

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL – UFRGS
DEPARTAMENTO DE GENÉTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR – PPGGBM

LIZANDRA JAQUELINE ROBE

**RELAÇÕES FILOGENÉTICAS ENTRE GRUPOS E ESPÉCIES NEOTROPICAIS DO
SUBGÊNERO *DROSOPHILA* INFERIDAS POR ANÁLISE MOLECULAR
DO GENE *AMD* (ALFA METIL-DOPA)**

Dissertação submetida ao Programa de
Pós-Graduação em Genética e Biologia Molecular
da UFRGS como requisito parcial
para a obtenção do grau de Mestre em Ciências.

Orientadora: Profa. Dra. Vera Lúcia da Silva Valente Gaiiesky

Co-Orientador: Prof. Dr. Élgion Lúcio da Silva Loreto

Porto Alegre, fevereiro de 2004.

O presente trabalho foi desenvolvido junto ao
Laboratório 210 da UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
e ao LabDros da UNIVERSIDADE FEDERAL DE SANTA MARIA,
com financiamento do CNPq, FAPERGS e PROPESQ-UFRGS.

BANCA EXAMINADORA

Prof. Dr. Louis Bernard Klaczko

Profa. Dra. Vera Nisaka Solferini

Profa. Dra. Karen Luisa Haag

*“Como todas as coisas estão em permanente mudança,
nenhum conhecimento é possível.”*
(Heráclito de Éfeso)

*“A classificação por descendência não pode ser inventada por biólogos,
ela pode apenas ser descoberta.”*
(Theodosius Dobzhansky)

Dedico este trabalho aos meus pais,
à minha irmã e ao meu esposo,
pela confiança, pelo apoio, pelo auxílio
e pelo amor que sempre me prestaram!

AGRADECIMENTOS

Ao Élgion, pela contínua orientação ao longo dos últimos seis anos, pelo apoio nos momentos de dificuldades técnicas e pelas freqüentes e sábias conversas amigas, enfim, por me ensinar a maior parte do pouco que, hoje, sei e sou como pesquisadora.

À Verinha, pela orientação e por me receber de braços abertos em seu Laboratório, propiciando sempre um ambiente de trabalho agradabilíssimo a todos. Agradeço também pelo entusiasmo e amabilidade com que, nos momentos difíceis, conseguia me incentivar a prosseguir em frente.

À Lenira, pelo exemplo de ética e profissionalismo sempre presente no LabDros.

Ao Elmo e a Helen, pelos mais diversos préstimos.

À Liliana, pelos valiosos empréstimos nos momentos de dificuldades, pela fundamental ajuda no uso do PAUP e, principalmente, pela amizade.

Ao Paulo, por estar sempre disposto a me ajudar e por continuar sendo este grande amigo que sempre foi.

Ao Marco, pelo auxílio na identificação das moscas.

Ao Ronaldo, pelas reconfortantes conversas e desabafos pessoais.

À Adriana, ao Newton e ao Daniel, por me “socorrerem”, sempre que necessário e pela inestimável amizade que construímos ao longo do último ano.

Ao Marcos, pelos freqüentes ensinamentos de bioinformática, pela companhia e pelas boas conversas nos finais-de-semana.

À Maríndia, pela sólida amizade, e pelo exemplo de persistência e dedicação ao trabalho.

À Rosane, à Ana, à Mônica, à Marícia e ao Fabiano, pela amizade e pela indispensável ajuda nos momentos de aperto na reta final.

À Fabi, pelos inspiradores momentos de discussão sobre filogenias e pelos sábios conselhos.

Aos demais colegas do Laboratório 210: Norma, André, Adri, Luís, Chirlei, Marisa, Cláudia, Tiago e Rodolfo, obrigada pelos ótimos momentos que juntos passamos.

À Nina e à Érika, por me confortarem e fazerem rir até mesmo nos momentos difíceis.

À Amanda e à Manuela, pelo auxílio na lavagem dos vidros para manutenção dos estoques de *Drosophila*.

À Carol, à Juline, à Juliana e à Vanessa, pela companhia no Labdros.

Aos meus colegas de Mestrado: Ana Luiza, Camila, Paulo, Andréia, Ana Ayala e Martín, pelas boas conversas nos breves intervalos e pelo forte companheirismo sempre presente. Ao Nico, pelos fundamentais auxílios nos computadores do departamento.

À Giovana e à Juliana, por tornarem um pouco menos difícil o meu primeiro ano em Porto Alegre.

Aos meus demais amigos que, mesmo distantes, me incentivam a sempre prosseguir: Luciane, Deise, Vívian, Irinês, Valquíria, Joelma, Giselda, Maguida e Raquel. Vocês são simplesmente D+.

Aos meus sogros, por muito me auxiliarem e por me acolherem de braços abertos em sua casa.

Ao meu sobrinho que, apesar das discussões, continua me apoiando em todos os momentos.

À minha irmã, por ser esta pessoa maravilhosa, que sempre me ajudou e torceu muito por minhas vitórias.

Aos meu pais, pelo amor, pelo exemplo, e pelo apoio incondicional que sempre me prestaram.

Ao Charles, por me acompanhar fielmente durante a elaboração desta dissertação, e por me fornecer momentos de entretenimento quando isto parecia bastante difícil.

Ao Sandro, pelo entendimento das faltas e pelo amor e incentivo nas presenças.

**ENFIM, Á TODOS AQUELES QUE, DE UMA FORMA OU DE OUTRA,
CONTRIBUÍRAM NA ELABORAÇÃO DESTA TRABALHO,
MEU SINCERO MUITO OBRIGADA!**

SUMÁRIO

Resumo	10
Abstract.....	11

Capítulo I

Introdução.....	12
I Sistemática Filogenética.....	13
1 Filogenias	15
1.1 A importância dos grupos-externos.....	16
1.2 Taxas evolutivas	19
2 Dificuldades da inferência filogenética	19
2.1 Variação em taxas evolutivas	20
2.2 Homoplasias	20
3 Sistemática Molecular	22
3.1 Fundamentos e limitações	24
3.1.1 Homoplasias	24
3.1.2 Taxas evolutivas	26
3.2 Marcadores Moleculares	27
3.2.1 Eletrofores de Isozimas	31
3.2.2 Citogenética Molecular.....	32
3.2.3 Hibridização DNA-DNA.....	32
3.2.4 RFLP (<i>Restriction Fragment Length Polymorphism</i>).....	34
3.2.5 PCR (<i>Polymerase Chain Reaction</i>) e Técnicas Relacionadas.....	34
3.2.6 SSCP (<i>Single-Strand Conformation Polymorphisms</i>), DGGE (<i>Denaturing Gradient Gel Electrophoresis</i>) e Análise de Heterodúplex	37
3.2.7 Sequenciamento.....	37
3.2.5.1 Que seqüência usar?	38
3.2.5.2 Isolamento das seqüências-alvo	39
3.2.5.3 Técnicas de sequenciamento	40

3.2.5.4 Alinhamento	41
3.2.5.5 Métodos de reconstrução filogenética	42
3.2.5.6 Testes de confiança em topologias	48
4 A Sistemática Filogenética e os Processos Evolutivos.....	51
4.1 Árvore das OTU´s X Árvore dos genes.....	53
II A Família Drosophilidae.....	55
1 O gênero <i>Drosophila</i>	56
2 Filogenia de Drosophilidae.....	65
2.1 Estudos Morfológicos.....	65
2.1.1 Throckmorton (1975)	65
2.1.2 Grimaldi (1990)	68
2.2 Sistemática Molecular	68
2.2.1 Os problemas na filogenia da família Drosophilidae	69
III O gene <i>Amd</i>	77
1 Tempo de Expressão.....	81
2 Utilização no desenvolvimento de hipóteses filogenéticas	81
Objetivos.....	84
Capítulo II	
Relações filogenéticas entre grupos e espécies Neotropicais do subgênero <i>Drosophila</i> (Drosophilidae: Diptera) inferidas com base em análise molecular do gene nuclear <i>Amd</i> (alfa metil-dopa)	85
Capítulo III	
Conclusão	105
Referências Bibliográficas.....	108

RESUMO

O gênero *Drosophila* é composto, tradicionalmente, por 15 subgêneros e mais de 1.400 espécies, perfazendo, sozinho, quase metade das espécies da família Drosophilidae. Esta diversidade, entretanto, ao mesmo tempo que fornece inúmeras oportunidades de investigação acerca de padrões evolutivos, impõe uma série de desafios taxonômicos e filogenéticos. De modo geral, pode-se dizer que a sistemática do gênero *Drosophila* ainda permanece, na maior parte dos casos, bastante controversa e incompleta, apesar de sua importância na elucidação da história evolutiva de inúmeros processos biológicos já evidenciados e caracterizados para o grupo. Este quadro é, ainda, agravado pela certa escassez de estudos moleculares direcionados à resolução filogenética de seus representantes Neotropicais, pertencentes, principalmente, ao subgênero *Drosophila*. Visando suprir algumas destas deficiências, realizou-se, neste trabalho, a análise filogenética de parte da região codificante do gene nuclear *Amd* (alfa metil-dopa) para um total de 49 espécies, representantes, em sua maioria, das radiações *immigrans-tripunctata* e *virilis-repleta* do subgênero *Drosophila*. Trata-se, portanto, da primeira abordagem molecular para muitas das espécies consideradas. Entre os resultados obtidos, destaca-se a corroboração de uma série de hipóteses de parafilia dentro da família Drosophilidae, entre as quais a do próprio gênero e do subgênero *Drosophila*. Novos estudos são, entretanto, necessários para o estabelecimento de um panorama mais concreto da filogenia destas espécies e, caso alguns dos resultados aqui apresentados se confirmem, uma revisão taxonômica tornar-se-á, possivelmente, indispensável.

ABSTRACT

The *Drosophila* genus is traditionally composed of 15 subgenus and more than 1.400 species, performing, alone, about half of the Drosophilidae family species. However, at the same time that this diversity provides numberless opportunities of investigation, concerning patterns of evolution for example, it imposes a set of taxonomic and phylogenetic challenges. Usually, we can say that the systematics of the *Drosophila* genus still remain fully controverse and incomplete, despite its importance on the elucidation of the evolutive history of numberless biological processes already known and characterized for the group. This picture is still worse given the scarcity of molecular studies seeking to resolve the phylogenetic relationships of the Neotropical members of the *Drosophila* subgenus. Aiming to supply some of these deficiencies we accomplished, in this work, the phylogenetic analysis of 513 base pairs of the codificant region of the *Amd* (alpha methyl-dopa) nuclear gene, for a whole of 49 taxa, mainly of which, are representatives of *immigrans-tripunctata* and *virilis-repleta* radiations of the *Drosophila* subgenus. Hence, this is the first molecular approach for a lot of the species here considered. Among the obtained results, we emphasize the corroboration of a set of hypothesis concerning merophyly inside de Drosophilidae family, among which we point out the paraphyletic status presented by the *Drosophila* genus and subgenus. We also corroborated many of Throckmorton (1975) radiations, as is the case for the *immigrans-Hirtodrosophila*, *tripunctata*, *virilis-repleta* and *repleta* radiations. Nevertheless, new studies are necessary to definitively establish the *Drosophila* subgenus phylogeny, and if some of the results here presented will come to be confirmed, a taxonomic revision of Drosophilidae will, certainly, become indispensable.

CAPÍTULO I

INTRODUÇÃO

I A SISTEMÁTICA FILOGENÉTICA

A aceitação da Teoria da Evolução proposta por Charles Darwin em 1859 levou à compreensão de que, subjacente à diversidade e aos padrões de ocorrência fenotípica, o fluxo genealógico conecta os organismos em diferentes escalas (DINIZ-FILHO, 2000). Com isso tem-se que, na maior parte das vezes, a expressão de um determinado caráter reflete, em maior ou menor grau, as relações de parentesco entre as espécies. Paradoxalmente, entretanto, durante muito tempo ainda, o uso destas relações na obtenção de uma classificação hierárquica, consistente e única dos organismos permaneceu na obscuridade. A classificação biológica continuava estruturada segundo as bases aristotélicas, seguidas por Linnaeus, sendo erigida com base em critérios simples de grau de semelhança (AMORIM, 1997).

Foi apenas em 1950 que HENNIG desenvolveu os procedimentos básicos para a reconstrução da história evolutiva entre grupos de organismos, contribuindo decisivamente para a criação de uma nova escola taxonômica, a “Sistemática Filogenética”, também conhecida como “Análise Cladística”. Dentre os conceitos essenciais adotados por esta nova Escola, destaca-se a compreensão adequada do termo “homologia”, como uma referência à estruturas, semelhantes ou não, encontradas em indivíduos diferentes e derivadas, muitas vezes com modificações, a partir de seu ancestral comum mais recente. A inferência de homologia deve ser feita a partir de comparações de ordem morfológica, topológica e/ou ontológica entre as estruturas apresentadas pelos grupos envolvidos, ou ainda, através da análise de filogenias dos mesmos (AMORIM, 1997). A partir desta noção inicial, pode-se propor a “série de transformação” que a estrutura sofreu, tornando-se sucessivamente mais derivada, sendo que a condição pré-existente de uma estrutura que foi modificada em uma condição nova foi chamada de “plesiomorfia” por HENNIG (1966), enquanto que a condição mais recente de uma estrutura, surgida por modificação de uma condição anterior foi chamada de “apomorfia”. Essa ordem de transformações pode ser linear (quando as séries são restritas a apenas duas condições ou mais de duas condições sucessivamente apomórficas) ou ramificada (quando as condições apomórficas originaram-se independentemente a partir da mesma condição plesiomórfica original) (MIYAKI *et al.*, 2001). O compartilhamento de estados plesiomórficos e apomórficos são denominados, respectivamente, “simplesiomorfias” e “sinapomorfias”; neste sentido, “autapomorfias”

são caracteres apomórficos para um único ramo terminal em um cladograma e “arqueomorfias” são condições apomórficas presentes em um grupo, mas que são sinapomórficas para um grupo mais amplo que aquele em foco (AMORIM, 1997).

Segundo HENNIG, a análise dos padrões de diferenciação interespecífica, em que descendentes compartilham condições apomórficas derivadas de caracteres plesiomórficos ancestrais, poderia refletir de modo inequívoco as relações de parentesco entre os grupos classificados. E é exatamente essa a essência do método filogenético: sinapomorfias como indícios de ancestralidade comum exclusiva, e conseqüentemente, de monofilia. HENNIG ainda propõe que as classificações biológicas deveriam constituir-se de uma expressão literal do conhecimento acumulado até então sobre as relações evolutivas entre os grupos, sendo que qualquer táxon estabelecido deveria apresentar uma origem estritamente monofilética.

Entretanto, a cladística somente passou a ser a escola taxonômica dominante a partir de 1980, permanecendo pouco difundida no mundo científico nos primeiros anos de sua existência (DINIZ-FILHO, 2000). Neste período houve muita controvérsia acerca de qual o Sistema de Classificação mais apropriado, sendo que por fim tanto adeptos da Escola Tradicional (que entendem que não deve haver um substrato filosófico subjacente à atividade classificatória) quanto feneticistas (que defendem que o Sistema de Classificação deve resultar da análise quantitativa de um grande número de caracteres cuja média de semelhança total leva à reunião ou separação de táxons) e gradistas (que entendem que o Sistema de Classificação deve ser construído para expressar os grados da evolução dos grupos) acabaram por perder esta batalha (AMORIM, 1997).

De um modo geral, a Sistemática Filogenética é o campo da Biologia Evolutiva com origem e aplicação interdisciplinar que visa, além da classificação e organização do conhecimento sobre a diversidade biológica, a partir das relações históricas entre os organismos, o entendimento da evolução e dos padrões de variação de caracteres morfológicos, fisiológicos, ecológicos, comportamentais, citogenéticos e moleculares (MIYAKI *et al.*, 2001). Segundo FUTUYMA (1997) a Sistemática Filogenética fornece ainda subsídios para a análise, interpretação e resolução de uma série de indagações características da Biologia, uma vez que, mesmo os fenômenos biológicos atuais são produtos do processo evolutivo, nos quais os eventos passados determinaram a situação atual. Neste sentido, o uso de filogenias interespecíficas pode auxiliar no estudo de tópicos

como especiação, biogeografia, hibridização, etc. (HILLIS *et al.*, 1996). Hipóteses filogenéticas podem ser convertidas, ainda, em informações tecnicamente importantes (AMORIM, 1997). Assim, por exemplo a descoberta do nível de surgimento de uma determinada característica humana fornece subsídios importantes para questões relativas a testes de reação em espécies semelhantes, ou até decisões referentes à transplantes de órgãos entre espécies. Mais do que isto, a confirmação da hipótese de monofiletismo para um grupo de parasitas pode levar à descoberta de sinapomorfias que podem ser alvo de ações farmacológicas eficazes contra todo o grupo. Assim, o método filogenético oferece múltiplas possibilidades, todas elas fundamentais para o completo entendimento da evolução biológica.

1 Filogenias

As Filogenias são diagramas que representam a história evolutiva dos organismos (MIYAKI *et al.*, 2001), expressando, portanto, as relações de ancestralidade entre os grupos analisados. Embora alguns autores ignorem divergências conceituais, AMORIM (1997) defende que, tecnicamente, existem diferenças importantes entre dendrogramas, filogenias, cladogramas e árvores filogenéticas. Segundo ele, “dendrograma” é qualquer diagrama ramificado que conecte elementos. Desta forma, a “filogenia”, de modo geral, é um dendrograma cujos elementos são táxons e cujas conexões indicam ancestralidade comum exclusiva. Neste contexto, o “cladograma” é uma filogenia que expressa relações filogenéticas apenas entre táxons terminais, sem incluir indicações de espécies ancestrais, enquanto que a “árvore filogenética” é uma filogenia que expressa relações filogenéticas tanto entre táxons terminais, quanto entre espécies ancestrais e espécies descendentes. Com isso tem-se que, se os cladogramas incluem como táxons terminais apenas grupos recentes, são topologicamente idênticos às suas árvores filogenéticas correspondentes; entretanto, se incluem espécies-fósseis entre os táxons terminais, as mesmas podem ser posicionadas como ancestrais em uma árvore filogenética do mesmo grupo, desde que não apresentem qualquer autapomorfia contraditória. AMORIM introduz ainda a expressão “cenário evolutivo” para filogenias mais complexas, que contém informações sobre idade absoluta dos táxons, distribuição das espécies ancestrais, bem como outros dados adicionais.

Segundo MIYAKI *et al.* (2001), graficamente, uma Filogenia consiste de nós (internos e terminais) ligados por ramos (internos e periféricos), de forma que, enquanto os nós terminais apresentam os táxons sobre os quais se busca inferir as relações filogenéticas

(também chamadas OTU's – *Operational Taxonomic Units*), os nós internos representam o ancestral comum mais recente desses táxons. Filogenias bem resolvidas são tipicamente dicotômicas – apresentando, para cada nó interno, um ramo que chega e dois ramos que partem. Quando ocorrem politomias, elas geralmente refletem ligações não resolvidas devido à incapacidade do método e/ou algoritmo utilizado, embora às vezes possa ser resultado de eventos de especiação múltipla (DINIZ-FILHO, 2000).

Em qualquer filogenia, aqueles grupos que apresentam um ancestral comum mais próximo entre si do que com os outros clados são chamados “grupos-irmãos”. Os clados representam “grupos monofiléticos”, também chamados “holofiléticos”, que são definidos como um conjunto de organismos que apresentam um ancestral comum exclusivo (incluindo tanto o táxon ancestral quanto todos os seus descendentes). Grupos não-monofiléticos podem ser chamados de “merofiléticos”, sendo que este termo inclui dois conceitos subjacentes: o de grupos “parafiléticos”, que apresentam um único ancestral comum não exclusivo (há espécies descendentes deste mesmo ancestral que não estão incluídas no grupo) e o de grupos polifiléticos, que reúnem representantes de dois ou mais grupos monofiléticos. (AMORIM, 1997; FUTUYMA, 1997; MIYAKI *et al.*, 2001)

De modo geral, segundo SWOFFORD *et al.* (1996), árvores totalmente dicotômicas, enraizadas ou não-enraizadas, apresentam, tipicamente, as seguintes características:

	Não-enraizada	Enraizada
Número de nós terminais	T	T
Número de nós internos	T – 2	T – 1
Número de ramos	2T – 3	2T – 2
Número de diferentes filogenias possíveis	$N = \frac{(2T - 3)!}{2^{T-2}(T - 2)!}$	$N = \frac{(2T - 5)!}{2^{T-3}(T - 3)!}$

Onde: T = número de OTU's

1.1 A importância dos grupos externos

Partindo-se da premissa de que a raiz da Filogenia representa a espécie ancestral de todo o grupo que se busca analisar (o grupo interno ou *ingroup*), uma Árvore Filogenética pode ser apresentada de forma enraizada ou não-enraizada. As relações de parentesco entre as espécies, que caracterizam a topologia das ligações, podem ser inferidas em ambos os casos, entretanto, o sentido temporal da evolução pode ser atribuído apenas a uma árvore enraizada (MIYAKI *et al.*, 2001). Para a adição de uma raiz à Filogenia, é necessária a inclusão de um grupo externo (ou *outgroup*) na análise. Este grupo externo deve ser uma

OTU próxima, mas que necessariamente divergiu antes do processo de divergência do grupo interno.

O grupo externo costuma ainda contribuir na definição das condições plesiomórficas para o grupo analisado, permitindo a polarização dos estados dos caracteres e o estabelecimento da direção das modificações evolutivas, ou seja, a série de transformação do carácter. Esta afirmação parte do pressuposto de que, se uma condição é apomórfica dentro de um grupo, então ela surgiu necessariamente depois da origem deste grupo. Na base da evolução do grupo, o que existia era a condição plesiomórfica, herdada de níveis ainda anteriores. As espécies atuais descendentes desses níveis anteriores que não fazem parte do grupo interno devem, portanto, apresentar a condição plesiomórfica daquela série de transformação (mas podem e devem conter apomorfias de outras séries de transformação) (AMORIM, 1997).

Este método foi proposto originalmente por HENNIG (1966), tendo sido gradualmente aperfeiçoado. Entretanto, é recomendada muita cautela em sua utilização, principalmente no que se refere à fundamental monofilia do grupo-interno. Caso esta premissa fundamental não seja seguida, ocorre um acúmulo de dados inconsistentes e incongruentes (AMORIM, 1997). Segundo o mesmo autor, na maioria das vezes o grupo-irmão é o grupo-externo que fornece dados mais aproveitáveis para a análise de polarização de caracteres, uma vez que é o mais comparável. Entretanto, qualquer outro grupo pode ser utilizado como grupo externo. Na verdade, recomenda-se a inclusão de vários grupos externos em estudos filogenéticos, de modo a evitar que homoplasias em um dos grupos externos resultem em interpretações equivocadas sobre a evolução de caracteres no grupo interno (MIYAKI *et al.*, 2001). Afinal, de acordo com a ordem de transformação evolutiva aceita, são impostas restrições à construção da árvore filogenética final.

Quando a análise do grupo-externo não fornece indícios concretos da série de transformação para uma característica qualquer, algumas alternativas plausíveis, apresentadas por AMORIM (1997), podem ser utilizadas. O procedimento de “otimização”, por exemplo, pode ser aplicado em casos de ocorrência de variação semelhante dentro e fora de um grupo sabidamente monofilético. Esta opção, embasada no princípio de máxima parcimônia, busca otimizar a distribuição de caracteres dentro de uma filogenia com um nível mais abrangente (que inclui também as relações entre os grupos externos),

possibilitando a inferência da condição plesiomórfica mais provável para o grupo-interno. Outra alternativa importante são os “grupos-externos funcionais”, que fornecem soluções ao problema de polarização de um caráter qualquer a partir da reconstrução inicial de alguns níveis da filogenia com base em outros caracteres. Neste caso, a polarização depende mais uma vez da comparação de grupos-externos a um grupo monofilético, mas agora estando ambos situados dentro do grupo de estudo.

A análise dos dados pode ainda ser feita seguindo um critério não-ordenado, em que, na ausência de informações adicionais, todas as seqüências entre as condições conhecidas são aceitas (MIYAKI *et al.*, 2001). Entretanto, este critério deve ser utilizado com cautela, uma vez que, em algumas situações pode permitir séries de transformação muito pouco prováveis (AMORIM, 1997).

Na inferência de filogenias moleculares, como será visto mais detalhadamente a seguir, os grupos-externos possuem uma utilidade maior no enraizamento das árvores reconstruídas (SWOFFORD *et al.*, 1996). Isto ocorre porque, principalmente no caso de seqüências de nucleotídeos, torna-se bastante arriscado admitir um determinado estado como plesiomórfico ou apomórfico em um sítio qualquer, pois isto implicaria em ausência de evolução nesse sítio a partir do ancestral comum até os seus representantes atuais (MIYAKI *et al.*, 2001). Assim, quando os métodos de reconstrução filogenética produzem topologias não-enraizadas, a posição do grupo-externo pode definir a raiz da filogenia com respeito ao grupo-interno. Alternativamente, segundo SWOFFORD *et al.* (1996), quando uma uniformidade de taxas evolutivas é assumida, a raiz pode ser situada na metade do ramo que conecta as duas OTU's mais divergentes, estratégia esta conhecida como “enraizamento no ponto médio” ou *midpoint-rooting*. Este método assume, entretanto, que estas taxa mais divergentes evoluem a taxas iguais, o que pode não ser verdadeiro.

TARRÍO *et al.* (2000) destacam, entretanto, que o enraizamento é, frequentemente, a etapa mais precária em qualquer reconstrução filogenética. Isto porque, o(s) grupo(s)-externo(s) podem causar problemas na designação da filogenia correta quando o *outgroup* adotado for muito extremo ou quando este apresentar diferenças, em relação ao grupo-interno, quanto aos processos de substituição nucleotídica. DESALLE (1992a) afirma que, no primeiro caso, se o grupo-externo adotado for muito distantemente relacionado ao grupo-interno (devido a um longo tempo de divergência ou à ocorrência de taxas evolutivas aceleradas), além da ocorrência do fenômeno conhecido como “atração de

ramos longos”, o comprimento dos ramos do grupo-interno tornam-se excessivamente longos quando relacionados ao grupo-externo. Este viés, causado pela alta frequência de homoplasias ocorrendo em sítios pouco restritos funcionalmente, pode ser combatido pelo emprego de métodos de reconstrução que levem em conta a variação na taxa de substituição apresentada pelos diferentes sítios presentes nas seqüências. Entretanto, é preciso ter cuidado na remoção ou consideração desigual de determinados sítios, uma vez que isso pode acarretar a perda de informações filogenéticas importantes.

Além disso, outro cenário que traz complicações no enraizamento filogenético envolve casos em que os grupos-externos divergiram em sua constituição nucleotídica com relação ao grupo-interno (TARRÍO *et al.*, 2000). Isto pode levar a uma grande instabilidade no posicionamento da raiz, dependendo dos critérios de reconstrução adotados, sendo que, muitas vezes, modelos mais simples podem superar modelos mais realísticos na inferência da árvore verdadeira. Neste caso, segundo TARRÍO *et al.* (2000), uma correção que relaxe a hipótese estacionária de constância nas frequências de nucleotídeos, pode fornecer o modelo evolutivo mais adequado. De qualquer forma, é importante que, quando possível, grupos-externos com composição nucleotídica similar ao grupo-interno sejam adotados.

1.2 Taxas Evolutivas

Muitas filogenias apresentam ainda o comprimento dos ramos, que pode ser expresso em tempo real (milhões de anos, por exemplo) ou relativo (dado por distâncias genéticas) (DINIZ-FILHO, 2000). Entretanto, segundo PEREIRA *et al.* (2001), estas atribuições estão geralmente embasadas em um relógio molecular atuante, ou seja, taxas evolutivas semelhantes entre as linhagens, de forma que devem ser utilizadas com cautela.

2 Dificuldades da Inferência Filogenética

O conhecimento filogenético não é estável, ele evolui. Apesar de apresentar como ponto de referência a tentativa de reconstruir uma história única, as hipóteses filogenéticas não podem ser provadas com certeza absoluta. Pelo contrário, elas são constantemente modificadas e aperfeiçoadas, não apenas devido à expansão da base de dados sobre o grupo (pela introdução de novas espécies ou caracteres) mas também por uma reinterpretação dos dados disponíveis. Assim, cladogramas são, essencialmente, inferências em permanente transformação, que aproximam-se da verdade à medida que o conhecimento da diversidade biológica de táxons aumenta (AMORIM, 1997). Entretanto, é

preciso ter em mente que as filogenias são conjuntos de suposições, de forma que cada hipótese individual pode ser correta ou não e a alteração de uma delas não invalida, necessariamente, as demais.

Dois processos podem ser apontados como os principais responsáveis pelo acréscimo de dificuldades na inferência de filogenias: a variação nas taxas evolutivas e as homoplasias. Entretanto, é importante notar que estes fatores não atribuem limitações aos princípios de Sistemática Filogenética, apenas impõem dificuldades que diferentes métodos buscam superar.

2.1 Variação em taxas evolutivas

Se cada linhagem divergisse das demais numa taxa constante, o grau de diferença entre duas OTU's quaisquer seria diretamente proporcional ao tempo transcorrido desde sua divergência. Assim, a filogenia poderia ser inferida a partir do grau total de diferenças entre os grupos. Entretanto, devido a variações no tamanho populacional e na amplitude das pressões seletivas, muitas vezes taxas inconstantes de evolução estão atuando, o que vêm afetar a correspondência entre a diferença observada e o tempo de divergência. Na verdade, além de diferenças nas taxas de evolução entre linhagens, frequentemente observa-se uma inconstância temporal e espacial de taxas evolutivas dentro de uma mesma linhagem. Até mesmo dentro da linhagem, diferentes características apresentam taxas evolutivas adversas, o que caracteriza a “evolução em mosaico”. (FUTUYMA, 1997)

Além de afetar as relações inferidas dentro do grupo de interesse, como já visto anteriormente, estas diferenças podem influenciar no comprimento dos ramos apresentados na filogenia.

2.2 Homoplasias

O termo “homoplasia” refere-se a casos de semelhança adquirida independentemente na evolução de um grupo. Com isso, dois ou mais grupos apresentam uma mesma condição apomórfica, sendo que sua espécie ancestral comum mais recente não portava esta condição. As homoplasias podem surgir de três maneiras distintas: (1) em duas espécies próximas, uma mesma condição plesiomórfica é alterada gerando condições apomórficas semelhantes, o que se constitui em um evento de “paralelismo”; (2) em duas espécies diferentes (mais afastadas que em 1), condições plesiomórficas diversas são alteradas resultando em condições apomórficas semelhantes (geralmente a semelhança é apenas superficial), o que caracteriza a “evolução convergente” ou “convergência

evolutiva”; (3) em uma determinada espécie uma característica arqueomórfica é alterada resultando em uma condição apomórfica final semelhante à condição plesiomórfica original, o que representa uma “reversão evolutiva”. Em situações ideais, na evolução convergente as apomorfias evoluem independentemente através de diferentes linhas de desenvolvimento, enquanto na evolução paralela o desenvolvimento das estruturas homoplásicas é similar. (AMORIM, 1997; FUTUYMA, 1997)

Com isso, casos de compartilhamento de apomorfias não podem ser vistos como indícios definitivos de monofilia, o que pode gerar conflito entre os dados filogenéticos e até problemas para a análise. Entretanto, diversos métodos já foram desenvolvidos para detectar e posteriormente superar as dificuldades inerentes à presença de homoplasias, validando apenas sinapomorfias verdadeiras.

O teste de congruência, cujo resultado pode ser expresso na forma de diagramas ou tabelas, é capaz de revelar a presença ou ausência de homoplasias. A congruência entre os caracteres é verificada quando as relações entre os táxons são apenas de inclusividade completa ou de exclusão mútua. Neste caso, diz-se que os caracteres são congruentes, sendo sua presença explicada apenas por sinapomorfias. Contrariamente, caracteres incongruentes evidenciam a presença de pelo menos um caráter homoplásico. Entretanto, o fato de um determinado carácter ser homoplásico não implica que ele não seja útil nas inferências filogenéticas em um nível taxonômico mais restrito, onde pode ser um indicador de monofilia (AMORIM, 1997).

HENNIG (1966) recomenda que, em casos de incongruência, a construção de filogenias se dê com base em “caracteres únicos”, ou seja, aqueles que não fossem conhecidos em grupos-externos. Entretanto, a solução mais utilizada atualmente para o conflito de caracteres baseia-se no princípio da máxima parcimônia, que irá minimizar a ocorrência de homoplasias e maximizar a ocorrência de sinapomorfias, através da escolha da(s) filogenia(s) que apresente(m) o menor número de eventos de surgimento de apomorfias (passos evolutivos). AMORIM (1997) recomenda ainda, como forma eficaz de aumentar a confiabilidade sobre a filogenia proposta, estender o número de caracteres analisados, o que tende a diminuir o efeito de eventuais homoplasias.

É importante destacar que autapomorfias são sempre congruentes com todos os caracteres, uma vez que, como autapomorfias não são informativas sobre as relações entre táxons terminais, não podem ser incongruentes com nenhuma outra informação que diz

respeito à relações entre táxons terminais (AMORIM, 1997). Desta forma, esses caracteres são importantes para demonstrar que os táxons terminais são monofiléticos, mas não ajudam a esclarecer as relações filogenéticas entre os membros do grupo-interno.

3 Sistemática Molecular

Ao mesmo tempo que a Sistemática Filogenética se desenvolvia, adquirindo novos adeptos, a Biologia Molecular também passava por fortes transformações. Novos métodos que permitiam a análise da estrutura molecular de ácidos nucleicos e proteínas eram descobertos, e sua adoção pelos evolucionistas possibilitou um enorme avanço na Sistemática Molecular. Esta sub-área da sistemática é caracterizada exatamente pelo uso de marcadores moleculares na elaboração de inferências sobre processos populacionais e filogenéticos, criando para isso um substancial banco de dados referente a proteínas ou genes específicos (MORITZ & HILLIS, 1996). Entretanto, o contexto teórico, os métodos gerais de análise e a maneira de lidar com os resultados filogenéticos não diferem, independente da fonte de informação ser morfológica ou molecular.

Há que se destacar, neste ponto, que com a ampla difusão da Sistemática Molecular, os estudos filogenéticos começaram a crescer exponencialmente (pelo menos para alguns táxons). Logo os dados moleculares passaram a ser utilizados, não apenas para esclarecer relações filogenéticas difíceis de abordar por outros métodos, mas também para questionar visões filogenéticas propostas anteriormente. Isto não quer dizer, no entanto, que as filogenias fundamentadas no uso de caracteres morfológicos tenham sido refutadas ou que a análise destes não caracterize uma ferramenta adequada para análise filogenética.

