

Universidade Federal do Rio Grande do Sul

Análise filogenômica de transferência horizontal em genomas de *Drosophila*

Marcos Oliveira de Carvalho

Tese submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para a obtenção do grau de Doutor em Ciências (Genética e Biologia Molecular)

Orientador: Élgion Lúcio da Silva Loreto

Porto Alegre, Abril de 2013

Apoio

O presente trabalho foi realizado no Laboratório de Biologia Molecular de *Drosophila* da Universidade Federal de Santa Maria e contou com o apoio das seguintes instituições:

- CAPES
- CNPq
- FAPERGS
- CESUP/UFRGS – Centro de Supercomputação da Universidade Federal do Rio Grande do Sul

Dedicatória

Dedico este trabalho para as quatro mulheres que inspiram e apoiam diligentemente todos meus empreendimentos: Minha mãe, minha irmã, minha sobrinha e a minha esposa.

Agradecimentos

É impossível completar um trabalho de doutorado sem contar com o apoio de diversas pessoas. Apoio ao qual credito o sucesso da finalização desta tese.

Também é difícil listar todas as pessoas com as quais interagi e que de alguma forma contribuíram para este trabalho, contudo, agradeço sinceramente a todos que me apoiaram durante minha passagem pelo PPGBM.

Àqueles que estiveram de forma próxima, gostaria de agradecer especialmente:

Ao meu orientador, Prof. Élgion Loreto, pela confiança e oportunidade de desenvolver este trabalho sob sua orientação, pela paciência, compreensão e consideração assim como pela determinação em apoiar a realização desta tese. Não obstante, eu gostaria de agradecer de forma ampla aos nossos mais de 12 anos de convivência, onde a oportunidade de trabalhar no LabDros da UFSM, desde a minha graduação, proporcionando a convivência com um grupo fantástico e a formidável oportunidade de trabalhar em um tema tão apaixonante como bioinformática e análise genômica, moldou de forma definitiva minha carreira.

Aos meus colegas do LabDros, presentes e passados, nossa convivência, mesmo que esparsa nos últimos momentos, sempre será lembrada como um dos melhores momentos da minha carreira acadêmica.

Agradeço especialmente à Coordenadora do PPGBM, Professora Márcia Margis pelo apoio e compreensão durante a finalização desta tese, sem os quais este trabalho não teria

sido realizado.

Ao Elmo e demais funcionários do PPGBM, pelo trabalho exemplar e dedicação que servem de inspiração para todos os alunos do nosso curso. Também ao Elmo por todo apoio aos projetos que desenvolvi durante minha participação junto à Comissão de Representação Discente do PPGBM.

Aos ex-coordenadores do PPGBM, Professora Maria Cátira Bortoloni e Professor Claiton Bau.

Aos meus colegas do PPGBM, especialmente aos colegas que formaram em conjunto comigo a Representação Discente do PPGBM, Luiz Felipe Valter de Oliveira, Caio Cezar, Paula Rohr e Camila Zanella.

Aos professores do PPGBM cuja dedicação é responsável pela qualidade incontestada do nosso curso.

Ao professor Rogério Margis pelo apoio e orientação durante a participação no Prêmio Santander de Empreendedorismo e Prêmio Ibero-Americano de Empreendedorismo.

À professora Vera Gaiesky pelo constante apoio.

Aos meus sócios Luis Felipe Miranda Ramos, Luiz Felipe Valter, Luiz Fernando Valter e Ronaldo Mota pelo apoio incondicional à realização desta tese. Sem o suporte de vocês este trabalho não seria possível.

A toda equipe da Neoprospecta pelo apoio e dedicação.

À CAPES pelo apoio na forma de uma bolsa de doutorado e ao CESUP pela oportunidade de utilizar os recursos de supercomputação da UFRGS.

À minha “grande família”, tios, primos, cunhados e a minha sogra e sogro pela compreensão e suporte nos mais diferentes momentos ao longo da realização do meu doutorado. Especialmente à minha tia Nelcy, Tia Marlene, Tio Jango, Tio Oscar, Tia Alba e Tio Daniel e meus primos Fabiana Costa, João Victor Costa, Luciano Carvalho e Fernando Carvalho. Muito obrigado pelo carinho e apoio de vocês.

Por fim, agradeço especialmente à minha mãe, Maria de Fátima e minha irmã Clarissa que sempre acreditaram e de forma incontestemente apoiaram a minha formação desde a graduação até a finalização desta tese. Impossível é descrever o quanto eu sou grato pelo apoio e amor de vocês. Agradeço à minha sobrinha, que apesar de ainda não compreender o que é o doutorado do tio muitas vezes me alegrou com o seu sorriso. Agradeço especialmente à minha esposa Kely, que durante semanas aguentou de forma leve e alegre minha rotina inconstante, fazendo todo o possível para que eu tivesse tranquilidade e conforto. A ti o meu agradecimento e amor incondicionais.

Sumário

Apoio.....	2
Dedicatória.....	3
Agradecimentos.....	4
Sumário.....	7
Lista de abreviaturas, símbolos e unidades.....	8
Resumo.....	11
Abstract.....	13
Capítulo 1.....	15
Introdução.....	15
Os elementos de transposição.....	15
Transposons e o genoma dinâmico.....	25
Análise genômica em larga escala.....	27
Objetivos.....	38
Capítulo 2.....	39
Materiais e métodos.....	39
Capítulo 3.....	46
Resultados e Discussão.....	46
Análise em larga escala de transferência horizontal.....	46
Análise comparativa entre 12 genomas de <i>Drosophila</i> e seus elementos de transposição.....	50
Perspectivas futuras.....	141
Referências.....	143
Anexos.....	156
Anexo 1: Methods for detection of horizontal transfer of transposable elements in complete genomes.....	157
Anexo 2 - Dados sumarizados das comparações de Tes e genes nucleares.....	165

Lista de abreviaturas, símbolos e unidades

A3 - Porcentagem de nucleotídeos A na terceira posição do códon da sequência de nucleotídeo

Aromo - Frequência de aminoácidos aromáticos

avg_nc - Média do número efetivo de códons entre sequências de um mesmo par de comparação

BLAST - Basic Local Alignment Tool

C3 - Porcentagem de nucleotídeos C na terceira posição do códon da sequência de nucleotídeo

Dana - *Drosophila ananassae*

Dere - *Drosophila erecta*

Dgri - *Drosophila grimshawi*

diff_nc - Razão entre o maior e o menor valor do número efetivo de códons entre sequências de um mesmo par de comparação

Dmel - *Drosophila melanogaster*

Dmoj - *Drosophila mojavensis*

DNAId - Identidade da sequência de nucleotídeos em %

Dper - *Drosophila persimilis*

Dpse - *Drosophila pseudobscura*

Dsec - *Drosophila sechellia*

Dsim - *Drosophila simulans*

Dvir - *Drosophila virilis*

Dwil - *Drosophila willistoni*

Dyak - *Drosophila yakuba*

ENC - Número efetivo de códons

G3 - Porcentagem de nucleotídeos G na terceira posição do códon da sequência de nucleotídeo

GC - Porcentagem de nucleotídeos GC na da sequência completa de nucleotídeo

GC3 - Porcentagem de nucleotídeos G e C na terceira posição do códon da sequência de nucleotídeo

Gravy - Média do índice de hidropaticidade para cada aminoácido

Index - Índice de similaridade

Ka - Taxa de mutação não-sinônima

Ka/Ks - Relação entre Ka e Ks

Kimura - Distância Kimura medida a partir de sequência de aminoácidos

Ks - Taxa de mutação sinônima

L_aa - Tamanho da sequência em aminoácidos

Logdet - Distância logdet medida a partir de sequencia de nucleotídeos

L_sym - Número de códons sinônimos ao longo da sequência de nucleotídeos

Mite - Miniature inverted-repeat transposable element

Nca - Valor do número efetivo de códons para a primeira sequência do par de ortólogos

Ncb - Valor do número efetivo de códons para a segunda sequência do par de ortólogos

NCBI - National Center for Biotechnology Information

piRNA - RNA associado à proteína Piwi

ProtId - Identidade da sequência de aminoácidos em %

RAM - Random-access memory

T3 - Porcentagem de nucleotídeos T na terceira posição do códon da sequência de nucleotídeo

TE - Transposable element

tRNA - RNA transportador

Resumo

A descoberta dos elementos genéticos móveis na década de 40 representa a formação de um novo paradigma no campo da genética, onde existe a formação de uma abordagem dinâmica à interpretação do genoma de organismos eucarióticos e procarióticos. A presença de elementos de transposição em praticamente todos os seres vivos indica a sua origem ancestral e indica a importância dos processos de transposição como mecanismo evolutivo. Apesar de em muitos casos os elementos de transposição apresentarem-se como agentes de processos deletérios, sua atividade também está ligada a diversos eventos adaptativos incluindo a formação de novos genes, regulação gênica e constituição estrutural do genoma. Com o recente desenvolvimento de novas tecnologias de sequenciamento, que possibilitaram a expansão massiva do número de genomas completos sequenciados, acumulou-se um grande número de dados de sequência que permitem a comparação direta entre diferentes genomas e assim a realização de análises específicas em larga escala sobre a evolução dos elementos de transposição. Este trabalho utilizou 12 genomas completos de espécies de *Drosophila* para comparações em larga escala entre genes e transposons, analisando diferentes variáveis evolutivas. Através dessas comparações foi possível identificar padrões específicos de sequências de elementos de transposição em relação às sequências gênicas do genoma hospedeiro. Especificamente, um grande número de sequências de elementos de transposição apresentam valores de K_s significativamente inferiores aos valores de K_s encontrados em sequências gênicas, ao mesmo tempo que apresentam alta similaridade da sequência de DNA. Contudo, em média

a sequência de proteína destas mesmas sequências tende a possuir similaridade menor que as sequências gênicas da respectiva comparação. É possível distinguir dois tipos de padrão evolutivo nas sequências de elementos de transposição, o primeiro relacionado a sequências com alta similaridade, baixo valor de K_s e K_a . Tais sequências são provavelmente derivadas de eventos de transposição recente. Um segundo padrão é caracterizado por sequências com alta similaridade, baixo valor de K_s e valor de K_a superior ao encontrado no genoma, gerando sequências com grande valor de K_a/K_s , mas fortemente conservadas, indicando uma possível seleção positiva em sequências de elementos de transposição sujeitos a recente transferência horizontal.

Abstract

The discovery of transposable elements in the 1940s represents the constitution of a new paradigm in the field of genetics, where is consolidated a new and dynamic interpretation of eukaryotic and prokaryotic genomes. The existence of transposable elements as ancestral components of the genome, indicates the importance of transposition processes as evolutionary mechanisms. Although transposable elements could cause a range of deleterious effects in the host, its activity is also connected to a broad number of adaptative events, including the constitution of new genes, gene regulation and the structural integrity of the genome. With the recent development of new sequencing technologies and massive increase in the number of complete sequenced genomes a new opportunity window was created for the large scale evolutionary analysis of transposable elements in different species. This work analysed 12 complete genomes of *Drosophila* to conduct large scale comparative analysis between transposable elements and nuclear genes, analysing a number of different variables. Trought this comparisons was possible to identify specific evolutionary patterns of transposable elements and *Drosophila* nuclear genes. A great number of transposition element sequences presented Ks values significantly lower than the Ks value found for nuclear gene sequences, also presenting a high DNA sequence similarity in relation to the genome sequences. However, the average protein sequence similarity of the transposable element sequences is lower than the nuclear genes protein sequence in the same comparison. It is possible to identify two evolutionary patterns in the transposable element sequences. The first one is characterized by sequences

with high similarity and low K_s and K_a values. These sequences are probably subject to a recent horizontal transfer event. The second pattern is characterized by sequences with high similarity, low K_s values and K_a values above the average genome nuclear gene sequences K_a values, producing high values of K_a/K_s . These second pattern sequences are highly conserved, although with a high K_a/K_s value, probably indicating positive selection over transposable element sequences subject to horizontal transfer.

Capítulo 1

Introdução

Os elementos de transposição

Em 1951 a edição anual do Simpósio de Biologia Quantitativa do Cold Spring Harbor teve como tema “Genes e Mutações”. Nesse encontro a Dra. Barbara McClintock realizou pela primeira vez uma palestra pública sobre suas descobertas a respeito do sistema *Ac-Ds* e o genoma dinâmico (McClintock, 1951).

Ao final desse encontro o conhecido geneticista de *Drosophila*, Alfred Sturtevant, teria observado: “Não entendi uma palavra, mas se a Barbara falou, deve ser verdade” (Comfort, 1995).

Barbara McClintock iniciou seu trabalho como citogeneticista na década de 20 do século passado, utilizando milho como organismo modelo no estudo de padrões de quebra cromossômica (Rhoades, 1984). Durante essa época ela identificou uma linhagem de milho triploide que serviria depois como base para seus primeiros experimentos com inovadoras técnicas de citogenética, técnicas estas que permitiriam, anos depois, a descoberta dos elementos de transposição (McClintock, 1929). Dos anos 20 até meados dos anos 40, Barbara McClintock tornou-se uma respeitada pesquisadora na área de citogenética, sendo responsável pela descoberta dos telômeros, centrômeros, a primeira caracterização visual do crossing-over e a identificação do ciclo de fusão-quebra-ponte em cromossomos de milho (Comfort, 1995).

Em 1944, instalada como pesquisadora no Cold Spring Harbor Laboratory, instituição onde permaneceria até o final da sua carreira, Barbara McClintock ao estudar o ciclo de fusão-quebra-ponte em milho, realizou os primeiros experimentos que dariam origem à caracterização do sistema *Ac-Ds* e a descoberta dos primeiros elementos de transposição (Comfort, 2001).

Inicialmente interessada no estudo dos rearranjos cromossômicos identificados em cromossomos que apresentavam ausência de telômeros, Barbara McClintock observou características de variegação em sementes de milho nos espécimes utilizados em seus experimentos de citogenética (Fedoroff, 2012). Estes padrões de coloração atípicos, com ausência de coloração ocorrendo ao longo das sementes com áreas com presença de pigmento indicavam que mutações somáticas aconteciam após a fecundação dupla e posterior formação do endosperma triploide da semente. O estudo da variegação em plantas já havia sido descrito para milho, onde outros pesquisadores haviam sugerido explicações à luz do pensamento mendeliano vigente (Emerson, 1917).

