do Ceará, Fortaleza, 2007.

Referências Digitais

BACKES, A. R.; BRUNO, O. M. **Fractal and Multi-Scale Fractal Dimension Analysis: A Comparative Study of Bouligand-Minkowski Method**.

Disponível em <http://www.dcc.ufla.br/infocomp/>, Acessado em 10 de setembro de 2009.

BACKES, A. R.; BRUNO O. M. **Técnicas de Estimativa da Dimensão Fractal: Um Estudo Comparativo**. Disponível em www.dcc.ufla.br/infocomp/artigos/v.4.3art07. Acessado em 07 de Março de 2008.

BLACK, P. E. **Manhattan Distance**

Disponível em <http://www.nist.gov/dads/HTML/manhattanDistance.html>

Acessado em 11 de Novembro 2008.

FREITAS, J.A.B. **Análise de Cluster da Lisozima**

Disponível em www.uspleste.usp.br/rvicente/BioInfo_jefferson_PIBIC05.pdf

Acessado em 03 de Novembro de 2008.

MACEDO, M. et al. **Cálculo da Dimensão Fractal: Método de Box Counting.**

Disponível em <http://www.cbpf.br/~maysagm/>. Acessado em 15 de dezembro de 2007.

DE MARCO, R. **Análise Filogenética baseada em Alinhamento de Domínios.**

Disponível em <http://www.ifsc.usp.br/~rdemarco/FFI0760/filogenia.pdf>

Acessado em 13 de Novembro de 2008.

MOMBACH. J.C.M. et al. **Bioinformatics analysis of mycoplasma metabolism:**

Important enzymes, metabolic similarities, and redundancy. Disponível em www.intil.elsevierhelth.com/journals/cobm. Acessado em 20 de abril de 2008.

MOREIRA, C. M. **Fractais.**

Disponível em http://br.geocities.com/projeto_caos_ufg/fractais/fractais.htm.

Acessado em 02 de Outubro de 2008.

WEISSETEIN, E. W. **Capacity Dimension.** Disponível em <http://mathworld.wolfram.com/CapacityDimension.html>. Acessado em 13 de maio de 2008.

APÊNDICES

APÊNDICE A – Algoritmo de Caminhada Aleatória

dna=lista de bases de nucleotídeos

```
caminhada[1]={0,0};
```

```
Steps={{1,0},{-1,0},{0,1},{0,-1}};
```

```
Bases={"A","T","G","C"};
```

```
caminho=Table[
```

```
    caminhada[i+1]=caminhada[i]+Steps[[
```

```
        Part[ Flatten[ Position[Bases, StringTake[dna, {i}]] ] , 1]
```

```
    ]],
```

```
    {i,1,StringLength[dna]}];
```

```
ListPlot[caminho, PlotJoined→True, Axes→True,PlotStyle→Hue[0.0]]
```

APÊNDICE B– Algoritmo de Cálculo de Dimensão Fractal

```

BoxCount[list_,delta0_,scale_,min_]:=
  Block[ {tmp,delta,nmax,l,i},
    tmp={};
    delta=delta0;
    nmax=Abs[Ceiling[Log[min/delta0]/Log[scale]]];
    Do[
      {
        l=Length[Complement[Ceiling[list/delta],{}]];
        AppendTo[tmp,{delta,l}];
        delta=delta*scale}, {i,1,nmax}];
    Return[tmp];
qqq[{x_,y_}]:=N[{-Log[10,x],Log[10,y]}]
res2=Map[qqq,res]
lplot=ListPlot[res2,PlotRange->All]
pfit=Fit[res2,{1,x},x]

```

Para que a rotina inicie o cálculo da D_F precisa-se entrar com os seguintes parâmetros:

List: nome da lista que se pretende calcular a D_F , no caso será a lista de coordenadas geradas pela caminhada aleatória.

delta0: tamanho inicial de caixa, pode ser escolhido como o menor tamanho possível de caixa ou então o maior, o que vai diferenciar é que devemos escolher o parâmetro de variação de caixa.

Scale: parâmetro de variação da caixa. Deve ser escolhido inicialmente, se quisermos que a caixa inicial seja a menor devemos escolher um parâmetro de variação maior do que 1, para que as caixas possam ir aumentando. Se pelo contrário, escolhermos que essa caixa seja a maior devemos escolher um parâmetro de variação menor do que 1.

Min: tamanho mínimo de caixa, para o caso de começar com a caixa maior, ou maior, no caso de começar com a caixa menor.