Na verdade, embora em alguns casos dados moleculares tenham revelado limitações da análise morfológica (pela deficiência de variações entre certos clados, excesso de homoplasias, ou até mesmo complexidades relacionados ao seu padrão evolutivo), ambos enfoques tem suas vantagens e desvantagens, e deveriam ser utilizadas conjuntamente. Por exemplo, dados moleculares costumam ter uma base genética clara e uma simplicidade de padrões evolutivos, sendo que homologias podem ser buscadas com maior segurança. Por outro lado, caracteres morfológicos podem ser prontamente obtidos de ancestrais fósseis (quando estes estão disponíveis), apresentando ainda o poder de interpretação em um contexto ontogenético (MORITZ & HILLIS, 1996). Entretanto, segundo MIYAKI *et al.* (2001), ao refutar pressupostos simplistas ou excessivamente reducionistas, percebe-se que a herdabilidade da maior parte dos caracteres morfológicos não é absoluta,

e como os caracteres devem ser herdados estavelmente para que apresentem um significado evolutivo, espera-se que enfoques moleculares produzam resultados mais confiáveis. Além disso, os caracteres moleculares de modo geral costumam ser mais variáveis, devido à ocorrência de mutações neutras e à concomitante atuação da deriva genética. A heterogeneidade de taxas evolutivas apresentada, ainda, por diferentes marcadores, faz com que as estratégias moleculares se adaptem, perfeitamente, aos mais variados propósitos e possibilitem a extração de informações filogenéticas para níveis taxonômicos próximos e/ou divergentes. Todas estas vantagens fazem os eventuais contratempos apresentados pela sistemática molecular (homoplasias, saturação, heterogeneidade de taxas evolutivas entre diferentes OTU's ou até mesmo diferentes sítios dentro de uma seqüência, diferenças entre filogenia do marcador e filogenia da espécie, etc.) parecerem meros detalhes, desde que devidamente superados.

Recentemente, entretanto, diversos autores (DESALLE & GRIMALDI, 1992; O'GRADY *et al.*, 1998; O'GRADY, 1999; REMSEN & O'GRADY, 2002) tem destacado a importância central das análises combinadas de nível morfológico-molecular, ou até mesmo molecular de diferentes loci, na compreensão da verdadeira história evolutiva das espécies. Entretanto, existem diversas escolas de pensamento acerca de qual é o melhor tratamento na análise das diferentes partições. Segundo alguns autores (revisão em O'GRADY *et al.*, 1998), a análise simultânea de diferentes caracteres maximiza a eficiência descritiva e o poder filogenético exploratório dos dados, uma vez que conduz à superação dos desvios tendenciosos e ao acúmulo de apoio a diversos nós internos, mesmo que algumas partições acrescentem incongruências em outros níveis. Muitas vezes, inclusive, relações não apresentadas na análise individual de qualquer uma das partições, são reveladas mediante a construção de uma árvore de evidências totais (O'GRADY *et al.*, 1998), ou então, mesmo mediante a presença de incongruências entre as partições, a combinação pode gerar um acréscimo de suporte a uma determinada hipótese filogenética que individualmente era defendida por apenas um ou alguns dos dados (GLEASON *et al.*, 1998). Entretanto, diversos outros pesquisadores (MIYAMOTO & FITCH, 1995) acreditam que é preferível apresentar um consenso das diferentes análises individuais a fazer uma “mistura de caracteres”, muitos dos quais incongruentes entre si e apresentando diferentes histórias evolutivas. Uma posição intermediária, chamada combinação condicional ou concordância primária, propõe que as partições sejam sujeitas a um teste de congruência e

incluídos na análise combinada apenas se devidamente congruentes. Neste caso pode-se optar pelo critério estrito, segundo o qual apenas aquelas partições congruentes com todas as outras são incluídas, ou pelo critério permissivo, que permite a inclusão de todas as partições parcialmente congruentes com as demais (revisão em O'GRADY, 1999). CUNNINGHAM (1997) demonstrou, entretanto, que apenas em casos de incongruências com valores P menores do que 0,001, a combinação das partições pode reduzir a acurácia filogenética.

Neste contexto, mesmo que não se faça a combinação, pode-se facilmente inferir o grau de concordância e discordância entre os diferentes conjuntos de dados, comparar os índices de consistência obtidos em análises distintas e/ou conjuntas, bem como restringir uma das análises à hipótese atribuída pela outra. Por fim, com base nestes resultados, pode-se assumir com maior segurança qual a hipótese que melhor representa a evolução das OTU's em questão e, a partir disto, como se deu a evolução de cada um dos seus caracteres em particular. Porém, como estes procedimentos só podem ser realizados mediante o acúmulo de inferências filogenéticas de âmbito morfológico e molecular, estas últimas análises continuam a constituir o ponto de partida para qualquer estudo evolutivo.

3.1 Fundamentos e Limitações

3.1.1 Homoplasias

Em geral, como visto anteriormente, a homologia deve ser a preocupação primordial do sistemata na reconstrução de qualquer tipo de filogenia. Esta, no entanto, deve ser inferida com base em herança a partir do ancestral comum mais recente e não em termos de similaridade empírica, como é muitas vezes utilizada. Afinal, não apenas a ancestralidade comum pode afetar a semelhança de duas moléculas quaisquer, como também as convergências, paralelismos e reversões, ou ainda a conversão gênica. A inferência quanto à homologia de estruturas é, pois, uma hipótese qualitativa e nunca quantitativa (RUSSO, 2001).

Existem ainda diferentes formas de homologia que devem ser diferenciadas entre si, quando marcadores moleculares são considerados. Se duas seqüências homólogas forem derivadas de um ancestral comum através de eventos de especiação, diz-se que as mesmas são “ortólogas”; por outro lado, se elas se originaram através de um evento de duplicação de um loco ancestral, diz-se que são seqüências “parálogas” (FUTUYMA, 1997; AMORIM, 1997). Quando seqüências parálogas não evoluem independentemente, mas em concerto,

devido a processos de conversão gênica ou *crossing-over* desigual, elas são denominadas seqüências “plerólogas” (MORITZ & HILLIS, 1996). Seqüências homólogas podem ainda estar relacionadas a eventos de transferência gênica horizontal, sendo então chamadas de “xenólogas” (MORITZ & HILLIS, 1996).

Esta distinção é fundamental, uma vez que apenas seqüências ortólogas devem ser utilizadas para fins de inferência filogenética entre organismos. O uso de seqüências parálogas, neste caso, pode resultar em uma estimativa correta da filogenia das moléculas envolvidas, mas esta não refletiria as relações entre os organismos dos quais as mesmas foram amostradas (MORITZ & HILLIS, 1996). Tipicamente, quando se utiliza seqüências parálogas para fins de reconstrução filogenética de grupos de organismos, o tempo de divergência entre duas OTU's quaisquer é superestimado, pois o evento de duplicação costuma ser anterior ao evento de especiação (RUSSO, 2001). Entretanto, quando a taxa de evolução em concerto em seqüências plerólogas é alta, a distinção entre seqüências ortólogas e parálogas fica obscurecida, de forma que as últimas podem ser utilizadas para inferir as relações entre táxons (desde que a taxa de homogeneização seja maior que a taxa de especiação dentro do grupo de interesse) (MORITZ & HILLIS, 1996; HILLIS *et al.*, 1996).

Ainda segundo MORITZ & HILLIS (1996), o uso da reação de polimerização em cadeia (PCR) tem aumentado os riscos de problemas práticos associados à paralogia. Isto ocorre não apenas devido à possibilidade de amplificação de pseudogenes ou outras cópias parálogas funcionais do gene de interesse, mas também devido à amplificação de produtos de recombinação *in vitro* entre alelos diferentes, parálogos funcionais ou pseudogenes. Para evitar este tipo de problema é sempre recomendada a análise em gel dos produtos do PCR, bem como a confirmação da presença da seqüência amplificada no organismo de interesse.

Quando duas seqüências são sabidamente ortólogas outra etapa limitante, que deve ser administrada com muita cautela, diz respeito ao alinhamento das moléculas, quando é fundamental que seja preservada a homologia posicional entre sítios individuais (MORITZ & HILLIS, 1996; RUSSO, 2001). Esta tarefa é muitas vezes dificultada pela presença de *indels* (inserções e/ou deleções entre sítios) ou alta heterogeneidade entre as seqüências (MIYAKI *et al.*, 2001). Entretanto, se o alinhamento gerado não é confiável, provavelmente a filogenia proposta com base no conjunto de dados alinhados também não será correta (RUSSO, 2001). Além disso, uma vez feito o alinhamento correto das seqüências, é

necessário estabelecer as relações de homologia entre os estados de carácter presentes em cada sítio. Nesta etapa, dificuldades relacionadas à evolução paralela e às reversões, levando à homoplasias, bem como às substituições múltiplas, tudo isso conduzindo à saturação dos dados, podem causar erros na inferência das relações evolutivas (NEI & KUMAR, 2000).

3.1.2 Taxas evolutivas

A variabilidade apresentada por qualquer molécula constitui-se em um balanço entre as mutações que surgem e suas restrições estruturais e funcionais. A esperada constância relativa destas interações entre grupos taxonômicos (HILLIS *et al.*, 1996) e as indicações prévias da correlação entre a estimativa de divergência entre seqüências e o tempo de divergência entre os organismos dos quais as mesmas foram amostrados (ZUCKERKANDL & PAULING, 1962) levaram ao estabelecimento da hipótese do relógio molecular, segundo a qual comparações moleculares poderiam fornecer indícios do tempo de divergência entre táxons na ausência de fósseis auxiliares. Entretanto, experimentos mais recentes tem demonstrado não apenas a variação nas taxas de substituição entre locus e até mesmo sítios dentro de um mesmo genoma, como também a heterogeneidade nas taxas de substituição de um mesmo locus entre linhagens diferentes (CALCAGNOTTO, 2001).

Estas averigüações acarretam uma série de implicações na Sistemática Molecular. Por exemplo, a constância nas taxas evolutivas, além de ser uma premissa básica da Teoria Neutra de Evolução Molecular, é assumida em uma série de métodos filogenéticos e tem sido largamente usada na estimativa do tempo de divergência entre linhagens, se pelo menos a idade de um determinado evento de cladogênese tiver sido previamente estimada (PEREIRA *et al.*, 2001). Nessas situações, o comprimento dos ramos de uma árvore refletiriam o número de substituições de cada linhagem.

Existem atualmente diferentes maneiras de se testar a hipótese do relógio molecular em um conjunto de dados. Entre estas destaca-se teste de taxas relativas, segundo o qual a diferença entre o número de substituições ocorridas em diferentes linhagens a partir do seu ancestral comum mais recente não deve ser estatisticamente diferente de zero (PEREIRA *et al.*, 2001). Entretanto, para avaliar isto deve-se incluir uma linhagem-externa, que tenha divergido anteriormente à separação das linhagens consideradas. Neste caso, o número de

substituições ocorrido em cada linhagem pode ser obtido diretamente a partir das seqüências, da seguinte forma:

$$\begin{aligned}K_{xa} &= (K_{xe} + K_{xy} - K_{ye})/2 \\K_{ya} &= (K_{ye} + K_{yx} - K_{xe})/2 \\K_{ea} &= (K_{ex} + K_{ey} - K_{xy})/2\end{aligned}$$

- Onde:
- K = número de substituições
 - x, y = linhagens-irmãs
 - a = ancestral comum mais recente entre x e y
 - e = linhagem-externa

Assim, se as seqüências estiverem evoluindo a taxas semelhantes, a diferença entre K_{xa} e K_{ya} não deve ser estatisticamente diferente de zero. Programas como o PHYLTEST executam esta análise de taxas relativas (PEREIRA *et al.*, 2001). Já o programa conhecido como LINTREE encontra-se embasado no teste de taxas relativas aplicado a diferenças de comprimento dos ramos entre grupos ou seqüências-irmãs, no modelo denominado “árvore linearizada”. O “teste de razão de verossimilhança”, também pode ser aplicado com esta finalidade, quando são comparadas as probabilidades de uma mesma árvore com e sem a restrição do relógio molecular.

Existem diferentes modelos que incorporam o relaxamento do relógio molecular à análise do conjunto de dados disponíveis (PEREIRA *et al.*, 2001). Entre estes, o método de máxima verossimilhança estima o comprimento de cada ramo de modo a maximizar a probabilidade da árvore. Outros modelos, como o de CUTLER (2000) admitem variações limitadas nas taxas de substituição entre diferentes linhagens. Recentemente, ainda, alguns autores descreveram novos modelos, segundo os quais a variação nas taxas evolutivas entre OTU's é diretamente proporcional à distância existente entre elas.

3.2 Marcadores Moleculares

Marcadores moleculares são locus gênicos que apresentam alguma variabilidade no escopo do problema a ser estudado. O uso de técnicas que forneçam marcadores moleculares é capaz de detectar, então, a variabilidade genética subjacente, o que se constitui, por fim, na matéria-prima da evolução. Portanto, é pela análise desta variabilidade genética que podemos, não apenas analisar, mas também entender o processo evolutivo como um todo. Porém, diferentes marcadores moleculares exibem taxas evolutivas distintas, bem como vantagens e limitações particulares, que devem ser cuidadosamente analisadas no momento da escolha do marcador apropriado.

Uma característica distintiva dos marcadores atuais diz respeito ao seu padrão de herança: marcadores dominantes, como aqueles obtidos por RAPD e AFLP são menos úteis em análises populacionais, pois requerem a pressuposição de que a população esteja em equilíbrio para o cálculo de suas frequências gênicas, o que é frequentemente violado (SOLÉ-CAVA, 2001). Para este tipo de estudo evolutivo, recomenda-se, então, a escolha de marcadores co-dominantes, como isozimas ou microssatélites.

Outra diferença importante entre os marcadores moleculares disponíveis são suas taxas evolutivas. Por exemplo, microssatélites e RAPDs evoluem muito rapidamente, ao contrário das isozimas. Este critério também deve ser cuidadosamente considerado na escolha do marcador apropriado para diferentes abordagens, de modo que o mesmo seja capaz de fornecer um nível suficiente de variação para possibilitar uma análise evolutiva, mas não esteja saturado de homoplasias (DOWLING *et al.*, 1996). SOLÉ-CAVA (2001) recomenda a adequação do grau de variabilidade do marcador molecular escolhido ao nível de divergência que se deseja estudar, enquanto DOWLING *et al.* (1996) acrescenta que o marcador ideal é aquele que varia entre, mas, preferencialmente, não dentro dos grupos de estudo. Assim, marcadores com altas taxas evolutivas são úteis para o estudo de indivíduos, famílias e populações, enquanto que marcadores com baixas taxas evolutivas são indicados para estudos de espécies ou táxons supra-específicos. Caso contrário, quando se utiliza um marcador que evolui muito lentamente para o nível estudado encontra-se uma saturação de plesiomorfias, de modo que, ao final, o marcador não será informativo; por outro lado, quando se utiliza um marcador que evolui muito rapidamente para o nível escolhido, encontra-se uma saturação de homoplasias, o que obscurece o sinal filogenético.

Além das considerações acerca da taxas evolutivas dos diferentes marcadores, é importante que o seu modo de herança também seja analisado durante o processo de escolha da seqüência apropriada ao problema em questão. Neste sentido, pode-se optar por seqüências nucleares, mitocondriais ou até mesmo de cloroplasto, em plantas.

O DNA mitocondrial, por exemplo, fornece excelentes marcadores para a inferência de filogenias maternas, principalmente entre espécies próximas, devido às suas peculiares propriedades de herança uniparental clonal e ausência de recombinação (ARIAS *et al.*, 2001). Em geral, essas moléculas apresentam uma alta taxa evolutiva, devido, principalmente, à larga ocorrência de transições. Neste caso, pode-se optar pela análise de

diferentes lócus, considerando que, em geral, genes codificantes de proteínas apresentam uma menor restrição evolutiva em relação a seus análogos para RNA. O isolamento do DNA mitocondrial, bem como sua posterior análise, são amplamente facilitados em decorrência de suas propriedades de homoplasma individual e elevado número de cópias por célula (PALUMBI, 1996). Segundo este autor, entretanto, a maior limitação na utilização do DNA mitocondrial em análises evolutivas diz respeito à potencial possibilidade de comparação entre cópias parálogas, obtidas a partir da amplificação de pseudogenes mitocondriais presentes no genoma nuclear. DOWLING *et al.* (1996) destaca, ainda, que a evolução extremamente rápida de alguns lócus mitocondriais resulta em muitas homoplasias, o que pode confundir a inferência das relações filogenéticas entre espécies distantemente relacionadas. Em *Drosophila*, por exemplo, a forte tendência em direção a altos valores de A + T conduzem a um efeito de saturação, que além de causar uma aparente diminuição nas taxas evolutivas à medida que o tempo de divergência entre os taxa aumenta, também reduz o poder discriminatório que a análise de DNA mitocondrial apresenta na reconstrução das relações filogenéticas entre espécies não proximamente relacionadas (BARRÍO *et al.*, 1992). Por outro lado, segundo AVISE *et al.* (1984), a amostragem inadequada de polimorfismos intraespecíficos conduz à filogenias errôneas quando espécies proximamente relacionadas estão sendo comparadas, uma vez que é provável, nestes casos, que algumas das linhagens mitocondriais presentes em ambas espécies sejam anteriores ao evento de especiação.

O DNA de cloroplasto, por outro lado, tem sido pouco utilizado em estudos evolutivos, provavelmente devido ao fato de apresentar uma baixa taxa de substituição, grande tamanho e padrão de herança variável (ARIAS *et al.*, 2001). Uma característica peculiar deste genoma diz respeito à quase invariabilidade de ordem e arranjo de seus genes, bem como à sua transmissão essencialmente clonal, mesmo em casos de herança biparental (DOWLING *et al.*, 1996). Mais uma vez, uma amostragem populacional inadequada pode conduzir à erros, embora, neste caso, este problema esteja grandemente reduzido devido aos baixos níveis de polimorfismo. Por outro lado, o potencial fluxo gênico e a eventual hibridização entre espécies de plantas representam um grande problema para a análise filogenética destes marcadores.

Já o DNA nuclear representa a maior e mais poderosa fonte de marcadores genéticos, cuja variabilidade de taxas evolutivas permite análises de nível intra e

interespecífico (ARIAS *et al.*, 2001). Trabalhando com este genoma, pode-se optar por utilizar seqüências de cópia única ou com padrão repetitivo, presentes em regiões codificadoras ou não codificadoras. Assim, por exemplo, seqüências VNTR (*Variable Number of Tandem Repeats*), que incluem micro e minissatélites, são tipicamente hipervariáveis, apresentando fundamental importância para análises comparativas de nível populacional, apesar de suas limitações referentes às possíveis homoplasias (DOWLING *et al.*, 1996). Por outro lado, a análise de seqüências codificadoras de cópia única pode fornecer substrato para comparações de nível intra e inter-específico, mediante análise dos íntrons, que costumam ser particularmente variáveis, e até de éxons, cuja variabilidade varia de acordo com a restritividade funcional do gene em questão. Seqüências repetidas, como o rDNA, também costumam fornecer tanto marcadores conservados, encontrados nas regiões codificadoras, quanto marcadores mais variáveis, referidos como espaçadores transcritos e não-transcritos.

DOWLING *et al.* (1996) destaca, ainda, a importância de se considerar a taxa de rearranjos no momento da escolha da seqüência a ser analisada. Isto porque os rearranjos, a menos que cuidadosamente caracterizados, costumam levar à superestimativas de divergência entre as seqüências consideradas. Por outro lado, uma vez identificados, estes rearranjos podem fornecer informações fundamentais para a inferência filogenética, devido à sua raridade e conseqüente ausência de homoplasias. Neste sentido, é importante considerar que seqüências mitocondriais e de cloroplasto costumam ser mais estáveis que as nucleares.

Outra opção que deve ser analisada com cuidado diz respeito à escolha entre a análise a nível de nucleotídeos ou aminoácidos, quando seqüências codificadoras estão em foco. Neste caso, deve-se considerar que, embora a seqüência protéica seja menos informativa, sua seqüência gênica codificadora pode, eventualmente, estar sujeita à heterogeneidade de taxas evolutivas, encontrando-se saturada ou ainda exposta ao viés composicional nas terceiras posições dos códons (SWOFFORD *et al.*, 1996). Assim, quando a divergência entre as OTU's é pequena, recomenda-se a análise a nível de nucleotídeos, cujos sítios informativos podem restringir-se às substituições sinônimas ocorridas nas terceiras posições dos códons. Por outro lado, quando existe uma substancial divergência entre as seqüências, pode ser recomendável limitar à análise às primeiras e segundas

posições de cada códon, nestes casos mais informativas. Uma terceira alternativa seria inferir a seqüência protéica a partir do gene e utilizá-la para as análises posteriores.

Estes são apenas alguns dos critérios mais importantes, que devem ser analisados no momento da escolha do marcador a ser usado. Poder-se-ia citar, ainda, a importância do tipo do material disponível, que deve ser compatível com as necessidades apresentadas pela técnica, e o custo geral da metodologia.

Uma série de diferentes técnicas são rotineiramente utilizadas na obtenção de marcadores moleculares diversos. De modo geral, os dados gerados por estas técnicas podem ser divididos em duas categorias: caracteres discretos, que fornecem dados sobre OTU's ou seqüências individuais; dados de distância ou similaridade, que representam comparações quantitativas de pares de OTU's (SWOFFORD *et al.*, 1996). Segundo o mesmo autor, os primeiros podem ainda ser subdivididos em caracteres qualitativos, binários ou multiestado, ou quantitativos (neste caso referindo-se às frequências de alozimas ou haplótipos mitocondriais).

Apresentar-se-á, a seguir, algumas das características principais das técnicas utilizadas pela Sistemática Molecular, enfatizando seus princípios, aplicações e limitações.

3.2.1 Eletroforese de Isozimas

Esta técnica, embasada nos princípios de variação eletroforética e coloração citoquímica, foi aplicada pela primeira vez na década de 60, por dois grupos independentes (HARRIS, 1966; LEWONTIN & HUBBY, 1966). Desde então, as isozimas (enzimas que atuam sobre o mesmo substrato), têm sido fundamentais em diversos estudos de Biologia Evolutiva, possibilitando um grande avanço na compreensão dos processos micro e macroevolutivos. Para uma revisão sobre a metodologia empregada, ver MURPHY *et al.* (1996) e SOLFERINI & SCHEEPMAKER (2001).

Segundo ARIAS *et al.*, (2001), as isozimas constituem-se marcadores extremamente úteis, uma vez que, embora fenotípicos, não são profundamente afetados pela interação com o ambiente. Entretanto, estes marcadores apresentam profundas limitações, como por exemplo as baixas taxas evolutivas, que prejudicam e até mesmo impossibilitam estudos evolutivos com organismos de menor nível de divergência (SOLÉ-CAVA, 2001). Por outro lado, MURPHY *et al.* (1996) destaca que, algumas vezes, a divergência entre alozimas é tamanha que poucos eletromorfos são encontrados, sendo estes ainda, em sua maior parte, homoplásicos. É necessário lembrar, ainda, que, através do uso de isozimas, os níveis de

variabilidade encontram-se possivelmente subestimados, uma vez que mutações no material genético nem sempre resultam em alterações nas proteínas codificadas e nem toda alteração na seqüência de aminoácidos de uma proteína resulta em diferenças de migração (SOLFERINI & SCHEEPMAKER, 2001). Esta última propriedade, por si só, impõe sérias limitações a esta técnica, pelo fato de que eletromorfos homoplásicos não são, na maioria das vezes, detectados. Além disso, esta técnica apresenta problemas referentes ao reduzido número de locos amostráveis, bem como deficiências na detecção de determinados alelos segregantes, como os alelos nulos e os produtos co-migrantes de isolocus (MURPHY *et al.*, 1996). Por fim, o mesmo autor destaca, ainda, que muitas isozimas conformacionais podem ser produzidas a partir de modificações epigenéticas pós-traducionais de um único polipeptídeo, o que pode levar a novos equívocos na inferência do número de locos codificantes no sistema.

3.2.2 Citogenética Molecular

A citogenética molecular fundamenta-se na análise de aspectos microscopicamente visíveis da estrutura molecular dos cromossomos. Sua aplicação em estudos evolutivos expande-se desde a análise de parâmetros populacionais até a obtenção de filogenias. Uma revisão metodológica pode ser encontrada em SESSIONS (1996).

POWELL & DESALLE (1995) destacam que este enfoque é bastante útil quando executado em espécies próximas, cuja estrutura cromossômica tenha permanecido suficientemente similar a ponto de que inferências acerca de homologias possam ser feitas com maior segurança. Uma análise recente, executada por O'GRADY *et al.* (2001) demonstrou que, para grupos de espécies de Drosophilidae, a comparação citogenética é enormemente informativa, fornecendo valores de suporte corrigidos bem maiores que a análise molecular, devido, principalmente, a seus altos índices de consistência. Por outro lado, segundo os mesmos autores, o enfoque citogenético produz uma menor resolução filogenética, levando a um maior número de politomias, devido ao seu reduzido número de caracteres informativos.

3.2.3 Hibridização DNA-DNA

Os primeiros estudos utilizando hibridização de DNA e sua cinética de renaturação foram realizados por WETMUR & DAVIDSON (1968). Entretanto, esta técnica só passou a ser utilizada em estudos filogenéticos em 1981, por Sibley e Ahlquist (revisão em WERMAN *et al.*, 1996).

Como pode ser visto em WERMAN *et al.* (1996), esta técnica apresenta uma série de dificuldades metodológicas, além de algumas limitações. Entre estas, destaca-se a dificuldade prática de separar efetivamente o DNA cópia única das seqüências repetitivas (o que deve afetar as distâncias obtidas principalmente na análise de seqüências muito divergentes) e o controle criterioso das condições de estringência necessárias para uma hibridização envolvendo exclusivamente seqüências ortólogas quando do estabelecimento dos heterodúplex. WERMAN *et al.* (1996) destaca, ainda, que é importante considerar a variação nas taxas evolutivas entre linhagens e a ausência de aplicação da hipótese de aditividade entre as distâncias obtidas (uma vez que homoplasias, pareamento entre seqüências parálogas ou xenólogas, ampla divergência entre as seqüências impossibilitando o estabelecimento de heterodúplex, etc., podem afetar este último parâmetro).

Ainda que ambos os problemas apresentados não afetem amplamente a precisão dos resultados obtidos, quando adequadamente controlados, muitas outras críticas são dirigidas à aplicação desta técnica. A principal delas diz respeito à ausência de informações concretas, seja na forma de caracteres individuais (nucleotídeos) ou conjuntos de seqüências, uma vez que os dados obtidos são sempre apresentados na forma de índices de divergência (POWELL & DESALLE, 1995), o que limita bastante a amplitude de análises possíveis. Também questiona-se os resultados obtidos em situações de grandes diferenças de tamanho entre os genomas das espécies analisadas, ou quando a análise envolve espécies muito próximas (em que os polimorfismos intraespecíficos não deviam ser ignorados) ou muito divergentes. Além disso, a técnica apresenta um custo relativamente alto e exige grandes quantidades de DNA.

Por outro, os defensores da técnica garantem que qualquer limitação é compensada pela amplitude da amostra utilizada, capaz de suprimir qualquer erro de amostragem, de forma que as medidas obtidas sejam inteiramente representativas da realidade. Portanto, mesmo que a divergência genética de um locus qualquer não reflita acuradamente a filogenia das espécies, devido à razões estocásticas ou a eventos de polimorfismo ancestral, quando a medida da divergência sobre todo o genoma é considerada, a filogenia correta das espécies deve aflorar (POWELL, 1991).

3.2.4 RFLP (Restriction Fragment Length Polymorphism)

Estes marcadores foram descobertos no fim da década de 60, por dois grupos independentes, que identificaram enzimas capazes de reconhecer seqüências específicas de 4-8 pb, presentes na dupla-hélice do DNA, e hidrolisar uma ligação fosfodiéster de ambas as fitas em pontos específicos dentro desta região, ou seja, as enzimas de restrição (LINN & ARBER, 1968; MESELSON & YUAN, 1968). Para uma revisão sobre RFLP, ver ARIAS *et al.* (2001) e DOWLING *et al.* (1996).

MORITZ & HILLIS (1996) e SWOFFORD *et al.* (1996) destacam que o uso inadequado desta técnica pode levar a muitos erros filogenéticos, causados pela má interpretação dos resultados obtidos. Entre estes equívocos, pode-se destacar o uso dos fragmentos de restrição como caracteres potencialmente homólogos, ao invés dos sítios de restrição, bem como a ausência de considerações referentes à assimetria de probabilidades entre ganho e perda de sítios de restrição.

Entretanto, segundo DOWLING *et al.* (1996) e ARIAS *et al.* (2001), esta técnica também apresenta alguns outros problemas, entre os quais pode-se citar a impossibilidade de identificação da origem da variação detectada; a freqüente ausência de independência entre os caracteres gerados por diferentes enzimas, quando eventos mutacionais únicos conduzem a uma alteração correlacionada (o que pode causar erros na estimativa da divergência de caracteres); a eventual ausência de fidelidade de transmissão devido a variações no estado de metilação da molécula de DNA, o que pode resultar em situações de homoplasia aparente; a possível não-reprodutibilidade quando as condições de reação não são rigidamente controladas. Por outro lado, segundo ARIAS *et al.* (2001), RFLP são marcadores sensíveis, independentes da ação ambiental, que se encontram ainda, na maioria das vezes, livres de efeitos pleiotrópicos, uma vez que a maior parte da variação detectada é fenotipicamente neutra.

3.2.5 PCR (*Polymerase Chain Reaction*) e Técnicas Relacionadas

PCR, esta poderosa técnica de amplificação de DNA, foi concebida por MULLIS (1987). O desenvolvimento desta técnica aumentou em muito a eficiência de detecção de polimorfismos, tanto a nível de DNA quanto de RNA, devido não apenas à redução de tempo e complexidade na execução dos experimentos, mas principalmente devido à sua enorme sensibilidade. PALUMBI (1996), FERREIRA & GRATTAPAGLIA (1998) e MATIOLI *et al.* (2001) apresentam uma revisão metodológica desta técnica.

Trata-se de uma técnica complexa, sujeita a uma série de artefatos que devem ser corretamente identificados e superados. A produção de produtos recombinantes, por exemplo, é um dos maiores problemas desta técnica, produzido quando uma sequência parcialmente estendida em um determinado ciclo age como *primer* de outro alelo, ou mesmo outro locus, em um ciclo subsequente (SWOFFORD *et al.*, 1996). Isto pode levar a uma série de problemas posteriores, uma vez que para a Biologia Evolutiva é fundamental que os segmentos amplificados e analisados não sejam apenas homólogos, como também, ortólogos. Entretanto, suas desvantagens são enormemente superadas pela sua utilidade em estudos evolutivos, entre as quais se pode destacar a possibilidade de amplificação de DNA-fóssil.

O princípio básico da técnica de PCR foi incorporado à vários métodos alternativos bastante úteis na detecção de polimorfismos, entre os quais destaca-se:

3.2.5.1 Combinação de PCR com RFLP

Esta técnica é embasada na utilização do PCR para amplificar o fragmento de interesse, e do RFLP para detectar sua variabilidade (MATIOLI *et al.*, 2001).

3.2.5.2 STR (*Short Tandem Repeats*) ou SSR (*Simple Sequence Repeats*)

STR ou SSR é uma técnica muito difundida para a análise de polimorfismos em regiões repetitivas. A natureza da variação, neste caso, é o número de repetições de unidades de 2 a 5 pb, também conhecidas como microssatélites (DOWLING *et al.*, 1996).

A utilização destes marcadores em estudos filogenéticos é, entretanto, desaconselhada por PALUMBI (1996) uma vez que, segundo ele, como apenas um reduzido número de estados é possível, estes caracteres são excessivamente homoplásicos. Além disso, segundo DOWLING *et al.* (1996), muitas vezes estes marcadores não são fielmente transmitidos, devido à sua alta taxa evolutiva.

3.2.5.3 RAPD (*Random Amplified Polymorphic DNA*)

Esta técnica foi desenvolvida simultânea e independentemente por dois grupos dos EUA: WILLIAMS *et al.* (1990), que patentearam a tecnologia com o nome mais comumente utilizado, RAPD; WELSH & MCCLELLAND (1990), que propuseram uma denominação mais apropriada para a técnica, chamando-a de AP-PCR (*Arbitrarily Primed-PCR*). RAPD é, portanto, apenas uma variação da técnica de PCR, em que *primers* únicos, com seqüências arbitrárias de cerca de 10-12 nucleotídeos, são utilizados. Uma breve revisão sobre esta

técnica pode ser encontrada em FERREIRA & GRATTAPAGLIA (1998), em MATIOLI *et al.* (2001), e em PALUMBI (1996).

Entre as maiores limitações desta metodologia, destacam-se a falta de homologia entre muitos fragmentos de igual tamanho (SOLÉ-CAVA, 2001) e a ausência de reprodutibilidade entre seus resultados, que variam não apenas de acordo com o grau de competitividade dos sítios de iniciação que geram determinados fragmentos mas também de acordo com as condições de amplificação, que devem ser, portanto, muito bem controladas (DOWLING *et al.*, 1996).

3.2.5.4 SPAR (*Single Primer Amplification Reaction*) ou ISSR (*Inter Simple Sequence Repeats*)

SPAR ou ISSR é uma técnica recente, desenvolvida por GUPTA *et al.* (1994), que consiste em uma variação do RAPD. Neste caso, porém, emprega-se um *primer* único de tamanho maior, formado por repetições. Como as repetições são abundantes no genoma, o aumento de tamanho dos *primers* reduz o problema de reprodutibilidade, mas não diminui a probabilidade de produção de bandas (MATIOLI *et al.*, 2001).

O maior problema enfrentado por esta técnica é o reduzido repertório de *primers* possíveis, uma vez que a maioria das seqüências repetitivas estariam sujeitas ao artefato de dimerização do *primer*.

3.2.5.5 AFLP (*Amplified Fragment Length Polymorphism*)

A técnica de AFLP foi inicialmente descrita por VOS *et al.* (1995) e combina a confiabilidade do RFLP com a conveniência e versatilidade do RAPD. A análise por AFLP resulta da amplificação diferencial de determinados fragmentos de restrição ligados a adaptadores, pelo uso de *primers* que apresentam complementaridade à estas seqüências adaptadoras, contendo ainda alguns nucleotídeos seletivos adicionais à sua extremidade 3'.

Trata-se de uma técnica robusta e informativa na detecção de um grande número de marcadores moleculares dispersos pelo genoma. SOLÉ-CAVA (2001) destaca, entretanto, que assim como nos marcadores RAPD, muitas vezes os fragmentos co-migrantes no gel não apresentam relação de homologia entre si.