Ao realizar a análise citogenética do tecido das sementes variegadas, Barbara McClintock percebeu a existência de quebras cromossômicas no cromossomo 9 que se repetiam com uma frequência inesperada. Utilizando uma série de marcadores citogenéticos específicos para o cromossomo 9, ela pôde deduzir a existência de um locus que seria responsável pelo processo de quebra, denominando-o de *Ds* (de “dissociação”). Experimentos posteriores comprovaram que o locus *Ds* estava sob controle de outro locus, ativado durante a formação do endosperma da semente. A este loci McClintock deu o nome de *Ac* (de “ativador”). Assim McClintock descrevia um novo sistema de controle

gênico (à mesma época onde Monod descrevia o sistema operon), contudo, ainda não existia a concepção de mobilidade dentro do sistema *Ac-Ds*. Experimentos posteriores realizados utilizando o mesmo conjunto de marcadores citogenéticos confirmaram a mobilidade do loci *Ds*, sempre em conjunto com a presença do loci *Ac*. Atualmente sabe-se que o loci *Ds* é em fato um elemento transponível *Ac* sem capacidade própria de transposição que se mobiliza via expressão de transposase de um elemento *Ac* ativo (Bai e Brutnell, 2011). No caso dos experimentos realizados por McClintock, ao mobilizar-se o elemento *Ds* promovia a quebra do cromossomo 9 e dessa forma inativava o gene responsável pela coloração da semente, causando o padrão fenotípico variegado (Jones, 2005).

A aceitação da existência de componentes móveis presentes no genoma no entanto não foi imediata, dada a ideia revolucionária que esta representava à época em que foi apresentada e também por dificuldades de comunicação e culturais enfrentadas por Barbara McClintock. Contudo, a descoberta da existência das sequências de inserção em bactérias (Shapiro, 1969) e trabalhos posteriores em organismos modelos como *Drosophila* nos anos 80 e 90 levaram à consolidação do conceito de um genoma dinâmico e da existência de sequências com capacidade de transposição (Daniels *et al.*, 1990; Finnegan and Fawcett, 1986; Galas e Chandler, 1981; Green, 1980; Kidwell, 1992).

O trabalho de Barbara McClintock forma portanto a base não só da existência dos elementos de transposição mas também das atuais discussões em torno de processos regulatórios regidos pela influência dos transposons.

De forma ampla, podemos definir elementos de transposição como sequências de

DNA com capacidade de transposição própria ou não (podendo ser mobilizadas por outro elemento de transposição). A transposição pode ocorrer, de forma geral, tanto por um processo de cópia (no caso de elementos com mobilidade mediada por um intermediário de RNA) ou por um processo de translocação da sequência do transposon de uma posição do genoma a outra. Esta diferença no processo de mobilização forma a base para a classificação dos diferentes elementos de transposição.

Por razões históricas convencionou-se denominar todas as formas de elementos genéticos móveis como transposons, contudo atualmente a complexidade de formas existentes de elementos de transposição requer sua organização em uma classificação mais formal. Contudo ainda não existe uma convenção de classificação e nomenclatura dentro da comunidade científica dedicada ao estudo de elementos de transposição (como uma ontologia com definições formais de anotação).

Em 2007 foi proposto um sistema hierárquico de classificação de elementos de transposição (Wicker *et al.*, 2007) onde mantinha-se a definição de que classes de transposons são definidas pela existência ou não de um intermediário de RNA para a transposição mas criava outros táxons como Ordem (dependente de características estruturais e enzimáticas), redefinia as classificações de subclasse de acordo com a existência ou não de excisão do elemento durante a transposição e estipulava critérios de similaridade estrutural e funcional para a inclusão de sequências de transposons em diferentes táxons. Este sistema foi criticado (Seberg e Petersen, 2009) por não levar em consideração diretamente as relações filogenéticas existentes entre os elementos, como por exemplo, a classificação de MITEs em diferentes subfamílias dos seus elementos

originários, permitindo a existência de grupos parafiléticos e definindo classificações arbitrárias que não correspondem à história evolutiva dos diferentes tipos de elementos de transposição (Seberg e Petersen, 2009). Um sistema paralelo de classificação (Kapitonov e Jurka, 2008) foi implementado no processo de desenvolvimento do banco de dados Repbase (Jurka *et al.*, 2005) (que também é amplamente usado como biblioteca de busca para o software RepeatMasker (Smith *et al.*, 1996)). Esta classificação segue a mesma convenção de determinação de classe em acordo com o intermediário de transposição mas difere da proposta de Wicker *et al.* (2007) no número de táxons empregados e na forma em que ocorre a definição da inserção dos diferentes elementos nos táxons.

Na conferência internacional “Impacto Genômico dos Elementos de Transposição Eucarióticos” realizada em Asilomar no ano de 2006 foi definido um comitê internacional para a determinação de regras específicas para a classificação de elementos de transposição (ICCTE – International Committee on Classification of Transposable Elements). Tal comitê representa para os elementos de transposição o que o ICTV (International Committee on Taxonomy of Viruses) representa para a classificação de vírus. Contudo, até o momento não existem definições específicas de classificação emitidas pelo ICCTE, sendo a classificação descrita abaixo o consenso existente na literatura revisada (Blomberg *et al.*, 2009; Feschotte *et al.*, 2009; Finnegan, 1985; Kapitonov e Jurka, 2008; Robertson e Lampe, 1995; Rouault *et al.*, 2009; Wicker *et al.*, 2007).

Pode-se assim classificar os elementos de transposição da seguinte forma:

Transposons classe I

Os elementos pertencentes à classe I, também denominados retrotransposons,

realizam o seu processo de transposição através de um intermediário de RNA, onde a sequência do transposon é transcrita pela maquinaria do genoma hospedeiro, convertida em DNA pela transcriptase reversa codificada pelo elemento e posteriormente inserida em um diferente locus do genoma através de uma integrase também codificada pelo retroelemento.

Certos elementos da Classe I possuem como característica peculiar a presença de longas repetições diretas (LTRs) que flanqueiam a sequência do retrotransposon. Contudo a presença de LTRs não é ubíqua, sendo alguns elementos desta classe desprovidos de tais repetições, sendo denominados como retroelementos non-LTR.

Elementos da Classe I que possuem LTR são subdivididos em subcategorias, notadamente Gypsy/Ty1, Copia-Ty3 e Pao-Bel. Estas subcategorias por sua vez são subdivididas em famílias seguindo geralmente o critério de que uma família de elementos de transposição consiste de sequências com ao menos 80% de similaridade (Capy *et al.*, 1996). Tal definição é arbitrária e debate ainda existe dentro da comunidade científica em relação à melhor forma de realizar a classificação de famílias de transposons dada a grande variabilidade de elementos encontrada. Esta observação tem sido corroborada por estudos recentes que tem se valido de dados de sequência de genomas completos.

Dentro da classe I pode-se distinguir o grupo de elementos genéticos móveis que não possuem repetições terminais longas, sendo este grupo subdividido em elementos chamados LINEs e SINEs.

LINEs são retrotransposons autônomos (possuem capacidade codificante própria, permitindo a sua transposição independente de outros elementos) que contém ao menos uma transcriptase reversa e uma nuclease codificadas em sua sequência. Estes elementos

variam amplamente de tamanho, podendo alcançar vários quilobases e são encontrados em todos os eucariotos.

SINEs são retroelementos de pequeno tamanho e não-autônomos. Contudo SINEs diferem-se de outros transposons não-autônomos pelo fato de não serem derivados de transposons autônomos, mas sim formados pela retrotransposição acidental de transcritos da Polimerase III (Pol III). Também diferentemente de pseudo-genes retroprocessados, os SINEs possuem um promotor interno para Pol III, permitindo a sua expressão. Contudo, para sua transposição estes elementos são dependentes de LINEs que realizam este processo em *trans*. Alguns SINEs são generalistas em relação a grupo de LINE que pode realizar sua transposição, enquanto outros possuem um tipo específico de LINE com o qual a transposição pode ser realizada.

Transposons classe II

Transposons de Classe II são definidos como aquelas sequências com capacidade de mobilização através de um mecanismo mediado por uma enzima específica denominada transposase. Esta enzima codificada pelo elemento (no caso de elementos autônomos) reconhece repetições específicas que delimitam o elemento do restante do genoma. Tais repetições apresentam-se invertidas uma em relação a outra, sendo denominadas assim Repetições Terminais Invertidas. A transposição inicia com um passo de clivagem e excisão da sequência do elemento de transposição a partir do sítio onde este estava previamente inserido. Esta clivagem e excisão são catalizadas pela enzima transposase, que também é responsável pela inserção da sequência do elemento de transposição em uma diferente posição do genoma.

Este processo gera um dano localizado em um dos cromossomos (considerando uma célula diploide) que eventualmente é reparado pela maquinaria de reparo de DNA, gerando o aumento do número de cópias do elemento transposto.

Tal processo contrasta com o mecanismo de transposição dos elementos de Classe I, onde o aumento do número de cópias independe da existência de posterior reparo do cromossomo onde originalmente localizava-se o elemento de transposição. Costuma-se referir-se ao mecanismo de transposição de transposons de Classe I como “cópia-e-cola” enquanto que para os elementos de Classe II e Helitrons refere-se como “corta-e-cola”.

Helitrons são elementos genéticos móveis também denominados transposons círculo-rolante em referência ao seu mecanismo de transposição. Este mecanismo é realizado basicamente por duas enzimas, uma RCp (rolling-circle protein) responsável pelo processo de inserção da sequência do Helitron em uma nova posição no genoma e uma helicase, responsável por facilitar o processo de invasão durante a síntese da nova fita de DNA do sítio alvo pelas polimerases do genoma hospedeiro. Apesar de recentemente descritos, os Helitrons são elementos presentes em um grande número de espécies e responsáveis por numerosos eventos de captura de genes durante o seu processo de transposição. Uma característica própria destes elementos é a ausência de formação de duplicações do sítio alvo de inserção.

Estes elementos inicialmente considerados como outra classe de transposons, atualmente são classificados dentro da Classe II, em conjunto com os Polintons.

Polintons ou Mavericks são elementos encontrados em todos animais, protistas e fungos e constituem um subgrupo especial dos elementos de transposição Classe II. Estes

elementos são também conhecidos como transposons auto-sintetizáveis, uma vez que são capazes de realizar sua duplicação a partir de uma cópia excisada do genoma. Constituem-se de elementos com tamanho muito superior à média dos elementos de transposição, possuindo centenas de quilobases. Sua sequência codifica para uma série de proteínas necessárias ao seu processo de transposição, incluindo uma DNA polimerase B, integrase retroviral, proteases e atpases. Polintons podem existir como cópias autônomas ou não, a exemplo de outros elementos de Classe II. Sua origem provável deu-se através de um plasmídeo que adquiriu uma integrase retroviral há aproximadamente 1 bilhão de anos.

Assim como o elemento *Ds* inicialmente descoberto em conjunto com o elemento *Ac*, muitos elementos genéticos móveis existem na forma de um transposon sem capacidade própria de transposição, possuindo apenas as características estruturais mínimas necessárias à ligação da enzima responsável pela transposição, como as repetições terminais invertidas.

Um tipo específico desse tipo de elemento é denominado MITE (do inglês, “miniature inverted-repeat transposable element”). Estes elementos possuem tamanho inferior ao transposon originalmente ativo, variando de 300nt a 500nt e são encontrados em alto número de cópias em diferentes genomas eucarióticos. Descrevem-se casos de MITEs originários de transposons como Elemento P, mariner e Hobo. Interessantemente, se considerarmos os elementos de transposição como “parasitas genômicos” uma vez que estes se valem do genoma hospedeiro e de suas funções celulares para sua manutenção, temos os MITEs como parasitas dos elementos de transposição, uma vez que estes utilizam transposases, que primariamente serviriam aos transposons autônomos, para sua própria

transposição (González e Petrov, 2009).

Devido as suas características como elementos genéticos móveis, os transposons apresentam-se como módulos genéticos especialmente sujeitos a sofrer um fenômeno evolutivo denominado como transferência genética horizontal (Loreto, Carareto e Capy, 2008). Este fenômeno consiste na transferência de material genético entre espécies distintas sem que exista um processo de hibridação sexual. Pode-se considerar como transferência horizontal os processos comumente encontrados em procariotos como conjugação, transdução e transformação (Juhas *et al.*, 2009). Em eucariotos, desprovidos das características estruturais e funcionais existentes em procariotos e que facilitam a troca de material genética via transferência horizontal, este fenômeno é catalizado por agentes externos como parasitas (que possivelmente carregam moléculas de DNA em seu aparelho bucal ao alimentarem-se de diferentes espécies) e vírus (Loreto, Carareto e Capy, 2008; Piskurek e Okada, 2007).

Por possuírem uma capacidade intrínseca de duplicação os elementos de transposição são especialmente propensos a, uma vez introduzidos em um hospedeiro novo através de um vetor de transferência horizontal, iniciar um processo de colonização e manutenção dentro do novo genoma. É importante notar que mesmo que ocorra o processo de transferência de material genético entre espécies separadas através de um vetor específico (seja um vírus ou um parasita) é necessário que ocorra a fixação da sequência recém introduzida de forma que esta seja posteriormente herdada verticalmente dentro da linhagem do novo genoma hospedeiro. Para isso dois requisitos precisam ser cumpridos. Primeiro a introdução do material genético na nova espécie deve ocorrer em tecido que

originará células gaméticas e em segundo lugar a sequência deverá ou apresentar uma vantagem evolutiva ao hospedeiro (caso contrário a mesma rapidamente deverá sofrer degradação em poucas gerações ou diminuir em frequência na população até a sua perda por deriva genética) ou possuir capacidade própria de duplicação, como no caso dos elementos genéticos móveis. Contudo, no caso dos transposons, a dinâmica entre o número de cópias de elementos de transposição em um determinado genoma, resultado do equilíbrio entre a taxa de transposição e a taxa de degradação dos transposons é ainda objeto de debate (Hua-Van *et al.*, 2005, 2011; Kofler, Betancourt e Schlötterer, 2012; Lockton e Gaut, 2010; Rouzic, Le e Capy, 2006, 2009; Rouzic, Le e Deceliere, 2005; Rouzic, Le, Dupas e Capy, 2007)

Transposons e o genoma dinâmico

Uma característica importante da estrutura e evolução dos genomas dos artrópodes (e de outros organismos de forma geral) encontra-se na presença e atividade de elementos genéticos móveis (Feschotte e Pritham, 2007; Frost *et al.*, 2005) .

Antes incluídos na classe “DNA lixo” (Zuckerandl, 1992), atualmente os transposons são considerados elementos importantes em uma série de processos evolutivos e regulatórios em genomas de eucariotos e procariotos (Alzohairy *et al.*, 2013; Blot, 1994; Kazazian, 2004; Rouzic, Le, Boutin e Capy, 2007; Rouzic, Le e Capy, 2005; Schneider e Lenski, 2004). Ademais, sua característica capacidade de mobilização tem sido explorada como útil ferramenta de análise genômica (Izsvák *et al.*, 2010).

Recentemente, claras evidências tem sido apresentadas de níveis muito altos de

adaptação de elementos genéticos móveis ao contexto do genoma hospedeiro, como no caso da manutenção dos telômeros em *Drosophila melanogaster* (Abad *et al.*, 2004; Villasante *et al.*, 2007) e dos insuladores gypsy (Labrador *et al.*, 2008). Ainda é possível citar como exemplo de integração dos transposons ao contexto genômico hospedeiro a possibilidade de recrutamento de sequências de elementos de transposição como novos genes e promotores (Iwashita *et al.*, 2006; Romanish *et al.*, 2007).

Como parte integral de grande parte dos genomas eucarióticos os elementos de transposição participam não apenas como sistemas “egoístas” como inicialmente teorizado. Atualmente existem exemplos de elementos de transposição como agentes cruciais no desenvolvimento de características adaptativas específicas em organismos eucarióticos (Alzohairy *et al.*, 2013; Flajnik e Kasahara, 2009; Lynch *et al.*, 2011).

Mesmo agindo no papel de mutagênicos (onde o processo de transposição intragenômica pode originar quebras cromossômicas, rearranjos e duplicações) os transposons podem ser considerados como um “motor evolutivo” para a espécie que os possui (Kazazian, 2004). Contudo, muitas espécies desenvolveram a capacidade de regular tal atividade (Lozovskaya, 1995).

Os transposons possuem uma evolução complexa, devido à sua capacidade inata de transposição, deriva genética, efeitos reguladores do hospedeiro e transferência horizontal (Loreto, Carareto e Capy, 2008; Rouzic e Capy, 2005). Dessa forma, tais elementos apresentam-se como indutores de maleabilidade genômica (Lim e Simmons, 1994), podendo em um momento ser considerados como entidades individuais e outrora como pertencentes a um sistema genômico definido (Capy *et al.*, 2000; Kidwell e Lisch, 2001).

Essas características únicas dos elementos transponíveis os colocam como interessantes alvos de estudos evolutivos, e principalmente, de estudos comparativos com genes nucleares constitutivos, a fim de esclarecer os mecanismos envolvidos na sua adaptação como genes móveis.

Análise genômica em larga escala

O ano de 1995 marca a publicação do primeiro genoma completo de um organismo celular, a bactéria *Haemophilus influenza* (Fleischmann *et al.*, 1995). Este trabalho não só representa o início da genômica como área científica constituída, mas também o começo de uma nova etapa dentro das ciências biológicas, onde podemos encontrar uma crescente tendência de análise quantitativa de dados moleculares (Rechenmann, 2000).

O grande propulsor de tal mudança de paradigma encontra-se no desenvolvimento de métodos automáticos de sequenciamento, que possibilitaram a produção de dados de sequências de nucleotídeos em grande escala (Birney, Clamp e Hubbard, 2002). Esta mudança continua a estabelecer-se com a formação de centros de sequenciamento dedicados aliados à nova geração de métodos de sequenciamento (Mardis, 2008), possibilitando que atualmente existam 23469 projetos genomas completos e em andamento, sendo 3334 de organismos eucarióticos (GOLD database, www.genomesonline.org, consulta em 2013), totalizando centenas de gigabases em dados de seqüência.

No esteio dessa grande quantidade de dados é possível observar a crescente

necessidade de análise e interpretação de tais informações (Dudley e Butte, 2010) efetivamente transformando dados em conhecimento biológico (Kováč, 2007; Milburn, 2001).

A suprir esta necessidade, a bioinformática agregou ao cotidiano de pesquisa em biologia molecular métodos originalmente utilizados em outros domínios científicos (Buldyrev *et al.*, 1998; Cohen, 2004; Haimovich *et al.*, 2006), que no entanto não descartam a presença ativa do analista orientado à interpretação biológica (Claverie, 2000). Estes métodos de análise incluem *wavelets* (Thurman, Noble e Stamatoyannopoulos, 2008), modelos ocultos de Markov (Boitard, Schlotterer e Futschik, 2009; Kumar e Cowen, 2009), redes neurais (Maszczyk e Duch, 2010; Nath e Mitchell, 2012; Yu *et al.*, 2011), análise de grafos (Fatakia, Costanzi e Chow, 2009; Pell *et al.*, 2012; Peterlongo e Chikhi, 2012), árvores de sufixo (Askitis e Sinha, 2010; Becher, Deymonnaz e Heiber, 2009; Simpson e Durbin, 2010) e amostragem Gibbs (Newberg *et al.*, 2007; Schultheiss *et al.*, 2009). Adicionalmente, novos métodos criados especificamente para a análise de dados biológicos foram desenvolvidos, como os algoritmos Needleman-Wunsch (Needleman e Wunsch, 1970) e Smith-Waterman (Smith e Waterman, 1981).

A união interdisciplinar entre biologia molecular e ciência da computação não é recente com softwares desenvolvidos para análise genômica inicialmente na década de 70 (Staden, 1977). No entanto, o fenômeno de escalonamento na produção de dados biológicos promove uma oportunidade para a análise integrada de genomas sem precedente (Stupka, 2002).

Esta oportunidade tem se refletido em um grande número de trabalhos de análise

genômica comparativa com crescente impacto na forma como hoje se define conceitos previamente cristalizados, como o de famílias gênicas (Chothia *et al.*, 2003; Koonin, Wolf e Karev, 2002; Ranea *et al.*, 2004), regulação gênica (Wang e Stormo, 2005), espécie (Coenye *et al.*, 2005; Gollery *et al.*, 2006; Konstantinidis, Ramette e Tiedje, 2006; Konstantinidis e Tiedje, 2005) e propriamente gene (Gerstein *et al.*, 2007). Da mesma forma, conceitos como o de genoma mínimo (Wegrzyn, 2001) tem sido explorados através de métodos comparativos.

Particularmente, a análise evolutiva molecular tem se beneficiado da atual disponibilidade de dados genômicos para ampliar análises antes realizadas com conjuntos de informações limitadas. Como exemplo cita-se o trabalho de (Jordan *et al.*, 2005), que expande observações sobre tendências de substituição de aminoácidos realizadas por Zuckerkandl *et al* (1971). De forma emblemática, tal trabalho também iniciou intenso debate sobre a possibilidade de existência de uma preponderância de substituições, relativas a codons recentemente incorporados ao código genético e adaptação dos resultados à teoria neutralista (Goldstein e Pollock, 2006; Hurst, Feil e Rocha, 2006).

Em 1998 foi cunhado o termo “filogenômica” em trabalho que descrevia o uso de técnicas filogenéticas para anotação em larga escala de genomas (Eisen, 1998). Atualmente o termo ampliou-se para definir a intersecção entre análise genômica comparativa e filogenética (Eisen e Fraser, 2003). Dessa forma, a filogenômica visa a integração de dados genômicos, sejam estes na forma de sequências de nucleotídeos/aminoácidos, propriedades de sequência (como conteúdo GC, tendência de uso de códon, entropia e classificação funcional), propriedades genômicas (como ordem gênica, rearranjos, abundância relativa

de domínios) ou dados derivados de outros procedimentos experimentais (como microarranjos e dados de interação) com o arcabouço evolutivo provido pela filogenética (Korbel *et al.*, 2002; Lin e Gerstein, 2000; Sjölander, 2004).

A abordagem filogenômica em larga escala apresenta como principal vantagem a menor probabilidade em erros de análise derivados de fatores estocásticos, uma vez que a possibilidade de uso de um grande conjunto de dados (muitas vezes representado por milhares de genes e centenas de milhares de sítios), aumenta consideravelmente a consistência estatística de reconstruções e testes filogenéticos (Gee, 2003).

De fato, diversos estudos mostram que o uso de dados genômicos em larga escala beneficiam a resolução de métodos filogenéticos analíticos (Baptiste *et al.*, 2002, 2008; Murphy *et al.*, 2001; Qiu e Palmer, 1999; Rokas *et al.*, 2003; Soltis, Soltis e Chase, 1999). Em adição, a determinação de genes ortólogos entre diferentes espécies só pode ser completamente resolvida utilizando-se métodos que empreguem dados de genomas completos. Tal requisito é necessário pois a tentativa de inferência de ortólogos usando-se conjuntos de dados genômicos incompletos pode induzir a determinação de out-parálogos (Sonnhammer e Koonin, 2002) como ortólogos verdadeiros (Fulton *et al.*, 2006).

A presença de repetições no genoma inicialmente apresentou-se como um grande desafio aos sistemas responsáveis pela montagem de genomas. Clones oriundos de repetições pertencentes a diferentes loci genômicos podem agrupar-se como se pertencessem ao mesmo locus, gerando portanto quimeras e gaps durante a montagem genômica. A abordagem inicial consistia sempre na extração das repetições antes do processo de montagem, excluindo leituras de sequenciamento que representavam

repetições. Com isso, praticamente toda informação relativa a elementos de transposição era perdida durante o pré-processamento da montagem genômica, dificultando a detecção em larga escala de elementos de transposons. Uma alternativa a essa perda de informação consistia em separar as leituras de sequenciamento que continham informações de repetição e incluir as repetições posteriormente ao processo de montagem principal. Contudo tal procedimento nem sempre era realizado. Ademais, a região heterocromática de muitos genomas de organismos modelos não era incluída no projeto de sequenciamento, excluindo um rico repositório de elementos de transposição do processo de análise genômica posterior.

Com a adoção das tecnologias de sequenciamento em larga escala através de pirosequenciamento ou síntese, devido às peculiaridades do preparo da amostra e do processo de sequenciamento em si, as regiões heterocromáticas ganharam representatividade dentro dos projetos genoma (Hoskins *et al.*, 2007).

Contudo, tais tecnologias de sequenciamento produzem leituras com tamanho variando de 400 a 100 pares de base (Metzker, 2009), o que dificulta a montagem de repetições pertencentes a diferentes posições genômicas (Chaisson, Brinza e Pevzner, 2009), gerando portanto grandes falhas ao longo da montagem e inviabilizando a totalização do número de diferentes loci pertencentes a elementos de transposição (Treangen e Salzberg, 2012).

Apesar de tais dificuldades, o sequenciamento em larga escala de diferentes genomas, e a quantidade de genomas sequenciados permite atualmente a realização de screenings genômicos com um nível de detalhe impossível de ser realizado com técnicas

tradicionais de biologia molecular .

Dois métodos principais podem ser utilizados para a detecção de elementos de transposição: A detecção “ab initio” de repetições , com posterior refinamento dos resultados para diferenciar transposons de outras repetições genômicas e a busca por similaridade utilizando bibliotecas de sequência específicas para elementos de transposição.

As técnicas de detecção “ab initio” de elementos de transposição utilizam algoritmos específicos para a detecção de repetições nos dados genômicos, tratando a sequência de DNA como um sistema de informação onde estruturas como genes, regiões intergênicas, transposons e outras definições biológicas não são consideradas especificamente. Tais métodos geralmente empregam algoritmos que são otimizados para a determinação de repetições idênticas ou com alta similaridade. Apesar de úteis para a identificação de famílias de elementos de transposição que apresentam pouca divergência, algumas implementações (como no software Piler) apresentam dificuldade para identificar famílias de transposons mais divergentes (mais do que 90%). Contudo, se configurados para realizar a busca por repetições com parâmetros estridentes, as implementações de busca por repetições geralmente apresentam boa performance de busca, podendo identificar repetições em um grande número de sequências com pequenas exigências de memória e CPU. Tal característica é proeminente em softwares que se utilizam de árvores de sufixo como estrutura de dados para realizar a busca por repetições. Outras implementações que se valem da metodologia de busca por repetições para a identificação de elementos de transposição são Reputer (Kurtz e Schleiermacher, 1999), Recon (Bao e Eddy, 2002) e

Mummer (Delcher *et al.*, 2002).

A técnica de busca de similaridade utiliza bibliotecas específicas de sequências de elementos de transposição que são utilizadas por ferramentas de busca de sequências. Como exemplo notável de implementação desta metodologia temos o software RepeatMasker (Smit *et al.*, 1996). Contudo, é possível implementar essa técnica de forma independente apenas utilizando-se uma biblioteca específica de sequências de elementos de transposição, tanto construída a partir de buscas em bancos de dados genéricos (como o NCBI) ou bancos específicos como o RepBase. A partir desses bancos basta a implementação local do software de busca BLAST onde o banco de dados alvo deverá ser constituído pelo genoma no qual se deseja realizar a anotação de transposons e o banco de dados de transposon é utilizado como origem de busca. Após realizada a busca é necessária a interpretação da saída da busca de similaridade para identificar quais sequências encontradas possuem a necessária relevância para serem consideradas como uma sequência de transposon. Este método tem como vantagem a identificação e anotação de elementos já conhecidos e permite identificar elementos muito degenerados em relação às referências existentes no banco de dados de origem de busca. Do ponto de vista prático os requisitos de memória e CPU não são muito superiores aos de busca por repetições, contudo estes requisitos variam de acordo principalmente com o tamanho do banco de dados alvo. A atual implementação do software BLAST carrega todo o banco de dados alvo na memória RAM do computador antes da busca, logo, um banco de dados alvo maior do que a memória do computador onde será realizada a busca poderá impactar negativamente a performance de busca uma vez que o computador necessitará utilizar a memória swap para

executar a busca, sendo que a memória swap é ordens de magnitude mais lenta que a memória RAM principal.