3.2.6 SSCP (*Single-Strand Conformation Polymorphisms*), DGGE (*Denaturing Gradient Gel Electrophoresis*) e Análise de Heterodúplex

Estas são técnicas poderosas e eficientes, capazes de detectar pequenas alterações na conformação ou estabilidade dos fragmentos gerados por PCR, RAPD, ou RFLP, através de sua mobilidade eletroforética. Uma revisão sobre as mesmas pode ser encontrada em FERREIRA & GRATTAPAGLIA (1998) e DOWLING *et al.* (1996).

Entretanto, a consistência e resolução destas técnicas depende largamente da padronização de suas condições. Caso contrário, padrões não repetíveis podem ser gerados. Além disso, DOWLING *et al.* (1996) destaca que estas técnicas são mais apropriadas e eficientes para fragmentos menores que 600 pb.

3.2.7 Seqüenciamento

Embora o seqüenciamento de ácidos nucleicos seja uma técnica relativamente recente (MAXAM & GILBERT, 1977; SANGER *et al.*, 1977), seu grande poder informativo tem garantido uma ampla aplicação em estudos filogenéticos. Segundo HILLIS *et al.* (1996) dentre seus principais atrativos destacam-se o imenso potencial informativo (considerando que os nucleotídeos são as unidades básicas de informação dos organismos), a possibilidade de incorporar dados acerca dos processos de evolução molecular subjacentes e a aplicabilidade à diferentes estudos evolutivos, incluindo o âmbito micro e macroevolutivo.

Entretanto, sua efetividade na obtenção de resultados confiáveis depende, como em qualquer outra metodologia aplicada à esses fins, da existência de homologia, não apenas entre as seqüências analisadas como também entre cada um dos sítios individuais considerados, que devem estar, portanto, corretamente alinhados (SWOFFORD *et al.*, 1996). É importante, ainda, que os caracteres multiestado gerados por esta técnica sejam analisados mediante critérios não-ordenados.

Qualquer estratégia que vise a obtenção de informações filogenéticas a partir de seqüências nucleotídicas pode ser subdividida em quatro etapas principais: (1) identificação da seqüência apropriada ao problema filogenético em questão; (2) isolamento e purificação de muitas cópias da seqüência-alvo nas OTU's de interesse; (3) seqüenciamento do DNA purificado; (4) alinhamento das seqüências homólogas. A seguir examinaremos cada uma destas etapas com maiores detalhes.

3.2.7.1 Que seqüência usar?

A escolha do gene adequado ao problema filogenético em questão é crucial, uma vez que já foi comprovado que árvores construídas para um mesmo grupo de organismos com base em dados diferentes podem ser altamente incongruentes (RUSSO *et al.*, 1996). Segundo a mesma autora, a escolha do gene é muito mais importante que a escolha do método de reconstrução filogenética, uma vez que, se o gene adotado for adequado, qualquer método de reconstrução filogenética pode recuperar a filogenia verdadeira.

Um bom indicador de quão adequado é o gene utilizado para problemas filogenéticos específicos é o resultado final do alinhamento das seqüências. Neste caso, RUSSO (2001) recomenda que a proporção de diferenças entre duas seqüências alinhadas não ultrapasse 20-30%, pois a partir destes valores as diferenças estariam sendo demasiadamente subestimadas, devido à ampla ocorrência de substituições múltiplas, reversas e paralelas. Por outro lado, alinhamentos perfeitos, em que a maioria dos sítios estão conservados entre as seqüências, costumam ser pouco informativos para a reconstrução filogenética, uma vez que não há informação para resolver muitos pontos na filogenia (RUSSO, 2001).

Neste sentido, HILLIS *et al.* (1996) afirma que o uso do percentual de divergência entre seqüências homólogas na estimativa da utilidade do gene para a resolução do problema em questão pode ser desastrosa. Assim, se uma pequena proporção dos sítios encontra-se livre de pressões seletivas, enquanto que os demais estão altamente restritos, a comparação pode ser uma combinação não informativa de sítios invariáveis e sítios saturados de mutações, embora o percentual de divergência esteja em uma faixa considerada razoável.

Quando as seqüências nucleotídicas apresentam alguma saturação, costuma ser mais adequada a utilização das seqüências de aminoácidos na análise, uma vez que estas apresentam uma taxa evolutiva mais lenta (RUSSO, 2001). Entretanto, esta alternativa não apresenta muita utilidade em casos de presença de muitos *indels* entre as seqüências. Da mesma forma, ela não é recomendada para os casos em que as seqüências nucleotídicas por si só apresentam-se bastante conservadas no grupo de interesse, uma vez que sua utilização acarretaria perda de informação importante.

Para evitar problemas posteriores é importante que seja feita uma análise prévia do gene de interesse antes do início do seqüenciamento. Neste caso, é importante avaliar sua

variabilidade, se possível dentro do grupo de interesse ou em grupos próximos, mas se necessário, fazendo extrapolações a partir de grupos com tempo de divergência similar. De modo geral, como já foi dito, genes com altas taxas evolutivas são úteis para o estudo de indivíduos, famílias e populações, enquanto que aqueles que evoluem mais lentamente são melhor utilizados no estudo de espécies ou táxons supra-específicos (SOLÉ-CAVA, 2001). Portanto, a utilidade do gene é determinada diretamente pelo problema filogenético em questão.

a) Qual o tamanho adequado da seqüência a ser analisada?

Esta é uma questão bastante complicada, e que não pode ser ignorada no momento da escolha do(s) gene(s) ou do(s) segmento(s) que será(ão) analisado(s), uma vez que ambas as abordagens limítrofes podem causar conseqüências desastrosas (RUSSO, 2001). Desta forma, ao trabalhar com seqüências pequenas, existe o risco de que erros de amostragem afetem a filogenia obtida. Por outro lado, quanto maior for a seqüência, maior a probabilidade de se incluir na análise regiões com taxas evolutivas heterogêneas, o que também pode causar sérios problemas.

3.2.7.2 Isolamento das seqüências-alvo

O isolamento da seqüência-alvo com fins sistemáticos pode ser procedido de diferentes maneiras, de acordo com o objetivo, tempo e materiais envolvidos. Pode-se, por exemplo, utilizar a tecnologia do DNA recombinante para clonar a seqüência de interesse, obtida através da construção de um banco genômico ou através da amplificação *in vitro* por PCR (quando as regiões flanqueadoras da seqüência de interesse são conhecidas). Alternativamente, a seqüência amplificada por PCR pode ser seqüenciada diretamente, o que garante uma enorme economia de tempo e esforço. Por fim, se a seqüência-alvo for um gene ativamente transcrito em determinados tecidos, pode-se purificar seu RNA, que será, então, analisado.

Segundo HILLIS *et al.* (1996), a pureza e homogeneidade do DNA a ser seqüenciado constitui-se em um item de importância primária para a obtenção de seqüências corretas e interpretáveis. Neste sentido, muitas vezes o DNA amplificado por PCR é suficientemente homogêneo a fim de garantir um bom seqüenciamento. Entretanto, outras vezes é aconselhável a clonagem prévia dos produtos obtidos nesta etapa, de forma a purificar as seqüências heterogêneas e, desta forma, facilitar o seqüenciamento e sua posterior análise.

Outra etapa limitante na confiabilidade dos resultados obtidos diz respeito à fidelidade da replicação do DNA (*in vivo* ou *in vitro*). Assim, principalmente quando se utiliza PCR, é fundamental a utilização de polimerases com baixas taxas de erro ou o seqüenciamento de mais de um isolado, quando polimerases com altas taxas de erro, como a *Taq*, são utilizadas. Em casos de erros na replicação, seria preferível seqüenciar diretamente o DNA amplificado, quando os mesmos poderiam ser identificados como ambigüidades, o que não é possível quando o seqüenciamento sucede a clonagem.

3.2.7.3 Técnicas de Seqüenciamento

Dois técnicas principais constituem o embasamento teórico-prático de toda e qualquer estratégia de seqüenciamento: o método de clivagem química, desenvolvido por MAXAM & GILBERT (1977); e o método enzimático de terminação de cadeia, desenvolvido por SANGER *et al.* (1977).

O método de clivagem química inicia-se pela desnaturação do DNA a ser seqüenciado, seguida pela marcação, por uma polinucleotídeo quinase, dos segmentos unifilamentares obtidos. Estes são, a seguir, divididos em quatro subamostras, cada qual é tratada com reagentes químicos específicos que clivam o DNA em um dos quatro nucleotídeos, em condições nas quais, em média, ocorra uma única quebra por cadeia. Os fragmentos gerados a partir de cada um destes tratamentos são, então, separados por eletroforese em gel desnaturante e, em seguida, visualizados em um padrão de bandas que revela a seqüência diretamente. (HILLIS *et al.*, 1996; STRYER, 1996)

O método enzimático de Sanger está fundamentado no uso de uma DNA-polimerase para copiar a seqüência-alvo, a partir de um *primer*, na presença dos quatro desoxirribonucleosídeos trifosfatados bem como de um ou dos quatro análogos 2',3'-didesoxi. Como a incorporação deste didesoxiribonucleotídeo leva à interrupção da extensão enzimática, fragmentos de diferentes tamanhos são gerados, e estes, quando submetidos à eletroforese, permitem a obtenção da seqüência de bases subjacente (HILLIS *et al.*, 1996; STRYER, 1996). Esta técnica constitui a base dos seqüenciadores automáticos. A estratégia de Sanger pode ainda ser desenvolvida em sucessivos ciclos de desnaturação, reanelamento e síntese, o que resulta na amplificação dos diferentes fragmentos e potencializa os resultados obtidos, mesmo na presença de pequenas quantidades de DNA-alvo.

Ambas as metodologias apresentam algumas limitações significativas. O método químico, por exemplo, pode ser utilizado apenas com DNA e requer o conhecimento prévio do mapa de restrição da seqüência-alvo, que deve ser clivada em fragmentos manipuláveis. Por outro lado, o método enzimático não requer nenhum conhecimento prévio da seqüência a ser analisada (pois os seus *primers* podem ser sintetizados com base nas seqüências do vetor ou mesmo do *primer* utilizado na reação de PCR), mas é amplamente afetado pela presença de estruturas secundárias na seqüência de interesse.

3.2.7.4 Alinhamento

O objetivo do alinhamento é fazer com que a posição de cada sítio individual comparado entre o conjunto de seqüências apresente uma relação de homologia com os demais (RUSSO, 2001). Para isto, algumas vezes é necessário a inserção de intervalos nas seqüências de algumas espécies que perderam trechos ou que simplesmente não os ganharam (PHILLIPS *et al.*, 2000). De modo geral, os algoritmos de alinhamento são designados para maximizar o percentual de similaridade entre as seqüências enquanto minimiza o número de *indels* ocorridos (HILLIS *et al.*, 1996).

Embora alguns autores não recomendem o alinhamento de seqüências por computador, existem atualmente uma série de programas eficazes neste procedimento, dentre os quais destaca-se o programa Clustal. Este, e segundo RUSSO (2001), a maior parte dos outros programas, executa o alinhamento em duas etapas principais. Na primeira etapa uma matriz de pontos, cujos eixos representam as duas seqüências comparadas, é construída para cada par de dados. Neste caso, cada posição da primeira seqüência é comparada com todas as posições da segunda seqüência, e em casos de similaridade, um ponto é colocado no gráfico. Esta matriz fornece então a base para o cálculo de similaridade entre as duas seqüências para cada alinhamento possível, onde o número de substituições é reduzido do número de bases idênticas, e do valor obtido reduz-se ainda o número de intervalos multiplicado pela penalidade de intervalos. Por fim, a similaridade máxima para cada par de seqüências é obtida, quando parte-se, então, para a segunda etapa. Nesta fase, os dados obtidos na primeira etapa são utilizados para a construção de um dendrograma a partir do qual o alinhamento final é realizado (partindo dos nós mais internos até os terminais). Desta forma, os programas de alinhamento testam um grande número de possibilidades, escolhendo por fim aquela mais parcimoniosa.

Os programas de alinhamento tornam-se especialmente necessários quando se trabalha com seqüências de tamanhos diferentes. No entanto, RUSSO (2001) recomenda que os alinhamentos obtidos sejam sempre revistos e, se possível, que as regiões de alinhamento duvidoso sejam retiradas. SWOFFORD *et al.* (1996) enfatiza que regiões do alinhamento que apresentam muitos *gaps* (ou intervalos) devem ser omitidas na análise, uma vez que a homologia posicional nestes pontos é extremamente incerta.

3.2.7.5 Métodos de Reconstrução Filogenética

A inferência de relações filogenéticas a partir de dados moleculares atuais requer, por fim, a escolha de métodos analíticos apropriados. De modo geral, estes métodos podem ser divididos em dois grupos principais (SWOFFORD *et al.*, 1996): no primeiro grupo situam-se os métodos estritamente algorítmicos, que definem uma seqüência específica de passos (um algoritmo), cuja aplicação resulta diretamente em uma árvore filogenética final; no segundo grupo encontram-se aqueles métodos embasados em critérios de otimização, cujos algoritmos levam à avaliação de um conjunto, total ou parcial, de árvores possíveis e determinam a escolha da(s) árvore(s) final(is), de acordo com o critério adotado. Portanto, neste último caso, os problemas de construção de árvores ótimas e de avaliação das mesmas encontram-se separados: especificamente, quando o conjunto de árvores geradas inclui todas aquelas possíveis, o algoritmo que as construiu é dito exato, e exemplos são fornecidos pelo algoritmo de busca exaustiva e pelo “*branch-and-bound*”; por outro lado, quando apenas um subconjunto delas é reconstruído, o algoritmo é dito heurístico, como é o caso do “*stepwise addition*”, da decomposição da politomia, do “*branch-swapping*”. MIYAKI *et al.* (2001) destacam, entretanto, que algoritmos heurísticos podem levar à escolha de uma árvore subótima segundo o critério escolhido. Os métodos estritamente algorítmicos incluem as formas de algoritmos geométricos fundamentados em análise de *cluster* (ou agrupamento), como UPGMA (*Unweighted Pair Group Method with Arithmetic means*) e NJ (*Neighbor Joining*), por exemplo, enquanto que os métodos de otimização abrangem todos os variantes da máxima parcimônia, da máxima verossimilhança e da aditividade de distâncias.

Existe ainda uma variedade de técnicas de inferência filogenética formuladas com base em hipóteses acerca do processo evolutivo. Neste caso, quanto mais o modelo evolutivo escolhido se aproximar do processo real, maior a consistência do método utilizado. Por outro lado, alguns métodos, entre os quais destaca-se o de máxima

parcimônia, não dependem de modelos precisos, mas apenas da ausência de determinados cenários evolutivos (SWOFFORD *et al.*, 1996).

Principais Métodos de Reconstrução Filogenética

a) Critério de Otimização I: Método de Máxima Parcimônia (MP)

A parcimônia, dentro de um contexto filogenético, é um método probabilístico que implica em um modelo evolutivo através de uma simplificação de natureza estatística (Miyaki *et al.*, 2001). De modo geral, os métodos de parcimônia visam à seleção da(s) árvore(s) com menor comprimento total, ou seja, aquelas que requerem o menor número de passos para explicar um determinado conjunto de caracteres (Swofford *et al.*, 1996). Neste sentido, obviamente, a árvore que minimiza o número total de passos também minimiza a ocorrência de homoplasias.

Neste tipo de análise, antes de mais nada, os sítios informativos devem ser selecionados. Isto porque nem toda variabilidade é útil para este método (afinal, sítios que apresentam variação em apenas uma seqüência, adicionarão um passo para qualquer uma das árvores consideradas). De modo geral, para ser informativo, o sítio deve dividir as OTU's em pelo menos dois grupos, cada qual com no mínimo dois representantes.

A próxima etapa consiste em inferir, para cada uma das árvores possíveis e de acordo com o algoritmo escolhido, o número total de substituições em cada sítio informativo e somá-las entre si. Neste ponto, aquela árvore que apresentar o menor comprimento é escolhida como sendo a mais parcimoniosa. Muitas vezes, entretanto, mais de uma árvore pode apresentar o comprimento mínimo e, neste caso, MIYAKI *et al.* (2001) recomenda a consideração conjunta na interpretação final dos resultados.

Uma crítica constante aos métodos de parcimônia refere-se à sua ausência de modelos evolutivos explícitos. Entretanto, a parcimônia adota determinadas hipóteses, que, uma vez violadas, afetam enormemente a consistência deste método (SWOFFORD *et al.*, 1996). Entre estas situações destacam-se aqueles casos em que o conjunto de dados apresenta uma quantidade relativamente grande de homoplasias, quando comparada ao número de sítios informativos. Outro problema freqüente envolve a ocorrência de taxas evolutivas heterogêneas entre os ramos periféricos, quando OTU's com taxas evolutivas semelhantes tendem a ser agrupadas. MIYAKI *et al.* (2001) sugerem, para estes casos, a adição de novas OTU's cujos ramos se liguem aos ramos maiores e os subdividam. Uma outra limitação deste método diz respeito ao fato de que, a menos que a taxa de

substituições apresentada pelas seqüências seja realmente pequena, o número total de mudanças evolutivas será subestimado, uma vez que eventos de substituição múltipla não são considerados (SWOFFORD *et al.*, 1996).

b) Critério de Otimização II: Métodos Probabilísticos ou Método de Máxima Verossimilhança (MV)

Métodos probabilísticos são aqueles métodos de inferência filogenética que buscam reconstituir a árvore com maior verossimilhança, a partir de um determinado conjunto de caracteres (PEREIRA *et al.*, 2001). Para isto, diversos parâmetros, como taxa de substituição entre nucleotídeos, frequência de bases, etc., são estimados diretamente a partir dos dados e utilizados na escolha do modelo evolutivo apropriado, bem como no cálculo da probabilidade de que uma determinada história evolutiva tenha dado origem ao conjunto de caracteres observados (NEI & KUMAR, 2000). Assim, a topologia e o comprimento dos ramos de uma árvore são individualmente maximizados, de forma que a história mais consistente é, finalmente, adotada.

Este método apresenta algumas vantagens importantes em relação aos demais, como o fato de apresentar menor variância e ser, desta forma, menos afetado por erros de amostragem. Além disso, trata-se de um método muitas vezes mais consistente que a parcimônia, uma vez que considera a variação no tamanho dos ramos durante a avaliação das diferentes árvores (SWOFFORD *et al.*, 1996). Os métodos probabilísticos possibilitam, ainda, o emprego de diversos modelos evolutivos diferentes no cálculo das probabilidades de substituição entre os nucleotídeos e na estimativa do comprimento dos ramos da árvore. Assim, existem modelos que consideram apenas a taxa de substituição geral, enquanto outros consideram as taxas de transição e transversão, a frequência de cada nucleotídeo, e até mesmo a heterogeneidade nas taxas evolutivas de diferentes sítios (PEREIRA *et al.*, 2001). A escolha e a aplicação destes modelos, entretanto, requerem a inferência de determinados parâmetros, o que deve ser feito a partir das próprias seqüências. Este procedimento requer um cuidado adicional, uma vez que a consistência do método de máxima verossimilhança é enormemente limitada pela veracidade do modelo adotado (SWOFFORD *et al.*, 1996), que pode ser escolhido mediante o uso de programas como o ModelTest (SCHNEIDER, 2002).

Para calcular a probabilidade da árvore como um todo, é necessário considerar as probabilidades de ocorrência de cada um dos possíveis estados ancestrais de caracteres em

cada nó interno em particular, como uma função da topologia e do tamanho dos ramos apresentados (SWOFFORD *et al.*, 1996). Segundo PEREIRA *et al.* (2001), este cálculo é feito, geralmente, segundo um critério não-ordenado, e de acordo com o modelo evolutivo escolhido.

Portanto, Máxima Parcimônia e Máxima Verossimilhança são métodos análogos entre si, uma vez que ambos buscam reconstruir a árvore que melhor representa a história evolutiva de um determinado conjunto de OTU's. Suas maiores diferenças, neste caso, dizem respeito ao fato de que, contrariamente à verossimilhança, a parcimônia desconsidera o comprimento dos ramos durante a estimativa do custo de uma determinada mudança, ao mesmo tempo que analisa apenas a solução mais simples, e não o conjunto de soluções (SWOFFORD *et al.*, 1996).

Uma variante desta análise de MV, recentemente introduzida no cenário filogenético, é a Análise Bayesiana. Neste caso, novamente o modelo evolutivo é escolhido e o programa busca as melhores árvores que sejam consistentes, tanto com o modelo, quanto com os dados brutos (HOLDER & LEWIS, 2003). Neste caso, entretanto, ao invés de buscar a árvore que maximiza a probabilidade de observar-se os dados, dada a árvore, busca-se a árvore que maximiza a probabilidade de observar-se a árvore, dados os dados e o modelo evolutivo (HALL, 2001). Outra diferença marcante em relação à MV, segundo este mesmo autor, é que a Análise Bayesiana procura o melhor conjunto de árvores, ao invés da melhor árvore, percorrendo, assim, o cenário das melhores árvores possíveis. Seu resultado é, portanto, um conjunto de árvores que o programa repetidamente encontrou em suas buscas independentes, o que constitui o método MCMC (Markov Chain Monte Carlo). Para isto, a partir de uma árvore inicial, novos estados (mudanças de posição ou comprimento de ramos) são propostos, e suas probabilidades, com relação aos estados anteriores, calculadas (o que constitui a probabilidade *a posteriori*), em cada geração do MCMC; um número aleatório entre 0 e 1 é tomado e, se o número for menor que a probabilidade calculada, o novo estado é aceito, enquanto que, se o número for maior, o novo estado é descartado. Isto faz com que árvores mais verossímeis tenham uma maior probabilidade de ser aceitas. À medida que o número de gerações aumenta, o processo converge em um pequeno conjunto de árvores com probabilidades muito semelhantes, cuja opção de escolha ou rejeição é meramente aleatória. Finalmente, quando o processo termina, pode-se optar pela construção da árvore consenso entre todas as árvores

analisadas, a partir de um determinado limiar de árvores com menor probabilidade, ou considerar a frequência com que uma determinada árvore foi encontrada como um sério indicativo de sua probabilidade, escolhendo, por fim, a árvore mais provável.

c) Critério de Otimização III: Método de Evolução Mínima (EM)

O método de Evolução Mínima utiliza-se das distâncias apresentadas pelas diferentes seqüências para a obtenção da melhor filogenia, que é avaliada segundo critérios de minimização da soma dos valores absolutos dos comprimentos dos ramos (S) (SWOFFORD *et al.*, 1996).

Na prática, entretanto, o cálculo do somatório do comprimento dos ramos para todas as árvores possíveis é muito trabalhoso. Para diminuir este problema, este critério de otimização é, geralmente, associado a algoritmos heurísticos, como NJ (*Neighbor Joining*), o que acaba produzindo ótimos resultados. Alternativamente, pode-se promover rearranjos a partir da árvore obtida por NJ, de forma a obter árvore ainda melhores.

d) Métodos geométricos ou Métodos de Distância

Os métodos geométricos abrangem todos aqueles algoritmos fundamentados na comparação quantitativa de seqüências alinhadas de nucleotídeos ou aminoácidos. Esta quantificação é, então, expressa na forma de distâncias, que podem ser utilizadas para a inferência de filogenias. Portanto, os métodos geométricos reduzem a informação armazenada nas seqüências a estimativas de distância entre elas e trabalham com essas distâncias para a reconstrução direta da árvore final.

Determinadas técnicas, como por exemplo a hibridização DNA-DNA, não oferecem alternativas ao uso deste método. Para os demais dados, entretanto, os métodos de distância constituem-se em alternativas eficazes para o uso de modelos evolutivos, sem as complicações práticas associadas aos métodos probabilísticos (SWOFFORD *et al.*, 1996). Sua maior limitação, neste caso, diz respeito à perda de informações implicada na transformação dos dados em medidas de distância.

Encontram-se disponíveis atualmente um grande número de modelos para o cálculo das distâncias existentes entre seqüências de DNA ou de aminoácidos. Entretanto, a escolha do modelo adequado deve ser cuidadosamente executada de forma que se possa esclarecer não apenas a história filogenética das OTU's em questão (recuperando sua topologia), como também, quando for o caso, o real tempo de divergência ou tamanho dos ramos existentes entre as seqüências. Assim, em alguns casos, é recomendável a utilização

de distâncias simples, como a distância p , que considera apenas a proporção de posições em que as duas seqüências diferem. Quando o valor de p ultrapassa o limite de 0,2 (NEI & KUMAR, 2000), entretanto, pode ser aconselhável a utilização de modelos mais complexos, que levem em conta, por exemplo, as substituições múltiplas, a razão de transições e transversões, a frequências de bases apresentadas pelas diferentes seqüências, a taxa de transições entre purinas e entre pirimidinas, e a heterogeneidade entre as taxas evolutivas apresentadas por diferentes sítios.

RUSO *et al.* (2001) recomenda que, para a estimativa de topologias, se utilize o modelo mais simples possível, uma vez que a variância associada à distância é proporcional ao número de parâmetros considerados. Por outro lado, os mesmos autores defendem o uso de uma distância não viciada, fundamentada no modelo evolutivo real, para a estimativa do tamanho dos ramos.

As maiores dificuldades na inferência de distâncias ocorrem na presença de *gaps* de alinhamento ou mesmo variações de comprimento terminal entre as seqüências analisadas. Muitas vezes os *gaps* são considerados como a quinta base ou o vigésimo-primeiro aminoácido. Este procedimento, entretanto, não é apropriado, pelo fato de que os eventos responsáveis pela substituição de bases são diferentes daqueles que levam a inserções e deleções. Alternativamente, pode-se optar pela omissão destes dados na análise. Nestes casos, a deleção pode ser executada apenas nas comparações afetadas pelos *gaps*, ou em todas as seqüências consideradas. SWOFFORD *et al.* (1996) recomenda a utilização do primeiro procedimento em casos de *gaps* curtos e distribuídos aleatoriamente, e do segundo quando a ocorrência de *gaps* é concentrada em determinadas regiões que, por serem menos restritas seletivamente, apresentam taxas evolutivas mais altas. Particularmente, neste último caso, a utilização diferencial da informação armazenada nestes pontos pode enviesar a análise. Para as ocorrências de variação de comprimento terminal entre as seqüências, o segundo tratamento descrito deve ser utilizado.

Após a obtenção das distâncias apropriadas, é necessário optar por diferentes algoritmos de reconstrução que as utilizem na obtenção de uma filogenia. O algoritmo mais utilizado, neste caso, é o NJ (*Neighbor Joining*).

NJ é um algoritmo heurístico, fundamentado na análise de dados aditivos, que sigam ou não o relógio molecular (SWOFFORD *et al.*, 1996). Segundo RUSO *et al.* (2001), este método assume o critério de evolução mínima, no qual a árvore com menor soma total

dos ramos é procurada. Neste caso, a partir de uma politomia inicial, é procurado o par de vizinhos que minimize a soma total dos ramos da árvore. Este par representa, então, a primeira bifurcação da árvore, passando a ser tratado como uma OTU única. A seguir, o próximo par que minimize a soma total dos ramos da árvore é procurado, até que, por fim, a filogenia se encontre completamente resolvida. Embora este algoritmo seja um pouco mais lento que o UPGMA, ele apresenta a grande vantagem de não depender da ocorrência de taxas constantes de evolução.

Segundo KUMAR (1996), este método quase sempre produz a mesma topologia que a Evolução Mínima, desde que as diferenças entre as seqüências sejam suficientemente grandes e o número de nucleotídeos em cada seqüência seja razoável (maior do que 500 pb). NEI *et al.* (1998) mostraram, ainda, que este método produz, geralmente, conclusões muito similares àquelas obtidas com algoritmos temporalmente mais dispendiosos, desde que a divergência entre as OTU's seja pequena.

3.2.7.6 Testes de Confiança em Topologias

Após o cladograma ter sido construído, é importante obter uma estimativa da confiabilidade inerente a cada um dos seus agrupamentos. Existem uma série de testes que buscam mensurar a informação contida nos cladogramas, de forma a fornecer resultados que permitam escolher entre filogenias alternativas. Entre estes destacam-se:

a) Índice de Consistência (*ci*)

Este teste busca obter uma estimativa do número de eventos homoplásicos para um determinado caráter ou cladograma (AMORIM, 1997), através da seguinte equação:

$$ci = \frac{M}{S}$$

Onde: - m = número mínimo de passos que um caráter ou conjunto de caracteres pode exibir em um cladograma

- s = número efetivo de passos ocorridos durante a evolução do caráter ou presentes no cladograma

Assim, se o índice de consistência do caráter ou cladograma considerado for igual a 1,0, não ocorreram homoplasias no decorrer de sua evolução. Por outro lado, segundo AMORIM (1997), se o índice de consistência for menor que 0,4, a possibilidade de ocorrência de erros durante a análise dos dados deve ser considerada.

A maior limitação deste teste consiste no fato de que seus valores não são alterados com a posição que as condições apomórficas ocupam no cladograma, ou seja, se são

sinapomorfias ou autapomorfias. Outro problema freqüente é que, em geral, os valores de *ci* decrescem com o aumento no número de OTU's consideradas para cada caracter (O'GRADY *et al.*, 2001), embora este efeito não tenha sido detectado por estes autores mediante a utilização de dados citogenéticos.

b) Índice de Retenção (*r*)

Este índice busca mensurar a fração de sinapomorfias aparentes em relação às sinapomorfias verdadeiras, levando em consideração o número máximo possível de surgimentos de um caráter em um determinado cladograma (*g*) (AMORIM, 1997). Sua fórmula é expressa pela equação:

$$R = \frac{g - s}{g - m}$$

Assim, $r = 0$ quando todos os eventos correspondem a autapomorfias, homoplásicas ou não ($g = s$), enquanto que $r = 1$ quando não ocorreram homoplasias e nem autapomorfias ($s = m$). Por outro lado, quando pelo menos um caráter for sujeito à homoplasias e pelo menos um de seus aparecimentos for sinapomórfico para algum nível, $0 < r < 1$, sendo que r é inversamente proporcional à amplitude de ocorrência de eventos homoplásicos.

DESALLE (1992a) destaca que os índices de retenção são mais importantes que os índices de consistência na verificação dos níveis de homoplasia, embora também haja uma correlação negativa de r com o número de OTU's envolvidos na análise.

c) *Bootstrap*

Trata-se de um teste de confiança que revela a consistência interna dos dados, cuja base consiste em reamostragens pseudoaleatórias com reposição dos dados originais (RUSSO *et al.*, 2001). A cada reamostragem, o número total de dados mantém-se constante, e, a partir destes, uma árvore-réplica é construída. Os valores de *bootstrap* são as porcentagens de casos nos quais cada grupo apareceu nas árvores resultantes. Com base nestes resultados, o programa PAUP mostra a topologia consenso de todas as árvores-réplicas, enquanto o programa MEGA apenas acrescenta à árvore original (construída com 100% dos dados) o percentual com que cada agrupamento apareceu nas réplicas. Alternativamente, neste último caso, pode-se optar por apresentar a filogenia consenso de *bootstrap* sob a regra da maioria dos 50%, na qual todos os ramos com valor de

bootstrap menor do que 50% são colapsados em uma politomia basal (O'GRADY *et al.*, 1998).

Este teste apresenta resultados positivos quando aplicado tanto à árvores de distância, quanto de parcimônia ou verossimilhança. Entretanto, segundo RUSSO *et al.* (2001), a confiabilidade a ser depositada neste teste depende da consistência do modelo utilizado na construção da topologia. Segundo esta mesma autora, se o método for consistente, os valores de *bootstrap* podem ser considerados uma subestimativa da probabilidade do ramo considerado ser verdadeiro, devido à comum ocorrência de sítios “contraditórios”.

Segundo RUSSO *et al.* (1995), entretanto, medidas de *bootstrap* levam, geralmente, a uma subestimativa da confiança estatística de um dado *cluster* quando um grande número de seqüências é considerado, principalmente em casos de análise de OTU's proximamente relacionadas. Além disso, segundo PÉLANDAKIS *et al.* (1991), quando as árvores são construídas com base em seqüências de nucleotídeos relativamente curtas, este teste de confiança pode revelar-se muito severo, uma vez que embasa-se em uma redundância informativa. Outras vezes, segundo o mesmo autor, os baixos valores de *bootstrap* podem ocorrer devido à presença de ramos internos muito curtos, não claramente significativos, que ocorrem devido à dicotomização sucessiva e rápida ou até mesmo devido à eventos de especiação múltipla.

d) Nível de Confiança

Este parâmetro, amplamente utilizado para árvores de distância e menos conservativo que o teste anterior, é calculado com base no erro padrão do ramo anterior ao nó considerado, de forma a medir a probabilidade de que o tamanho deste ramo seja maior do que zero. Ele é obtido da seguinte forma:

$$CN = (1 - a)$$

Segundo RUSSO *et al.* (2001), esta técnica é muito eficiente quando o modelo evolutivo adequado é utilizado. Por outro lado, como seus valores dependem largamente do erro padrão obtido, o aumento no número de parâmetros considerados leva a uma subestimativa da confiança para aquele nó.

e) Índice de Decaimento de Bremer - DI (Bremer, 1994)

Este índice indica o número de mutações necessárias em uma árvore de máxima parcimônia para desarranjar um determinado clado, ou seja, examina o número de passos

adicionais ao comprimento da árvore mais parcimoniosa em que o clado permanece resolvido (SCHNEIDER, 2003). Desta forma, valores de DI altos indicam, geralmente, um elevado suporte para o nó em questão.

f) Probabilidade Posterior dos Clados em Análise Bayesiana

Este índice mede a proporção das árvores analisadas pela Análise Bayesiana em que um determinado agrupamento apareceu (HALL, 2001). Entretanto, este mesmo autor aponta que estes valores podem ser considerados uma superestimativa da confiabilidade dos ramos, de forma que não deveriam ser considerados isoladamente.

4 A Sistemática Filogenética e os Processos Evolutivos

A evolução biológica ocorre principalmente através dos mecanismos de anagênese (mudança direcional dentro de qualquer ramo filético ocasionada pelos fatores evolutivos), de cladogênese (ramificação filogenética devida a eventos de especiação causados principalmente pelos processos de vicariância e dispersão) e de extinção (FUTUYMA, 1997; AMORIM, 1997). Neste panorama evolutivo, a importância da evolução reticulada, que surge através da migração entre populações, hibridização entre espécies próximas ou pela transferência horizontal de material genético entre linhagens de organismos, tem-se revelado cada vez maior (DOOLITTLE, 1999; LORETO *et al.*, 2001; HERÉDIA, 2002). A maioria dos cladistas, entretanto, costuma ignorar os processos de evolução anagenética e reticulada, o que tem causado, em alguns casos, limitações conceituais e até prejuízos metodológicos à Sistemática Filogenética (AMORIM, 1997).