A partir da identificação e anotação de diferentes sequências de elementos de transposição é possível utilizar tais dados para realizar experimentos e análises de inferência de padrões de evolução de diferentes componentes do genoma ou transferência horizontal em larga escala.

Assim como para a identificação de elementos de transposição, a análise de transferência horizontal apresenta dois principais tipos de métodos (Carvalho, de e Loreto, 2012). O primeiro tipo refere-se a identificação de transferência horizontal “ab initio” utilizando-se dados de composição de sequência para inferir padrões anômalos e assim determinar a presença de transferência horizontal. Também chamados de métodos indiretos, tais análises procuram identificar regiões genômicas onde a composição da sequência significativamente difere do restante do genoma em análise. Tal premissa é baseada na hipótese de que existe a manutenção de uma taxa constante e direcional de mutação ao longo do genoma de uma espécie, sendo esta taxa responsável por manter a uniformidade de composição em variáveis como porcentagem de nucleotídeos GC e de viés de uso de códons. Assim, regiões genômicas que apresentem valores de GC ou viés de uso de códon diferenciados em relação ao restante do genoma provavelmente foram adquiridas de outra espécie, submetida a outros padrões evolutivos, e portanto deve ter sido introduzida via transferência horizontal. Apesar de tal método ser útil para a identificação de ilhas de patogenicidade em genomas procariotos e também para realizar análises em espécies onde não existem dados suficientes para uma análise comparativa mais detalhada,

os métodos indiretos são especialmente sujeitos a retornar resultados falso positivo, uma vez que diferenças de composição podem acontecer em regiões específicas do genoma sem que estas tenham sofrido transferência horizontal, mas devido a outros processos evolutivos. Da mesma forma, sequências transferidas horizontalmente tornam-se sujeitas à mesma taxa mutacional do genoma hospedeiro e adquirem composição similar às sequências vizinhas em um fenômeno conhecido como amelioração (Wang, 2001).

Um segundo tipo de metodologia é a realização de análises filogenéticas específicas para a inferência de transferência horizontal. Estes métodos filogenéticos apresentam maior eficiência na detecção de transferência horizontal por serem baseados em premissas comparativas e na análise direta da presença de transferência horizontal de genes em relação à filogenia das espécies hospedeiras (Poptsova e Gogarten, 2007)

Um método filogenético relativamente robusto para a identificação de transferência horizontal é encontrado nas análises de incongruência de árvore, onde realiza-se a comparação direta entre a topologia da árvore de genes onde suspeita-se a existência de transferência horizontal e a árvore das espécies relativas aos genes em análise.

Contudo este método ainda não possui implementação para análise em larga escala, impossibilitando a identificação de eventos de transferência horizontal em comparações de vários genomas completos. Adicionalmente, efeitos de “lineage sorting” devidos deriva genética podem fazer com que incongruências na árvore filogenética de genes específicos sejam identificadas como transferência horizontal, resultando em falso positivos (Than *et al.*, 2006, 2007).

Neste trabalho apresentamos um um terceiro tipo de detecção de transferência

horizontal, através da análise comparativa em larga escala de índices filogenéticos. Este método une as vantagens dos métodos “ab initio” por permitir a análise de uma grande quantidade de sequências simultaneamente ao mesmo tempo que incorpora a robustez das análises filogenéticas ao trabalhar com índices derivados de análise comparativa. Esta análise introduz o conceito de identificação de transferência horizontal em larga escala pela comparação de variáveis filogenéticas anexadas a cada par de comparação de genes com similaridade mínima entre duas espécies (pertencentes ao um grupo de espécies em análise), sendo que para a realização do cálculo dessas variáveis filogenéticas é utilizado o conjunto completo de genes do grupo de espécies em análise, trazendo robustez ao processo de cálculo final. Este processo é especialmente relevante para a determinação de dados como K_a e K_s .

A teoria de evolução neutra (Kimura, 1979) preconiza que a divergência dentro de uma espécie é causada majoritariamente por efeitos de deriva genética de mutações neutras (Kimura, 1983). Neste cenário, a determinação das taxas de substituição sinônimas (K_s) e não-sinônimas (K_a) constitui-se uma importante ferramenta para a compreensão da dinâmica de evolução molecular de sequências (Yang e Nielsen, 2008). Contudo a estimativa dessas variáveis não é um problema simples (Anisimova, Bielawski e Yang, 2001), onde seja considerada a simples contagem de substituições sinônimas ou não-sinônima em determinada comparação de sequências. Processos como substituições múltiplas e viés de códon agem efetivamente interferem com a real estimativa dos valores de K_a e K_s ao violar as premissas utilizadas na determinação por simples contagem (Bierne e Eyre-Walker, 2003). Assim, atualmente aceita-se o método de cálculo através de

modelagem por verossimilhança, onde uma função de distribuição de substituições sinônimas e não-sinônimas é aplicada aos dados obtidos por alinhamento, permitindo uma aproximação mais realista das taxas de K_a e K_s (Zhang e Yu, 2006)

Objetivos

O objetivo desse trabalho foi realizar uma análise em larga escala dos elementos de transposição presentes em 12 genomas de *Drosophila* em comparação com os genes nucleares com o objetivo de identificar limitações e vantagens deste tipo de abordagem para a detecção de transferência horizontal em larga escala.

Especificamente, i) utilizar variáveis evolutivas como K_a e K_s para testar a hipótese de que elementos de transposição que tenham sofrido transferência horizontal apresentam consistentemente valores de K_s inferiores ao restante do genoma, ii) comparar os padrões evolutivos das regiões codificantes (CDS) entre genes ortólogos anotados nos genomas com as CDS dos elementos transponíveis.

Capítulo 2

Materiais e métodos

As sequências de 12 genomas de espécies de *Drosophila* (Tabela 1) foram obtidos via download a partir do banco de dados Flybase (Crosby *et al.*, 2007). Para cada genoma foi obtido em separado os arquivos de anotação GFF e as sequências de aminoácidos e nucleotídeos para cada gene anotado. Estes 12 genomas constituíram o banco de dados inicial a partir do qual foram realizadas tanto a determinação de ortólogos dos genes nucleares quanto a anotação de elementos transponíveis. Assim, para cada genoma foi construído um banco de dados relativos às sequências gênicas de nucleotídeos e outro banco de dados com as sequências gênicas traduzidas. Estes foram utilizados posteriormente para a determinação de ortólogos e demais análises. Define-se aqui o banco de dados do total de sequências gênicas de nucleotídeo dos 12 genomas de *Drosophila* como *dros-gnuc-1* e o banco de dados de sequências gênicas de aminoácidos como *dros-gpep-1*. O banco de dados de sequências de nucleotídeos correspondente ao total do genoma das 12 espécies de *Drosophila* é definido aqui como *dros-chromossomes-nuc*.

A fim de identificar os elementos de transposição nos genomas de *Drosophila* em análise com vistas ao posterior cálculo de variáveis como Ka e Ks que exigem o alinhamento códon a códon de sequências, foram construídos dois bancos de dados a partir do banco *dros-chromossomes-nuc*. Estes dois bancos foram criados pela extração de todas as ORFs maiores que 300 nucleotídeos, utilizando-se o software *getorf* do pacote EMBOSS (Rice,

Longden e Bleasby, 2000), presentes nos 12 genomas de *Drosophila* e sua posterior tradução. Estes dois bancos de dados são definidos aqui como *dros-orf-nuc-1* e *dros-orf-pep-1*. Ambos foram utilizados posteriormente para a anotação de elementos de transposição e análises evolutivas.

Tabela 1 – Sigla e designação das espécies utilizadas neste trabalho

Sigla	Espécie
Dana	<i>Drosophila ananassae</i>
Dere	<i>Drosophila erecta</i>
Dgri	<i>Drosophila grimshawi</i>
Dmel	<i>Drosophila melanogaster</i>
Dmoj	<i>Drosophila mojavensis</i>
Dper	<i>Drosophila persimilis</i>
Dpse	<i>Drosophila pseudobscura</i>
Dsec	<i>Drosophila sechellia</i>
Dsim	<i>Drosophila simulans</i>
Dvir	<i>Drosophila virilis</i>
Dwil	<i>Drosophila willistoni</i>
Dyak	<i>Drosophila yakuba</i>

Para a realização da anotação dos elementos de transposição foi empregada a abordagem de busca de similaridade baseada em uma extensa biblioteca de sequências de transposons previamente formatada. Esta biblioteca se constituiu pelo concatenamento do banco de dados de elementos de transposição canônicos determinados pelo Flybase (Crosby *et al.*, 2007) , pelo banco de dados Repbase (Jurka *et al.*, 2005) e por sequências obtidas a partir do banco de dados NT do NCBI via buscas específicas usando as palavras-chave “transposon”, “mobile genetic element”, “transposable element”, “retroelement” e “retrotransposon”. Ao final do processo de concatenação foi obtido um banco de dados com 925,311 sequências que foram submetidas a um processo de remoção de redundância utilizando-se o software CD-HIT (Fu *et al.*, 2012). Entradas com mais de 80% de

identidade ao longo de 80% da sequência foram removidas sendo substituídas por uma sequência consenso representante. Após o processo de remoção de redundância a biblioteca de elementos de transposição continha 246,839 sequências de nucleotídeo referentes a diferentes elementos de transposição de classe I e II.

Esta biblioteca de elementos de transposição foi utilizada como banco de dados para uma busca de similaridade utilizando o software BLAST contra o banco de sequências de ORFs dos 12 genomas de *Drosophila* *dros-orf-pep-1*. Os resultados desta busca que apresentaram ao menos 70% de identidade e valor e menor do que $1e-10$ foram considerados como positivos. Utilizando a referência das sequências com resultado positivo na busca supracitada foi realizado um mapeamento para identificar as respectivas sequências de nucleotídeo no banco de dados *dros-orf-nuc-1*. As sequências com identificação positiva como similares a sequências de elementos de transposição presentes na biblioteca usada para a busca foram extraídas tanto do banco *dros-orf-nuc-1* bem como do banco *dros-orf-pep-1* constituindo-se assim dois novos bancos de dados denominados *dros-te-orf-nuc-1* e *dros-te-orf-pep-1*.

Paralelamente foi conduzida a análise de determinação de ortólogos de genes derivados das anotações dos 12 genomas de *Drosophila*. Esta análise foi realizada utilizando-se o software Ortholuge (Fulton *et al.*, 2006), sendo utilizado como dado de entrada uma busca de comparação realizada através do software BLASTp onde todas as sequências de aminoácido de todos os genes anotados nos 12 genomas foram comparadas entre si. Ao final desta análise de ortologia foram selecionados 2,330 genes que correspondiam aos genes mais conservados presentes nos 12 genomas e com cópia única. Este resultado gerou

portanto o conjunto de “genes núcleo” de *Drosophila*, contendo aqueles genes que potencialmente são os mais conservados e sob maior pressão de seleção ao longo da linhagem das espécies em análise. Dois novos bancos de dados foram gerados a partir da extração das sequências dos genes ortólogos dos bancos de dados *dros-gnuc-1* e *dros-gpep-1*, denominados *dros-ortho-gnuc-1* e *dros-ortho-gpep-1*. Estes bancos de dados de sequências foram utilizados para a determinação dos valores de Ks, Ka, Ka/Ks, identidade de aminoácido, identidade de nucleotídeos e valores de viés de códon para cada sequência e comparação entre os 12 genomas. A determinação de Ka, Ks e Ka/Ks foi realizada utilizando-se o software *codeml* do pacote PAML (Yang, 2007). Para esta análise foi realizado o alinhamento entre cada grupo de ortólogo entre os 12 genomas utilizando-se o software *Muscle* (Edgar, 2004) inicialmente para as sequências de aminoácidos. Estes alinhamentos serviram como base para o alinhamento das sequências de nucleotídeos codon a codon utilizando-se o software *tranalign* do pacote EMBOSS (Rice, Longden e Bleasby, 2000). Este alinhamento final códon a códon foi então utilizado em conjunto com o software PAML para a determinação dos dados de Ka, Ks e Ka/Ks. Os dados de identidade de aminoácidos e de nucleotídeos foram extraídos dos alinhamentos prévios entre as sequências que constituíam os grupos de ortólogos. Os dados de viés de códon calculados através do índice de uso efetivo de códon ENC foram calculados utilizando-se o software *condonW* (Peden, 1999), onde também foram calculados os dados de porcentagem de GC, GC na terceira posição do códon, aromaticidade e hidropaticidade.

O mesmo procedimento foi realizado para as sequências de elementos de transposição anotadas nos 12 genomas de *Drosophila*. Contudo, considerando que a determinação de

ortólogos verdadeiros em elementos de transposição é ainda uma metodologia falha, optou-se por realizar um agrupamento de sequências utilizando-se o método estatístico implementado no software MCL (Enright, Dongen, Van e Ouzounis, 2002). Este método tem sido extensamente utilizado na determinação de grupos de proteínas relacionadas, contudo não é uma metodologia para a determinação de ortólogos verdadeiros, permitindo apenas um controle específico sobre a sensibilidade na determinação do tamanho dos grupos de sequências que integrarão os grupos finais de sequências similares. A partir desses dados de agrupamento de sequências de elementos de transposição realizado entre os 12 genomas foi executado o mesmo procedimento de cálculo de K_a , K_s , K_a/K_s , identidade de aminoácidos, identidade de nucleotídeo e os cálculos composicionais como viés de códon e porcentagem de GC na sequência

Todos os resultados das análises prévias sempre foram referenciados ao identificador de cada sequência, o que permitiu a construção de um banco de dados relacional na plataforma MySQL (<http://www.mysql.com>) que era alimentado a medida que os resultados de cada análise eram finalizados.

Após a finalização das análises acima descritas e construção do banco de dados relacional, um novo passo de anotação foi realizado onde os resultados prévios foram verificados manualmente em busca de sequências anotadas erroneamente como falso positivo. Esse segundo passo de anotação utilizou o software Google Refine (Huynh *et al.*, 2011) para identificar padrões de anotação específico na descrição de cada sequência que potencialmente poderiam indicar um erro de atribuição de família de elemento de transposição.