Um exemplo claro deste afastamento cladista em relação aos eventos populacionais é o conceito de sinapomorfia, que não se aplica a caracteres que são polimórficos entre ou mesmo dentro de populações. Considerando os dois processos envolvidos na evolução destes caracteres (mutação e fixação por seleção ou deriva genética), AMORIM (1993) criou os termos “sintreptia” e “sinapousia”, para se referir respectivamente ao compartilhamento de um alelo apomórfico entre OTU’s descendentes, independentemente deste estar fixado ou não e ao compartilhamento da ausência de um alelo plesiomórfico, tendo em vista uma condição apomórfica conhecida surgida por mutação em uma espécie ancestral. Mais do que esta limitação conceitual, amostragens parciais e incompletas dentro das espécies analisadas levam à uma subestimativa do grau de polimorfismo apresentado, o que representa sempre uma perda de informações, pois tanto as sintreptias quanto as sinapousias correspondem a indícios legítimos de ancestralidade comum exclusiva entre as

espécies. Além do mais, a presença de polimorfismos está geralmente envolvida em padrões evolutivos de grupos recentes, uma vez que quanto mais antiga for a espécie ancestral na qual um alelo apomórfico surgiu, mais provável deve ser que suas espécies descendentes tenham fixado um ou outro de seus alelos. Com isso, ao ignorar a distinção entre sinapousias e sintreptias, a tendência é interpretar os casos de fixação independente de alelos iguais como homoplasias. (revisão em AMORIM, 1997)

Assim, DOWLING *et al.* (1996) recomenda que a amostra coletada para cada uma das OTU's em foco inclua, se possível, todos os variantes morfológico/genéticos e geográficos possíveis. Entretanto, quando não há disponibilidade de informação sobre a variação intrataxonômica, e a busca desta é inviável (devido a dificuldades operacionais, por exemplo), aconselha-se a inclusão de um volume significativo de dados, o que tende a reduzir estes desvios metodológicos.

Um outro aspecto filogenético que não pode ser explicado adequadamente senão à luz de um entendimento imparcial da evolução biológica é o padrão de assimetria característico da maior parte dos cladogramas, em que ocorre um número desigual de espécies descendentes em cada um dos ramos de um par de grupos-irmãos, sendo que o grupo menor é também menos variável morfolologicamente. A compreensão destas características é simples, quando embasada em um modelo de vicariância heterométrica, segundo o qual o grau de assimetria da filogenia é proporcional ao grau de heterometria das fragmentações geográficas, à velocidade de divisões secundárias em cada uma das áreas menores e à taxa de extinção diferencial entre os ramos. As populações pequenas e periféricas, que apresentam pouca variabilidade, possuem uma maior probabilidade de passar por um processo que resulte em extinção. Inversamente, a população maior e que apresenta maior variabilidade tem uma maior probabilidade de gerar, a partir de eventos de cladogênese posterior, ramos subordinados muito diferentes entre si. (revisão em AMORIM, 1997)

Estes são exemplos da importância da extensão e aplicação do raciocínio anagenético em questões filogenéticas. Entretanto, casos de evolução reticulada tendem a trazer uma complexidade ainda maior a esse contexto. Normalmente, este mecanismo evolutivo pode camuflar processos de divergência entre linhagens, ou até ocasionar uma subestimativa do tempo de divergência e da distância entre espécies. Casos de hibridação interespecífica, por exemplo, podem ocasionar a existência de sinapomorfias semelhantes e

simultâneas dentro de táxons diferentes, o que traz dificuldades à análise filogenética (AMORIM, 1997). A discrepância entre a filogenia do gene e a dos organismos, neste caso, deve ser ainda maior quando são analisados segmentos não-recombinantes, como aqueles pertencentes ao genoma de organelas (MORITZ & HILLIS, 1996). Como, em geral, os métodos filogenéticos assumem uma evolução ramificada, a aplicação dos mesmos para a análise de casos envolvendo uma evolução reticulada produz resultados inesperados e até desastrosos, se não forem considerados criticamente.

4.1 Árvore das OTU's X Árvore dos genes

Uma árvore filogenética construída a partir de seqüências de DNA de um determinado locus (a árvore dos genes) não concorda, necessariamente, com a árvore que representa a verdadeira história evolutiva das espécies envolvidas (a árvore das espécies). Entre as principais causas que levam a estas incongruências, pode-se destacar os erros amostrais (introduzidos quando o número de linhagens ou nucleotídeos considerados for pequeno), a sobrevivência estocástica de linhagens gênicas não-recombinantes, a duplicação e/ou perda estocástica de genes, a transferência horizontal e o polimorfismo ancestral (AVISE *et al.*, 1984; PAMILO & NEI, 1988; POWELL, 1991; PAGE & CHARLESTON, 1997a e b). Todos estes fatores contribuem de forma decisiva, tanto na alteração da verdadeira topologia das relações existentes entre as espécies, quanto na inferência do comprimento dos seus ramos. Desta forma, segundo estes autores, as substituições nucleotídicas deveriam ser consideradas como caracteres apresentados pelos genes, e não pelos organismos, afinal, dependendo do conjunto de dados utilizados para construir a árvore, diferentes relações podem aflorar (POWELL, 1991). Nenhuma árvore pode, neste caso, ser considerada como certa ou errada, uma vez que ela simplesmente reflete diferentes aspectos da história das mesmas OTU's quando diferentes dados são utilizados.

A sobrevivência estocástica de linhagens gênicas não-recombinantes pode ser fielmente ilustrada pela extinção diferencial de linhagens de DNA mitocondrial em populações animais, mediante uma variedade de cenários demográficos (AVISE *et al.*, 1984). Neste caso, segundo estes autores, a probabilidade de sobrevivência de duas ou mais linhagens independentes ao longo de G gerações depende largamente do tamanho da população fundadora, do tamanho da população em sua capacidade suporte, e da distribuição das freqüências de progênie sobrevivente (sua média e variância). Em geral, esta probabilidade é maior do que 0,9 apenas se o número de gerações (G) for menor do

que $0,5n$ (em que n refere-se ao tamanho da população fundadora), enquanto que a mesma será menor do que $0,1$ se G for maior do que $4n$. Assim, se mais do que $4n$ gerações transcorreram desde que uma dada espécie ramificou-se, é provável que toda a divergência apresentada por seu DNA mitocondrial seja posterior ao evento de especiação; por outro lado, em uma espécie derivada a menos do que $0,5n$ gerações, algumas linhagens mitocondriais certamente serão anteriores à sua separação do estoque ancestral, de modo que, em comparações com outras taxa, é possível que algumas destas OTU's apresentem-se poli ou parafiléticas (AVISE *et al.*, 1984). Isto acrescenta problemas ao uso do DNA mitocondrial no estudo das relações filogenéticas entre espécies proximamente relacionadas.

Quando existe polimorfismo alélico dentro das espécies, uma árvore construída a partir de seqüências de DNA de um dado locus pode ser bastante diferente da árvore das espécies que o portam, desde que os eventos de surgimento dos diferentes alelos sejam anteriores aos eventos de especiação (POWELL, 1991). Este quadro é agravado, ainda, nos casos em que o tempo de divergência entre as diferentes espécies é curto ou quando o tamanho efetivo de suas populações é grande (PAMILO & NEI, 1988).

Desta forma, P , a probabilidade de que a árvore gênica apresente a mesma topologia que a árvore das espécies, cresce proporcionalmente ao tempo evolutivo implicado desde o início da separação das espécies, de forma que, se o tempo de divergência entre os taxa envolvidos for grande em relação ao seu tamanho populacional, a probabilidade da árvore gênica representar a árvore das OTU's aproxima-se da unidade (DESALLE, 1992a). Entretanto, se o comprimento de um dos ramos internos for pequeno, P não pode ser substancialmente aumentado, mesmo que o(s) valor(es) apresentado(s) pelos demais ramos seja(m) maior(es) (PAMILO & NEI, 1988). Portanto, se a árvore das espécies apresentar um único ramo interno curto, a árvore construída a partir de um único gene pode ser bastante diferente da árvore real das espécies.

Esta probabilidade não pode, ainda, ser substancialmente aumentada pelo acréscimo de outros alelos deste mesmo locus, mas sim pelo uso de seqüências de muitos outros locus que tenham evoluído independentemente (PAMILO & NEI, 1988).

II A FAMÍLIA DROSOPHILIDAE

A família Drosophilidae é apenas uma das mais de 120 famílias de insetos que compõem a Ordem Diptera, mas encontra-se entre as mais diversas e amplamente distribuídas. WHEELER (1986) assim sumariza a posição desta família dentro dos Diptera:

- Ordem Diptera
 - Subordem Nematocera
 - Subordem Brachycera
 - Subordem Cyclorrhapha
 - Divisão Aschiza
 - Divisão Schizophora
 - Seção Calyptratae
 - Seção Acalyptratae
 - Superfamília Ephydroidea
 - Família Drosophilidae
 - Família Curtonotidae
 - Família Canaceidae
 - Família Diastatidae
 - Família Camilidae
 - Família Chloropidae
 - Família Tethinidae
 - Família Ephydridae
 - Família Cryptochetidae

A posição da família Drosophilidae dentro da superfamília Ephydroidea foi melhor analisada por GRIMALDI (1990), com base em caracteres morfológicos. Segundo este autor a família Drosophilidae é o clado-irmão da família Curtonotidae, e que este par compõe o grupo irmão dos demais Ephydroidea.

Internamente, a família Drosophilidae apresenta cerca de 3.000 espécies descritas, distribuídas em mais de 60 gêneros, e apenas 2 subfamílias, Steganinae e Drosophilinae (tabela 1) (WHEELER, 1986). A subfamília Drosophilinae é composta, atualmente, por 35 gêneros, dentre os quais destaca-se o gênero *Drosophila*.

Tabela 1 – Classificação de 52 dos 60 gêneros pertencentes à família Drosophilidae em suas respectivas subfamílias. O número de espécies apresentado por cada táxon está indicado entre parênteses. [Retirado de WHEELER (1986)].

Gêneros da subfamília Steganinae	Gêneros da subfamília Drosophilinae	
1 – <i>Acletoxenus</i> (4)	18 – <i>Ateledrosophila</i> (2)	36 – <i>Mycodrosophila</i> (73)
2 – <i>Amiota</i> (92)	19 – <i>Baeodrosophila</i> (4)	37 – <i>Neorhinoleucophenga</i> (4)
3 – <i>Cacoxenus</i> (11)	20 – <i>Calodrosophila</i> (1)	38 – <i>Neotanygastrella</i> (16)
4 – <i>Electrophortica</i> (1)	21 – <i>Celidosoma</i> (1)	39 – <i>Nesiodrosophila</i> (6)
5 – <i>Eostegana</i> (4)	22 – <i>Chymomyza</i> (33)	40 – <i>Nudidrosophila</i> (5)
6 – <i>Gitona</i> (16)	23 – <i>Cladochaeta</i> (20)	41 – <i>Paraliodrosophila</i> (4)
7 – <i>Leucophenga</i> (158)	24 – <i>Dettopsomyia</i> (10)	42 – <i>Paramycodrosophila</i> (10)
8 – <i>Luzonimyia</i> (1)	25 – <i>Diathoneura</i> (29)	43 – <i>Phorticella</i> (13)
9 – <i>Mayagueza</i> (1)	26 – <i>Dicladochaeta</i> (1)	44 – <i>Samoiaia</i> (7)
10 – <i>Paraleucophenga</i> (3)	27 – <i>Drosophila</i> (1467)	45 – <i>Scaptomyza</i> (226)
11 – <i>Pararhinoleucophenga</i> (1)	28 – <i>Drosophilella</i> (6)	46 – <i>Sphaerogastrella</i> (3)
12 – <i>Pseudiastata</i> (8)	29 – <i>Grimshawomyia</i> (2)	47 – <i>Sphyrnoceps</i> (1)
13 – <i>Pyrgometopa</i> (1)	30 – <i>Hypselothyrea</i> (17)	48 – <i>Styloptera</i> (3)
14 – <i>Rhinoleucophenga</i> (12)	31 – <i>Jeannelopsis</i> (1)	49 – <i>Tambourella</i> (3)
15 – <i>Soederbomia</i> (1)	32 – <i>Liodrosophila</i> (44)	50 – <i>Titanochaeta</i> (11)
16 – <i>Stegana</i> (65)	33 – <i>Lissocephala</i> (20)	51 – <i>Zaprionus</i> (22)
17 – <i>Trachyleucophenga</i> (1)	34 – <i>Marquesia</i> (2)	52 – <i>Zygothrica</i> (65)
	35 – <i>Microdrosophila</i> (33)	

Obs.: Oito dos sessenta gêneros da família não estão incluídos nesta tabela, uma vez que sua localização permanece incerta.

1 O Gênero *Drosophila*

O gênero *Drosophila* composto por 15 subgêneros e mais de 1.400 espécies, perfaz, sozinho, quase metade das espécies da família, apresentando uma ampla e diversa distribuição (tabela 2) (WHEELER, 1986). Esta diversidade, entretanto, ao mesmo tempo que fornece diversas oportunidades de investigações acerca de padrões evolutivos, impõe uma série de desafios taxonômicos e filogenéticos.

Tabela 2 – Apresentação dos quinze subgêneros que compõem o gênero *Drosophila*, com os respectivos números de espécies situados entre parênteses. [Retirado de WHEELER (1986)].

Subgêneros do gênero <i>Drosophila</i>
1 – <i>Antopocerus</i> (15)
2 – <i>Chusqueophila</i> (1)
3 – <i>Dichaetophora</i> (7)
4 – <i>Dorsilopha</i> (1)
5 – <i>Drosophila</i> (759)
6 – <i>Dudaica</i> (1)
7 – <i>Engiscaptomyza</i> (6)
8 – <i>Hirtodrosophila</i> (112)
9 – <i>Lordiphosa</i> (13)
10 – <i>Phloridosa</i> (7)
11 – <i>Psilodorha</i> (2)
12 – <i>Scaptodrosophila</i> (184)
13 – <i>Siphlodora</i> (2)
14 – <i>Sophophora</i> (233)
15 – <i>Spinodrosophila</i> (1)

Dentre estes 15 subgêneros, pode-se destacar, pela diversidade apresentada, os subgêneros *Sophophora* e *Drosophila*. O subgênero *Sophophora*, estabelecido por STURTEVANT (1939), compreende 233 espécies, subdivididas em 7 grupos, dentre os quais destacam-se os grupos *melanogaster*, *obscura*, *saltans* e *willistoni*.

O grupo *melanogaster* é formado por 174 espécies (ASHBURNER *et al.*, 1984), classificadas em doze subgrupos: grupo *ananassae*, *montium*, *elegans*, *ficuspshila*, *eugracilis*, *takahashii*, *suzuki*, *denticulata*, *flavohirta*, *nipponica*, *rhopaloo* e *melanogaster*. Todos estes táxons agrupam-se, segundo ASHBURNER *et al.* (1984) e KASTANIS *et al.* (2003) (mediante análise cromossômica e de DNA mitocondrial, respectivamente), em três clados principais: (1) subgrupo *ananassae*, apresentado por KASTANIS *et al.* (2003) como o mais primitivo dentro do grupo; (2) subgrupo *montium*; (3) subgrupo *melanogaster* e demais subgrupos Orientais analisados (*elegans*, *eugracilis*, *ficuspshila*, *suzukii* e

takahashii). Dentre todos estes representantes, *D. ananassae* (grupo *ananassae*), *D. kikkawai* (grupo *montium*), *D. simulans* e *D. melanogaster* (grupo *melanogaster*) apresentam uma distribuição quase que cosmopolita. A origem deste grupo ocorreu, provavelmente, na região sul da Ásia (THROCKMORTON, 1975).

O grupo *obscura* é constituído por um total de 35 espécies descritas (BARRIO & AYALA, 1997), tradicionalmente divididas entre os subgrupos *obscura*, com representantes Neárticos e Paleárticos, e *affinis*, com espécies primariamente Neárticas (STURTEVANT, 1942). Alguns trabalhos (LATORRE *et al.*, 1988; BARRIO *et al.*, 1992; BARRIO & AYALA, 1997), entretanto, defendem a posição de que os representantes Neárticos do subgrupo *obscura* deveriam ser elevados de categoria, constituindo o subgrupo por eles denominado *pseudoobscura*, filogeneticamente próximo do subgrupo *affinis* (BARRIO & AYALA, 1997). Adicionalmente, BARRIO *et al.* (1992, 1994), BARRIO & AYALA (1997) e O'GRADY (1999) afirmaram que o restante do subgrupo *obscura* também formaria um grupo polifilético, constituído de espécies largamente divergentes, que deveriam ser subdivididas em dois taxa [irmãos entre si, segundo O'GRADY (1999)]: um mantendo o nome do subgrupo e a maioria das espécies; o outro compondo um novo clado, o subgrupo *subobscura*. Além disso, segundo alguns autores (BARRIO *et al.*, 1992), um grande número de espécies africanas recentemente descritas (TSACAS *et al.*, 1985) constituiriam o quinto subgrupo do grupo *obscura*, o subgrupo *microlabis*, formado por espécies exclusivamente Afrotropicais. THROCKMORTON (1975) aponta evidências de que este grupo tenha se originado nos trópicos do Velho Mundo, a partir de uma linhagem *proto-melanogaster*, seguida de uma adaptação a ambientes temperados.

Os grupos *saltans* e *willistoni* representam a radiação do subgênero *Sophophora* ocorrida no Novo mundo. O grupo *saltans* é formado por um total de 21 espécies descritas, divididas em cinco subgrupos listados em ordem ascendente a partir do mais primitivo, segundo as relações estabelecidas por THROCKMORTON & MAGALHÃES (1962) - a partir de uma variedade de caracteres morfológicos - e O'GRADY *et al.* (1998) - com base na análise molecular combinada de quatro loci: *cordata*, *elliptica*, *sturtevanti*, *parasaltans* e *saltans*. Seu provável local de origem são os trópicos da América do Norte (THROCKMORTON, 1975), onde os subgrupos *cordata* e *elliptica* são encontrados. Entretanto, segundo RODRIGUEZ-TRELLES *et al.* (1999), cuja análise molecular do gene *Xdh* foi particularmente informativa, o subgrupo *parasaltans* ocuparia a posição mais basal dentro do grupo

saltans, seguido, em sucessão progressiva, pelos grupos *sturtevanti*, *elliptica*, *cordata* e *saltans*, o que transfere o local de origem do grupo para os trópicos da América do Sul. Já o grupo *willistoni* é formado por 25 espécies, divididas em dois subgrupos: *willistoni* e *bocaiensis*. Dentro do subgrupo *willistoni* encontram-se agrupadas as seis espécies crípticas de *D. willistoni*, que são: *D. willistoni*, *D. equinoxialis*, *D. paulistorum*, *D. pavlovskiana*, *D. tropicalis* e *D. insularis*. Além disso, este grupo é, ainda, caracterizado pela presença de seis semispécies de *D. paulistorum*, de forma que, ao final, pode-se dizer que o grupo *willistoni* é um complexo de diversos níveis taxonômicos. As relações filogenéticas dentro do grupo, especificamente dentro do subgrupo *willistoni*, foram raramente estudadas. Entre as exceções, pode-se destacar o trabalho de AYALA *et al.* (1974), com base em diferenciação aloenzimática, e a análise molecular de GLEASON *et al.* (1998), a partir dos genes *Per*, *Adh* e *COI*. As evidências apresentadas neste último trabalho apontam a seguinte ordem de divergência dentro do subgrupo *willistoni*, a partir da espécie mais primitiva: *D. insularis*, *D. tropicalis*, *D. willistoni*, *D. paulistorum* e *D. equinoxialis*. Neste caso, embora o conceito biológico de espécie situe *D. pavlovskiana* como uma espécie diferente, filogeneticamente, esta taxa seria mesclado a *D. paulistorum*. THROCKMORTON (1975) argumenta que a origem deste grupo deve ter ocorrido na América do Sul, talvez após uma migração a partir da América do Norte.

O subgênero *Drosophila*, por outro lado, é o maior representante da família Drosophilidae em regiões Neotropicais, com 25 grupos de espécies (VAL *et al.*, 1981). MALOGOLOWKIN (1953) e THROCKMORTON (1975) dividiram as espécies do subgênero *Drosophila* em duas linhagens principais: a seção *virilis-repleta* e a seção *quinaria-tripunctata*, além do grupo *funebria*, basal em relação a ambas as seções. A seção *virilis-repleta* é formada por 15 grupos: *polychaeta*, *virilis*, *melanica*, *carbonaria*, *flavopilosa*, *bromeliae*, *xanthopallescens*, *repleta*, *mesophragmatica*, *castanea*, *canalinae*, *dreyfusi*, *annulimana*, bem como os grupos desérticos *nannoptera* e *peruviana*. Já a seção *quinaria-tripunctata* é constituída por 9 grupos, a maioria dos quais altamente abundante em regiões Neotropicais: *quinaria*, *immigrans*, *macroptera*, *tripunctata*, *cardini*, *guarani*, *calloptera*, *pallidipennis* e *rubrifrons*.

O grupo *mesophragmatica*, pertencente à seção *virilis-repleta* do subgênero *Drosophila*, foi estabelecido por BRNCIC & KOREF-SANTIBAÑEZ (1957). Este grupo, proximamente relacionado aos grupos *castanea*, *canalinae* e *dreyfusi* (THROCKMORTON,

1975), ou aos grupos *repleta*, *dreyfusi* e *canalinaea* (TATARENKOV & AYALA, 2001) compreende atualmente 9 espécies predominantemente andinas (VAL *et al.*, 1981). Todas estas espécies são morfologicamente bastante similares e algumas delas, como *D. gaucha* e *D. pavani*, são espécies crípticas, indistinguíveis mediante a inspeção de caracteres morfológicos, inclusive análise da genitália externa (BRNCIC, 1958). Segundo este mesmo autor, estas duas espécies, quando inter cruzadas, produzem, geralmente, híbridos estéreis, enquanto que do cruzamento entre *D. altiplanica* e *D. orkui*, eventualmente, fêmeas híbridas férteis são produzidas. NACRUR (1958) dividiu este grupo em dois subgrupos, com base em sua morfologia externa e na estrutura da sua genitália masculina: subgrupo a, constituído por *D. pavani*, *D. gaucha*, *D. mesophragmatica*, *D. altiplanica* e *D. orkui*; e subgrupo b, formado por *D. viracochi*. Entretanto, BRNCIC *et al.* (1971), através da análise de placas metafásicas e cromossomos politênicos, sugere três caminhos evolutivos diversos dando origem aos atuais representantes do grupo: o primeiro originando *D. mesophragmatica*, *D. brncici* e *D. gasici*, o segundo envolvendo *D. viracochi* e o terceiro levando à *D. gaucha* e *D. pavani*.

O grupo *flavopilosa*, outro membro da seção *virilis-repleta*, foi originalmente estabelecido por WHEELER *et al.* (1962), como incluindo *D. flavopilosa* e mais 13 novas espécies exclusivamente Neotropicais. De acordo com THROCKMORTON (1975), este grupo é derivado da radiação *repleta* ou, então, da radiação que levou aos grupos *bromeliae* e *peruviana*, grupos estes que, segundo ele, formam uma ligação entre as linhagens *nannoptera* e *repleta*. As espécies do grupo *flavopilosa* caracterizam-se por apresentarem uma ecologia bastante restrita, muitas vezes associada, necessariamente, a flores do gênero *Cestrum* (Solanaceae) (BRNCIC, 1978). Este grupo é composto, atualmente, por 11 espécies descritas (VILELA, 1984), tradicionalmente subdivididas em dois subgrupos (WHEELER *et al.*, 1962): subgrupo *flavopilosa*, que apresenta *D. acroria*, *D. crossoptera*, *D. lauta*, *D. cordeiroi*, *D. cestri* e *D. flavopilosa*; subgrupo *nesiota*, que inclui *D. incompta*, *D. melina*, *D. nesiota*, *D. gentica* e *D. mariahelena*.

Outros grupos de sumária importância pertencentes à seção *virilis-repleta* são os grupos *virilis* e *repleta*. O grupo *virilis* é composto por 13 espécies, das quais apenas *D. virilis* apresenta uma distribuição que se estende até a região Neotropical (VAL *et al.*, 1981). Este grupo encontra-se, atualmente, subdividido em quatro clados, que segundo NURMINSKY *et al.* (1996) equivalem à categoria de subgrupos de espécies: o clado *virilis*,

que inclui *D. virilis*, *D. lummei*, *D. novamexicana*, *D. texana* e *D. americana*; o clado *montana* composto por *D. montana*, *D. lacicola*, *D. flavomontana* e *D. borealis*; o clado *littoralis*, formado por *D. littoralis*, *D. ezoana* e *D. canadiana*; e o clado monotípico *kanekoi*, composto apenas por *D. kanekoi*. Já o grupo *repleta*, um dos maiores dentre todos os grupos do gênero *Drosophila*, é composto, atualmente, por 91 espécies (DURANDO *et al.*, 2000), cuja distribuição é tipicamente Neotropical (VAL *et al.*, 1981). THROCKMORTON (1975) associou a diversificação do grupo com o desenvolvimento dos desertos Americanos. Este grupo é subdividido em 5 subgrupos, listados em ordem ascendente a partir do mais primitivo (TATARENKOV & AYALA, 2001): *fasciola*, *mercatorum*, *repleta*, *mulleri* e *hydei*. Cada um destes subgrupos foi, ainda, subdividido em complexos e *clusters* de espécies. O subgrupo *mulleri*, por exemplo, é composto por 5 complexos de espécies (*mulleri*, *buzzatii*, *eremophila*, *meridiana* e *anceps*). Entretanto, DURANDO *et al.* (2000) não foi capaz de suportar a monofilia do subgrupo *mulleri*, que apresentou-se como um agrupamento parafilético em sua análise.

Dentro da outra seção do subgênero *Drosophila*, a seção *quinaria-tripunctata*, um lugar de destaque é reservado ao grupo *tripunctata* (MALLOGOLOWKIN, 1953 e THROCKMORTON, 1975). Este grupo foi proposto por STURTEVANT (1942), como composto por *D. tripunctata* e, duvidosamente, por *D. histrio*. Desde então, muitas outras espécies foram incluídas no grupo, a maioria das quais descritas por FROTA-PESSOA (1954). Atualmente, o grupo *tripunctata* encontra-se constituído por 64 espécies (VILELA & BÄCHLI, 2000), todas devidamente classificadas em 1 dos 4 subgrupos propostos por FROTA-PESSOA (1954), com base em caracteres morfológicos: (1) subgrupo I (7 espécies): *D. angustibucca*, *D. medioobscurata*, *D. setula*, *D. rostrata*, *D. mediocris*, *D. platitarsus* e *D. neoguaramunu*; (2) subgrupo II (12 espécies): *D. paraguayensis* (= *D. medionotata*, segundo BÄCHLI *et al.*, 2000), *D. mediosignata* (revalidada por BÄCHLI *et al.*, 2000), *D. cuaso* (descrita por BÄCHLI *et al.*, 2000), *D. bodemanna*, *D. facialba*, *D. prosimilis*, *D. unipunctata*, *D. roehrae*, *D. mediodelta*, *D. mediopunctata*, *D. medioimpressa* e *D. morena*; (3) subgrupo III (23 espécies): *D. addisoni*, *D. bipunctata*, *D. trapeza*, *D. blumelae*, *D. fairchildi*, *D. johnstonae*, *D. converga*, *D. mediopicta*, *D. mediopictoides*, *D. mediodstriata* (= *D. crocina*, segundo VILELA & BÄCHLI, 1990), *D. paramediodstriata*, *D. nigricincta*, *D. bandeirantum*, *D. divisa*, *D. triangula*, *D. triangulina*, *D. bifilum*, *D. curvapex*, *D. mesostigma*, *D. fragilis*, *D. frotapessoai*, *D. trifilum* e *D. trifiloides*; (4)

subgrupo IV (17 espécies): *D. tripunctata*, *D. loewi*, *D. metzii*, *D. albirostris*, *D. albescens*, *D. albicans*, *D. semialba*, *D. greerae*, *D. whartoniae*, *D. mediodiffusa*, *D. medioparva*, *D. pellewae*, *D. tristriata*, *D. mediovittata*, *D. pruinifacies*, *D. leticiae* e *D. spinatermina*. Adicionalmente, segundo VILELA (1992) e, posteriormente, VILELA & BÄCHLI (2000), as seguintes 5 espécies também foram designadas ao grupo *tripunctata*, mas ainda não foram corretamente classificadas: *D. argenteifrons*, *D. hansonii*, *D. hansonioides*, *D. mcclintockae* e *D. cundinamarca*. Exceto por *D. tripunctata*, que ocorre em regiões Neárticas, todas as demais espécies do grupo *tripunctata* são endêmicas da região Neotropical, onde o grupo é superado, com relação à diversidade de espécies, apenas pelo grupo *repleta* (VILELA, 1992). São, entretanto, as espécies dominantes nas florestas Neotropicais chuvosas (VILELA & VAL, 1985). Segundo THROCKMORTON (1975), o grupo *tripunctata* encontra-se proximamente relacionado aos grupos *rubrifrons*, *pallidipennis*, *sticta*, *macroptera*, *cardini*, *guarani* e *calloptera*, todos representantes de sua proposta radiação *tripunctata*. De acordo com este autor, este é um grupo a partir do qual, em diferentes períodos e níveis, outros grupos emergiram.

O grupo *cardini*, pertencente a esta mesma seção, foi estabelecido por STURTEVANT (1942). Atualmente compreendendo 16 espécies de ampla distribuição geográfica nas Américas (VILELA *et al.*, 2002), este grupo teve suas relações analisadas por HEED E RUSSEL (1971), que, com base em inversões cromossômicas, o dividiram em dois subgrupos: subgrupo *cardini*, amplamente distribuído no continente Americano, apresentando um *cluster* constituído por *D. cardinoides*, *D. cardini*, *D. procardinoides*, *D. parthenogenetica* e *D. bedicheki* e outro *cluster* constituído por *D. polymorpha*, *D. neomorpha* e *D. neocardini*; e subgrupo *dunni*, que apresenta *D. similis* bem como as seguintes espécies exclusivamente insulares: *D. arawakana*, *D. belladunni*, *D. dunni*, *D. negrodunni*, *D. antillea* e *D. caribiana*. Este grupo é filogeneticamente aparentado aos grupos *calloptera*, *guarani* e *tripunctata* (THROCKMORTON, 1975), sendo que os limites entre eles são bastante confusos.

Outro grupo de ampla e restrita distribuição Neotropical, pertencente à seção *quinaria-tripunctata* do subgênero *Drosophila*, é o grupo *guarani*. Segundo VILELA & PEREIRA (1993), este grupo, originalmente reconhecido por DOBZHANSKY & PAVAN (1943), compreende, atualmente, 13 espécies distintas: *D. alexandrei*, *D. araucana*, *D. griseolineata*, *D. guaraja*, *D. guaru*, *D. huilliche*, *D. limbinervis*, *D. maculifrons*, *D.*

ornatifrons, *D. peruensis*, *D. subbadia*, *D. tucumana* e *D. urubamba*. KING (1947) atesta que o grupo *guarani*, então com apenas 6 espécies, encontra-se naturalmente dividido em dois subgrupos: o subgrupo *guaramunu* constituído por *D. maculifrons* (= *D. guaramunu*), *D. griseolineata* e *D. guaraja*; e o subgrupo *guarani* que contém *D. ornatifrons* (= *D. guarani*), *D. guaru* e *D. subbadia*. Dentro deste último subgrupo, especificamente, KING (1947) relata a obtenção de híbridos, cujos machos são estéreis, nos cruzamentos envolvendo *D. subbadia* e *D. guaru*, enquanto KASTRITSIS (1969), descreve a produção de híbridos estéreis em cruzamentos entre *D. subbadia* e *D. ornatifrons*. Neste último trabalho e em KASTRITSIS *et al.* (1970), entretanto, os autores defendem que estes subgrupos do grupo *guarani* devem ser elevados taxonomicamente à categoria de grupos, uma vez que *D. griseolineata* e *D. maculifrons* parecem ser muito mais próximas de alguns membros do grupo *tripunctata* do que aos representantes do subgrupo *guarani*. Recentemente, ROBE *et al.* (2002) atestaram, molecular e morfológicamente, a monofilia do subgrupo *guarani*, mas não conseguiram eficientemente comprová-la para os membros do subgrupo *guaramunu*.

Pode-se reservar, ainda, um papel de destaque dentro da seção *quinaria-tripunctata* do subgênero *Drosophila* ao grupo *immigrans*, que, embora apresente uma única espécie (*D. immigrans*) cuja distribuição atinge a região Neotropical (VAL *et al.*, 1981), possui uma frequência bastante ampla em diversos estados brasileiros. Este grupo é tradicionalmente subdividido em 6 subgrupos: subgrupo *quadrilineata*, subgrupo *nasuta*, subgrupo *lineosa*, subgrupo *immigrans*, subgrupo *hypocausta* e mais um sexto subgrupo com espécies de classificação incerta (WILSON *et al.*, 1969). Entretanto, segundo THROCKMORTON (1975), as cerca de 70 espécies “nominais” já descritas para o grupo, apresentam, ainda, uma taxonomia bastante confusa.

Ecologicamente, todas estas espécies apresentam uma importância fundamental nas cadeias saprofíticas dos mais diversos ecossistemas, uma vez que seus estágios imaturos são, na maior parte dos casos, dependentes de organismos fermentadores (THROCKMORTON, 1975). Alguns representantes da família Drosophilidae são parasitas (gênero *Cladochaeta* e alguns membros do gênero *Drosophila*) e até mesmo predadores de insetos em seu estágio larval (REMSEN & O'GRADY, 2002), embora a grande maioria se desenvolva a partir de fungos, flores, cactos e, principalmente, frutos (NOOR & AKASHI, 2001). Desta forma, não apenas a sua variação morfológica, mas também sua variabilidade

ecológica tem atraído a atenção de inúmeros pesquisadores ao longo dos anos. Tudo isto, aliado ainda às suas facilidades de manutenção e manipulação, ao seu ciclo de vida relativamente curto, e à riqueza de dados biológicos já acumulados tem transformado estas espécies em organismos modelo ideais para uma série de estudos biológicos, principalmente no âmbito genético e evolutivo. Pode-se dizer, inclusive, que dos muitos milhões de espécies de animais e plantas existentes atualmente, nenhum organismo tem recebido maior atenção como organismo experimental quanto *Drosophila*, que tem desempenhado um papel central na elucidação de inúmeros processos biológicos, entre os quais destacam-se aspectos genéticos, evolutivos, moleculares e ontogenéticos. Sem exagero, pode-se afirmar com convicção que o gênero *Drosophila* tem sido fundamental na origem e no subsequente desenvolvimento da genética evolutiva, inaugurando estudos de biologia evolutiva com interface em ecologia e genética e estreitando o abismo antes existente entre os estudos fenotípicos e genotípicos, como é claramente mostrado por NOOR & AKASHI (2001).