Por fim, tabelas específicas foram extraídas do banco de dados relacional e utilizadas para a realização de análises estatísticas implementadas através da linguagem R (R Development Core Team, 2011) ou SPSS (<http://www-01.ibm.com/software/analytics/spss/>). A manipulação de sequências e estruturação do *pipeline* foi realizada com o auxílio de *scripts* específicos em Python (Rossum, Van e Et Al., 2010) e BASH (Burtch, 2004).

Uma representação em forma de fluxograma das análises desenvolvidas neste trabalho pode ser encontrada na Figura 1.

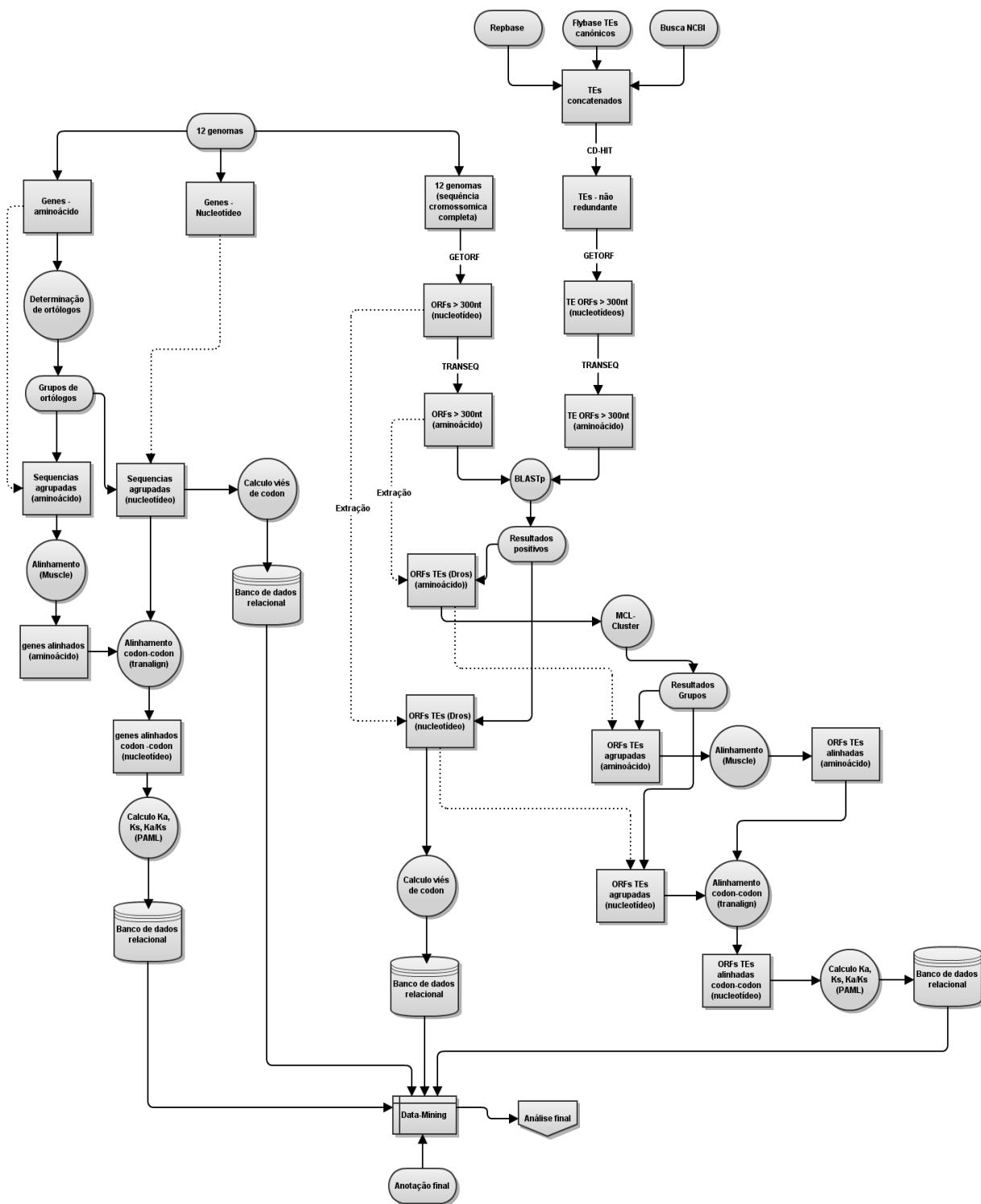


Figura 1 – Fluxograma das atividades realizadas para a obtenção do conjunto final de dados

Capítulo 3

Resultados e Discussão

Análise em larga escala de transferência horizontal

A presença de elementos de transposição no genoma de praticamente todos os organismos vivos indica a sua relevância como constituintes genômicos, tanto por serem promotores de efeitos deletérios (Johnson, 2007) quanto efetores de processos adaptativos (Oliver e Greene, 2012). Sua atividade em diferentes genomas também tem aplicações práticas uma vez que elementos de transposição são classificados como causa de ao menos 50 patologias em humanos, incluindo câncer, (Burns e Boeke, 2012; Kaer e Speek, 2013; Wilkins, 2010), são vetores importantes na aquisição de resistência a antibióticos em procariotos (Poirel *et al.*, 2009) e também ferramentas importantes dentro da emergente área de terapia gênica (Feschotte, 2006; Ohlfest, Ivics e Izsvák, 2009).

Considerando tais fatores, acompanhados da crescente disponibilidade de sequências derivadas de amplos projetos de sequenciamento, destaca-se a necessidade de análise compreensiva dos processos evolutivos e estruturais nos quais os elementos de transposição são participantes.

Em grande parte dos projetos genomas os elementos de transposição não são considerados o foco da análise, e pode-se verificar que o processo de anotação desse tipo de sequência geralmente não é conduzido de forma sistemática, prevalecendo apenas a anotação de outros genes.

A fim de explorar de forma efetiva a oportunidade de análise apresentada pelos dados

disponíveis é necessária a constituição de um sistema de colaboração entre a comunidade dedicada ao estudo dos elementos de transposição, de forma que a anotação de tais sequências seja realizada adequadamente. Tal esforço possibilitaria um incremento na eficiência de comunicação de resultados e evitaria a duplicação de esforços relativos a anotação e análise de sequências em diferentes genomas.

Um problema semelhante surgiu durante a estruturação dos primeiros projetos genoma em larga escala. A ausência de um sistema formalizado para nomeação e anotação de genes causava grande dificuldade para a comunicação efetiva de resultados de análises genômicas comparativas e também para o estabelecimento de novas metodologias de análise, especialmente com o crescente uso de técnicas de análise de expressão em larga escala (Rinaldi, 2009). Estas técnicas, como microarranjos, necessitavam de uma forma consistente de atribuir categorias funcionais e outras classificações aos diferentes genes que estavam sendo descobertos através do sequenciamento de novos genomas (Plessis, du, Skunca e Dessimoz, 2011)

Assim criou-se o projeto Gene Ontology (GO) (Ashburner *et al.*, 2000) com o objetivo de desenvolver um vocabulário controlado específico na forma de uma ontologia que possibilitaria a formalização do processo de anotação gênica entre diferentes projetos genoma e a transferência efetiva de conhecimento entre trabalhos de anotação. Juntamente com o sistema GO surgiu uma série de novos *softwares* e algoritmos que permitem a análise integrada de uma grande quantidade de genes, valendo-se da consistência do sistema de classificação permitido pela ontologia (Chen *et al.*, 2009; Kofler e Schlötterer, 2012; Liu, Hu e Wu, 2005; Tchagang *et al.*, 2010).

As facilidades do sistema de ontologia para a classificação de entidades biológicas mostraram-se adequadas não só a genes mas também a outros tipos de dados, de forma que hoje existem diferentes ontologias para áreas como biologia de sistemas, proteômica e rotas metabólicas (Ai, Smith e David, 2010; Hoehndorf, Kelso e Herre, 2009; Jensen e Bork, 2010; Lambrix *et al.*, 2007).

Como vantagens do uso de um sistema de ontologias destaca-se a possibilidade de atribuir-se relações de dados tanto hierárquicas quanto horizontais, o que permitiria resolver problemas típicos relativos a classificação de elementos de transposição, como a relação entre elementos autônomos e MITEs (Jensen e Bork, 2010). Como estruturas de dados adicionais, é possível através de uma ontologia também modelar eventos, como por exemplo, a existência de transferência horizontal específica para elementos de transposição.

É importante ressaltar que um forte apoio da comunidade científica dedicada à área da ontologia em desenvolvimento é crucial para seu sucesso e adoção (Hoehndorf *et al.*, 2009). Diferentemente de sistemas de classificação estanques, uma ontologia é dinâmica e evolui a medida que novas descobertas e paradigmas são desenvolvidos. Tal dinâmica é própria da modelagem de sistemas em ontologias e suas estruturas de dados são apropriadas para serem descontinuadas e ligadas à novas classes de entidades biológicas sem que seja necessário reescrever continuamente novos sistemas de classificação (Liu, Hu e Wu, 2005).

Uma vez estabelecida uma ontologia que tenha suporte da comunidade métodos de análise de transposons em larga escala se beneficiarão amplamente da troca de informações

através de um sistema consistente.

Contudo, um segundo desafio ainda precisa ser vencido no caso específico de análise de transferência horizontal. Como apresentado durante a introdução, os métodos indiretos de medida de transferência horizontal, que permitem realizar análises em escala genômica de transferência horizontal de forma relativamente acessível, são especialmente sujeitos a produzir resultados falso positivo (Azad e Lawrence, 2011). Já os métodos filogenéticos, apesar de mais robustos, também estão sujeitos a apresentar resultados falso positivos ao serem alimentados com dados de genes sujeitos a “incomplete lineage sorting” (Than *et al.*, 2006). Da mesma forma, tais métodos são difíceis de implementar em larga escala. Tanto os métodos indiretos e os métodos filogenéticos possuem diferentes implementações e algoritmos, uma revisão a respeito dos mesmos foi produzida durante a realização deste trabalho (Carvalho, de e Loreto, 2012) e pode ser encontrada como Anexo 1.

Análise comparativa entre 12 genomas de Drosophila e seus elementos de transposição

O sequenciamento completo de 12 genomas do gênero *Drosophila* oferece uma oportunidade de análise de padrões das propriedades evolutivas de sequências em uma escala sem precedentes. Esta escala oferece uma visão global do comportamento de processos evolutivos e auxilia na geração de novas hipóteses e a confirmação de forma ampla de hipóteses previamente testadas em uma quantidade limitada de dados (Singh *et al.*, 2009).

Neste trabalho a análise comparativa entre 2,330 genes de *Drosophila* e 342,047 sequências de elementos de transposição, gerou o total 169,087 pares de ortólogos e 368,823 pares de elementos de transposição (Figura 2).

A exceção é para a análise de viés de códon, uma vez que a variável ENC, que representa o número efetivo de códons e não é uma medida derivada de comparação, foi calculada para cada sequência independentemente. A fim de acoplar um valor de viés de códon a cada comparação, foi calculada a média da variável ENC para cada par de sequências analisadas.

O conjunto final de dados corresponde a 537,910 pontos de dados que em conjunto com as 20 variáveis (Tabela 2) totalizam 10,758,200 valores.

Um sumário dos dados obtidos para cada variável é apresentado nas Tabelas 7,8,9 e 11, assim como na tabela incluída no material anexo e na tabela eletrônica que pode ser acessada pelo endereço <http://goo.gl/pLkob> (forma expandida do link:

<https://www.dropbox.com/s/x0nxiwftsckul76/marcodecarvalho-doutorado-summarydata.xl>

s)

Tabela 2 – Variáveis utilizadas neste trabalho e sua respectiva definição.

<i>Variáveis</i>	<i>Definição</i>
Ka	Taxa de substituição não-sinônima
Ks	Taxa de substituição sinônima
Ka/Ks	Relação entre Ka e Ks
ProtId	Identidade da sequência de aminoácidos em %
DNAId	Identidade da sequência de nucleotídeos em %
Logdet	Distância logdet medida a partir de sequência de nucleotídeos
Kimura	Distância Kimura medida a partir de sequência de aminoácidos
T3	Porcentagem de nucleotídeos T na terceira posição do códon da sequência de nucleotídeo
C3	Porcentagem de nucleotídeos C na terceira posição do códon da sequência de nucleotídeo
A3	Porcentagem de nucleotídeos A na terceira posição do códon da sequência de nucleotídeo
G3	Porcentagem de nucleotídeos G na terceira posição do códon da sequência de nucleotídeo
GC3	Porcentagem de nucleotídeos G e C na terceira posição do códon da sequência de nucleotídeo
GC	Porcentagem de nucleotídeos GC na da sequência completa de nucleotídeo
L sym	Número de codons sinônimos ao longo da sequência de nucleotídeos
l aa	Tamanho da sequência em aminoácidos
Nca	Valor do número efetivo de códons para a primeira sequência do par de ortólogos
Ncb	Valor do número efetivo de códons para a segunda sequência do par de ortólogos
Aromo	Frequência de aminoácidos aromáticos
Gravy	Média do índice de hidropaticidade para cada aminoácido
Index	Índice de similaridade

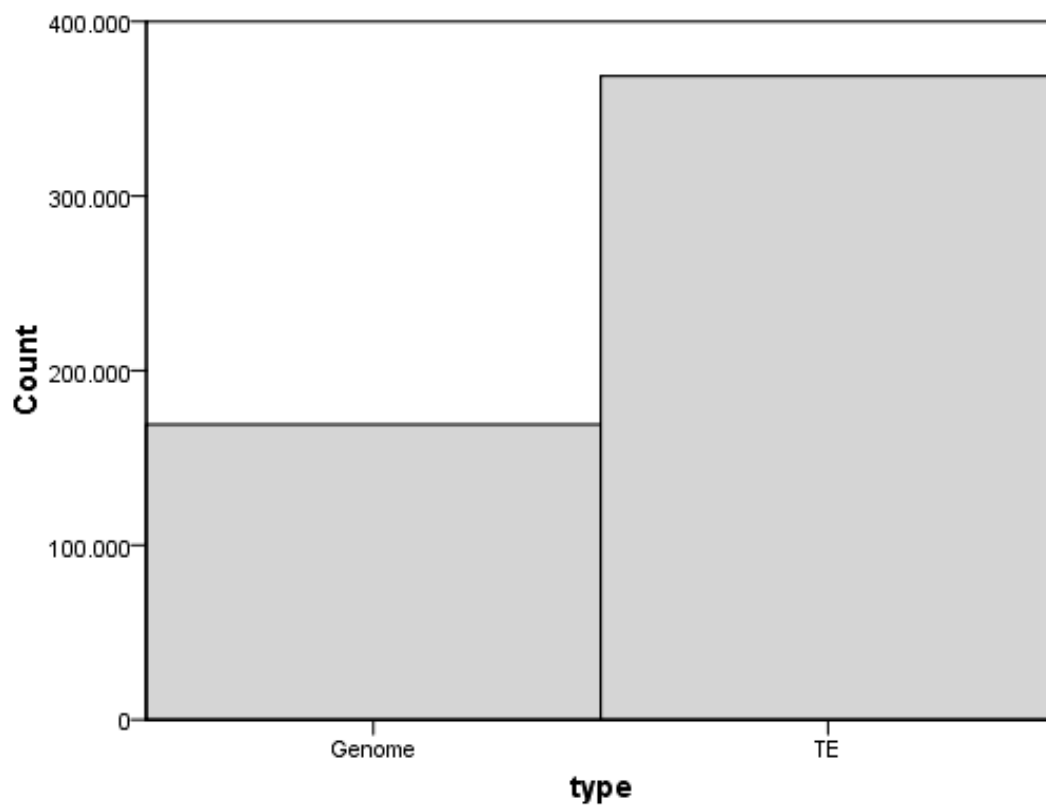


Figura 2 – Histograma da distribuição do número de comparações entre elementos transponíveis e genes para 12 espécies de *Drosophila* em análise.

As comparações de elementos de transposição entre genomas dos grupos *Melanogaster* e entre os genomas do grupo *Obscura* apresentam-se em maior número dentro do conjunto de dados final (Tabela 3 e Tabela 4), conforme é possível observar na Figura 3.. Esta preponderância de comparações segue a maior proporção de sequências anotadas dentro do Grupo *Melanogaster* e também a maior proximidade entre os genomas do grupo *Obscura*. É importante salientar que pares de comparação entre elementos de transposição com similaridade de DNA menor do que 70% foram excluídos da análise após a determinação dos grupos de similaridade entre genomas.

Esta medida foi adotada para evitar que erros devidos a saturação sejam inseridos assim como prevenir falsos positivos no processo de anotação. Seis grupos de comparações não estão presentes na análise final de elementos de transposição uma vez que as sequências com similaridade encontradas não satisfaziam o critério de possuírem ao menos 70% de similaridade entre si. São estes, dgri-dere, dgri-dmoj, dgri-dwil, dgri-dyak, dmoj-dere, dwil-dere. Nota-se que nenhuma das espécies pertencentes a estes pares de comparação possuem sobreposição geográfica entre si (Singh *et al.*, 2009).

Tabela 3 – Número de comparações para genoma e elementos de transposição

		Tipo	
		Genoma	TE
Grupo	Melanogaster_Group	12821	6995
	Melanogaster_Subgroup	25717	284147
	Melanogaster-Hawaiian	2557	128
	Melanogaster-Obscura	5117	6860
	Melanogaster-Repleta	2560	5479
	Melanogaster-Virilis	2560	1545
	Melanogaster-Willistoni	2560	7645
	MelanogasterSubgroup-Hawaiian	12795	3
	MelanogasterSubgroup-Obscura	25613	2521
	MelanogasterSubgroup-Repleta	12804	1485
	MelanogasterSubgroup-Virilis	12799	365
	MelanogasterSubgroup-Willistoni	12811	2177
	Obscura Group	2561	24636
	Obscura-Hawaiian	5111	2
	Repleta-Hawaiian	2562	0
	Repleta-Obscura	5114	5090
	Virilis-Hawaiian	2560	2
	Virilis-Obscura	5112	2021
	Virilis-Repleta	2563	2232
	Willistoni-Hawaiian	2558	0
Willistoni-Obscura	5115	7659	
Willistoni-Repleta	2559	6802	
Willistoni-Virilis	2558	1029	

Tabela 4 – Número de comparações entre sequências de elementos de transposição para cada par de genoma analisado.

<i>Comparação</i>	<i>Quantidade</i>	<i>Comparação</i>	<i>Quantidade</i>	<i>Comparação</i>	<i>Quantidade</i>	<i>Comparação</i>	<i>Quantidade</i>
dana__dere	3723	dgri__dsim	1	dmoj__dvir	2232	dpse__dyak	58
dana__dgri	128	dgri__dvir	2	dmoj__dwil	6802	dsec__dere	9396
dana__dmel	1783	dmel__dere	12008	dmoj__dyak	157	dsec__dsim	45329
dana__dmoj	5479	dmel__dmoj	794	dper__dere	2	dsec__dvir	90
dana__dper	6516	dmel__dper	791	dper__dpse	24636	dsec__dwil	772
dana__dpse	344	dmel__dpse	121	dper__dsec	752	dsec__dyak	7163
dana__dsec	800	dmel__dsec	155609	dper__dsim	105	dsim__dere	3535
dana__dsim	206	dmel__dsim	36725	dper__dvir	1273	dsim__dvir	28
dana__dvir	1545	dmel__dvir	215	dper__dwil	7075	dsim__dwil	121
dana__dwil	7645	dmel__dwil	996	dper__dyak	619	dsim__dyak	2456
dana__dyak	483	dmel__dyak	6403	dpse__dere	1	dvir__dere	27
dgri__dmel	1	dmoj__dper	4678	dpse__dsec	59	dvir__dwil	1029
dgri__dper	1	dmoj__dpse	412	dpse__dsim	13	dvir__dyak	5
dgri__dpse	1	dmoj__dsec	448	dpse__dvir	748	dwil__dyak	288
dgri__dsec	1	dmoj__dsim	86	dpse__dwil	584	dyak__dere	5523

O número de comparações para cada par de genes ortólogos no entanto é aproximadamente constante (11 comparações por gene) uma vez que a determinação de ortólogos foi restrita a genes de cópia única e presentes nas 12 espécies de *Drosophila* em análise. Contudo, o número final total de comparações (Tabela 5) entre genomas apresenta variações devido a certos pares de genes apresentarem divergência de sequência de DNA menor do que 50%. Estes pares de comparação foram excluídos para evitar erros no cálculo das variáveis evolutivas devido a saturação (Smith e Smith, 1996).

O gráfico de barras da Figura 4 apresenta a distribuição do número de comparações para cada grupo de espécies.

Analisando o número de comparações de acordo com cada família de elemento de transposição anotada é possível verificar que elementos de Classe II são o mais representados, sendo que os elementos R1, FW e Tabor são os que possuem o maior número de comparações no conjunto de dados final (Tabela 6). No total foram identificadas 116 diferentes anotações de elementos de transposição.

Considerando que o número de comparações é diretamente ligado à similaridade de sequência dentro de determinada família de elemento de transposição bem como em relação ao número de cópias existente, a diferença observada está em acordo com trabalhos prévios que indicam maior taxa de transposição e número de cópias em elementos de Classe I (Scheifele *et al.*, 2009).



Figura 4 –Gráfico de barras do número de comparações de elementos de transposição e de genes ortólogos de acordo com cada grupo de espécies.

Tabela 5 - Número de comparações entre sequências de genes para cada par de genoma analisado.

Comparação	Quantidade	Comparação	Quantidade	Comparação	Quantidade
dana___dere	2564	dmel___dmoj	2579	dper___dyak	2556
dana___dgri	2557	dmel___dper	2579	dpse___dere	2562
dana___dmel	2581	dmel___dpse	2584	dpse___dsec	2562
dana___dmoj	2560	dmel___dsec	2588	dpse___dsim	2547
dana___dper	2557	dmel___dsim	2586	dpse___dvir	2558
dana___dpse	2560	dmel___dvir	2577	dpse___dwil	2559
dana___dsec	2562	dmel___dwil	2585	dpse___dyak	2562
dana___dsim	2552	dmel___dyak	2589	dsec___dere	2562
dana___dvir	2560	dmoj___dere	2561	dsec___dsim	2560
dana___dwil	2560	dmoj___dper	2556	dsec___dvir	2561
dana___dyak	2562	dmoj___dpse	2558	dsec___dwil	2558
dgri___dere	2559	dmoj___dsec	2560	dsec___dyak	2563
dgri___dmel	2582	dmoj___dsim	2546	dsim___dere	2559
dgri___dmoj	2562	dmoj___dvir	2563	dsim___dvir	2544
dgri___dper	2552	dmoj___dwil	2559	dsim___dwil	2545
dgri___dpse	2559	dmoj___dyak	2558	dsim___dyak	2559
dgri___dsec	2556	dper___dere	2558	dvir___dere	2559
dgri___dsim	2542	dper___dpse	2561	dvir___dwil	2558
dgri___dvir	2560	dper___dsec	2558	dvir___dyak	2558
dgri___dwil	2558	dper___dsim	2545	dwil___dere	2564
dgri___dyak	2556	dper___dvir	2554	dwil___dyak	2559
dmel___dere	2587	dper___dwil	2556	dyak___dere	2564

Tabela 6 – Número de comparações de elementos de transposição de acordo com cada família anotada.

<i>Familia</i>	<i>Quantidade</i>	<i>Familia</i>	<i>Quantidade</i>	<i>Familia</i>	<i>Quantidade</i>	<i>Familia</i>	<i>Quantidade</i>
ACCORD	2251	DOC2	579	M4DM	21	TAHRE	42
AURORA	1193	DOC3	982	MARINER	4814	TART	433
BAGGINS	3330	FROGGER	6	MAX	19193	TC1	12
BATUMI	220	FW	61301	MDG1	4954	TE50134	97
BEL	7084	FW2	65	MDG3	1157	TE50157	61
BICA	365	G3-JOCKEY	6	MICROPIA	489	TE50164	5
BILBO	316	G5	4788	MINIME	311	TE50200	12
BLASTOPIA	2477	GALILEO	11	NETR1B	619	TE50218	2
BLOOD	3997	GATE	22	NINJA	2487	TE50219	25
BS	16	GTWIN	6098	NOBEL-BEL	3848	TE50222	8
BS2	3258	GYPSY	21719	NOMAD	1041	TE50237	1
BS3	50	HARBINGER	8	OPUS	3	TE50238	30
BURDOCK	453	HAT	91	OSVALDO	15	TELEMAC	15
BUT2	34	HELENA	49	PAO	183	TF001011	8
CHIMPO	1356	HELITRON	2157	P_ELEMENT	29	TF001012	15
CHOUTO	193	HETA	62	PIF	14	TIANG	6
CIRC	897	HMSBEAGLE	1151	PIFO	50	TINKER	170
CIRCE	63	HOANA	19	POLINTON	487	TIRANT	501
COPIA	4274	HOBO	798	PROTOP	33	TRAM-BEL	227
CR1	3541	HOBO-LIKE	3	QUASIMODO	3243	TRANSIB	181
DD37E	30	HOYAK	23	QUETZAL	10	TRANSPAC	214
DEC1	6	IDEFIX	504	R1	62430	TRIM	1323
DIVER	2963	INVADER	6527	R2	3783	ULYSSES	2
DIVER-BEL	6355	ISIS	1	ROO	2	WORF	3337
DM1731	3074	IVK	237	ROOA	9298	YOYO	27
DM2	1893	JOCKEY	113	ROVER	76	ZAM	334
DM412	7660	JOCKEY-LIKE	132	SPOCK	51		
DMR_DV	29	LIAN-AAL-LIKE	13	SPRINGER	3		
DMRT	25265	LINEJ1	446	STALKER	4611		
DNAREP1	1171	LINER1	90	TABOR	46665		

A distribuição dos valores de todas as variáveis analisadas mostraram-se significativamente diferentes entre o conjunto de dados relativos aos genes e as sequências de elementos de transposição (teste de amostras independentes Mann-Whitney U, $p < 0.05$).

Os dados derivados de relações entre genes foram sistematicamente menos esparsos quando comparados entre si, sendo possível inferir correlações entre as variáveis estudadas de forma mais sólida. Já os dados derivados de comparações entre elementos de transposição apresentaram considerável variação, sendo esta evidenciada quando realizada a comparação entre variáveis na forma de gráficos de diagrama de caixa e gráficos de dispersão (Figuras 5 a 11 e Figuras 24 a 38).

Como principal padrão observável na comparação entre sequências de elementos de transposição e as sequências dos genes nucleares dos genomas analisados é o menor valor de taxas de substituição sinônima em transposons em relação aos genes (Figuras 7, 14 e 21).

Espera-se que as sequências de elementos de transposição evoluam de forma mais neutra que os genes nucleares (Wallau, Ortiz e Loreto, 2012), uma vez que estes últimos estão de forma geral sob pressão de seleção negativa e genes de elementos de transposição não contribuem diretamente para o *fitness* do hospedeiro. Este padrão pode ser evidenciado na comparação entre as Figuras 12 e 13 onde temos uma quantidade inexpressiva de comparações com valores de Ka/Ks superior a 1 (seleção positiva) para sequências gênicas em relação às comparações de sequências de elementos de transposição.

Ao mesmo tempo, as sequências de elementos de transposição apresentam

similaridade média de sequências de DNA superior aos genes nucleares (Figuras 5, 14 e 23), porém com similaridade média das sequências de aminoácido inferior (Figura 6). Esta diferença é refletida nos valores diferenciados de distância Logdet (DNA) (Figura 22) e Kimura (proteína) que seguem o mesmo padrão dos dados de similaridade de sequência (Figuras 10 e 11).

A comparação entre as Figuras 19 e 20 mostra que de forma geral, os elementos de transposição possuem valores de K_s menores que o genoma ao mesmo tempo que possuem valores de K_a superiores ao genoma. Esta diferença entre as taxas de substituição sinônima e não-sinônima pode explicar o padrão de diferença de similaridade entre as sequências de DNA e proteína, uma vez que uma substituição não-sinônima reduz a similaridade da sequência de proteína três vezes mais que uma substituição sinônima reduz a similaridade de uma sequência de DNA.