Concomitante ao contínuo interesse em *Drosophila* como um organismo modelo, tem havido um crescente interesse na utilização da biologia comparativa para obtenção de um melhor entendimento de uma série de padrões e processos biológicos. Neste contexto, pode-se dizer que a compreensão das relações filogenéticas dentro da família Drosophilidae é fundamental, uma vez que propicia: 1) a análise da história evolutiva de uma grande diversidade de características já conhecidas e estudadas dentro do grupo; 2) a extrapolação ou não das eventuais conclusões obtidas a outros organismos, a partir da análise do contexto filogenético em questão; 3) a definição prévia das estratégias amostrais necessárias em qualquer tipo de projeto (POWEL & DESALLE, 1995).

A sistemática de Drosophilidae tem sido objeto de interesse de morfologistas desde STURTEVANT (1921, 1942). Porém, foi apenas mais recentemente que se obteve uma representação mais concreta da filogenia deste grupo, sendo que sua taxonomia permanece amplamente controversa.

2 Filogenia de Drosophilidae

2.1 Estudos Morfológicos

2.1.1 THROCKMORTON (1975)

THROCKMORTON (1975), com base na análise de caracteres reprodutivos internos, descreveu parte da filogenia, ecologia e biogeografia do gênero *Drosophila* dentro da família Drosophilidae. Segundo o mesmo autor, esta família faz parte da superfamília Acalypterae, tendo-se originado, provavelmente, há 50 milhões de anos atrás, em regiões tropicais do Velho Mundo, a partir de ancestrais que utilizavam fungos crescidos em folhço como principal recurso alimentar. Com isso, instala-se um quadro em que, a partir dos Drosophilidae mais primitivos, a evolução ocorreu no sentido de aumentar a diversidade de substratos utilizados, em uma variedade de habitats, atribuindo sempre uma maior ênfase à utilização de frutos. Este autor afirma, ainda, que a evolução da família ocorreu, principalmente, através de sucessivas radiações, onde se verifica o surgimento de muitas linhagens a partir de um mesmo nível filogenético ancestral. No total, teriam sido cinco radiações sucessivas, todas iniciadas antes do final do Eoceno e altamente influenciadas pela disjunção tropical, ocorrida no início do Oligoceno, e pela disjunção das Florestas Temperadas, situada no Mioceno Médio. Além disso, THROCKMORTON (1975) apresenta uma série de subgêneros e gêneros surgindo a partir do gênero *Drosophila*, que ocupa, portanto, uma posição basal dentro da família Drosophilidae e um caráter parafilético na taxonomia do grupo. Segundo este autor, as principais radiações dentro da família são:

1 Radiação Steganinae

Constituída pelos gêneros *Leucophenga*, *Pararhinoleucophenga*, *Paraleucophenga*, *Stegana*, *Steganina*, *Oxyphortica*, *Amiota*, *Phortica* e *Gitona*, entre outros, esta radiação tem sua principal distribuição em regiões tropicais, do Novo e do Velho Mundo, embora alguns de seus representantes (*Amiota* e *Gitona*) sejam cosmopolitas.

2 Radiação Drosophilinae

Dentre seus representantes, espécies do gênero *Microdrosophila* apresentam uma distribuição cosmopolita, enquanto que os membros da Radiação *Scaptodrosophila* são primariamente tropicais.

3 Radiação *Scaptodrosophila*

Provavelmente ocorrida na Ásia Tropical, esta radiação é constituída pelos grupos *latifasciaeformis*, *coracina*, *victoria* e *subtilis*, entre outros. Segundo THROCKMORTON (1975), estes dois últimos grupos, especificamente, apresentam diversas características apomórficas, de forma que sua distribuição temperada disjunta é considerada derivada em relação às demais.

4 Radiação *Sophophora*

Esta radiação é formada pelos grupos *saltans-willistoni* (típicos de regiões Neotropicais), *melanogaster* (presente nos trópicos do Velho Mundo) e *obscura* (distribuído através da zona temperada terrestre), bem como pelos gêneros *Chymomyza*, cosmopolita, e *Neotanygastrella*, pantropical. Como na radiação anterior, as linhagens do Novo Mundo são consideradas derivadas com relação às formas do Velho Mundo, de forma que a radiação *Sophophora* é referida como originada nos trópicos do Velho Mundo, a partir de ancestrais primitivos pertencentes à radiação *Scaptodrosophila*, sendo *obscura* e *melanogaster* os grupos-irmãos que primeiramente se diversificaram.

5 Radiação *Drosophila*

Esta radiação é composta primariamente de outras sub-radiações, além do grupo *funnebris*, de distribuição primariamente temperada, e de outras duas linhagens distintas e independentes, *Liodrosophila* e *Hypselothyrea*, restritas ao Velho Mundo. São elas:

5.1 Radiação *virilis-repleta*

Segundo THROCKMORTON (1975), esta radiação iniciou-se nos trópicos do Velho Mundo, expandindo-se, eventualmente, para o Novo Mundo, de forma que o grupo *repleta* acaba por tornar-se endêmico das Américas. Ainda a partir de migrações desta linhagem inicial, surgiram novos táxons, de distribuição temperada, entre os quais destacam-se os grupos *carsoni*, *virilis*, *melanica* e *robusta*. Assim, a radiação inicial encontra-se constituída de uma série de linhagens bastante heterogêneas entre si, entre as quais destacam-se: os grupos *polychaeta*, *carbonaria*, *tumiditarsus* e *annulimana*, o gênero *Dettopsomyiae*, o subgênero *Phloridosa*, a linhagem temperada já mencionada, e a linhagem formada pelos grupos *nannoptera*, *peruviana* e *bromeliae*, cujos ancestrais acabaram por dar origem à radiação *repleta*. Já esta inclui as seguintes linhagens: o grupo andino *mesophragmatica*, o grupo *castanea*, a linhagem constituída pelos grupos *dreyfusi*

e *canalinae*, e o grupo *repleta*, formado pelos subgrupos *fasciola*, *mulleri*, *melanopalpa*, *mercatorum* e *hydei*.

5.2 Radiação *immigrans-Hirtodrophila*

Esta radiação originou-se e diversificou-se nos trópicos do Velho Mundo, de onde veio a expandir-se, atingindo os trópicos do Novo Mundo através das linhagens incluídas na radiação *tripunctata* e *Hirtodrosophila* e, finalmente, as regiões temperadas, na forma dos grupos *quinaria*, *testacea* e *melanderi* (THROCKMORTON, 1975). O grupo *immigrans*, originário do Velho Mundo, e o gênero *Zaprionus*, tipicamente africano, entre outros, também constituem a base desta radiação.

Segundo THROCKMORTON (1975) o grupo *tripunctata* deu origem, em tempos e níveis diversos, a outros grupos de *Drosophila*, dentre os quais destaca-se o grupo *guarani*. Entre outros grupos originados a partir da radiação *tripunctata* é importante citar o grupo *cardini*, com representantes continentais e insulares no Novo Mundo, bem como os grupos *calloptera*, *pallidipennis*, *macroptera* e *rubrifrons*.

Já o subgênero *Hirtodrosophila*, que constitui-se em outra sub-radiação menor dentro desta radiação, é composto pelos grupos *duncani*, *pinicola* e *melanderi*, além dos gêneros *Zygothrica*, *Nesiodrosophila*, *Mycodrosophila*, *Paramycodrosophila*, *Scaptomyza*, do subgênero *Dorsilopha* e dos drosofilídeos havaianos.

Entretanto, os achados de THROCKMORTON (1975) são, muitas vezes, questionados, devido à inconsistência e às contradições de alguns de seus resultados quanto à análises anteriores (GRIMALDI, 1990), bem como devido à ausência de métodos explícitos e medidas de suporte para cada relação (REMSSEN & O'GRADY, 2002). Outro ponto relevante é que Throckmorton mantém, convenientemente, uma série de táxons parafiléticos e, apesar de suas análises filogenéticas, não propõe nenhuma nova classificação para Drosophilidae com qualquer embasamento cladístico, de forma que a taxonomia do grupo permanece inconsistente com suas relações filogenéticas. Além disso, POWEL E DESALLE (1995), por exemplo, queixam-se que suas árvores são extremamente difíceis de interpretar, uma vez que não explicitam os padrões dicotômicos apresentados pelos cladogramas modernos.

2.1.2 GRIMALDI (1990)

A partir destas críticas, em 1990 GRIMALDI propôs uma nova classificação baseada na análise cladística de 217 caracteres morfológicos adultos em mais de 160 espécies pertencentes a 50 gêneros da família Drosophilidae. Neste trabalho, GRIMALDI diagnostica a monofilia da família Drosophilidae e mantém sua tradicional divisão em duas subfamílias. Entretanto, com base nos caracteres analisados, este autor atesta que o gênero *Drosophila* é, na verdade, derivado e monofilético e que seus subgêneros, propostos por THROCKMORTON (1975) e incorporados por WHEELER (1986), *Hirtodrosophila*, *Scaptodrosophila*, *Antopocerus*, *Engiscaptomyza* e *Lordiphosa* deveriam ser considerados como gêneros à parte. *Zaprionus*, neste caso, representaria o grupo-irmão do gênero *Drosophila*, cujos subgêneros *Sophophora* e *Drosophila* seriam clados-irmãos.

Segundo o mesmo autor, os drosofilídeos Hawaianos, formalmente introduzidos no subgênero *Drosophila*, também não pertenceriam ao gênero *Drosophila*. Mais do que isso, espécies hawaianas de *Drosophila* e *Scaptomyza* não formariam um grupo monofilético. Desta forma, ele translocou as espécies de *Drosophila* hawaianas para um gênero à parte, *Idiomyia*, que constituiria o taxa-irmão ao clado formado por gêneros micófagos, inclusive *Hirtodrosophila*, que ele nomeou de grupo genérico *Zygothrica*. Já *Scaptomyza*, incluindo suas espécies hawaianas, constituiria um grupo tipicamente monofilético, irmão ao clado constituído pelos gêneros *Drosophila* e *Zaprionus*.

Entretanto, até mesmo os trabalhos de GRIMALDI (1990) são constantemente criticados, devido às inúmeras incongruências apresentadas com relação, tanto a estudos anteriores quanto à análises posteriores. Além disso, REMSEN & O'GRADY (2002) afirmam que a metodologia empregada por Grimaldi foi altamente tendenciosa uma vez que, das cerca de 30.000 árvores igualmente parcimoniosas geradas por seus dados, ele considerou apenas cerca de 100, devido às restrições de memória e velocidade computacional então existentes. Porém, há que se ressaltar que o estudo de Grimaldi alcançou uma amostragem taxonômica ainda não superada por qualquer outra tentativa.

2.2 Sistemática Molecular

Pelo fato da família Drosophilidae ser altamente especiosa, em um grande número de casos há uma certa insuficiência de caracteres morfológicos e comportamentais com os quais trabalhar. Neste caso, dados moleculares tem-se mostrado extremamente valiosos,

uma vez que adicionam inúmeros novos caracteres filogeneticamente informativos, capazes de revelar relações previamente obscuras.

Porém, embora estas técnicas tenham resultado em alguma resolução, em geral elas têm sido incapazes de decifrar algumas das mais interessantes politomias filogenéticas, que foram inferidas com base em dados morfológicos e comportamentais. De modo geral e ao contrário do que se possa imaginar, há apenas poucas hipóteses filogenéticas para *Drosophilidae* baseadas em dados moleculares, sendo estas, em grande parte dos casos, incongruentes com análises prévias. Além disso, a grande maioria destas filogenias direcionam-se à resolução das relações entre espécies proximamente relacionadas dentro do gênero *Drosophila*, havendo ainda uma grande concentração de esforços para o subgênero *Sophophora*. Um exame detalhado deste quadro permite dizer que, enquanto algumas relações filogenéticas dentro de *Drosophilidae* são fortemente suportadas por todo um conjunto de dados, outras permanecem controversas em todos os níveis.

Desta forma, uma série de questões permanecem em aberto devido à não-abordagem ou até mesmo às incongruências apresentadas pelos diversos estudos que visam a resolução de uma dada questão. Isto se deve, principalmente, às diferenças entre a filogenia do gene x filogenia das espécies ou até mesmo devido às amostragens divergentes, que impedem a construção de uma estrutura filogenética mais confiável. Na verdade, nenhum único gene foi capaz, até o momento, de produzir uma filogenia inequívoca para a família *Drosophilidae*.

2.2.1 - Os Problemas na Filogenia da família *Drosophilidae*

1) Monofilia da subfamília *Steganinae*

REMSSEN & O'GRADY (2002), ao utilizar representantes da subfamília *Steganinae* (*Amiota*, *Stegana*, *Rhinoleucophenga* e *Gitona*) como grupo-externo à sua análise, comprovaram sua monofilia, confirmando o parecer de THROCKMORTON (1975) e GRIMALDI (1990).

2) Grupos-externos ao gênero *Drosophila*

Segundo PÉLANDAKIS *et al.* (1991), TATARENKOV *et al.* (1999a) e KWIATOWSKI & AYALA (1999), os gêneros *Scaptodrosophila* e *Chymomyza* divergiram antes dos demais *Drosophilinae*, mas nestes trabalhos a ordem correta de ramificação não pôde ser determinada. THROCKMORTON (1975) posicionou

Scaptodrosophila na condição ancestral, enquanto GRIMALDI (1990) deixou esta questão em aberto. Os trabalhos de DESALLE (1992a), POWELL & DESALLE (1995) e RUSSO *et al.* (1995) também apontam *Scaptodrosophila* como a linhagem mais basal, porém os dois primeiros propõem que *Hirtodrosophila* divergiu dos outros Drosophilidae antes de *Chymomyza*, um resultado altamente inconsistente com os demais estudos. Trabalhos mais recentes, entretanto, como o de REMSEN & O'GRADY (2002), têm apontado representantes do clado *Zygothrica* (composto por *Hirtodrosophila*, *Mycodrosophila* e *Paramycodrosophila*) juntamente com *Samoia* como o primeiro grupo de Drosophilinae a divergir, seguidos pelo clado *Formosa* (constituído por *Liodrosophila*, *Sphaerogastrella* e *Dettopsomyia*). Este resultado foi, de certa forma, apoiado por YOTOKO *et al.* (2003) em que o subgênero *Hirtodrosophila* emerge como grupo-irmão do gênero *Drosophila*. Entretanto, a ausência de um íntron do gene *Sod* em todos drosofilídeos com exceção de *Scaptodrosophila* e *Chymomyza* confirma a posição basal destes gêneros (KWIATOWSKI *et al.*, 1994).

3) Posicionamento de *Zaprionus*

O gênero *Zaprionus* foi apresentado como grupo-irmão ao clado formado pelos subgêneros *Sophophora/Drosophila* na filogenia de GRIMALDI (1990) e de DESALLE (1992a), em contraposição aos achados de THROCKMORTON (1975), segundo o qual *Zaprionus* faria parte da radiação *immigrans-Hirtodrosophila* do subgênero *Drosophila*. Posteriormente, *Zaprionus* foi, ainda, apresentado como estando entre os primeiros grupos a divergir depois de *Sophophora* (TATARENKOV *et al.*, 1999a; KWIATOWSKI & AYALA, 1999) - ou até mesmo como sendo o primeiro (PÉLANDAKIS *et al.*, 1993; RUSSO *et al.*, 1995; TATARENKOV *et al.*, 2001; YOTOKO *et al.*, 2003). REMSEN & O'GRADY (2002), entretanto, posicionam *Zaprionus* como taxa-irmão do subgênero *Dorsilopha*, sendo que o clado assim constituído agrupa-se, neste trabalho, ao clado formado pelo subgênero *Sophophora* e pelos gêneros *Scaptodrosophila/Chymomyza*.

4) Monofilia e Posicionamento do subgênero *Sophophora*

Segundo THROCKMORTON (1975), o subgênero *Sophophora* posiciona-se na base da filogenia de *Drosophila*, sendo o subgênero *Drosophila* filogeneticamente mais próximo de diversos gêneros e subgêneros (*Dorsilopha*, *Hirtodrosophila*,

Samoaia, *Liodrosophila*, *Engyscapteromyza*, *Scapteromyza* e *Zaprionus*) do que ao próprio subgênero *Sophophora*. De acordo com essa visão, a radiação *Sophophora* emerge a partir da radiação *Scaptodrosophila*, e compreende 5 linhagens principais: o grupo *obscura*; o grupo *melanogaster*; o grupo *saltans*; o grupo *willistoni*; e os gêneros *Chymomyza* e *Neotanygastrella*. Entretanto, no cladograma de GRIMALDI (1990), *Sophophora* é o grupo-irmão do subgênero *Drosophila*, sendo que estes, uma vez acrescentados ao subgênero *Dorsilopha*, constituiriam o gênero *Drosophila* em sua totalidade. Este último resultado foi, em parte, corroborado pelos achados de DESALLE (1992a), que apresentam o subgênero *Sophophora* como taxa-irmão ao clado composto pelo subgênero *Drosophila* e pelo gênero *Liodrosophila*.

Entretanto, a maior parte dos dados moleculares tem favorecido a hipótese de THROCKMORTON (1975), de que o subgênero *Sophophora* é o primeiro ramo a divergir de todas as demais espécies do gênero *Drosophila* (RUSSO *et al.*, 1995; KWIATOWSKI & AYALA, 1999; TATARENKOV *et al.*, 1999a e 2001;). RUSSO *et al.* (1995), inclusive, estimaram que esta separação se deu há aproximadamente 40 milhões de anos atrás. No trabalho de YOTOKO *et al.* (2003), porém, o subgênero *Sophophora* foi o segundo grupo de *Drosophila* a divergir, perdendo apenas para o subgênero *Hirtodrosophila*. Na análise de REMSEN & O'GRADY (2002), por outro lado, o subgênero *Sophophora*, monofileticamente definido, agrupa-se ao clado constituído por *Zaprionus/Dorsilopha*, *Scaptodrosophila/Chymomyza*, de forma a constituir o clado-irmão do subgênero *Drosophila*, também parafilético com relação a *Scapteromyza/Engyscapteromyza* e a *Siphlodora*.

Outra pendência importante foi trazida à tona no estudo molecular de PÉLANDAKIS *et al.* (1991) e PÉLANDAKIS *et al.* (1993), que levantaram a questão de que os grupos *saltans/willistoni* podem não formar um grupo monofilético com os grupos *melanogaster/obscura*, de forma que o subgênero *Sophophora* constitua um táxon parafilético. Neste último trabalho, por exemplo, o grupo *willistoni* posicionou-se externamente à todas as outras linhagens de *Drosophila*, inclusive ao subgênero *Sophophora*, agrupando-se com *Chymomyza*. Além disso, no trabalho de PÉLANDAKIS *et al.* (1991), o grupo *melanogaster* também é parafileticamente definido com relação ao grupo *obscura*.

A monofilia do subgênero *Sophophora* foi, porém, confirmada no estudo de TATARENKOV *et al.* (1999a), onde todas as suas espécies apresentam uma deleção de um único códon na região codificante do gene *Ddc*. Os problemas no correto posicionamento dos grupos *saltans-willistoni* são atribuídos, por diversos autores (RODRIGUEZ-TRELLES *et al.*, 1999; TATARENKOV *et al.*, 1999a; TARRÍO *et al.*, 2000), à evolução molecular atípica apresentada por estes grupos, caracterizada por altas taxas de substituição nucleotídica e baixa proporção de G/C. De modo geral, a maior parte dos trabalhos tem confirmado a proximidade evolutiva entre os grupos *melanogaster/obscura*, e os grupos *saltans/willistoni* (DESALLE, 1992a; POWELL & DESALLE, 1995; REMSEN & O'GRADY, 2002), sendo a ocorrência desta separação estimada para cerca de 36 milhões de anos atrás (RUSSO *et al.*, 1995). Entretanto, outros grupos tradicionalmente situados dentro deste subgênero (como *alagitans*, *fima*, etc.) permanecem por ser incluídos em análises envolvendo este clado.

5) Filogenia do subgênero *Drosophila*

Segundo THROCKMORTON (1975), o subgênero *Drosophila* consiste de dois cladogramas principais: a radiação *virilis-repleta* e a radiação *immigrans-Hirtodrosophila* (da qual deriva a radiação *tripunctata*), derivadas a partir de um ancestral comum existente há mais de 33 milhões de anos atrás (RUSSO *et al.*, 1995). Ainda segundo Throckmorton, o subgênero *Drosophila* deriva origem, em tempos remotos, a diversos outros gêneros e subgêneros, de forma que não se constitui em um grupo monofilético. DESALLE (1992b) e POWELL & DESALLE (1995), entretanto, favoreceram a monofilia do subgênero *Drosophila*, como sendo o grupo-irmão aos drosofilídeos hawaianos. Uma série de trabalhos posteriores, porém, têm indicado que o subgênero *Drosophila* é, na verdade, parafilético (TATARENKOV *et al.*, 1999a; KWIATOWSKI & AYALA, 1999; REMSEN & O'GRADY, 2002). No trabalho de KWIATOWSKI & AYALA (1999), por exemplo, *D. immigrans* aparece como mais proximamente relacionada ao subgênero *Hirtodrosophila* do que às demais espécies do subgênero *Drosophila* que foram consideradas, sendo que este clado, juntamente com *Zaprionus*, é um dos primeiros a ramificar após o subgênero *Sophophora*. Este parecer, na verdade, apenas confirma a visão de THROCKMORTON (1975). Recentemente, entretanto, YOTOKO *et al.* (2003) contradisse esta visão ao

apresentar um subgênero *Drosophila* monofilético para as espécies consideradas e estatisticamente distante ao subgênero *Hirtodrosophila*.

6) Filogenia dos membros da radiação *repleta*

As relações filogenéticas dos grupos e espécies na radiação *virilis/repleta* encontram-se bastante mal resolvidas na classificação de THROCKMORTON (1975). DURANDO *et al.* (2000) sugerem que o grupo *repleta* é, na verdade, parafilético com relação aos grupos *dreyfusi*, *canalinae* e *mesophragmatica*. Ainda neste trabalho, dos cinco subgrupos pertencentes ao grupo *repleta*, apenas *mercatorum* e *hydei* são vistos como monofiléticos (o subgrupo *mulleri* revela-se polifilético enquanto que os representantes do subgrupo *melanopalpa* aparecem como parafiléticos com relação ao subgrupo *mercatorum*). Entretanto, na análise de TATARENKOV & AYALA (2001), o grupo *repleta* constitui-se monofileticamente e, com exceção do grupo *tumiditarsus*, todos os demais membros da radiação *repleta* formam um *cluster* monofilético, proximamente relacionado aos drosofilídeos hawaianos. A parte mais derivada da filogenia apresentada neste trabalho contém todos os representantes do grupo *repleta*, cujas relações são altamente consistentes com os estudos prévios de THROCKMORTON (1975) e WASSERMANN (1982), com exceção do posicionamento correto dos grupos *mercatorum* e *melanopalpa*, que são apresentados como grupos-irmãos nas filogenias destes últimos.

D. gaucha (grupo *mesophragmatica*), *D. camargoi* (grupo *dreyfusi*) e *D. canalinae* (grupo *canalinae*) são os grupos mais próximos ao grupo *repleta*, sendo que os dois primeiros constituem grupos-irmãos entre si no trabalho de TATARENKOV & AYALA (2001), enquanto que, na análise de THROCKMORTON (1975), os grupos *canalinae* e *dreyfusi* apresentam-se mais proximamente relacionados entre si do que com o grupo *mesophragmatica*. Segundo REMSEN & O'GRADY (2002) o grupo *bromeliae* representa o táxon-irmão do grupo *mesophragmatica*. Outro clado bem definido no estudo de TATARENKOV & AYALA (2001) e REMSEN & O'GRADY é constituído pelos grupos *robusta* e *melanica*. Nos trabalhos anteriores de DESALLE (1992a), entretanto, representantes do grupo *melanica* agruparam-se com o clado constituído por *D. funebris*/*D. pinicola*, com *D. robusta* ramificando em uma posição intermediária da filogenia. O posicionamento correto do grupo *virilis* também permanece bastante confuso, sendo

que nas análises de POWELL & DESALLE (1995), REMSEN & DESALLE (1998) e de REMSEN & O'GRADY (2002), *D. virilis* posiciona-se mais próximo aos membros da radiação *repleta* do que a *D. melanica* e *D. robusta*, enquanto na análise de KATOH *et al.* (2000), o grupo *virilis* é agrupado com o grupo *robusta*.

7) Posicionamento de *Scaptomyza* e de *Drosophila* hawaianas

THROCKMORTON (1975) considerou os grupos hawaianos de *Drosophila* e *Scaptomyza/Engyscaptomyza* como clados-irmãos e derivados com relação aos demais membros do subgênero *Drosophila*, pertencentes à radiação *immigrans-Hirtodrosophila*. GRIMALDI (1990), por outro lado, propôs que as espécies hawaianas de *Drosophila* juntamente com *Hirtodrosophila* representariam uma ramificação basal na subfamília Drosophilinae, devendo ser elevados à categoria genérica (*Idiomya*), enquanto *Scaptomyza* constituiria o grupo-irmão do clado formado pelos gêneros *Drosophila* e *Zaprionus*. Já DESALLE (1992b) e POWELL & DESALLE (1995), descrevem as espécies hawaianas de *Scaptomyza* e *Drosophila* como uma ramificação basal e irmã do subgênero *Drosophila*. Mais recentemente, diversos trabalhos (RUSSO *et al.*, 1995; REMSEN & DESALLE, 1998; KWIATOWSKI & AYALA, 1999; TATARENKOV *et al.*, 1999a e 2001; REMSEN & O'GRADY, 2002) tem confirmado a idéia de que as espécies hawaianas de *Drosophila* e *Scaptomyza/Engyscaptomyza* formam um grupo monofilético, que é bastante próximo de espécies da linhagem *virilis-repleta*. Esta visão é consistente com a hipótese, defendida por THROCKMORTON (1975), de que todos os drosofilídeos hawaianos originaram-se a partir de uma única espécie ancestral. Este conjunto de espécies parece ser, ainda, mais aparentado ao subgênero *Hirtodrosophila* do que às demais espécies do subgênero *Drosophila* (TATARENKOV *et al.*, 1999a).

8) Filogenia dos membros da radiação *tripunctata*

Embora o grupo *tripunctata* represente um dos grupos de *Drosophila* mais especiosos da região Neotropical, onde apresenta 64 espécies identificadas (Vilela & Bächli, 2000), sua sistemática tem sido largamente negligenciada. THROCKMORTON (1975) incluiu o grupo *tripunctata* dentro da radiação *tripunctata*, juntamente com outros grupos: *rubrifrons*, *guarani*, *calloptera*, *pallidipennis*, *cardini*, *sticta* e *macroptera*. Segundo este autor, este grupo é pobremente definido, e proximamente relacionado aos outros grupos de sua radiação, servindo ainda

como base a partir da qual outros grupos surgem. Neste último ponto, Throckmorton concorda com FROTA-PESSOA (1954), que dividira este grupo em outros quatro subgrupos, apontando que, possivelmente, os subgrupos I e III eram transicionais a outros grupos, como *guarani* e *cardini*, respectivamente, de forma que o grupo constituiria um conjunto parafilético de espécies.

Na análise executada por TATARENKOV *et al.* (2001) *D. tripunctata* e *D. putrida* (grupo *testacea*) agrupam-se com *D. funebris* e *D. phalerata* (grupo *quinaria*) com alto suporte estatístico, sendo *D. immigrans* e *Samoaia* os grupos-irmãos mais próximos deste clado. Assim, neste estudo, especificamente, os membros do subgênero *Drosophila* constituintes da radiação *immigrans-Hirtodrosophila*, juntamente com *D. funebris* revelam-se monofiléticos. No trabalho de REMSEN & O'GRADY (2002) e na análise de YOTOKO *et al.* (2003) a linhagem *immigrans-tripunctata* também constitui-se monofileticamente, como clado-irmão à linhagem formada pela radiação *virilis-repleta/Drosophilidae* Hawaianos e mediante a exclusão de *Hirtodrosophila*, que aparece como o primeiro ramo a divergir dentro da subfamília Drosophilinae em ambos os trabalhos. YOTOKO *et al.* (2003) afirmam que a radiação *tripunctata* é, na verdade, parafilética, devido à inclusão de *D. funebris* e *D. putrida* (grupo *testaceae*) neste *cluster*, proximamente aparentado aos representantes dos grupos *quinaria* e *immigrans*. Já em REMSEN & O'GRADY (2002), o grupo *tripunctata* agrupa-se diretamente com o grupo *testaceae*, enquanto os grupos *funebris/pinicola* constituem o grupo-irmão do clado formado pelos grupos *guarani/calloptera/cardini* e pelos grupos *bizonata/macroptera/rubrifrons/quinaria*. De modo geral, estes resultados contradizem THROCKMORTON (1975), segundo o qual o grupo *testaceae* constituiria, juntamente com os grupos *quinaria* e *immigrans*, a parte basal da radiação *immigrans-Hirtodrosophila*, confirmando, ainda, os resultados de GRIMALDI (1990) e DESALLE (1992a), que conferiram uma posição derivada ao grupo *funebris*, que segundo THROCKMORTON (1975) ocuparia uma posição basal dentro da radiação *Drosophila*.

Por outro lado, o trabalho de YOTOKO *et al.* (2003) apresentou um grupo *tripunctata* parafilético e altamente descontínuo ao longo da filogenia, na qual alguns de seus membros mesclam-se com representantes dos grupos *calloptera*,

funebri, *testaceae*, *guarani* e *pallidipennis*, sendo impossível identificar qualquer grupo original, provavelmente devido a episódios de especiação rápida. Da mesma forma, dentre estes grupos, aqueles em que mais de uma espécie fora considerada (*calloptera* e *guarani*) revelaram-se polifiléticos.

Por tudo isso, a maior parte dos trabalhos considerados tem indicado que o gênero *Drosophila* é, na verdade, um grupo parafilético, que reúne representantes de diversos outros gêneros de Drosophilinae. Assim, embora GRIMALDI (1990) tenha defendido a hipótese de que o gênero *Drosophila* forme um grupo monofilético, derivado e irmão de outros gêneros de Drosophilinae, a visão de THROCKMORTON (1975), apresentando outros subgêneros e gêneros surgindo a partir do gênero *Drosophila* (que ocupa uma posição basal dentro da subfamília *Drosophilinae*) ainda parece ser a mais correta.

III O GENE *AMD*

O gene *Amd* (α -metil dopa) é um gene essencial para a viabilidade e a fertilidade de drosofilídeos, de forma que a maior parte de suas mutações ocasionam um fenótipo recessivo letal (LINDSLEY & ZIMM, 1992), e até mesmo um fenótipo dominante condicional letal (MARSH *et al.*, 1986). Este gene apresenta uma relação de paralogia com o gene *Ddc* (dopa-descarboxilase), estando ambos relacionados ao metabolismo de catecolaminas durante a diferenciação terminal de *Drosophila*, atuando especialmente no desenvolvimento cuticular. Catecolaminas servem como agentes de ligação cruzada para a esclerotização da cutícula, como precursores para a formação da melanina ou ainda como neurotransmissores.

O gene *Amd* codifica um transcrito maduro de 1779 pb [antes da adição da cauda poli-A mas após a remoção de seu único íntron, cujo tamanho é variável (TATARENKOV *et al.*, 2001)]. Segundo MARSH *et al.* (1986), uma única ORF (*open reading frame* – fase aberta de leitura) de 1395 pb pode ser encontrada neste transcrito, sendo que a mesma inicia-se com um códon para metionina a 284 pb *downstream* à extremidade 5' e estende-se até cerca de 100 pb *upstream* à extremidade 3' (figura 1). A tradução desta seqüência leva à produção de uma proteína de 464 aminoácidos, com mais de 50 kD (kilodáltons), que apresenta uma carga levemente negativa, devido ao excesso de resíduos de Asp e Glu. Esta proteína apresenta, ainda, um amplo domínio hidrofóbico, que participa, possivelmente, da ligação à membrana celular.

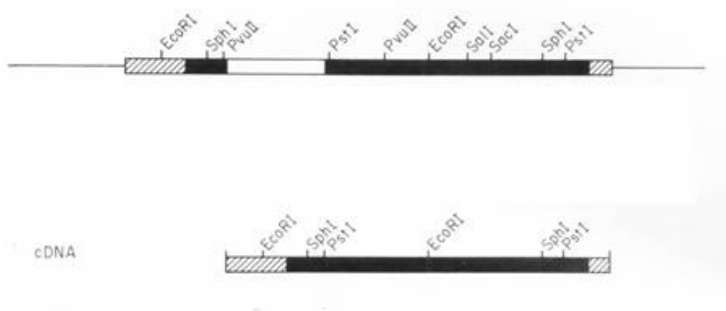


Fig. 1- Mapa do DNA genômico contendo o gene *Amd* e seu cDNA, com as regiões transcritas indicadas por retângulos: regiões codificantes indicadas por retângulos pretos, regiões não traduzidas do mRNA indicadas por retângulos hachurados e íntrons representados por retângulos vazios. (Retirado de MARSH *et al.*, 1986)

Inicialmente, a caracterização deste gene envolveu a descoberta de que mutantes, tanto para *Amd* quanto para *Ddc*, apresentam um desenvolvimento cuticular anormal. Logo demonstrou-se, entretanto, que enquanto o produto gênico *Amd*⁺ conferia resistência à administração de análogos de dopa, especialmente alfa-metildopa, a atividade de *Ddc* era inibida pela adição destes mesmos compostos (EVELETH & MARSH, 1986). Assim, ainda que envolvidos na mesma rota metabólica e codificando para descarboxilases que reconhecem substratos de estrutura similar, *Amd* e *Ddc* desenvolveram funções não sobrepostas, incluindo, segundo WANG *et al.* (1996), a produção de diferentes tipos de cutícula e, para *Ddc*, a produção dos neurotransmissores dopamina e serotonina. Segundo estes mesmos autores o gene *Amd* é particularmente ativo na produção de uma cutícula flexível e incolor durante o desenvolvimento do embrião, quando é expresso primordialmente na epiderme, no intestino anterior e posterior, no proventrículo e no sistema traqueal. Sua expressão é, ainda, requerida na linhagem germinativa e nas células foliculares femininas, onde é responsável pela produção de uma membrana vitelínica flexível e impermeável, característica dos ovos de *Drosophila*, bem como no sistema nervoso de larvas maduras. *Ddc*, por outro lado, especializou-se na esclerotização de cutícula pigmentada e na produção de neurotransmissores, sendo expresso apenas em estágios posteriores, após a cutícula já ter sido depositada.