Contudo, verifica-se que as sequências que possuem maiores taxas de substituição não-sinônima são também as sequências com maior similaridade entre as comparações, mostrando que existe um processo de rápida evolução das sequências de elementos de transposição (Figura 15). Esta rápida evolução é evidenciada pelo grande número de comparações que resultam em valores de K_a/K_s acima de 1, ao mesmo tempo que apresentam valores baixos de K_s (Figura 16), K_a (Figura 17) e significativa similaridade de sequência (Figura 18).

Tal resultado pode ser esperado tomando-se o modelo do ciclo de vida de elementos de transposição (Schaack, Gilbert e Feschotte, 2010), notadamente diferenciado dos genes nucleares, sendo sujeitos a uma série de processos de repressão de atividade (Johnson,

2008) bem como transferência horizontal (Loreto, Carareto e Capy, 2008). Assim, genes transferidos horizontalmente teriam valores de Ks significativamente inferiores aos genes nucleares do genoma hospedeiro em comparação ao genoma de origem, uma vez que os valores de Ks podem ser tomados como um indicador da taxa de evolução neutra e portanto do tempo de divergência entre as sequências analisadas (Lanfear, Welch e Bromham, 2010; Wallau, Ortiz e Loreto, 2012).

Contudo também observa-se uma subpopulação destas sequências de elementos de transposição com valores de Ks muito abaixo dos valores de Ks observados em genes nucleares juntamente com valores de Ka/Ks muito altos, o que indica um processo de seleção positiva (Figura 13). Considerando a escala encontrada de sequências de elementos de transposição sob seleção positiva e alta similaridade é possível sugerir que se trata de um padrão inerente ao ciclo de vida dos mesmos.

Apesar de processos de seleção positiva em elementos de transposição geralmente estarem relacionados à cooptação de genes de TEs e neofuncionalização, este é um processo raro (Kidwell e Lisch, 2000) e processos de seleção positiva como parte integral do ciclo de vida de um transposon já foram demonstrados para o elemento Copia (Setta, *et al.*, 2011).

Tabela 7 – Valores de Ks, Ka e Ka/Ks organizados de acordo com cada comparação entre genomas de *Drosophila* para sequências gênicas e de elementos de transposição.

		Genoma			TE		
		ks	ka	kaks	ks	ka	kaks
		Média				Média	
Melanogaster_Group	dana_dere	2.12	0.08	0.04	0.42	0.03	51.24
	dana_dmel	2.14	0.08	0.04	1.49	0.10	1.40
	dana_dsec	2.16	0.08	0.04	0.85	0.10	1.28
	dana_dsim	2.14	0.08	0.04	0.77	0.12	0.55
	dana_dyak	2.08	0.08	0.04	2.96	0.11	15.37
Melanogaster_Subgroup	dmel_dere	0.29	0.02	0.07	0.12	0.08	25.86
	dmel_dsec	0.14	0.01	0.08	0.11	0.03	43.05
	dmel_dsim	0.14	0.01	0.08	0.13	0.03	36.62
	dmel_dyak	0.31	0.02	0.06	0.19	0.05	32.94
	dsec_dere	0.28	0.02	0.08	0.21	0.08	20.54
	dsec_dsim	0.07	0.01	0.13	0.08	0.04	36.41
	dsec_dyak	0.30	0.02	0.07	0.12	0.05	24.85
	dsim_dere	0.28	0.02	0.08	0.20	0.07	16.60
	dsim_dyak	0.30	0.02	0.07	0.12	0.05	26.43
	dyak_dere	0.22	0.02	0.07	0.13	0.11	32.80
Melanogaster-Hawaiian	dana_dgri	6.27	0.14	0.03	0.83	0.12	0.14
Melanogaster-Obscura	dana_dper	3.60	0.10	0.04	3.43	0.18	11.83
	dana_dpse	3.48	0.10	0.04	1.40	0.11	92.51
Melanogaster-Repleta	dana_dmoj	6.58	0.14	0.03	1.05	0.11	43.85
Melanogaster-Virilis	dana_dvir	6.17	0.13	0.03	1.44	0.10	23.97
Melanogaster-Willistoni	dana_dwil	6.93	0.13	0.03	1.04	0.16	20.35
MelanogasterSubgroup-Hawaiian	dgri_dere	5.30	0.14	0.04			
	dgri_dmel	5.14	0.14	0.04	2.63	0.09	0.04
	dgri_dsec	5.28	0.14	0.04	2.95	0.09	0.03
	dgri_dsim	5.22	0.14	0.04	2.68	0.09	0.03
	dgri_dyak	5.17	0.14	0.04			
MelanogasterSubgroup-Obscura	dmel_dper	3.05	0.10	0.04	5.88	0.24	3.82
	dmel_dpse	2.94	0.09	0.04	2.39	0.11	0.20
	dper_dere	2.96	0.10	0.04	1.87	0.10	0.06
	dper_dsec	3.07	0.10	0.04	4.77	0.18	1.63
	dper_dsim	3.00	0.10	0.04	4.97	0.22	2.89
	dper_dyak	2.93	0.10	0.04	3.83	0.11	1.62
	dpse_dere	2.88	0.09	0.04	2.13	0.03	0.01
	dpse_dsec	2.98	0.10	0.04	2.65	0.12	0.16
	dpse_dsim	2.92	0.10	0.04	3.42	0.14	0.17
	dpse_dyak	2.86	0.09	0.04	4.10	0.08	0.04
MelanogasterSubgroup-Repleta	dmel_dmoj	5.64	0.14	0.04	0.74	0.08	86.09
	dmoj_dere	5.73	0.14	0.04			
	dmoj_dsec	5.83	0.14	0.04	0.92	0.12	67.36
	dmoj_dsim	5.73	0.14	0.04	0.76	0.09	109.10

	dmoj	dyak	5.67	0.14	0.04	1.90	0.18	0.12
MelanogasterSubgroup-Virilis	dme1	dvir	5.15	0.13	0.04	0.84	0.03	22.47
	dsec	dvir	5.20	0.14	0.04	0.84	0.04	23.70
	dsim	dvir	5.16	0.14	0.04	0.76	0.05	17.37
	dvir	dere	5.18	0.13	0.04	0.35	0.09	0.31
	dvir	dyak	5.04	0.13	0.04	2.29	0.24	0.10
MelanogasterSubgroup-Willistoni	dme1	dwil	6.26	0.13	0.03	1.17	0.19	82.38
	dsec	dwil	6.51	0.13	0.03	1.44	0.17	39.05
	dsim	dwil	6.51	0.13	0.03	1.13	0.18	75.22
	dwil	dere	6.44	0.13	0.03			
	dwil	dyak	6.24	0.13	0.03	2.90	0.17	0.07
Obscura Group	dper	dpse	0.04	0.01	0.16	0.11	0.07	32.37
Obscura-Hawaiian	dgri	dper	3.21	0.14	0.05	3.30	0.08	0.02
	dgri	dpse	3.09	0.13	0.05	3.29	0.08	0.02
Repleta-Hawaiian	dgri	dmoj	1.30	0.09	0.07			
Repleta-Obscura	dmoj	dper	3.44	0.14	0.05	3.54	0.25	14.67
	dmoj	dpse	3.32	0.13	0.05	0.70	0.10	44.60
Virilis-Hawaiian	dgri	dvir	0.98	0.07	0.07	1.36	0.12	0.09
Virilis-Obscura	dper	dvir	2.95	0.13	0.05	4.36	0.28	7.46
	dpse	dvir	2.87	0.13	0.05	0.24	0.09	26.68
Virilis-Repleta	dmoj	dvir	0.93	0.06	0.07	1.01	0.07	53.65
Willistoni-Hawaiian	dgri	dwil	3.87	0.15	0.05			
Willistoni-Obscura	dper	dwil	4.35	0.13	0.04	1.68	0.19	7.37
	dpse	dwil	4.19	0.12	0.04	0.65	0.23	28.64
Willistoni-Repleta	dmoj	dwil	4.69	0.14	0.04	0.82	0.23	61.25
Willistoni-Virilis	dvir	dwil	4.10	0.14	0.05	1.15	0.21	35.99

Tabela 8:- Valores de identidade de proteína (protid), identidade de nucleotídeo (dnaid), distancia logdet (logdet) e distancia kimura (kimura) organizados de acordo com cada comparação entre genomas de *Drosophila* para sequências gênicas e de elementos de transposição.

		Genoma				TE			
		protid	dnaid	logdet	kimura	protid	dnaid	logdet	kimura
		Média				Média			
Melanogaster_Group	dana__dere	85.55	78.85	0.27	0.16	93.69	96.15	0.08	0.07
	dana_dmel	85.60	78.68	0.27	0.16	75.64	79.64	0.37	0.29
	dana_dsec	85.17	78.56	0.29	0.17	75.75	79.25	0.37	0.30
	dana_dsim	85.23	78.60	0.31	0.17	75.65	78.97	0.37	0.30
	dana_dyak	85.65	78.86	0.27	0.16	76.96	80.78	0.42	0.28
Melanogaster_Subgroup	dmel_dere	95.22	92.86	0.09	0.05	85.07	89.62	0.20	0.15
	dmel_dsec	97.15	95.98	0.06	0.03	89.16	92.46	0.14	0.10
	dmel_dsim	97.06	95.88	0.09	0.03	89.32	92.60	0.15	0.10
	dmel_dyak	95.30	92.62	0.09	0.05	87.68	91.78	0.17	0.12
	dsec_dere	94.79	92.83	0.10	0.05	84.28	88.56	0.24	0.17
	dsec_dsim	97.87	97.52	0.08	0.02	90.64	94.05	0.13	0.10
	dsec_dyak	94.84	92.54	0.10	0.05	88.88	92.80	0.14	0.12
	dsim_dere	94.68	92.71	0.12	0.05	85.42	90.04	0.21	0.16
	dsim_dyak	94.78	92.46	0.12	0.05	88.56	92.75	0.15	0.12
	dyak_dere	96.15	94.31	0.07	0.04	85.59	90.76	0.18	0.17
Melanogaster-Hawaiian	dana_dgri	76.93	71.72	0.37	0.28	79.05	79.43	0.32	0.24
Melanogaster-Obscura	dana_dper	81.83	76.08	0.33	0.21	75.93	80.33	0.34	0.27
	dana_dpse	82.30	76.42	0.31	0.20	74.15	79.61	0.36	0.30
Melanogaster-Repleta	dana_dmoj	77.21	71.92	0.37	0.28	77.42	80.85	0.36	0.26
Melanogaster-Virilis	dana_dvir	78.03	72.42	0.36	0.26	75.72	79.82	0.37	0.29
Melanogaster-Willistoni	dana_dwil	78.76	71.04	0.38	0.25	76.68	81.51	0.34	0.26
MelanogasterSubgroup-Hawaiian	dgri_dere	76.95	72.16	0.37	0.28				
	dgri_dmel	76.98	72.17	0.37	0.28	76.47	72.75	0.40	0.29
	dgri_dsec	76.68	72.01	0.38	0.28	75.66	72.92	0.40	0.30
	dgri_dsim	76.79	72.09	0.40	0.28	75.95	73.12	0.41	0.30
	dgri_dyak	77.01	72.24	0.37	0.28				
MelanogasterSubgroup-Obscura	dmel_dper	82.07	76.45	0.32	0.21	74.47	79.13	0.34	0.31
	dmel_dpse	82.47	76.75	0.30	0.20	74.88	79.52	0.35	0.30
	dper_dere	82.06	76.67	0.32	0.21	77.91	78.41	0.33	0.27
	dper_dsec	81.74	76.41	0.34	0.21	74.39	79.64	0.34	0.31
	dper_dsim	81.83	76.48	0.35	0.21	74.58	79.50	0.34	0.31
	dper_dyak	82.14	76.70	0.32	0.21	77.53	79.10	0.42	0.27
	dpse_dere	82.48	76.97	0.30	0.20	85.39	79.85	0.28	0.16
	dpse_dsec	82.14	76.70	0.31	0.21	74.73	79.51	0.35	0.30
	dpse_dsim	82.28	76.80	0.33	0.20	76.40	79.97	0.34	0.29

	dpse	dyak	82.54	76.99	0.30	0.20	78.50	78.31	0.51	0.35
MelanogasterSubgroup-Repleta	dmel	dmoj	77.25	72.19	0.37	0.28	75.60	79.88	0.35	0.30
	dmoj	dere	77.22	72.28	0.37	0.28				
	dmoj	dsec	76.92	72.06	0.38	0.28	74.80	78.80	0.38	0.32
	dmoj	dsim	77.02	72.12	0.40	0.28	75.38	79.68	0.36	0.31
	dmoj	dyak	77.28	72.33	0.37	0.27	74.39	77.25	0.42	0.32
MelanogasterSubgroup-Virilis	dmel	dvir	78.04	72.78	0.36	0.26	73.69	78.05	0.37	0.33
	dsec	dvir	77.70	72.64	0.37	0.27	73.98	78.39	0.36	0.33
	dsim	dvir	77.83	72.71	0.39	0.27	75.40	79.93	0.33	0.30
	dvir	dere	78.03	72.84	0.36	0.26	81.71	87.03	0.15	0.21
	dvir	dyak	78.08	72.89	0.36	0.26	73.70	79.77	0.43	0.34
MelanogasterSubgroup-Willistoni	dmel	dwil	78.75	71.34	0.38	0.25	74.35	79.43	0.36	0.31
	dsec	dwil	78.43	71.11	0.39	0.26	74.01	79.11	0.39	0.32
	dsim	dwil	78.50	71.12	0.41	0.25	74.38	79.54	0.37	0.31
	dwil	dere	78.67	71.20	0.38	0.25				
	dwil	dyak	78.76	71.31	0.38	0.25	73.04	77.07	0.45	0.34
Obscura Group	dper	dpse	98.57	98.45	0.05	0.01	88.65	92.73	0.18	0.12
Obscura-Hawaiian	dgri	dper	77.28	73.36	0.37	0.27	75.35	72.55	0.38	0.30
	dgri	dpse	77.63	73.61	0.34	0.27	75.07	72.74	0.39	0.30
Repleta-Hawaiian	dgri	dmoj	84.10	79.28	0.27	0.18				
Repleta-Obscura	dmoj	dper	77.39	73.49	0.37	0.27	79.34	81.22	0.35	0.25
	dmoj	dpse	77.79	73.78	0.34	0.27	78.51	81.18	0.37	0.26
Virilis-Hawaiian	dgri	dvir	86.12	81.68	0.23	0.16	77.10	75.69	0.34	0.24
Virilis-Obscura	dper	dvir	78.32	74.21	0.35	0.26	77.37	81.57	0.34	0.27
	dpse	dvir	78.76	74.52	0.33	0.25	80.15	88.29	0.23	0.23
Virilis-Repleta	dmoj	dvir	87.87	82.94	0.22	0.14	81.69	85.22	0.27	0.22
Willistoni-Hawaiian	dgri	dwil	77.15	71.78	0.37	0.28				
Willistoni-Obscura	dper	dwil	79.02	72.20	0.38	0.24	75.38	81.09	0.35	0.29
	dpse	dwil	79.44	72.48	0.36	0.24	74.73	78.70	0.38	0.30
Willistoni-Repleta	dmoj	dwil	77.37	71.36	0.38	0.27	76.89	81.00	0.35	0.28
Willistoni-Virilis	dvir	dwil	78.28	72.10	0.37	0.26	73.34	78.68	0.38	0.33