Segundo TATARENKOV *et al.* (1999b), a duplicação que deu origem aos genes *Amd* e *Ddc* precedeu a divergência Diptera-Lepidoptera. Atualmente, a comparação de suas seqüências revela a existência de duas grandes regiões estruturalmente relacionadas, que restringem-se, principalmente, à seqüência encontrada em grande parte do éxon II de *Amd*, homóloga ao éxon III de *Ddc*, bem como à parte da seqüência 5' UTR de *Amd*, homóloga ao éxon II e à ambas as bordas flanqueadoras do íntron II de *Ddc* (EVELETH & MARSH, 1986). A posição e a seqüência das regiões intervenientes é altamente divergente. Dentro das regiões com homologia estabelecida, podem ser detectadas duas porções de extensa identidade entre *Amd* e *Ddc*, responsáveis pela maior parte do seu alto percentual de similaridade [61% em *D. melanogaster*, de acordo com WANG *et al.* (1996)]: uma localizada próxima ao segundo éxon de *Amd* (80% de identidade) e a outra situada 700 pb da extremidade 3' do transcrito de *Amd* (90% de identidade) (EVELETH & MARSH, 1986). Estes mesmos autores cogitam, ainda, a hipótese de que esta similaridade estrutural possa

levar à retenção funcional de dois domínios protéicos conservados: um provavelmente relacionado à dimerização da proteína, o outro atuando como possível sítio ativo de reação.

Amd e *Ddc* encontram-se separados por uma seqüência de apenas 2 kb, que contém outra unidade de transcrição - o gene *Cs*. TATARENKOV *et al.* (1999b) enfatizam que estes três genes estão entre os genes mais proximamente ligados de *D. melanogaster* (figura 2), a ponto de as extremidades 3' UTR convergentemente transcritas dos genes *Cs* e *Ddc* sobreporem-se por uma extensão de cerca de 90 pb, enquanto as extremidades 3' de *Amd* e 5' de *Cs* distanciam-se em menos de 650 pb. STATHAKIS *et al.* (1995) sugerem que, neste caso, o alto nível de expressão de um gene poderia suprimir a expressão de seu vizinho sobreposto através de uma "interferência transcricional", já que ambos os genes são expressos com um padrão espaço-temporal diferente. *Amd* e *Ddc* também são transcritos a partir de fitas opostas.

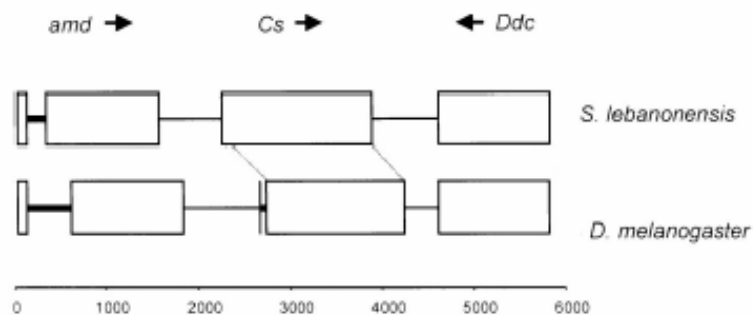


Fig. 2: Organização gênica e direção da transcrição de um segmento de DNA genômico abrangendo os genes *Amd*, *Cs* e *Ddc* em menos de 6 kb dos genomas de *Scaptodrosophila lebanonensis* e *Drosophila melanogaster*. Setas indicam a direção de transcrição, retângulos indicam suas regiões codificantes, linhas conectoras grossas representam íntrons enquanto linhas finas representam regiões não codificantes. Modificado de TATARENKOV *et al.* (1999b).

Outro indicativo surpreendente das peculiaridades dos genes *Amd* e *Ddc* diz respeito à sua organização estrutural de nível maior. Estes genes constituem, em *Drosophila melanogaster*, uma fração essencial de um *cluster* gênico único, chamado *Cluster Ddc*, localizado no braço esquerdo do cromossomo 2, próximo ao centrômero (STATHAKIS *et al.*, 1995). Este *cluster* gênico é composto por 18 grupos de complementação já caracterizados, mais três unidades de transcrição para as quais nenhuma mutação fenotípica foi, até o momento, detectada (TATARENKOV *et al.*, 1999b). Segundo STATHAKIS *et al.* (1995), estes genes abrangem uma região total de

aproximadamente 145 kb, que corresponde a um máximo de 11 bandas politêmicas. Entretanto, a maior parte dos 21 genes não se distribui aleatoriamente ao longo desta região, localizando-se preferencialmente em um de dois *subclusters* densamente organizados: *subcluster* distal, de aproximadamente 27,4 kb, constituído por 6 loci e duas unidades de transcrição; *subcluster* proximal, de 23 kb, que contém 8 loci (entre os quais situam-se *Amd* e *Ddc*) e mais uma unidade de transcrição (o gene *Cs*, que ocupa uma posição intermediária aos dois anteriores). Aproximadamente 53,3 kb de DNA separa os dois *subclusters*, região esta ocupada por outros dois genes. Os dois últimos genes ocupam uma área de 27,2 kb proximal ao *subcluster* proximal (figura 3). Surpreendentemente, mutações na maior parte destes genes resultam em defeitos morfo genéticos similares, que afetam a esclerotização ou a melanização da cutícula ou resultam em anormalidades no metabolismo de catecolaminas.

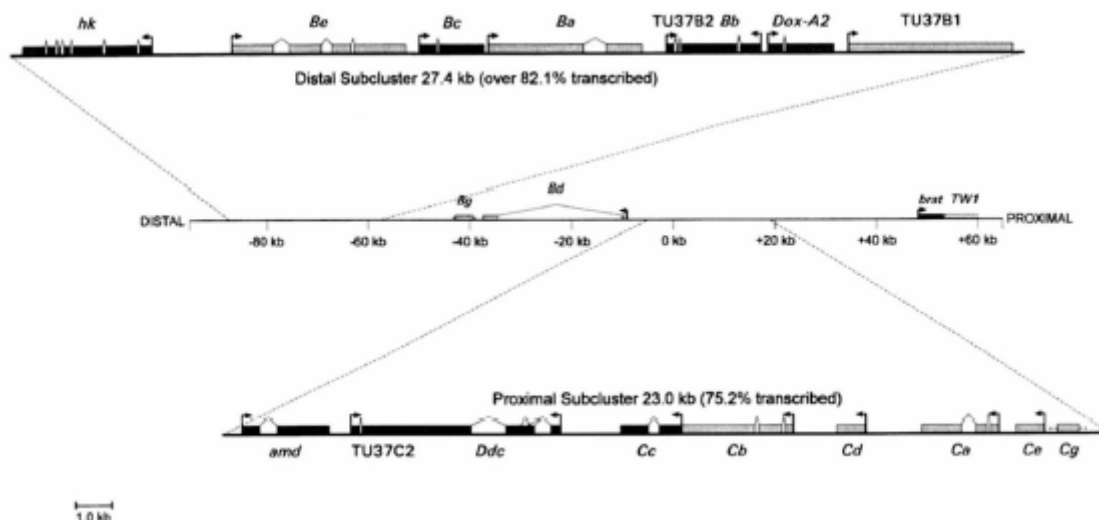


Fig. 3: Organização molecular do *cluster* gênico *Ddc* e sua divisão em *subclusters*. São apresentadas a localização, estrutura e direção transcripcional dos genes dentro de cada *subcluster*. Éxons estão representados por retângulos e íntros por linhas. Obtido a partir de STATHAKIS *et al* (1995).

Desta forma, este *cluster* gênico representa um exemplo incomum, para eucariotos superiores, de agrupamento funcional de genes cujo produto é requerido em uma rota bioquímica comum. Sua conservação ao longo do tempo serve como um importante indício de que, evolutivamente, esta configuração espacial é adaptativa. STATHAKIS *et al.* (1995) sugerem que, embora não tenham sido identificados até o momento, a operação de supostos elementos controladores globais pode ser uma importante fonte de manutenção,

já que a expressão temporal comum da maior parte destes genes pode ser experimentalmente verificada durante a embriogênese.

1 Tempo de Expressão

O transcrito de *Amd* começa a acumular-se em polissomos logo após a gastrulação, atingindo seu ápice em estágios posteriores da embriogênese, de 12 a 16 hs, mas persistindo em níveis menores durante a maior parte deste estágio (MARSH *et al.*, 1986). Estes autores reiteram, ainda, que baixíssimos níveis de expressão são detectados em adultos e larvas de 3º estágio, nas quais também pode-se observar a presença de um transcrito menor, aparentemente estágio específico.

Ddc, por outro lado, começa a acumular-se após cerca de 12 hs de embriogênese, atingindo seu máximo por volta de 18 hs. Porém, neste caso, a transcrição é reiniciada no estágio de pupa e logo após a eclosão dos adultos (EVELETH & MARSH, 1986).

2 Utilização no desenvolvimento de hipóteses filogenéticas

TATARENKOV *et al.* (1999a, 2001) e TATARENKOV & AYALA (2001) têm utilizado seqüências codificantes do gene *Amd* no endereçamento de diversas questões filogenéticas relacionadas ao gênero *Drosophila*, tendo obtido resultados altamente consistentes e congruentes entre si, que conduzem à uma filogenia bastante resolvida. Dentre as características essenciais deste gene, e que o tornam filogeneticamente confiável para este nível de divergência, pode-se destacar sua relativa restrição funcional, inferida a partir da baixa razão entre o número de substituições não-sinônimas e sinônimas encontradas ao longo dos seus sítios [da ordem de 0,066 entre *D. melanogaster* e *D. simulans* segundo TATARENKOV *et al.* (1999b)], contrastada pelos altos níveis de divergência encontrados na terceira posição de seus códons e por um ENC (número efetivo de códons) caracteristicamente alto para seqüências codificantes importantes (figuras 4 e 5, respectivamente), com um viés na utilização de códons significativamente baixo. A correlação negativa existente entre o viés na utilização de códons e a taxa de evolução molecular fornece fortes evidências de que este gene tem evoluído a taxas relativamente rápidas, de modo que pode parecer justificado utilizá-lo, ainda, para estimar o tempo de divergência entre as espécies analisadas. Para o gene *Adh*, a taxa evolutiva de substituições sinônimas, estimada a partir de espécies hawaianas, é de 15×10^{-9} por sítio sinônimo por ano (NURMINSKY *et al.*, 1996). Entretanto, como até as substituições sinônimas estão

sujeitas a uma certa extensão de seleção purificadora, pode ser mais adequada a utilização da estimativa geral de substituição de 10×10^{-9} por sítio por ano, como foi feito por RUSSO *et al.* (1995).

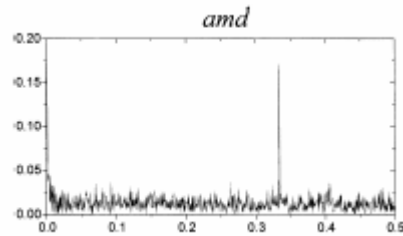


Fig. 4: Periodicidade de substituições apresentada pelos éxons do gene *Amd*, cujo padrão revela uma frequência de substituições dominante a cada três posições (1/3). Modificado de TATARENKOV *et*

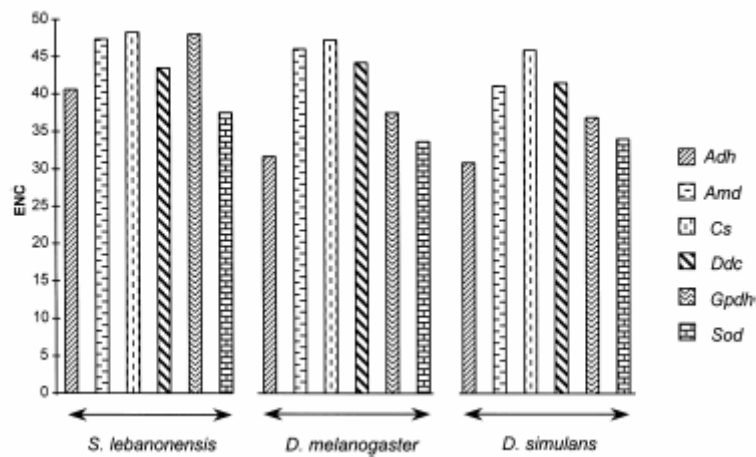


Fig. 5: Viés na utilização de códons em *S. Lebanonensis*, *D. melanogaster* e *D. simulans*, dado a partir do ENC (número efetivo de códons) apresentado por 6 diferentes genes. Um ENC menor indica um viés maior na utilização de códons sinônimos. Obtido em TATARENKOV *et al.* (1999b).

Entretanto, mesmo com esta alta frequência de substituições, a terceira posição dos códons de *Amd* permanece informativa e altamente congruente com os demais sítios de modo que, em geral, este gene não apresenta altos índices de saturação nem mesmo nestes pontos (TATARENKOV & AYALA, 2001). Além disso, o gene *Amd* apresenta uma distribuição heterogênea de substituições não-sinônimas, de modo que diferentes regiões do gene podem ser utilizadas de acordo com o problema filogenético em questão (figura 6). Trata-se, pois, de um gene cujas seqüências podem ser altamente informativas para a detecção de relações filogenéticas entre grupos próxima e medianamente relacionados.

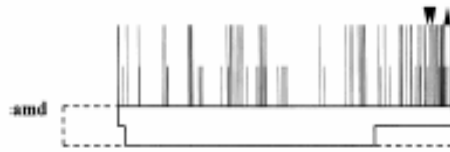


Fig. 6: Heterogeneidade na distribuição das substituições de aminoácidos ao longo de *amd*, cuja extremidade C-terminal destaca-se como a menos restrita funcionalmente. Dados obtidos pela comparação das seqüências do gene *Amd* de *S. lebanonensis* e *D. melanogaster*. Substituições são mostradas como linhas verticais: curtas quando a substituição é conservativa (obtido aminoácidos com propriedades bioquímicas similares) e longas quando ela for não conservativa. Setas indicam a ocorrência de gaps. Modificado a partir de TATARENKOV *et al.* (1999b).

OBJETIVOS

Dadas as circunstâncias anteriormente mencionadas, o presente trabalho visa a aplicação filogenética de seqüências nucleotídicas do gene *Amd* na resolução de importantes questões evolutivas relacionadas ao gênero *Drosophila* (Capítulo II), dentre as quais destaca-se:

- a) a resolução de questões relacionadas ao correto posicionamento evolutivo dos gêneros *Zaprionus*, *Liodrosophila* e *Samoia*, bem como dos subgêneros *Sophophora*, *Hirtodrosophila* e *Dorsilopha*, com relação ao gênero e ao subgênero *Drosophila*;
- b) a análise da monofilia do gênero e do subgênero *Drosophila*, bem como das suas radiações *virilis-repleta* e *immigrans-Hirtodrosophila*;
- c) a inferência das relações filogenéticas dentro e dentre grupos Neotropicais do subgênero *Drosophila* que, subrepresentados em análises anteriores, pertençam, tradicionalmente, a ambas as radiações: grupos *mesophragmatica* e *flavopilosa*, representantes da radiação *virilis-repleta*; e grupos *tripunctata*, *cardini*, *guarani*, *guaramunu* e *immigrans*, incluídos na radiação *immigrans-Hirtodrosophila*;
- d) a análise de alguns padrões evolutivos do gene *Amd*.

CAPÍTULO II

Trabalho em preparação para ser enviado à revista
Molecular Phylogenetics and Evolution.

**RELAÇÕES FILOGENÉTICAS ENTRE GRUPOS E ESPÉCIES NEOTROPICAIS DO SUBGÊNERO
DROSOPHILA (DROSOPHILIDAE – DIPTERA) INFERIDAS COM BASE EM ANÁLISE
MOLECULAR DO GENE NUCLEAR *AMD* (ALFA METIL-DOPA)**

Lizandra J. Robe¹, Vera L. S. Valente¹, Myriam Budnik², Élgion L. S. Loreto³

1 - Programa de Pós-Graduação em Genética e Biologia Molecular, Departamento de Genética – Universidade Federal do Rio Grande do Sul (UFRGS)

2 - Instituto de Ciências Biomédicas, Faculdade de Medicina - Universidade do Chile

3 - Departamento de Biologia – Universidade Federal de Santa Maria (UFSM)

RESUMO

O gênero *Drosophila* vem se destacando, ao longo dos últimos 100 anos, como sendo um organismo modelo ideal para os mais diversos estudos biológicos. Entretanto, a sua sistemática ainda permanece, na maior parte dos casos, bastante controversa e irresolvida. Este quadro é, ainda, agravado no que se refere às espécies Neotropicais pertencentes ao subgênero *Drosophila*, totalmente negligenciadas em muitos estudos. Foi, portanto, com o objetivo de contribuir para a resolução de algumas das questões evolutivas relacionadas a este subgênero que realizou-se a análise filogenética de 513 pares de bases da região codificante do gene *Amd* (alfa-metil dopa) em um total de 49 taxa. Entre os resultados obtidos, pode-se destacar a confirmação da parafilia do subgênero *Drosophila* com relação aos gêneros *Zaprionus*, *Samoia* e *Liodrosophila*, e ao subgênero *Hirtodrosophila*, bem como sua subdivisão em dois *clusters* principais, um agrupando espécies tradicionalmente posicionadas na radiação *virilis-repleta* e o outro reunindo representantes da radiação *immigrans-Hirtodrosophila*. Dentro do primeiro destes clados, pôde-se comprovar a monofilia do grupo *flavopilosa*, que aparece como clado-irmão do grupo *annulimana*, bem como a monofilia do grupo *mesophragmatica*, proximamente relacionado ao grupo *repleta*. Já na linhagem *immigrans-Hirtodrosophila*, os gêneros *Zaprionus* e *Samoia*, juntamente com *Hirtodrosophila* destacam-se como os primeiros a ramificar, seguidos pelos grupos *immigrans*, *quinaria*, *testaceae* e *funebri*, que confirmam sua posição basal com relação aos membros da radiação *tripunctata*, monofileticamente definida. Dentro deste clado menor, entretanto, o grupo *tripunctata* revelou-se parafilético com relação aos grupos *cardini*, *guarani* e *guaramunu*. Este trabalho, portanto, revela e/ou ratifica algumas incongruências referentes à taxonomia do gênero *Drosophila*.

1. INTRODUÇÃO

A família Drosophilidae apresenta cerca de 3.000 espécies, distribuídas em mais de 60 gêneros (Wheeler, 1986), dentre os quais destaca-se o gênero *Drosophila*. Segundo o mesmo autor, este gênero perfaz sozinho mais da metade das espécies da família, apresentando uma ampla e dispersa distribuição. Trata-se, portanto, de um gênero altamente especioso, cuja sistemática ainda permanece, na maior parte dos casos, controversa e incompleta, apesar de sua importância na elucidação de inúmeros padrões e processos biológicos, cuja base já se encontra previamente caracterizada para este organismo-modelo fundamental.

Taxonomicamente, o gênero *Drosophila* contém 15 subgêneros (Wheeler, 1986), dos quais sobressaem, pela diversidade apresentada, os subgêneros *Sophophora* e *Drosophila*. O subgênero *Drosophila* é o maior representante da família Drosophilidae em regiões Neotropicais (Val *et al.*, 1981), onde apresenta 25 grupos de espécies. Malogolowkin (1953) e Throckmorton (1975) subdividiram estas espécies em duas linhagens principais: a seção (ou radiação) *virilis-repleta*, composta por 15 grupos de espécies, e a seção *quinaria-tripunctata* (ou radiação *immigrans-Hirtodrosophila*), que agrupa outros 9 grupos Neotropicais. Além disso, este subgênero é composto pelo grupo *funbris*, que segundo Throckmorton (1975) ocuparia uma posição basal em relação a ambas radiações. Este mesmo autor defende, ainda, que o subgênero *Drosophila* constitui um clado parafilético, uma vez que dá origem, em locais e tempos diversos, a uma série de outros gêneros e subgêneros. Esta hipótese têm sido, recentemente, reiterada por uma série de novos trabalhos (Kwiatowski & Ayala, 1999; Tatarenkov *et al.*, 1999 e 2001, Remsen & O'Grady, 2002), ao mesmo tempo que é contradita em muitos outros (Grimaldi, 1990; DeSalle, 1992; Powell & DeSalle, 1995). Este quadro de obscuridade filogenética é, ainda, agravado, pelo fato de que muitos representantes deste subgênero tem sido largamente negligenciados na maior parte dos estudos evolutivos, principalmente no tange aos grupos de distribuição primariamente Neotropical.

Entre os representantes Neotropicais do subgênero *Drosophila* que encaixam-se neste quadro de subrepresentação pode-se destacar o grupo *tripunctata*, pertencente à seção *quinaria-tripunctata* de Malogolowkin (1953) e Throckmorton (1975). Este grupo, composto atualmente por 64 espécies (Vilela & Bächli, 2000), teve sua filogenia interna

raramente retratada, com exceção dos trabalhos morfológicos de Frota-Pessoa (1954) e da análise molecular recente executada por Yotoko *et al.* (2003). Segundo Throckmorton (1975) o grupo *tripunctata* constituiria um grupo parafilético, pertencente à sua “radiação *tripunctata*”, juntamente com outros grupos como *pallidipennis*, *rubrifrons*, *sticta*, *macroptera*, *calloptera*, *cardini* e *guarani*. Este último grupo, particularmente, foi proposto por Kastritsis (1969) como contendo alguns membros filogeneticamente mais próximos ao grupo *tripunctata* do que aos demais representantes de seu grupo, o que levou ao seu desmembramento em dois novos grupos: *guarani* e *guaramunu*.

Já a seção ou radiação *virilis-repleta* do subgênero *Drosophila*, embora tenha sido mais amplamente considerada em estudos filogenéticos (Nurminsky *et al.*, 1996; Durando *et al.*, 2000; Tatarenkov & Ayala, 2001), ainda apresenta uma série de lacunas no que se refere a alguns de seus grupos. O grupo andino *mesophragmatica*, por exemplo, ainda mantém suas relações intergrupais essencialmente conflitantes (Throckmorton, 1975; Tatarenkov & Ayala, 2001; Remsen & O’Grady, 2002) e, embora sua subdivisão interna já tenha sido morfológica e citogeneticamente estabelecida (Nacur, 1958; Brncic *et al.*, 1971) esta não instigou, até o momento, qualquer tratamento molecular. O grupo *flavopilosa*, por outro lado, teve sua derivação exata questionada até mesmo por Throckmorton (1975), mas não recebeu, desde então, qualquer menção evolutiva mais profunda.

Dadas as circunstâncias acima mencionadas, o presente trabalho tem como objetivo principal contribuir na elucidação das relações filogenéticas dentro e dentre grupos Neotropicais do subgênero *Drosophila* que, subrepresentados em análises anteriores, pertençam tradicionalmente a ambas seções: grupos *immigrans*, *tripunctata*, *cardini*, *guarani* e *guaramunu*, incluídos na seção *quinaria-tripunctata*, e grupos *mesophragmatica* e *flavopilosa*, representantes da seção *virilis-repleta*. Além disso, pretende-se contribuir com questões relacionadas à monofilia do gênero e do subgênero *Drosophila*, bem como de algumas de suas radiações. Por fim, visa-se, ainda, fornecer algum subsídio às questões relacionadas ao correto posicionamento dos gêneros *Zaprionus*, *Samoaia* e *Liodrosophila* e dos subgêneros *Hirtodrosophila* e *Dorsilopha* dentro da família *Drosophilidae*. Para isto, 513 nucleotídeos da região codificante do gene *Amd* (alfa-metil dopa) são filogeneticamente analisados.

2. MATERIAIS E MÉTODOS

2.1. Espécies

A Tabela 1 lista os 49 taxa de Drosophilidae utilizadas no presente estudo, bem como o respectivo número de acesso à sua seqüência de *Amd* no GenBank (NCBI site). As novas seqüências obtidas neste trabalho encontram-se sublinhadas. A espécie *Scaptodrosophila latifasciaeformis* foi utilizadas como grupo-externo.

Tabela 1 – Lista de espécies de *Drosophila* e de outros Drosophilidae utilizados neste estudo:

GÊNERO	SUBGÊNERO	SEÇÃO	GRUPO	ESPÉCIE	NÚMERO DE ACESSO	
<i>Drosophila</i>	<i>Drosophila</i>	<i>quinaria-tripunctata</i>	<i>guarani</i>	<i>D. ornatifrons</i> (= <i>D. guarani</i> Dobzhansky e Pavan, 1943)	<u>AY699250</u>	
				<i>D. subbadia</i> Patterson e Mainland, 1943	<u>AY699251</u>	
				<i>D. griseolineata</i> Duda, 1927	<u>AY699257</u>	
				<i>D. angustibucca sensu</i> Frota-Pessoa, 1954 1	<u>AY699248</u>	
				<i>D. angustibucca sensu</i> Frota-Pessoa, 1954 2	<u>AY699249</u>	
				<i>D. mediopunctata</i> Dobzhansky e Pavan, 1943 4	<u>AY699254</u>	
				<i>D. mediopunctata</i> Dobzhansky e Pavan, 1943 22	<u>AY699255</u>	
				<i>D. crocina</i> Patterson e Mainland, 1944 = <i>D. mediotriata</i> (Vilela & Bächli, 1990)	<u>AY699252</u>	
				<i>D. mediotriata</i> Duda, 1925	<u>AY699253</u>	
				<i>D. bandeirantorum</i> Dobzhansky e Pavan, 1943	<u>AY699256</u>	
		<i>cardini</i>	<i>D. tripunctata</i> Loew, 1862 ^a	AF293709		
			<i>D. cardinoides</i> Dobzhansky e Pavan, 1943	<u>AY699258</u>		
			<i>D. neocardini</i> Streisinger, 1946	<u>AY699260</u>		
			<i>D. polymorpha</i> Dobzhansky e Pavan, 1943	<u>AY699259</u>		
			<i>testaceae</i>	<i>D. putrida</i> Sturtevant, 1916 ^a	AF293723	
				<i>quinaria</i>	<i>D. phalerata</i> Meigen, 1830 ^a	AF293721
					<i>immigrans</i>	<i>D. immigrans</i> Sturtevant, 1921 1
			<i>D. immigrans</i> Sturtevant, 1921 ^a 2	AF293713		
			<i>virilis-repleta</i>	<i>funebis</i>	<i>D. funebris</i> (Fabricius), 1787 ^a	AF293709
					<i>mesophragmatica</i>	<i>D. gasici</i> Brncic, 1957 AR
		<i>D. gasici</i> Brncic, 1957 CA		<u>AY699240</u>		
		<i>D. gasici</i> Brncic, 1957 CO		<u>AY699241</u>		
		<i>D. brncici</i> Hunter e Hunter, 1964		<u>AY699238</u>		
		<i>D. mesophragmatica</i> Duda, 1927		<u>AY699243</u>		
		<i>D. gaucha</i> Jaeger & Salzano, 1953 ^b		AF324955		
		<i>D. pavani</i> Brncic, 1957		<u>AY699245</u>		
		<i>D. viracochi</i> Brncic e Santibañez, 1957		<u>AY699244</u>		
		<i>repleta</i>		<i>D. hydei</i> Sturtevant, 1921 ^a	AF293712	
			<i>D. buzzatii</i> Patterson & Wheeler, 1942 ^b	AF324947		
			<i>D. mercatorum</i> Patterson & Wheeler, 1942 ^b	AF324957		
			<i>dreyfusi</i>	<i>D. camargoi</i> Dobzhansky & Pavan, 1950 ^b	AF324951	
				<i>D. canalinea</i> Patterson & Mainland, 1944 ^b	AF324952	
		<i>canalineae</i>	<i>flavopilosa</i>	<i>D. cestri</i> Brncic, 1978	<u>AY699246</u>	
				<i>D. incompta</i> Wheeler & Takada, 1962	<u>AY699247</u>	
				<i>D. aracatacas</i> ^b	AF324949	
		<i>annulimana</i>	<i>virilis</i>	<i>D. virilis</i> Sturtevant, 1916 ^a	AF293729	
				<i>D. robusta</i> Sturtevant, 1916 ^a	AF293724	
				<i>robusta</i>	<i>D. nannoptera</i> Wheeler, 1949 ^b	AF324959
					<i>nannoptera</i>	<i>D. simulans</i> Sturtevant, 1919 1
				<i>D. simulans</i> Sturtevant, 1919 ^a 2		AF293726
		<i>Sophophora</i>	<i>melanogaster</i>	<i>D. erecta</i> Tsacas & Lachaise, 1974 ^a	AF293708	
<i>D. pseudoobscura</i> Frolova, 1929 ^a	AF293722					
<i>D. busckii</i> Coquillett, 1901 ^a	AF293707					
<i>D. pictiventris</i> Duda, 1925 ^a	AF293711					
<i>D. aerea</i> Okada, 1956 ^a	AF293715					
<i>Dorsilopha</i>	<i>Hirtodrosophila</i>	<i>D. leonensis</i> Wheeler & Kambysellis, 1966 ^a	AF293725			
		<i>Z. indianus</i> Gupta, 1970	<u>AY699263</u>			
		<i>Z. tuberculatus</i> Malloch, 1932 ^a	AF293731			
<i>Liodrosophila</i>			<i>S. latifasciaeformis</i> Duda, 1940	<u>AY699264</u>		
<i>Samoia</i>						
<i>Zaprionus</i>						
<i>Scaptodrosophila</i>						

Obs.: 1) Seguiu-se, nesta tabela, a classificação de Wheeler (1982), exceto para *Scaptodrosophila*, classificada como um subgênero de *Drosophila* pelo mesmo autor, mas elevada à categoria de gênero por Grimaldi (1990).

2) Para alguma espécie, mais de uma população foi amostrada: *D. mediopunctata* 4 = Morro Santana, Porto Alegre, RS, 22 = Itapuã, Porto Alegre, RS; *D. immigrans* 1 = Morro Santana, Porto Alegre, RS, 2 = população utilizada por Tatarenkov *et al.*, 2001; *D. gasici* CA = Campeche, Bolívia, CO = Bogotá, Colômbia, AR = Arica, Chile; *D. simulans* 1 = Morro Santana, Porto Alegre, RS, 2 = população utilizada por Tatarenkov *et al.*, 2001. Para *D. angustibucca*, os números 1 e 2 referem-se a duas isolinhagens estabelecidas a partir da mesma população inicial, coletada em Itapuã, Porto Alegre, RS.

^a As seqüências de *Amd* das espécies assim demarcadas são do trabalho de Tatarenkov *et al.* (2001).

^b As seqüências de *Amd* das espécies demarcadas com esta simbologia são do trabalho de Tatarenkov & Ayala, (2001).

2.2. Manipulação do DNA

O DNA total de cada uma das isolinhagens estabelecidas foi isolado mediante a utilização do protocolo descrito por Jowett (1986), com algumas modificações. Os *primers* degenerados Amd-un2 e Amd-bw, previamente descritos por Tatarenkov *et al.* (2001), foram, então, utilizados na amplificação de um fragmento de tamanho variável do gene *Amd*, abrangendo parte do primeiro e do segundo éxons, de acordo com o esquema representado na Figura 1. Os componentes finais em cada reação de 50µl foram: 5µl de tampão de reação, 2,5mM de MgCl², 0,25mM de dNTPs, 20pM de cada *primer*, 1 U de *Taq* Polimerase (todos da Invitrogen®) e 100ng de DNA. As condições de amplificação variaram de acordo com os grupos em estudo, segundo critérios de otimização local. Os fragmentos obtidos por PCR foram, por fim, visualizados em gel de agarose 0,8%, quando as bandas de tamanho esperado foram cortadas e purificadas com o kit Qiaquick DNA-Gel Extraction System (Qiagen®), segundo as especificações do fornecedor.

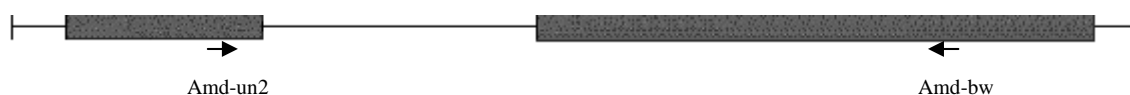


Fig. 1: Mapa genômico do gene *Amd*, cujos retângulos indicam os éxons e a linha intermediária representa o íntron. Os primers Amd-un2 e Amd-bw foram utilizados na amplificação de uma região do gene *Amd* abrangendo 92pb (de 464pb) do primeiro éxon, 991pb (de 1317pb) do segundo éxon, e mais um íntron de tamanho variável.

O seqüenciamento de DNA foi, então, realizado diretamente a partir dos produtos de PCR purificados, em um seqüenciador automático MegaBACE 500. A técnica de terminação de cadeia foi implementada mediante utilização do kit DYEnamic ET (Amersham®) e do *primer* Amd-bw.

Os produtos de PCR ainda foram, ainda, submetidos a uma eletroforese vertical longa, em gel de poliacrilamida 7%, para distinção dos pesos moleculares dos *amplicons* obtidos para as diferentes espécies. Estes géis foram analisados no programa LabImage 2.62a (Kapelan, Halle, Alemanha), mediante comparação com o padrão eletroforético apresentado pelo marcador 1 Kb Plus DNA Ladder (Invitrogen®).

2.4 – Análises Filogenéticas

As seqüências de nucleotídeos tiveram, inicialmente, sua identidade confirmada pelo uso do BlastN (NCBI site), sendo então alinhadas, no programa ClustalX 1.81 (Jeanmougin *et al.*, 1998), para cada uma das espécies em questão. Estes alinhamentos foram, em seguida, submetidos a uma análise de consenso de maioria, no programa GeneDoc versão 2.6.002 (Nicholas *et al.*, 1997). As seqüências consenso obtidas para todas as espécies foram, posteriormente, alinhadas pelo programa ClustalX 1.81, segundo os parâmetros “*default*” do sistema. Estas mesmas seqüências foram, ainda, submetidas a um teste de saturação no programa DAMBE (Xia & Xie, 2001).

As análises filogenéticas foram, finalmente, executadas mediante utilização de quatro métodos principais: 1) critério de Neighbor Joining (NJ) (Saitou & Nei, 1987), executado pelo MEGA 2.1 (Kumar *et al.*, 2001); 2) método de Máxima Parcimônia (MP), com busca heurística, realizado no PAUP 4.0b1 (Swofford, 1999); 3) método de Máxima Verossimilhança (MV), segundo o modelo indicado pelo ModelTest (Posada & Crandal, 1998) e mediante busca heurística com adição randômica de OTU’s, seguida de rearranjos de bissecção-reconexão da árvore (TBR – *tree bisection and reconnection*), executado no PAUP 4.0b1; 4) análise Bayesiana, segundo o modelo indicado pelo ModelTest, com avaliação de 1.000.000 de gerações, das quais 7.000 foram, ao final, consideradas, realizada através do programa MrBayes 3.0b4 (Huelsenbeck & Ronquist, 2001). A confiabilidade nas três primeiras análises foi medida pela utilização do teste de *bootstrap* (Felsenstein, 1985), com 1.000 replicações, enquanto que na análise bayesiana, a probabilidade posterior de cada um dos agrupamentos foi avaliada (Hall, 2001).

3. RESULTADOS E DISCUSSÃO

3.1. Análise filogenética das seqüências do gene *Amd*

O alinhamento das 49 seqüências do gene *Amd* mostrou 279 (de 513) sítios variáveis, dos quais 238 foram parcimoniosamente informativos. A razão média entre transições e transversões foi de 1,3, ou seja, mais de 2,5 vezes o esperado mediante uma distribuição randômica de substituições. As freqüências gerais dos nucleotídeos T, C, A e G foram de 25,2, 24,4, 20,2 e 30,2, respectivamente. Entretanto, este viés no conteúdo de G+C aumentou consideravelmente na terceira posição dos códons, quando atingiu limiares que chegaram a 75,1 em *D. griseolineata*. Realizando-se o teste de saturação (figura 2) de forma a avaliar a informação filogenética apresentada pelas seqüências (Schneider, 2002), pôde-se perceber que, embora na maior parte das substituições tanto transições quanto transversões permaneçam informativas, parte da curva de transições apresenta-se parcialmente saturada a partir de valores de cerca de 12% de divergência, o que fornece claros indícios da ocorrência de substituições múltiplas e homoplasia dentro desta classe de mutações. Entretanto, como a sua eliminação acarretaria a perda de informações fundamentais na resolução dos problemas propostos, apenas optou-se pela utilização de modelos evolutivos que considerassem as substituições múltiplas, bem como a razão entre transições e transversões e o conteúdo de G+C.

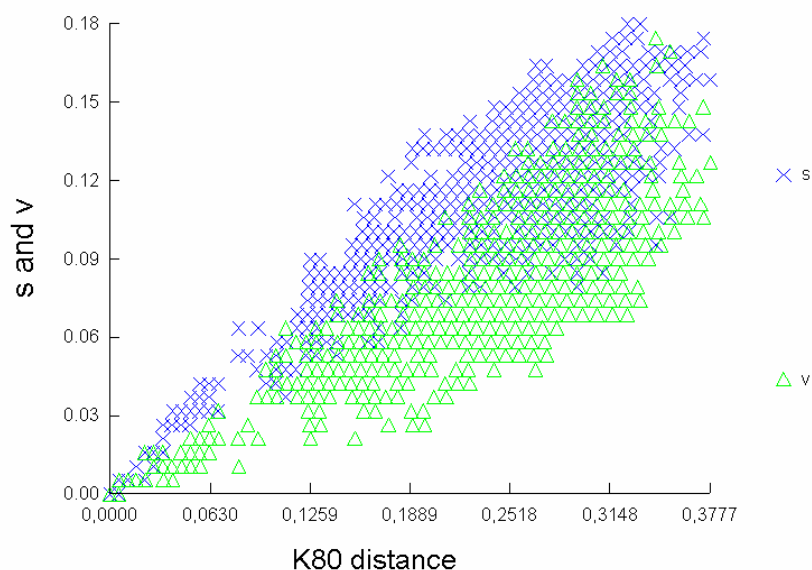


Fig. 2: Gráfico de transições (s) e transversões (v) versus a distância Kimura 2-Parâmetros (Kimura, 1980) do gene *Amd* em diferentes espécies de *Drosophilidae*.

Na análise por NJ a distância escolhida foi Tamura 3-Parâmetros (Tamura, 1992). Neste caso, as distâncias obtidas variaram desde 0,00 (para populações da mesma espécie) até 0,37 (entre *D. pseudoobscura* e *D. cardinoides*). Entretanto, como na maior parte das comparações os valores obtidos não ultrapassam o limiar de 0,30 [apresentado por Russo (2001) como sendo o limite a partir do qual as diferenças são demasiadamente subestimadas] optou-se por não utilizar as seqüências de aminoácidos. Já para as análises probabilísticas, o modelo indicado pelo teste de razão de verossimilhanças (LRT – *log-likelihood ratio test*), executado no ModelTest, foi o *General Time Reversible* (GTR) com considerações referentes à proporção de sítios invariáveis (I) e à distribuição gama dos sítios variantes (G). A análise de MP foi realizada segundo parâmetros não-ordenados, seguindo o modelo proposto por Fitch (1971).

A análise Bayesiana produziu 1.000.000 de árvores, das quais 6.153 foram consideradas na construção da árvore consenso apresentada na Figura 3, bem como no cálculo das probabilidades de cada ramo. As filogenias obtidas pelos demais métodos são essencialmente semelhantes, de forma que apenas sua confiabilidade, medida pelo teste de *bootstrap*, em relação aos clados apresentados nesta figura encontram-se indicados na Tabela 2. De modo geral, entretanto, como são poucos os agrupamentos divergentes entre as diferentes filogenias, embora o suporte estatístico de alguns destes clados não seja significativo (particularmente aqueles de nível mais inclusivo), sua reiteração pelos diferentes métodos de inferência podem ser tomados como um indício da alta consistência interna dos dados quando tomados em sua totalidade.

Como pode ser visto na Figura 3, ao utilizar-se *S. latifasciaeformis* como *outgroup*, o subgênero *Sophophora* posicionou-se na base da filogenia do gênero *Drosophila*, onde é precedido apenas pelo subgênero *Dorsilopha*, cujo único representante considerado, *D. busckii*, compôs uma politomia basal com *S. latifasciaeformis*. Embora o primeiro destes resultados seja altamente congruente com uma série de estudos anteriores (Throckmorton, 1975; Russo *et al.*, 1995; Tatarenkov *et al.*, 1999 e 2001; Kwiatowski & Ayala, 1999), o segundo é, no mínimo, inesperado, uma vez que a maior parte dos trabalhos tem posicionado *Dorsilopha* em clados relativamente mais derivados dentro do gênero, como divergindo depois de *Sophophora* (Throckmorton, 1975; Tatarenkov *et al.*, 1999 e 2001; Kwiatowski & Ayala, 1999).

Dado o posicionamento basal do subgênero *Sophophora*, estes mesmos trabalhos têm, por conseguinte, apontado que o subgênero *Drosophila* é filogeneticamente mais próximo de diversos outros subgêneros de *Drosophila*, e até mesmo de outros gêneros de Drosophilidae, do que do próprio subgênero *Sophophora*, o que é corroborado pelo presente estudo. Mais do que isso, na Figura 3, por exemplo, o subgênero *Hirtodrosophila*, bem como os gêneros *Zaprionus*, *Samoia* e *Liodrosophila* agrupam-se internamente ao subgênero *Drosophila*, resultados estes, muitos dos quais já anteriormente apresentados (Throckmorton, 1975; Kwiatowski & Ayala, 1999; Tatarenkov *et al.*, 1999, 2001), que transformam o subgênero *Drosophila*, juntamente com o seu próprio gênero, em agrupamentos parafiléticos, não naturais, de espécies. O subgênero *Drosophila* se subdivide, ainda, em dois clados principais: o primeiro constituído por membros da radiação *virilis-repleta* e o segundo incluindo os representantes da radiação *immigrans-Hirtodrosophila* com adição de *D. funebris*, ambos com constituições extremamente semelhantes àquelas apresentadas por Throckmorton (1975).

Na radiação *virilis-repleta* há que se destacar, pelo alto suporte obtido, a confirmação da monofilia do grupo *mesophragmatica*. Este grupo subdividiu-se internamente em três clados principais: o primeiro agrupando *D. gasici*, *D. brncici* e *D. mesophragmatica*; o segundo reunindo as espécies crípticas, *D. pavani* e *D. gaucha*; e o terceiro, mais basal e monotípico, constituído apenas por *D. viracochi*. Esta composição é altamente consistente com os subgrupos apresentados por Brncic *et al.* (1971) mediante análise citogenética, da mesma forma que o é a parafilia demonstrada pela espécie *D. gasici*, a qual é devida, provavelmente, a uma condição de elevado polimorfismo ancestral. Como grupo-irmão de *mesophragmatica* situa-se o grupo *repleta*, um resultado parcialmente congruente com algumas análises anteriores (Durando *et al.*, 2000; Tatarenkov & Ayala, 2001), mas que discorda dos achados de Remsen & O'Grady (2002). A este clado segue-se, sucessivamente, o agrupamento de *D. camargoi* (grupo *dreyfusi*) e de *D. canalinea* (grupo *canalinea*), de forma a compor uma radiação *repleta* monofilética, o que também já fora previamente detectado por Tatarenkov & Ayala (2001). Como clado-irmão desta radiação aparece o agrupamento constituído pelo grupo *flavopilosa* e *D. aracatacas* (grupo *annulimana*), que embora tenham sido situados como basais em relação aos demais membros da radiação *virilis-repleta* por Throckmorton (1975), ocupam uma posição relativamente derivada neste trabalho. *D. nannoptera* (grupo *nannoptera*) e o clado

constituído por *D. virilis* (grupo *virilis*) e *D. robusta* (grupo *robusta*) compõem, neste trabalho, os primeiros ramos a divergir dentro desta radiação.

Já dentro da radiação *immigrans-Hirtodrosophila* é o gênero *Zaprionus*, seguido pelo clado formado por *Samoia leonensis* e *Hirtodrosophila pictiventris*, que constituem as ramificações mais basais. *D. immigrans*, outro grupo indicado por Throckmorton (1975) como constituindo a base desta radiação, é o próximo clado a ramificar, no que é seguido por *Liodrosophila aerea*, *D. phalerata* (grupo *quinaria*), *D. putrida* (grupo *testaceae*) e *D. funebris* (grupo *funebris*). Este último grupo, juntamente com *Liodrosophila*, foi indicado por Throckmorton (1975) como ocupando uma posição basal em relação a ambas as radiações, o que não é confirmado pelo presente estudo, nem pelos achados de Tatarenkov *et al.* (2001), de Remsen & O'Grady (2002) e de Yotoko *et al.* (2003). O que resta, dentro deste clado maior que constitui a radiação *immigrans-Hirtodrosophila*, é uma radiação *tripunctata* monofilética, composta por um grupo *tripunctata* inteiramente parafilético em relação aos grupos *cardini*, *guarani* e *guaramunu*. Esta parafilia do grupo *tripunctata* já fora previamente apontada tanto por Frota-Pessoa (1954) e Throckmorton (1975), quanto por Yotoko *et al.* (2003). Neste último trabalho, entretanto, a radiação *tripunctata* foi parafileticamente definida com relação aos grupos *funebris* e *testaceae*, clados estes que compõem o grupo-irmão da radiação *tripunctata* no presente estudo.

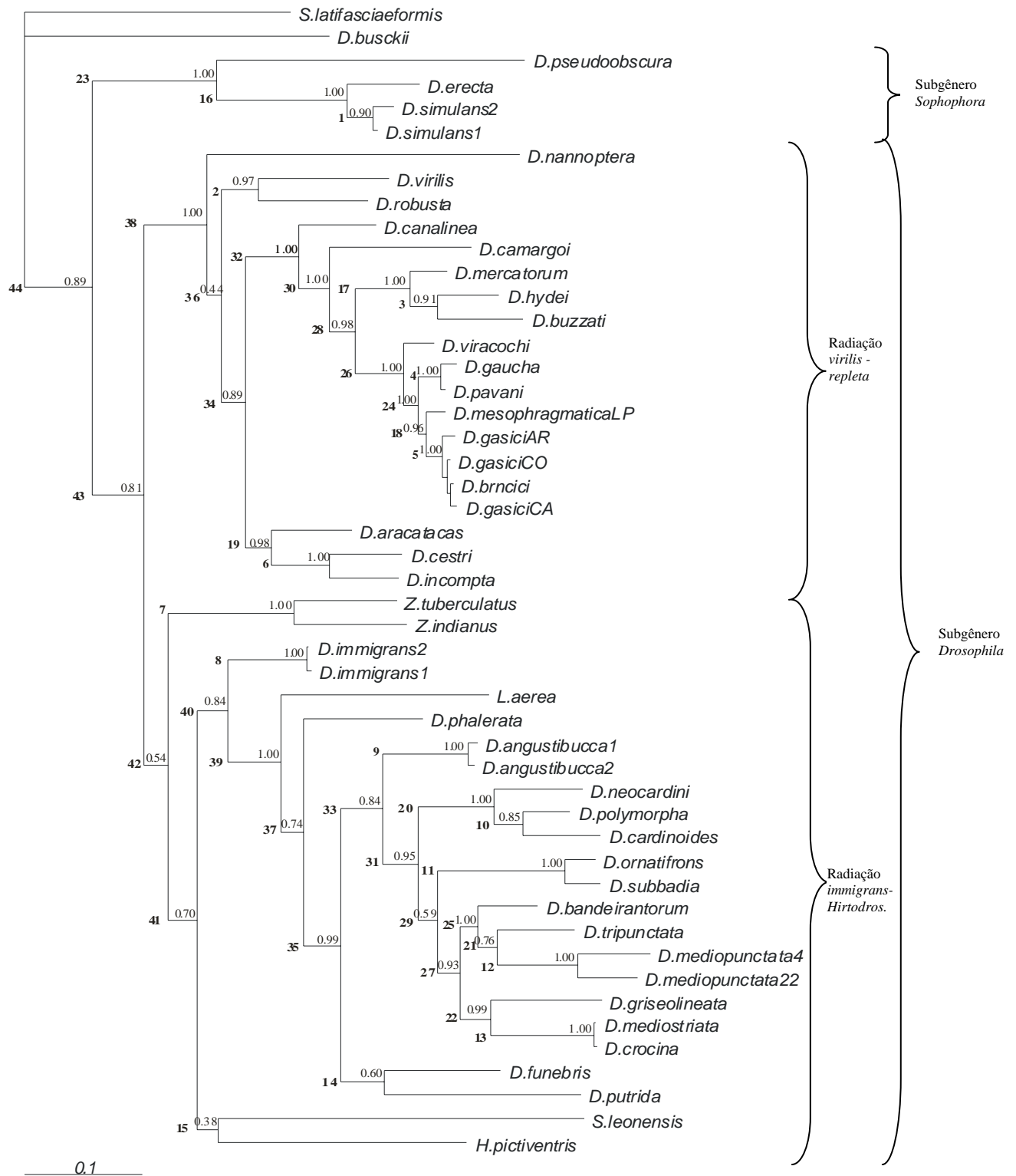


Fig. 3: Filogenia consenso obtida através de análise Bayesiana mediante utilização do modelo GTR + I + G, como proposto pelo LRT, para 49 seqüências de *Amd*. A probabilidade de cada clado encontra-se indicada acima do seu respectivo ramo interno. Ao lado dos ramos internos encontra-se, ainda, um número, arbitrariamente definido, que representa o agrupamento que o segue, para fins de interpretação da tabela 02.

Tabela 2 – Valores de *bootstrap* obtidos para cada um dos clados apresentados na Figura 3 mediante utilização de diferentes métodos de reconstrução filogenética:

Clado	Método de Reconstrução		
	Neighbor Joining	Máxima Parcimônia	Máxima Verossimilhança
1	100	97	51
2	69	49	42
3	75	64	56
4	100	97	94
5	71	94	98
6	99	97	95
7	99	98	99
8	100	100	99
9	100	100	99
10	98	68	71
11	100	100	99
12	100	99	99
13	100	100	100
14	32	-	39
15	-	-	-
16	100	100	97
17	99	92	91
18	75	83	70
19	82	48	51
20	99	95	84
21	60	25	42
22	36	36	61
23	95	81	93
24	54	86	73
25	64	30	41
26	100	99	90
27	26	30	11
28	76	65	56
29	-	20	-
30	-	51	46
31	-	27	23
32	91	75	52
33	71	49	28
34	59	35	30
35	-	27	38
36	39	29	-
37	66	40	30
38	72	40	44
39	25	22	28
40	36	9	20
41	-	7	-
42	-	-	-
43	32	18	13
44	78	37	42

Particularmente interessante no trabalho de Yotoko *et al.* (2003), realizado com base na análise filogenética do gene mitocondrial COII, foi a ausência de uma resolução bem suportada estatisticamente nas relações intergrupais e interespecíficas dentro da radiação *tripunctata*, cujos ramos internos apresentaram-se bastante curtos. No presente trabalho, principalmente quando os valores de *bootstrap* obtidos pelas demais análises (tabela 02) são considerados [já que as probabilidades da análise bayesiana podem levar a superestimativas da confiabilidade dos clados (Hall, 2001)], um padrão bastante semelhante torna-se evidente, o que dá suporte à visão, apresentada primeiramente por Throckmorton (1975), de que estes grupos surgiram a partir de episódios de especiação rápida e múltipla. Talvez seja exatamente por isso que não tenha sido possível, por fim, encontrar maiores similaridades entre os resultados morfológicos obtidos por Frota-Pessoa (1954) e a composição interna da radiação *tripunctata* aqui apresentada. Esta subdividiu-se em 5 clados principais, citados em ordem ascendente a partir do mais primitivo: o primeiro composto apenas por *D. angustibucca*, representante do subgrupo I de Frota-Pessoa (1954); o segundo apresentando um grupo *cardini* monofilético para as três espécies consideradas, cuja subdivisão interna discorda dos *clusters* apresentados por Heed & Russel (1971) para o subgrupo *cardini*, uma vez que agrupa *D. polymorpha* com *D. cardinoides*, ao invés de reuni-la com *D. neocardini*; o terceiro constituído por dois representantes do grupo *guarani*: *D. subbadia* e *D. ornatifrons*; o quarto reunindo três representantes do grupo *tripunctata*: *D. bandeirantum*, *D. tripunctata* e *D. mediopunctata*, situados, respectivamente, nos subgrupos III, IV e II de Frota-Pessoa (1954); e o quinto constituído por uma espécie do grupo *guaramunu* e uma espécie do grupo *tripunctata* (*D. mediotriata*, sendo que *D. crocina* = *D. mediotriata*, segundo Vilela & Bächli, 1990), membro do subgrupo III de Frota-Pessoa (1954).

3.2. Evolução do tamanho molecular do gene *Amd* ao longo da filogenia proposta

Como já mencionado anteriormente, o tamanho molecular dos fragmentos do gene *Amd* gerados por PCR variaram intensamente entre os grupos e espécies consideradas, devido à presença de um íntron na região amplificada. Esta variação pode ser visualizada na figura 04, que apresenta o padrão eletroforético, em géis de poliacrilamida, apresentado pelas diferentes espécies.

Ao interpretar-se estes géis dentro de uma perspectiva filogenética pode-se, facilmente, perceber que há uma forte tendência evolutiva de incremento no tamanho do

íntron de *Amd* dentro da radiação *immigrans-Hirtodrosophila*, cujos pesos moleculares do fragmento obtido por PCR variaram desde 1,55 kb em *Zaprionus indianus* até mais de 3 kb em *D. bandeirantorum* e *D. mediopunctata*. Mais do que isso, uma espécie pertencente à esta radiação, *D. paramediostriata*, obteve um padrão eletroforético com dois fragmentos intensamente amplificados, possivelmente devido a uma duplicação do gene *Amd*, hipótese esta que não pôde, entretanto, ser confirmada até o momento. Este padrão de variabilidade interna é contrastado pela alta homogeneidade apresentada pelos membros da radiação *virilis-repleta*, no que tange ao peso molecular do fragmento de *Amd*, que atinge um tamanho médio de 1,4 kb em todas suas espécies. Se o que tem causado esta variação entre os membros da radiação *immigrans-Hirtodrosophila* é a ação de elementos transponíveis ou a inserção de qualquer outro tipo de material genético, permanece, pois, por ser verificado, assim como o significado evolutivo destas alterações.

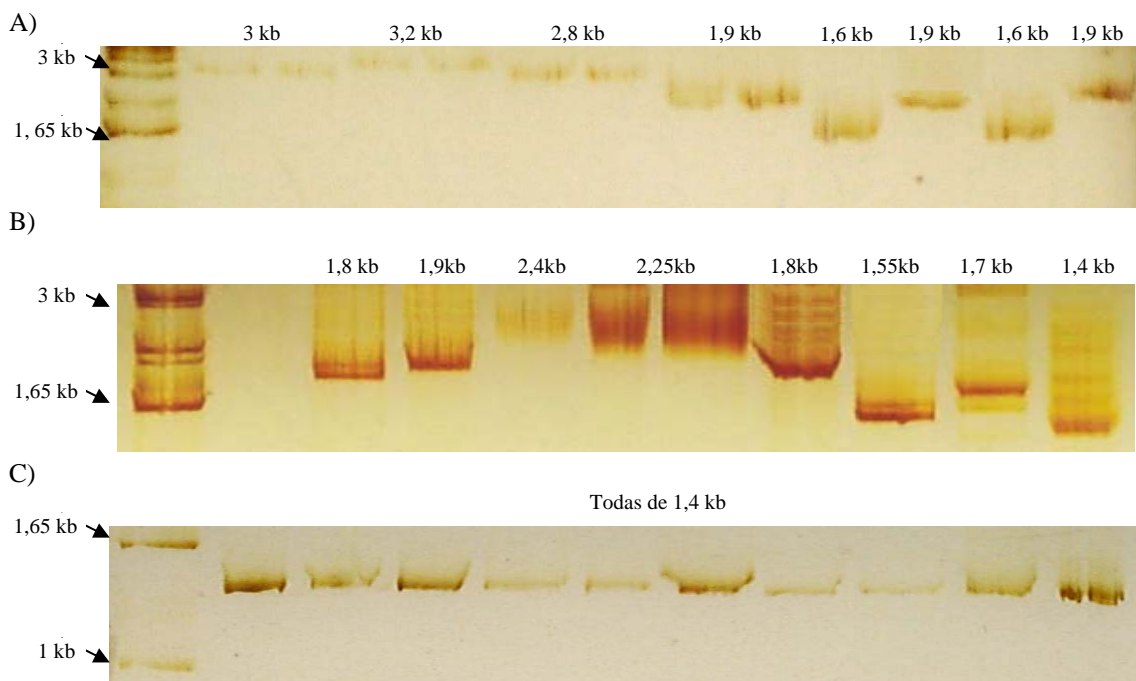


Fig. 04: Géis de Poliacrilamida demonstrando o padrão eletroforético do fragmento de *Amd* apresentado pelas diferentes espécies consideradas neste estudo, cujo tamanho aproximado encontra-se indicado acima da figura. Da esquerda para a direita: A) marcador de peso molecular 1 Kb Ladder Plus, *D. bandeirantorum*, *D. bandeirantorum*, *D. mediopunctata22*, *D. mediopunctata4*, *D. crocina*, *D. mediostriata*, *D. angustibucca1*, *D. angustibucca2*, fragmento menor de *D. paramediostriata*, fragmento maior de *D. paramediostriata*, fragmento menor de *D. paramediostriata*, fragmento maior de *D. paramediostriata*; B) marcador de peso molecular 1 Kb Ladder Plus, *D. griseolineata*, *D. ornatifrons*, *D. subbadia*, *D. cardinoides*, *D. neocardini*, *D. polymorpha*, *D. immigrans*, *Z. indianus*, *D. simulans*, *S. latifasciaeformis*; C) marcador de peso molecular 1 Kb Ladder Plus, *D. brncici*, *D. gasiciAR*, *D. gasiciCA*, *D. gasiciCO*, *D. mesophragmatica*, *D. mesophragmatica*, *D. pavani*, *D. viracochi*, *D. cestri*, *D. incompta*.

4. CONCLUSÕES

Taxonomicamente, a família Drosophilidae é um aglomerado de táxons merofiléticos, a começar pelo próprio gênero *Drosophila* e pelo seu subgênero mais diverso, o subgênero *Drosophila*. Gêneros emergindo a partir de subgêneros são um quadro comum na taxonomia do grupo, como pôde aqui ser verificado. Porém, apesar do acúmulo de evidências neste sentido, com o qual o presente trabalho vem contribuir, poucas foram as tentativas feitas até o momento, com o intuito de redimensionar a nomenclatura desta família. Entre as exceções, pode-se destacar o memorável trabalho de Grimaldi (1990), que propôs uma série de mudanças taxonômicas a partir de seus próprios resultados. Entretanto, como grande parte de suas conclusões não têm sido corroboradas pelas análises moleculares posteriores, muitas de suas proposições ficaram apenas na boa intenção.

Raros são os casos em que uma nova tendência taxonômica foi trazida à tona devido às suas incongruências com a filogenia de um grupo. Entre estes pode-se citar a elevação de categoria do antes subgênero *Scaptodrosophila*, proposta por Grimaldi (1990); a subdivisão do antigo subgrupo *obscura* em três novos subgrupos do grupo *obscura*, proposta desde Lakovaara & Saura (1982) até Barrío *et al.* (1994); e o redimensionamento dos antigos subgrupos do grupo *guarani* para os atuais grupos *guaramunu* e *guarani*, proposta desde Kastritsis (1969). É, entretanto, urgente, que novas revisões sejam feitas em inúmeros outros grupos artificiais, como é o caso do próprio grupo *tripunctata*, cujo caráter parafilético, aqui corroborado, já vem sendo apontado desde Frota-Pessoa (1954), que apresentou seus subgrupos I e III como sendo transicionais aos grupos *guarani* e *cardini*, respectivamente.

AGRADECIMENTOS

Os autores agradecem enormemente a colaboração do MSc. Marco Silva Gottshalk, pela gentil identificação de diversas espécies aqui utilizadas, e aos MSc. Liliana Essi e Paulo Pinto, pelo valiosos préstimos nos momentos de dificuldades. Ao CNPq, pela concessão de bolsas, e ao CNPq, FAPERGS e PROPESQ-UFRGS, pelos recursos de financiamento à pesquisa.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- BARRIO, E.; LATORRE, A.; MOYA, A. (1994) Phylogeny of the *Drosophila obscura* species group deduced from mitochondrial DNA sequences. *J. Mol. Evol.* 39: 478-488.
- BRNCIC, D; NAIR, P. S.; WHEELER, M. R. (1971) Cytotaxonomic relationships within the *mesophragmatica* species group of *Drosophila*. *Univ. Texas Publ.* 7103: 1-16.
- DESALLE, R. (1992). The phylogenetic relationships of flies in the family Drosophilidae deduced from mtDNA sequences. In: *Mol. Phyl. Evol.* 1: 31-40.
- DURANDO, C. M.; BAKER, R. H.; ETGES, W. J.; HEED, W. B.; WASSERMAN, M.; DESALLE, R. (2000) Phylogenetic analysis of the *repleta* species group of the genus *Drosophila* using multiple sources of characters. *Molecular Phylogenetics and Evolution* 16: 296-307.
- FELSENSTEIN, J. (1985) Confidence limits on phylogenies: an approach using bootstrap. *Evolution* 39: 783-791.
- FITCH, W. M. (1971) Toward defining the course of evolution: Minimal change for a specific tree topology. *Syst. Zool.* 20: 406-416.
- FROTA-PESSOA, O. (1954) Revision of the *tripunctata* group of *Drosophila* with description of fifteen new species. *Arq. Mus. Paranaense, Curitiba* 10: 253-304.
- GRIMALDI, D. A. (1990) A phylogenetic, revised classification of the genera in the Drosophilidae (Diptera). In: *Bull. Am. Mus. Nat. Hist.* 197: 1-139.
- HALL, B. G. (2001) *Phylogenetic trees made easy – A how to manual for molecular biologists.* 2. Ed. Massachusetts : Sinauer Associates.
- HEED, W. B.; RUSSELL, J. S. (1971) Phylogeny and population structure in island and continental species on the *cardini* group of *Drosophila* studied by inversion analysis. *Univ. Texas Publ.* 7103: 91-130.
- HUELSENBECK, J. P.; RONQUIST, F. (2001) MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754-755.
- JEANMOUGIN, F.; THOMPSON, J. D.; GOUY, M.; HIGGINS, D. G.; GIBSON, T. J. (1998) Multiple Sequence Alignment with ClustalX. *Trends Biochem. Sci.* 23: 403-405.
- JOWET, T. (1986) Preparation of nucleic acids, in *Drosophila: a practical approach*, edited by D. B. Roberts. Oxford: IRL Press.

- KASTRITSIS, C. D. (1969) The chromosomes of some species of the guarani group of *Drosophila*. In: J. Hered. 60: 50-57.
- KIMURA, M. (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16:111-120
- KUMAR, S.; TAMURA, K.; JAKOBSEN, I. B.; NEI, M. (2001) MEGA2: Molecular Evolutionary Genetics Analysis Software. Bioinformatics.
- KWIATOWSKI, J.; AYALA, F. J. (1999) Phylogeny of *Drosophila* and Related Genera: Conflict between Molecular and Anatomical Analyses. In: Mol. Phyl. Evol. 13(2): 319-328.
- LAKOVAARA, S.; SAURA, A. (1982) Evolution and speciation in the *Drosophila obscura* group. In: ASHBURNER, M.; CARSON, H. L.; THOMPSON, J. N. (Eds.) The genetics and biology of *Drosophila*. New York: Academic Press.
- MALOGOLOWKIN, C. (1953) Sobre a Genitália dos Drosofilídeos. IV. A genitália masculina no subgênero *Drosophila* (Diptera, Drosophilidae). Rev. Brasil. Biol. 13: 245-264.
- NACRUR, J. (1958) Genitália masculina de *Drosophila* do grupo *mesophragmatica* (Diptera). Ver. Brasil. Biol. 18: 243-249.
- NICHOLAS, K. B.; NICHOLAS, H. B. JR. (1997) GeneDoc: a tool for editing and annotating multiple sequence alignments. Distributed by the authors.
- NURMINSKY, D. I.; MORIYAMA, E. N.; LOZOVSKAYA, E. R.; HARTL, D. L. (1996) Molecular Phylogeny and Genome Evolution in the *Drosophila virilis* Species Group: Duplications of the *Alcohol Dehydrogenase* Gene. Mol. Biol. Evol. 13: 132-149.
- POSADA, C.; CRANDALL, K. A. (1998) MODELTEST: testing the model of DNA substitution. Bioinformatics 14: 817-818.
- POWELL J. R.; DESALLE, R. (1995) *Drosophila* Molecular Phylogenies and Their Uses. In: Evolutionary Biology, 28: 87-138.
- REMSEM, J.; O'GRADY, P. (2002) Phylogeny of Drosophilinae (Diptera: Drosophilidae), with comments on combined analysis and character support. Mol. Phyl. Evol. 24: 249-264.
- RUSSO, C. A. M.; TAKEZAKI, N.; NEI, M. (1995) Molecular Phylogeny and Divergence Times of Drosophilid Species. Mol. Biol. Evol. 12(3): 391-404.

- RUSSO, C. A. DE M. (2001) Como Escolher Genes para Problemas Filogenéticos Específicos. In: MATIOLI, S. R. (ed.) *Biologia Molecular e Evolução*. Ribeirão Preto: Holos, p. 130-136.
- SAITOU, N.; NEI, M. (1987) The Neighbor-Joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 9: 406-425.
- SWOFFORD, D. L. (1999) PAUP: Phylogenetic Analysis using Parsimony (and other methods), versão 4.0. Sinauer Associates, Sunderland, Massachusetts, USA.
- TAMURA, K. (1992) Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G + C-content biases. *Mol. Biol. Evol.* 9: 678–687.
- TATARENKOV, A.; KWIATOWSKI, J.; SKARECKY, D.; BARRIO, E.; AYALA, F. J. (1999) On the evolution of *Dopa decarboxylase (Ddc)* and *Drosophila* Systematics. *J. Mol. Evol.* (48) : 445-462.
- TATARENKOV, A.; ZUROVCOVÁ, M.; AYALA, F. J. (2001) *Ddc* and *Amd* Sequences Resolve Phylogenetic Relationships of *Drosophila*. *Mol. Phyl. Evol.* 20(2) : 321-325.
- TATARENKOV, A.; AYALA, F. J. (2001) Phylogenetic Relationships among Species Groups of the *virilis-repleta* Radiation of *Drosophila*. In: *Mol. Phyl. Evol.* 21(2): 327-331
- THROCKMORTON, L. H. (1975) The Phylogeny, Ecology and Geography of *Drosophila*. V. 3. In: *Handbook of Genetics*. New York: Plenum.
- VAL, F. C.; MARQUES, M. D.; VILELA, C. R. (1981) *Drosophilidae* of Neotropical Region. In: ASHBURNER, M.; CARSON, H. L.; THOMPSON, J. N., J. R. (Eds). *The Genetics and Biology of Drosophila*. V. 3a. Orlando, Flórida: Academic Press.
- VILELA, C. R.; BÄCHLI, G. (1990) Taxonomic studies on Neotropical species of seven genera of *Drosophilidae* (Diptera). *Mitteilungen der Schweizerischen Entomologischen Gesellschaft Bulletin de la Société Entomologique Suisse* 63 (Suppl.): 1-332.
- VILELA, C. R.; BÄCHLI, G. (2000) Five new species of Neotropical *Drosophila* (Diptera, *Drosophilidae*). *Mitteilungen der Schweizerischen Entomologischen Gesellschaft Bulletin de la Société Entomologique Suisse* 73: 49-65.
- XIA, X.; XIE, Z. (2001) DAMBE: Data analysis in molecular biology and evolution. *Journal of Heredity* 92: 371-373.

- YOTOKO, K. S. C.; MEDEIROS, H. F.; SOLFERINI, V. N.; KLACZKO, L. B. (2003) A molecular study of the systematics of the *Drosophila tripunctata* group and the *tripunctata* radiation. *Mol. Phyl. Evol.* 28: 614-619.
- WHEELER, M. R. (1986) The Drosophilidae: a taxonomic overview. In: ASHBURNER, M.; CARSON, H. L.; THOMPSON, J. N., J. R. (Eds.). *The Genetics and Biology of Drosophila*. V. 3a. Orlando, Flórida: Academic Press.

CAPÍTULO III

CONCLUSÃO

As hipóteses desenvolvidas pelo presente trabalho encontram-se, certamente, entre as primeiras a envolver uma abordagem molecular para grande parte das espécies e grupos considerados. O grupo *tripunctata*, por exemplo, fora anteriormente avaliado apenas com base em caracteres morfológicos (Frota-Pessoa, 1954; Throckmorton, 1975), com exceção da análise molecular recentemente executada por Yotoko *et al.* (2003). Outros pesquisadores, em análises de nível mais inclusivo, incluíram apenas uma de suas espécies (*D. tripunctata*) (Tatarenkov *et al.*, 2001; Remsen & O'Grady, 2002), negligenciando totalmente aqueles representantes com distribuição primariamente Neotropical. Outros grupos aqui retratados, como os próprios grupos *cardini*, *guarani* e *guaramunu* também têm sido largamente subrepresentados em análises anteriores, sendo sua composição interna molecularmente retratada apenas por Yotoko *et al.* (2003). A inclusão do grupo *flavopilosa* em uma abordagem molecular é, sem dúvida, inédita, da mesma forma que o é uma análise mais profunda das relações intra-grupais do grupo *mesophragmatica*.

Como esta análise visava, ainda, fazer algumas inferências de nível mais inclusivo, referentes, por exemplo, ao status taxonômico do próprio gênero e do subgênero *Drosophila*, optou-se pela utilização de um gene medianamente conservado, cuja variabilidade já tinha se mostrado adequada para estes níveis de divergência: o gene *Amd* (alfa metil-dopa) (Tatarenkov & Ayala, 2001; Tatarenkov *et al.* 2001). Embora alguns indícios de saturação tenham sido, inicialmente, detectados, fazendo necessária a sua correção mediante o uso de modelos evolutivos mais complexos (Nei & Kumar, 2000), a escolha deste gene se mostrou particularmente adequada. A consistência e forte congruência interna apresentada pelas mais diversas análises executadas, que foram desde o critério da Máxima Parcimônia até a recente Análise Bayesiana, permitem, pois, a atribuição de uma maior confiabilidade aos resultados obtidos.

A árvore apresentada pelo presente trabalho assemelhou-se, em muitos aspectos, ao esquema apresentado por Throckmorton (1975), como pode ser visualizado na Figura 3 do seu Capítulo II. Neste sentido, ela expôs um gênero *Drosophila* parafilético com relação aos gêneros *Zaprionus*, *Liodrosophila* e *Samoia*, bem como um subgênero *Drosophila* parafilético com relação a estes gêneros e ao subgênero *Hirtodrosophila*. O subgênero *Drosophila* também adquiriu, nesta análise, uma posição derivada com relação ao subgênero *Sophophora*, dividindo-se, internamente, nas duas radiações propostas por

Throckmorton (1975): radiação *virilis-repleta* e radiação *immigrans-Hirtodrosophila*, cujas composições também assemelham-se enormemente àquelas apresentadas por este autor. Nesta última radiação, particularmente, o grupo *tripunctata* apresentou, como também já adiantado por Throckmorton, um caráter parafilético com relação aos grupos *cardini*, *guarani* e *guaramunu*, de forma a compor, ainda, uma radiação *tripunctata* monofilética para as espécies consideradas.

A única região da presente filogenia cuja composição não concordou com qualquer análise prévia diz respeito, exatamente, às relações internas à radiação *tripunctata*. Analisando-se o suporte estatístico destes ramos, principalmente os valores de *bootstrap* apresentados na Tabela 2 do Capítulo II, pode-se perceber que, estatisticamente, esta é a região menos suportada da filogenia, cujos ramos, na análise por NJ destacam-se também como sendo os mais curtos. Considerando-se que este não é um problema que possa ser atribuído ao gene, uma vez que o mesmo foi capaz de resolver outras questões com nível de divergência similar, pode-se encontrar uma explicação plausível nos próprios escritos de Throckmorton (1962), recentemente aludidos por Yotoko *et al.* (2003), que relacionou a radiação *tripunctata* a eventos de especiação rápida e múltipla, ocorridos a partir de uma população ancestral grande e polimórfica.

Se esta hipótese for, realmente, confirmada, o estabelecimento das relações evolutivas entre os representantes da radiação *tripunctata* deve ser, futuramente, um dos maiores problemas dentro da sistemática de *Drosophila*, ávida por uma revisão taxonômica mais profunda também em outros táxons, como pôde aqui ser demonstrado. É claro que estas alterações causariam um grande impacto na comunidade científica, principalmente de Drosofilistas, mexendo com as bases de inúmeros estudos anteriores. Entretanto, tornou-se bastante complicado trabalhar com *Drosophila* em meio a esta Torre de Babel que virou a sua taxonomia. É preciso, pois, que a Sistemática Molecular de Drosophilidae se desenvolva ainda mais e que, mediante o acúmulo inquestionável de evidências em um determinado sentido, Drosofilistas do mundo inteiro se reúnam na elaboração de uma nova e fundamentada taxonomia para o grupo, de forma que esta venha a ser apenas mais um fator determinante, entre os inúmeros outros, na manutenção de *Drosophila* em seu ranking de organismo modelo ideal.

REFERÊNCIAS BIBLIOGRÁFICAS

- AMORIM, D. S. (1993) Elementos de Sistemática Filogenética. 2. Ed. Ribeirão Preto: Holos.
- AMORIM, D. S. (1997) Elementos de Sistemática Filogenética. 2. Ed. Ribeirão Preto: Holos.
- ARIAS, M. C.; INFANTE-MALACHIAS, M. E. (2001) RFLP: O Emprego de enzimas de restrição para a detecção de polimorfismos de DNA. In: MATIOLI, S. R. (ed.) *Biologia Molecular e Evolução*. Ribeirão Preto: Holos, p. 143-152.
- ASHBURNER, M.; BODMER, M.; LAMEUNIER, F. (1984) On the evolutionary relationships of *Drosophila melanogaster*. *Dev. Genet.* 4: 295-312.
- AVISE, J. C.; NEIGEL, J. E.; ARNOLD, J. (1984) Demographic influences on mitochondrial DNA lineage survivorship in animal populations. *J. Mol. Evol.* 20: 99-105.
- AYALA, F. J.; TRACEY, M. L.; HEDGECOCK, D.; RICHMOND, R. C. (1974) Genetic differentiation during the speciation process in *Drosophila*. *Evolution* 28: 24-41.
- BÄCHLI, G.; VILELA, C. R.; RATCOV, V. (2000) Morphological differences among *Drosophila paraguayensis* Duda, 1927 and its close relatives (Díptera, Drosophilidae). *Mitteilungen der Schweizerischen Entomologischen Gesellschaft Bulletin de la Société Entomologique Suisse* 73: 67-92
- BARRIO, E.; LATORRE A.; MOYA, A.; AYALA, F. J. (1992) Phylogenetic Reconstruction of the *Drosophila obscura* Group, on the Basis of Mitochondrial DNA. *Mol. Biol. Evol.* 9(4): 621-635.
- BARRIO, E.; LATORRE, A.; MOYA, A. (1994) Phylogeny of the *Drosophila obscura* species group deduced from mitochondrial DNA sequences. *J. Mol. Evol.* 39: 478-488.
- BARRIO, E.; AYALA, F. J. (1997) Evolution of the *Drosophila obscura* species group inferred from the *Gpdh* and *Sod* Genes. *Mol. Phyl. Evol.* 7(1): 79-93.
- BRNCIC, D.; KOREF SANTIBAÑEZ, S. (1957) The *mesophragmatica* group of species of *Drosophila*. *Evolution* 11: 300-310.

- BRNCIC, D. (1958) The *mesophragmatica* group as an example of the speciation phenomena in *Drosophila*. Proceedings of the X International Congress of Genetics 1: 420-433.
- BRNCIC, D.; NAIR, P. S.; WHEELER, M. R. (1971) Cytotaxonomic relationships within the *mesophragmatica* species group of *Drosophila*. Univ. Texas Publ. 7103: 1-16.
- BRNCIC, D. (1978) A note on the *flavopilosa* group of species of *Drosophila* in Rio Grande do Sul, Brazil, with the description of two new species (Díptera, Drosophilidae). Revista Brasileira de Biologia 38(3): 647-651.
- CALCAGNOTTO, D. (2001) Taxas de Evolução e o Relógio Molecular. In: MATIOLI, S. R. (ed.) Biologia Molecular e Evolução. Ribeirão Preto: Holos, p. 51-63.
- CUNNINGHAM, C. W. (1997) Can three incongruence tests predict when data should be combined? Mol. Biol. Evol. 14: 733-740.
- CUTLER, D. J. (2000) Estimating divergence times in the presence of overdispersed molecular clock. Mol. Biol. Evol. 17: 1647-1660.
- DESALLE, R. (1992a). The phylogenetic relationships of flies in the family Drosophilidae deduced from mtDNA sequences. In: Mol. Phyl. Evol. 1: 31-40.
- DESALLE, R. (1992b) Origin and possible time of divergence of the Hawaiian Drosophilidae: evidence from DNA sequences. Mol. Biol. Evol. 9(5): 905-916.
- DESALLE, R.; GRIMALDI, D. (1992) Characters and the Systematics of Drosophilidae. Journal of Heredity, 83: 182-188.
- DINIZ-FILHO, J. A. (2000) Métodos Filogenéticos Comparativos. Ribeirão Preto: Holos.
- DOBZHANSKY, T.; PAVAN, C. (1943) Studies on Brazilian species of *Drosophila*. Bol. Facul. Fil. Cien. Letr. Univ. S. Paulo 36.
- DOOLITTLE, W. F. (1999) Lateral genomics. TIG 15(12): 5-8.
- DOWLING, T. E.; MORITZ, C.; PALMER, J. D.; RIESEBERG, L. H. (1996) Nucleic Acids III: Analysis of Fragments and Restriction Sites. In: HILLIS, D. M.; MORITZ, C.; MABLE, B. K. Molecular Systematics. 2. Ed. Sunderland: Sinauer Associates, p. 249-320.
- DURANDO, C. M.; BAKER, R. H.; ETGES, W. J.; HEED, W. B.; WASSERMAN, M.; DESALLE, R. (2000) Phylogenetic analysis of the *repleta* species group of the genus *Drosophila* using multiple sources of characters. Molecular Phylogenetics and Evolution 16: 296-307.

- EVELETH, D. D.; MARSH, J. L. (1986) Evidence for evolutionary duplication of genes in the dopa decarboxylase region of *Drosophila*. *Genetics* 114 : 469-483.
- FELSENSTEIN, J. (1985) Confidence limits on phylogenies: an approach using bootstrap. *Evolution* 39: 783-791.
- FERREIRA, M. E.; GRATTAPAGLIA, D. (1998) Introdução ao uso de marcadores moleculares em análise genética. 3. Ed. Brasília: Embrapa.
- FITCH, W. M. (1971) Toward defining the course of evolution: Minimal change for a specific tree topology. *Syst. Zool.* 20: 406-416.
- FROTA-PESSOA, O. (1954) Revision of the *tripunctata* group of *Drosophila* with description of fifteen new species. *Arq. Mus. Paranaense, Curitiba* 10: 253-304.
- FUTUYMA, D. (1997) *Biologia Evolutiva*. 2. Ed. Ribeirão Preto: SBG/CNPq.
- GLEASON, J. M.; GRIFFITH, E. C.; POWELL, J. R. (1998) A molecular phylogeny of the *Drosophila willistoni* group: conflicts between species concepts? *Evolution* 52(4): 1093-1103.
- GRIMALDI, D. A. (1990) A phylogenetic, revised classification of the genera in the Drosophilidae (Diptera). In: *Bull. Am. Mus. Nat. Hist.* 197: 1-139.
- GUPTA, M.; CHYI, Y. S.; ROMERO-SEVERSON, J.; OWEN, J. L. (1994) Amplification of DNA markers from evolutionarily diverse genomes using single *primers* of simple-sequence repeats. *Theor. Appl. Genet.* 89: 998-1006.
- HALL, B. G. (2001) *Phylogenetic trees made easy – A how to manual for molecular biologists*. 2. Ed. Massachusetts : Sinauer Associates.
- HARRIS, H. (1966) Enzyme polymorphism in man. *Proc. Roy. Soc. London B* 164: 298-310.
- HEED, W. B.; RUSSELL, J. S. (1971) Phylogeny and population structure in island and continental species on the *cardini* group of *Drosophila* studied by inversion analysis. *Univ. Texas Publ.* 7103: 91-130.
- HENNIG, W. (1966) *Phylogenetic systematics*. Urbana, III. University of Illinois Press.
- HERÉDIA, F. (2002) *Evolução do Retroelemento gypsy em espécies de Drosophila e Zaprionus indianus: uma abordagem filogenética*. Tese de Doutorado, UFRGS, Brasil.

- HILLIS, D. M.; MABLE, B. K.; LARSON, A.; DAVIS, S. K.; ZIMMER, E. A. (1996) Nucleic Acids IV: Sequencing and Cloning. In: HILLIS, D. M.; MORITZ, C.; MABLE, B. K. Molecular Systematics. 2. Ed. Sunderland : Sinauer Associates, p. 321-381.
- HOLDER, M.; LEWIS, P. O. (2003) Phylogeny estimation: traditional and bayesian approaches. *Nature Reviews – Genetics* 4: 275-284.
- HUELSENBECK, J. P.; RONQUIST, F. (2001) MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754-755.
- JEANMOUGIN, F.; THOMPSON, J. D.; GOUY, M.; HIGGINS, D. G.; GIBSON, T. J. (1998) Multiple Sequence Alignment with ClustalX. *Trends Biochem. Sci.* 23: 403-405.
- JOWET, T. (1986) Preparation of nucleic acids, in *Drosophila: a practical approach*, edited by D. B. Roberts. Oxford: IRL Press.
- KASTANIS, P.; ELIOPOULOS, E.; GOULIELMOS, G. N.; TSAKAS, S.; LOUKAS, M. (2003) Macroevolutionary relationships of species of *Drosophila melanogaster* group based on mtDNA sequences. *Mol. Phyl. Evol.* 28: 518-528.
- KASTRITSIS, C. D. (1969) The chromosomes of some species of the guarani group of *Drosophila*. In: *J. Hered.* 60: 50-57.
- KASTRITSIS, C. D.; PASTEUR, G.; QUICK, J. (1970) Relationships of the polytene chromosomes of *Drosophila mediostriata* and *D. griseolineata*. *Can. J. Genet. Cytol.* 12: 952-959.
- KATOH, T.; TAMURA, K.; AOTSUKA, T. (2000) Phylogenetic position of the subgenus *Lordiphosa* of the genus *Drosophila* (Diptera: Drosophilidae) inferred from *Alcohol Dehydrogenase* gene sequences. *J. Mol. Evol.* 51: 122-130.
- KIMURA, M. (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16:111-120
- KING, J. C. (1947) Interspecific Relationships within the *guarani* Group of *Drosophila*. *Evolution* 1: 143-153.
- KUMAR, S. (1996) Patterns of nucleotide substitution in mitochondrial protein coding genes of vertebrates. *Genetics* 145: 537-548.
- KUMAR, S.; TAMURA, K.; JAKOBSEN, I. B.; NEI, M. (2001) MEGA2: Molecular Evolutionary Genetics Analysis Software. *Bioinformatics*.

- KWIATOWSKI, J.; SHARECKY, D.; BAILEY, K.; AYALA, F. J. (1994) Phylogeny of *Drosophila* and related genera inferred from the nucleotide sequence of the Cu, Zn, *Sod* gene. *J. Mol. Evol.* 38: 443-454.
- KWIATOWSKI, J.; AYALA, F. J. (1999) Phylogeny of *Drosophila* and Related Genera: Conflict between Molecular and Anatomical Analyses. In: *Mol. Phyl. Evol.* 13(2): 319-328.
- LAKOVAARA, S.; SAURA, A. (1982) Evolution and speciation in the *Drosophila obscura* group. In: ASHBURNER, M.; CARSON, H. L.; THOMPSON, J. N. (Eds.) *The genetics and biology of Drosophila*. New York: Academic Press.
- LATORRE, A.; BARRIO, E.; MOYA, A.; AYALA, F. J. (1988) Mitochondrial DNA Evolution in the *Drosophila obscura* group. *Mol. Biol. Evol.* 5(6): 717-728.
- LEWONTIN, R. C.; HUBBY, J. (1966) A molecular approach to the study of genic heterozygosity in natural populations. II. Amounts of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* 54: 595-609.
- LINDSLEY, D. L.; ZIMM, G. G. (1992) *The Genome of D. melanogaster*. Califórnia: Academic Press.
- LINN, S.; ARBER, W. (1968) Host specificity of DNA produced by *Escherichia coli*. X. *In vitro* restriction of phage fd replicative form. *Proc. Natl. Acad. Sci. USA.* 59: 1300-1306.
- LORETO, E. L. DA S.; VALENTE, V. L. DA S.; ZAHA, A.; SILVA, J. C.; KIDWELL, M. G. (2001) *Drosophila mediopunctata* P elements: A new example of horizontal transfer. *The Journal of Heredity* 92(5): 375-381.
- MALOGOLOWKIN, C. (1953) Sobre a Genitália dos Drosofilídeos. IV. A genitália masculina no subgênero *Drosophila* (Diptera, Drosophilidae). *Rev. Brasil. Biol.* 13: 245-264.
- MARSH, D. L.; ERFLE, M. P.; LEEDS, C. A. (1986) Molecular localization, developmental expression and nucleotide sequence of the *alpha-methyl-dopa hypersensitive* gene of *Drosophila*. *Genetics* 114 : 453-467.
- MATIOLI, S. R.; SANTOS E PASSOS BUENO, M. R. (2001) Métodos Baseados em PCR para a análise de polimorfismos de ácidos nucléicos. In: MATIOLI, S. R. (ed.) *Biologia Molecular e Evolução*. Ribeirão Preto: Holos, p. 153-161.

- MAXAM, A. M.; GILBERT, W. (1977) A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA 74: 560-564.
- MESELSON, M.; YUAN, R. (1968) DNA restriction enzyme from *E. coli*. Nature 217: 1110.
- MIYAMOTO, M. M.; FITCH, W. M. (1995) Testing species phylogenies and phylogenetic methods with congruence. Syst. Biol. 44: 64-76.
- MIYAKI, C. Y.; RUSSO, C. A. DE M.; PEREIRA, S. L. (2001) Reconstrução Filogenética: Introdução e o Método da Máxima Parcimônia. In: MATIOLI, S. R. (ed.) Biologia Molecular e Evolução. Ribeirão Preto: Holos, p. 97-107.
- MORITZ, C.; HILLIS, D. M. (1996) Molecular Systematics: Context and Controversies. In: HILLIS, D. M.; MORITZ, C.; MABLE, B. K. Molecular Systematics. 2. Ed. Sunderland: Sinauer Associates, p. 1-13.
- MULLIS, K. B.; FALOONA, F. A. (1987) Specific synthesis of DNA in vitro via a polymerase catalyzed chain reaction. Meth. Enzymol. 155: 335-350.
- MURPHY, R. W.; SITES Jr, J. W.; BUTH, D. G.; HAUFLE, C. H. (1996) Proteins: Isozyme Eletrophoresis. In: HILLIS, D. M.; MORITZ, C.; MABLE, B. K. Molecular Systematics. 2. Ed. Sunderland: Sinauer Associates, p. 51-120.
- NACRUR, J. (1958) Genitália masculina de *Drosophila* do grupo *mesophragmatica* (Diptera). Ver. Brasil. Biol. 18: 243-249.
- NEI, M.; KUMAR, S.; TAKAHASHI, K. (1998) The optimization principle in phylogenetic analysis tend to give incorrect topoloies when the number of nucleotides or amino acids is small. Proc. Natl. Acad. Sci. USA 95: 12390-12397.
- NEI, M.; KUMAR, S. (2000) Molecular Evolution and Phylogenetics. New York: Oxford.
- NICHOLAS, K. B.; NICHOLAS, H. B. JR. (1997) GeneDoc: a tool for editing and annotating multiple sequence alignments. Distributed by the authors.
- NOOR, M.; AKASHY, H. (2001) *Drosophila* evolutionary genetics. Encyclopedia of Life Sciences, Nature Publishing Group: 1-7.
- NURMINSKY, D. I.; MARIYAMA, E. N.; LOZOVSKAYA, E. R.; HARTL, D. L. (1996) Molecular phylogeny and genome evolution in the *Drosophila virilis* species group: duplications of the *Alcohol Dehydrogenase* Gene. Mol. Biol. Evol. 13(1): 132-149.
- O'GRADY, P. M.; CLARK, J. B.; KIDWELL, M. (1998) Phylogeny of the *Drosophila saltans* Species Group Based on Combined Analysis of Nuclear and Mitochondrial DNA Sequences. Mol. Biol. Evol. 15(6): 656-664.

- O'GRADY, P. M. (1999) Reevaluation of Phylogeny in the *Drosophila obscura* Species Group Based on Combined Analysis of Nucleotide Sequences. *Mol. Phyl. Evol.* 12(2): 124-139.
- O'GRADY, P. M.; BAKER, R. H.; DURANDO, C. M.; ETGES, W. J.; DESALLE, R. (2001) Polytene chromosomes as indicators of phylogeny in several species groups of *Drosophila*. *BMC Evolutionary Biology* 1: 1-6.
- O'GRADY, P. M.; KIDWELL, M. G. (2002) Phylogeny of the Subgenus *Sophophora* (Diptera: Drosophilidae) Based on Combined Analysis of Nuclear and Mitochondrial Sequences. *Mol. Phyl. Evol.* 22(3): 443-453.
- PAGE, R. D. M.; CHARLESTON, M. A. (1997a) From gene to organismal phylogeny: Reconciled trees and the gene tree / species tree problem. *Mol. Phyl. Evol.* 7(2): 231-240.
- PAGE, R. D. M.; CHARLESTON, M. A. (1997b) Reconciled trees and incongruent gene and species trees. In: Mirkin, B.; McMorris, F. R.; Roberts, F. S.; Rzhetsky, A. (Eds.). *Mathematical Hierarchies in Biology*. Providence, Rhode Island: American Mathematical Society. p. 57-70.
- PALUMBI, S. R. (1996) Nucleic Acids II: The Polymerase Chain Reaction. In: HILLIS, D. M.; MORITZ, C.; MABLE, B. K. *Molecular Systematics*. 2. Ed. Sunderland: Sinauer Associates, p. 205-247.
- PAMILO, P.; NEI, M. (1988) Relationships between gene trees and species trees. *Mol. Biol. Evol.* 5(5): 568-583.
- PÉLANDAKIS, M.; HIGGINS, D. G.; SOLIGNAC, M. (1991) Molecular phylogeny of the subgenus *Sophophora* of *Drosophila* derived from large subunit of ribosomal RNA sequences. *Genética* 84: 87-94.
- PÉLANDAKIS, M.; HIGGINS, D. G.; SOLIGNAC, M. (1993) Molecular phylogeny of *Drosophila* based on ribosomal RNA sequences. *Journal of Molecular Evolution* 37: 525-543.
- PEREIRA, S. L.; MIYAKI, C. Y.; RUSSO, C. A. DE M. (2001) Reconstrução Filogenética: Métodos Probabilísticos. In: MATIOLI, S. R. (ed.) *Biologia Molecular e Evolução*. Ribeirão Preto: Holos, p. 117-129.
- PHILLIPS, A.; JANIES, D.; WHEELER, W. (2000) Multiple sequence alignment in phylogenetic analysis. *Mol. Phyl. Evol.* 16: 317-330.

- POSADA, C.; CRANDALL, K. A. (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.
- POWELL, J. R. (1991) Monophyly / paraphyly / polyphyly and gene / species trees: an example from *Drosophila*. *Mol. Biol. Evol.* 8(6): 892-896.
- POWELL J. R.; DESALLE, R. (1995) *Drosophila* Molecular Phylogenies and Their Uses. In: *Evolutionary Biology*, 28: 87-138.
- REMSEM, J.; DESALLE, R. (1998) Character congruence of multiple data partitions and the origin of the Hawaiian Drosophilidae. *Mol. Phyl. Evol.* 9: 225-235.
- REMSEM, J.; O'GRADY, P. (2002) Phylogeny of Drosophilinae (Diptera: Drosophilidae), with comments on combined analysis and character support. *Mol. Phyl. Evol.* 24: 249-264.
- ROBE, L. J.; BASSO DA SILVA, L.; LORETO, E. L. DA S. (2002) Phylogenetic relationships among four species of the *guarani* group of *Drosophila* (Drosophilidae, Diptera) as inferred by molecular and morphological analyse. *Revista Brasileira de Entomologia*. 46(4): 515-519.
- RODRIGUEZ-TRELLES, F.; TARRÍO, R.; AYALA, F. J. (1999) Molecular Evolution and Phylogeny of the *Drosophila saltans* Species Group Inferred from the *Xdh* Gene. *Mol. Phyl. Evol.* 13(1): 110-121.
- RUSO, C. A. M.; TAKEZAKI, N.; NEI, M. (1995) Molecular Phylogeny and Divergence Times of Drosophilid Species. *Mol. Biol. Evol.* 12(3): 391-404.
- RUSO, C. A. M.; TAKEZAKI, N.; NEI, M. (1996) Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny. *Mol. Biol. Evol.* 13(3): 525-536.
- RUSO, C. A. DE M. (2001) Como Escolher Genes para Problemas Filogenéticos Específicos. In: MATIOLI, S. R. (ed.) *Biologia Molecular e Evolução*. Ribeirão Preto: Holos, p. 130-136.
- RUSO, C. A. DE M.; MIYAKI, C. Y.; PEREIRA, S. L. (2001) Reconstrução Filogenética: Métodos Geométricos. In: MATIOLI, S. R. (ed.) *Biologia Molecular e Evolução*. Ribeirão Preto: Holos, p. 108-116.
- SAITOU, N.; NEI, M. (1987) The Neighbor-Joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 9: 406-425.

- SANGER, F.; NICKLEN, S.; COULSON, A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci USA* 74: 5463-5467.
- SCHNEIDER, H. (2002) Métodos de análise filogenética – Um guia prático. 2. Ed. Ribeirão Preto: Holos e SBG.
- SESSIONS, S. K. (1996) Chromosomes: Molecular Cytogenetics. In: HILLIS, D. M.; MORITZ, C.; MABLE, B. K. *Molecular Systematics*. 2. Ed. Sunderland: Sinauer Associates, p. 121-168.
- SOLÉ-CAVA, A. M. (2001) Biodiversidade Molecular e Genética da Conservação. In: MATIOLI, S. R. (ed.) *Biologia Molecular e Evolução*. Ribeirão Preto: Holos, p. 171-192.
- SOLFERINI, V. N.; SCHEEPMAKER, D. S. (2001) Polimorfismos de isozimas. In: MATIOLI, S. R. (ed.) *Biologia Molecular e Evolução*. Ribeirão Preto: Holos, p. 137-142.
- STATHAKIS, D. G.; PENTZ, E. S.; FREEMAN, M. E.; KULLMAN, J.; HANKINS, G. R.; PEARLSON, N. J.; WRIGHT, T. R. F. (1995) The Genetic and Molecular Organization of the *Dopa decarboxylase* Gene Cluster of *Drosophila melanogaster*. *Genetics* (141) : 629-655.
- STRYER, L. (1996) *Bioquímica*. 4. Ed. Rio de Janeiro: Guanabara Koogan.
- STURTEVANT, A. H. (1921) *The North American species of Drosophila*. Carnegie Institution of Washington Publication 301.
- STURTEVANT, A. H. (1939) On the subdivision of the genus *Drosophila*. *Proc. Natl. Acad. Sci. USA* 25: 137-141.
- STURTEVANT, A. H. (1942) The classification of the genus *Drosophila* with descriptions of nine new species. *Univ. Texas Publ* 4213: 5-51.
- SWOFFORD, D. L.; OLSEN, G. J.; WADDELL, P. J.; HILLIS, D. M. (1996) Phylogenetic Inference. In: HILLIS, D. M.; MORITZ, C.; MABLE, B. K. *Molecular Systematics*. 2. Ed. Sunderland: Sinauer Associates, p. 407-514.
- SWOFFORD, D. L. (1999) PAUP: Phylogenetic Analysis usind Parsimony (and other methods), versão 4.0. Sinauer Associates, Sunderland, Massachusetts, USA.
- TAMURA, K. (1992) Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C - content biases. *Mol. Biol. Evol.* 9: 678–687.

- TARRÍO, R.; RODRÍGUEZ-TRELLES, F.; AYALA, F. (2000) Tree rooting with outgroups when they differ in their nucleotide composition from the ingroups: the *Drosophila saltans* and *willistoni* groups, a case study. *Mol. Phyl. Evol.* 16(3): 344-349.
- TATARENKOV, A.; KWIATOWSKI, J.; SKARECKY, D.; BARRIO, E.; AYALA, F. J. (1999a) On the evolution of *Dopa decarboxylase (Ddc)* and *Drosophila* Systematics. *J. Mol. Evol.* (48) : 445-462.
- TATARENKOV, A.; SÁEZ, A. G.; AYALA, F. J. (1999b). A compact gene cluster in *Drosophila*: the unrelated *Cs* gene is compressed between duplicated *Amd* and *Ddc*. *Gene* 231: 111-120.
- TATARENKOV, A.; ZUROVCOVÁ, M.; AYALA, F. J. (2001) *Ddc* and *Amd* Sequences Resolve Phylogenetic Relationships of *Drosophila*.. *Mol. Phyl. Evol.* 20(2) : 321-325.
- TATARENKOV, A.; AYALA, F. J. (2001) Phylogenetic Relationships among Species Groups of the *virilis-repleta* Radiation of *Drosophila*. In: *Mol. Phyl. Evol.* 21(2): 327-331
- THROCKMORTON, L. H.; MAGALHÃES, L. E. (1962) Changes with evolution of pteridine accumulations in species of the *saltans* group of the genus *Drosophila*. *Univ. Texas Publ.* 6205: 489-505.
- THROCKMORTON, L. H. (1962) The problem of phylogeny in the genus *Drosophila*. *Univ. Texas Publ.* 6205: 207-343.
- THROCKMORTON, L. H. (1975) The Phylogeny, Ecology and Geography of *Drosophila*. V. 3. In: *Handbook of Genetics*. New York: Plenum.
- TSACAS, L.; CARIU, M. L.; LACHAISE, D. (1985) Le groupe *Drosophila obscura* en Afrique de l'Est, description de trois nouvelles espèces (Diptera: Drosophilidae). *Ann. Soc. Ent. Fr. (NS)* 21: 413-424.
- VAL, F. C.; MARQUES, M. D.; VILELA, C. R. (1981) *Drosophilidae* of Neotropical Region. In: ASHBURNER, M.; CARSON, H. L.; THOMPSON, J. N., J. R. (Eds). *The Genetics and Biology of Drosophila*. V. 3a. Orlando, Flórida: Academic Press.
- VILELA, C. R. (1984) Occurrence of the *Drosophila flavopilosa* species group (Diptera, Drosophilidae) in the state of São Paulo (Brazil) with description of one new species. *Revista Brasileira de Zoologia* 2(2): 63-69.

- VILELA, C. R.; VAL, F. C. (1985) The male genitalia of types of six members of the *Drosophila tripunctata* species group (Diptera, Drosophilidae). *Revista Brasileira de Entomologia* 29(3/4): 503-513.
- VILELA, C. R.; BÄCHLI, G. (1990) Taxonomic studies on Neotropical species of seven genera of Drosophilidae (Diptera). *Mitteilungen der Schweizerischen Entomologischen Gesellschaft Bulletin de la Société Entomologique Suisse* 63 (Suppl.): 1-332.
- VILELA, C. R. (1992) On the *Drosophila tripunctata* species group (Diptera, Drosophilidae). *Revista Brasileira de Entomologia* 36(1): 197-221.
- VILELA, C. R.; PEREIRA, M. A. Q. R. (1993) A case of misidentification of a Neotropical species of *Drosophila* belonging to the *guarani* group. *Revista Brasileira de Entomologia* 37(4): 819-820.
- VILELA, C. R.; BÄCHLI, G. (2000) Five new species of Neotropical *Drosophila* (Diptera, Drosophilidae). *Mitteilungen der Schweizerischen Entomologischen Gesellschaft Bulletin de la Société Entomologique Suisse* 73: 49-65.
- VILELA, C. R.; SILVA, A. F. G.; SENE, F. de M. (2002) Preliminary data on the geographical distribution of *Drosophila* species within the morphoclimatic domains of Brazil. III. The *cardini* group. *Revista Brasileira de Entomologia* 46(2): 139-148.
- VOS, P.; HOGERS, R.; BLEEKER, M.; REIJANS, M.; VAN DE LEIE, T.; HORNES, M.; FRIJTERS, A.; POT, J.; PELEMAN, J.; KUIPER, M.; ZABEAU, M. (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* 23: 4407-4414.
- XIA, X.; XIE, Z. (2001) DAMBE: Data analysis in molecular biology and evolution. *Journal of Heredity* 92: 371-373.
- YOTOKO, K. S. C.; MEDEIROS, H. F.; SOLFERINI, V. N.; KLACZKO, L. B. (2003) A molecular study of the systematics of the *Drosophila tripunctata* group and the *tripunctata* radiation. *Mol. Phyl. Evol.* 28: 614-619.
- WANG, D.; MARSH, J. L.; AYALA, F. J. (1996) Evolutionary changes in the expression pattern of a developmentally essential gene in three *Drosophila* species. *Evolution* 93 : 7103-7107.
- WASSERMAN, M. (1982) Cytological evolution in the *Drosophila repleta* species group. *In*: BARKER, J. S. F.; STARMER, W. T. (Eds.). *Ecological Genetics and Evolution: The Cactus-Yeast-Drosophila model system*. New York: Academic Press. p. 49-64.

- WELSH, J.; MCCLELLAND, M. (1990) Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res.* 18: 7213-7218.
- WERMAN, S. D.; SPRINGER, M. S.; BRITTEN, R. J. (1996) *Nuclei Acids I: DNA-DNA Hybridization*. In: HILLIS, D. M.; MORITZ, C.; MABLE, B. K. *Molecular Systematics*. 2. Ed. Sunderland : Sinauer Associates, p. 169-203.
- WETMUR, J. G.; DAVIDSON, N. (1968) Kinetics of renaturation of DNA. *J. Mol. Biol.* 31: 349-370.
- WHEELER, M. R.; TAKADA, H.; BRNCIC, D. (1962) The *flavopilosa* species group of *Drosophila*. *Studies in Genetic*. II. 6205: 395-413.
- WHEELER, M. R. (1986) The Drosophilidae: a taxonomic overview. In: ASHBURNER, M.; CARSON, H. L.; THOMPSON, J. N., J. R. (Eds.). *The Genetics and Biology of Drosophila*. V. 3a. Orlando, Flórida: Academic Press.
- 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. *Nucleic Acids Res.* 18: 6531-6535.
- WILSON, F. D.; WHEELER, M. R.; HARGET, M; KAMBYSELLIS, M. (1969) Cytogenetic relations in the *Drosophila nasuta* subgroup of the *immigrans* group of species. *Stud. Genet. V Univ. Texas Publ.* 6918: 207-270.
- ZUCKERKANDL, E.; PAULING, L. (1962) Molecular disease, evolution and genic heterogeneity. p. 189-225. In: KASHA, M.; PULLMAN, B. (Eds.). *Horizonts in Biochemistry*. New York: Academic Press.