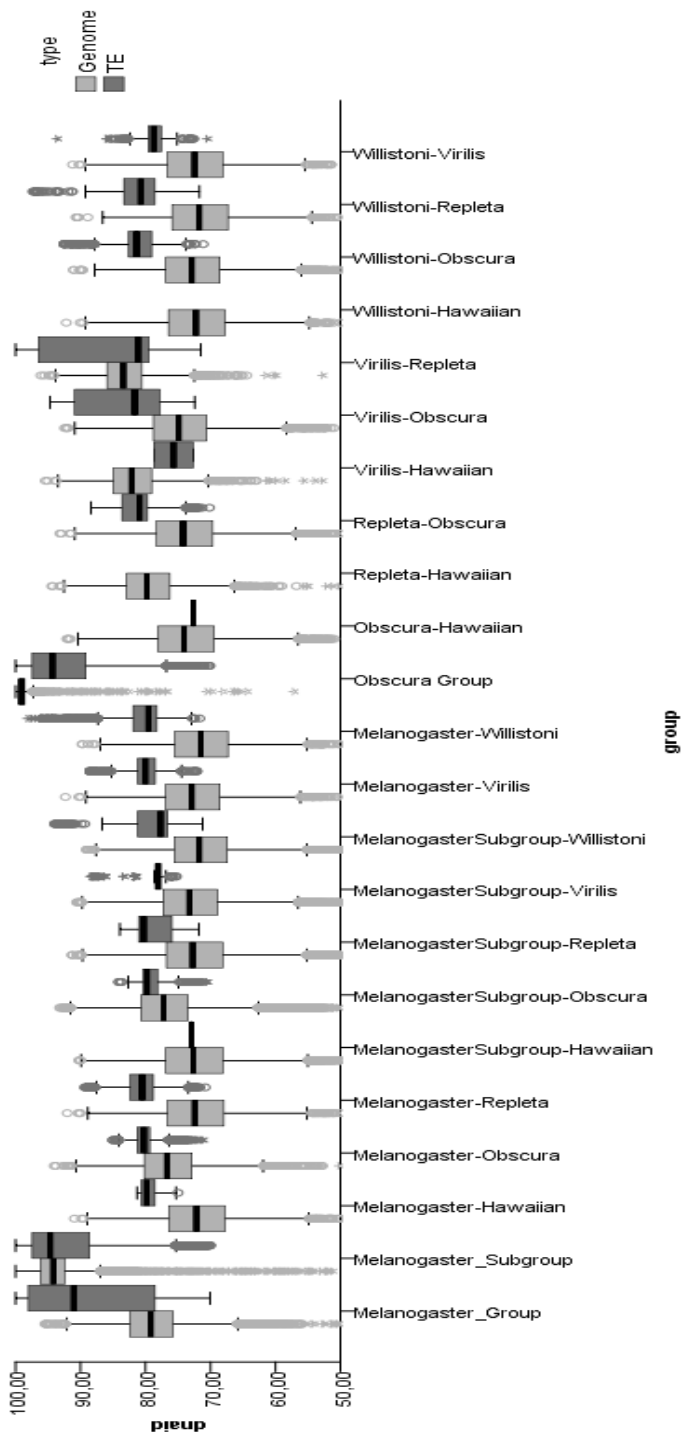


Figura 5 – Diagrama de caixa para dados de relação de sequências gênicas e elementos de transposição para a variável DNAid.

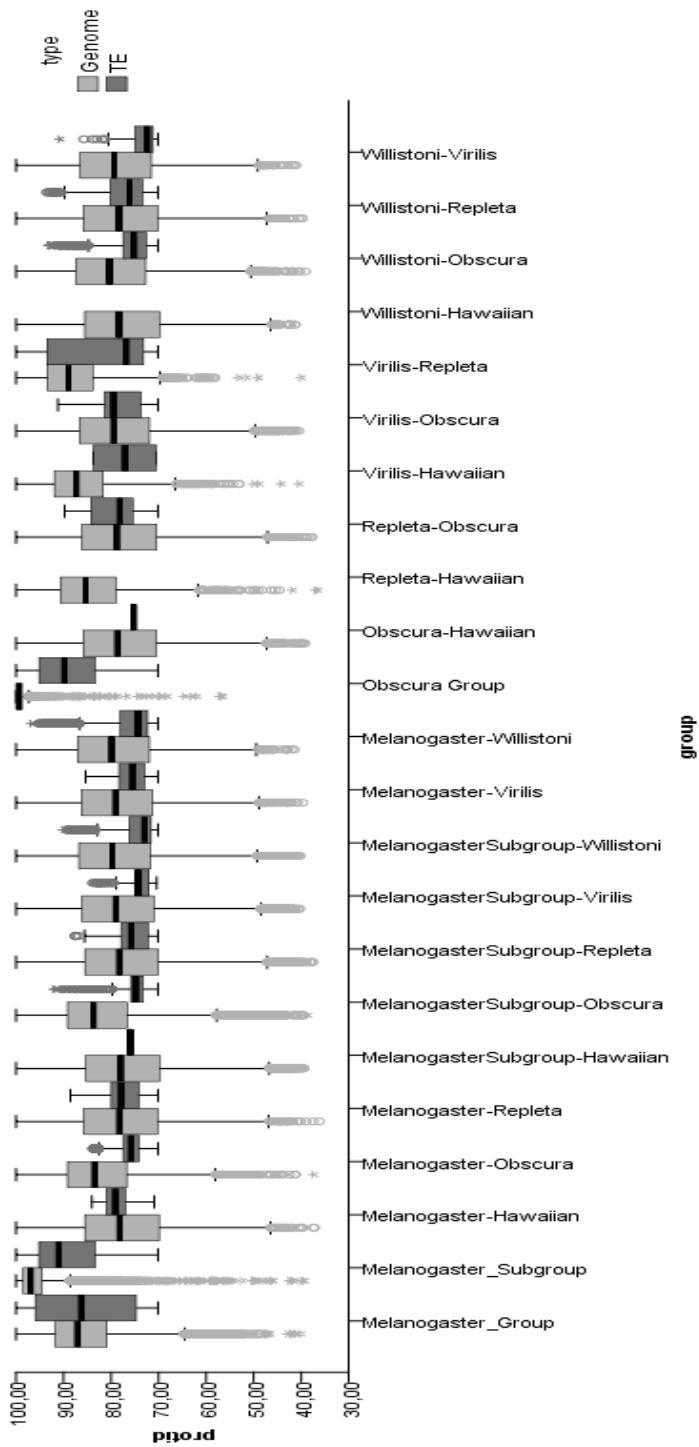


Figura 6 – Diagrama de caixa para dados de relação de sequências gênicas e elementos de transposição para a variável ProtId.

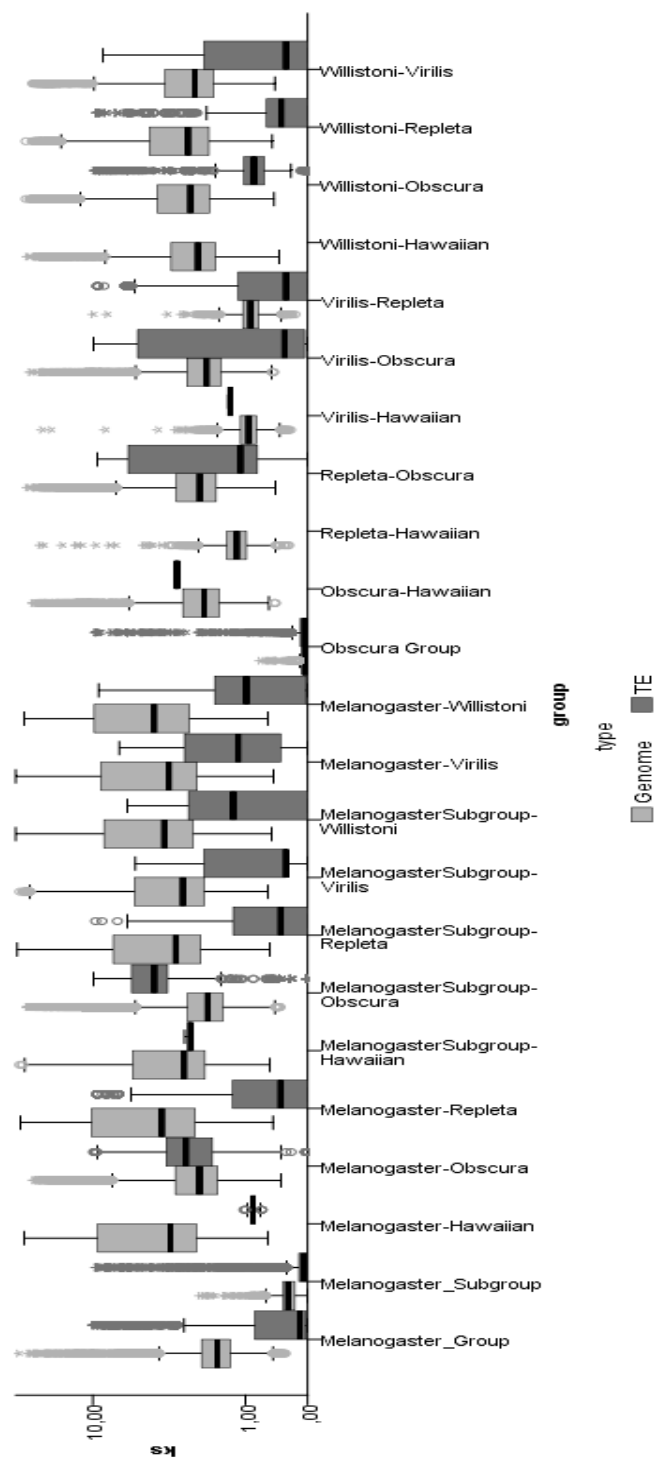


Figura 7 – Diagrama de caixa para dados de relação de sequências gênicas e elementos de transposição para a variável Ks.

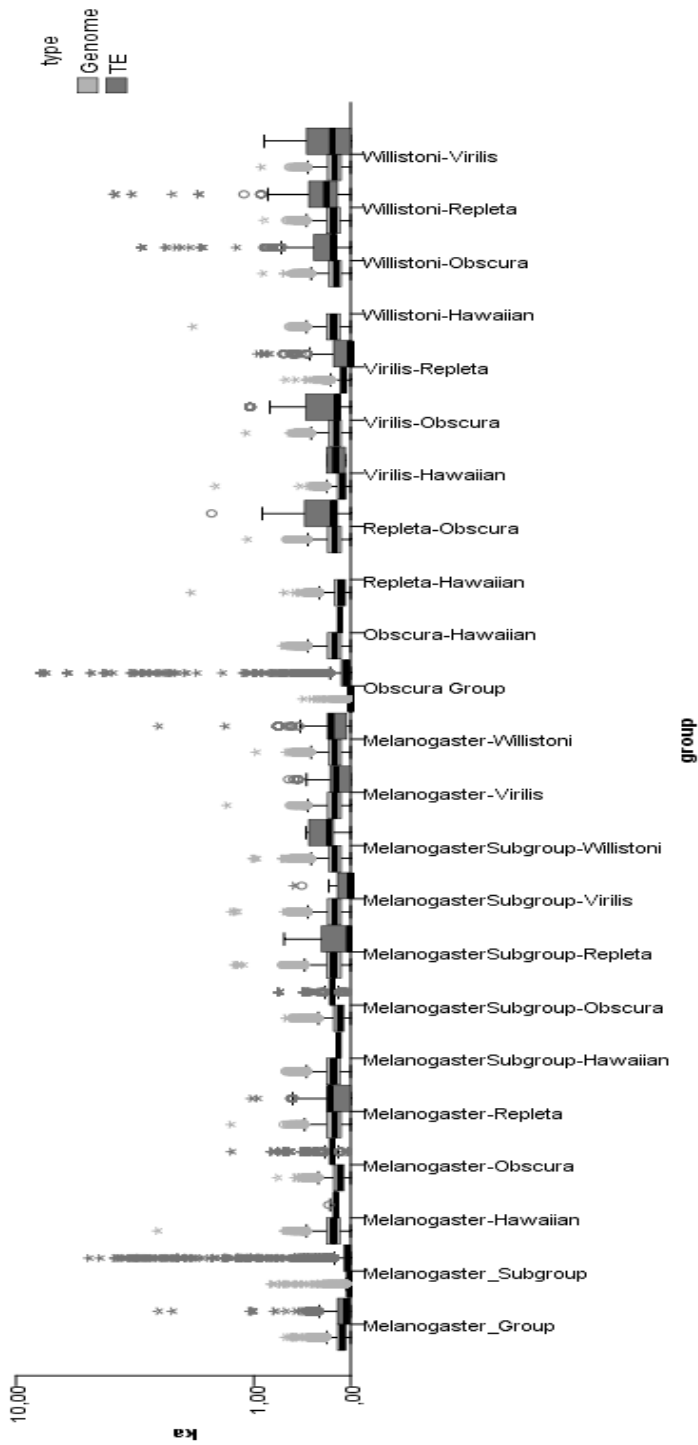


Figura 8 – Diagrama de caixa para dados de relação de sequências gênicas e elementos de transposição para a variável Ka.

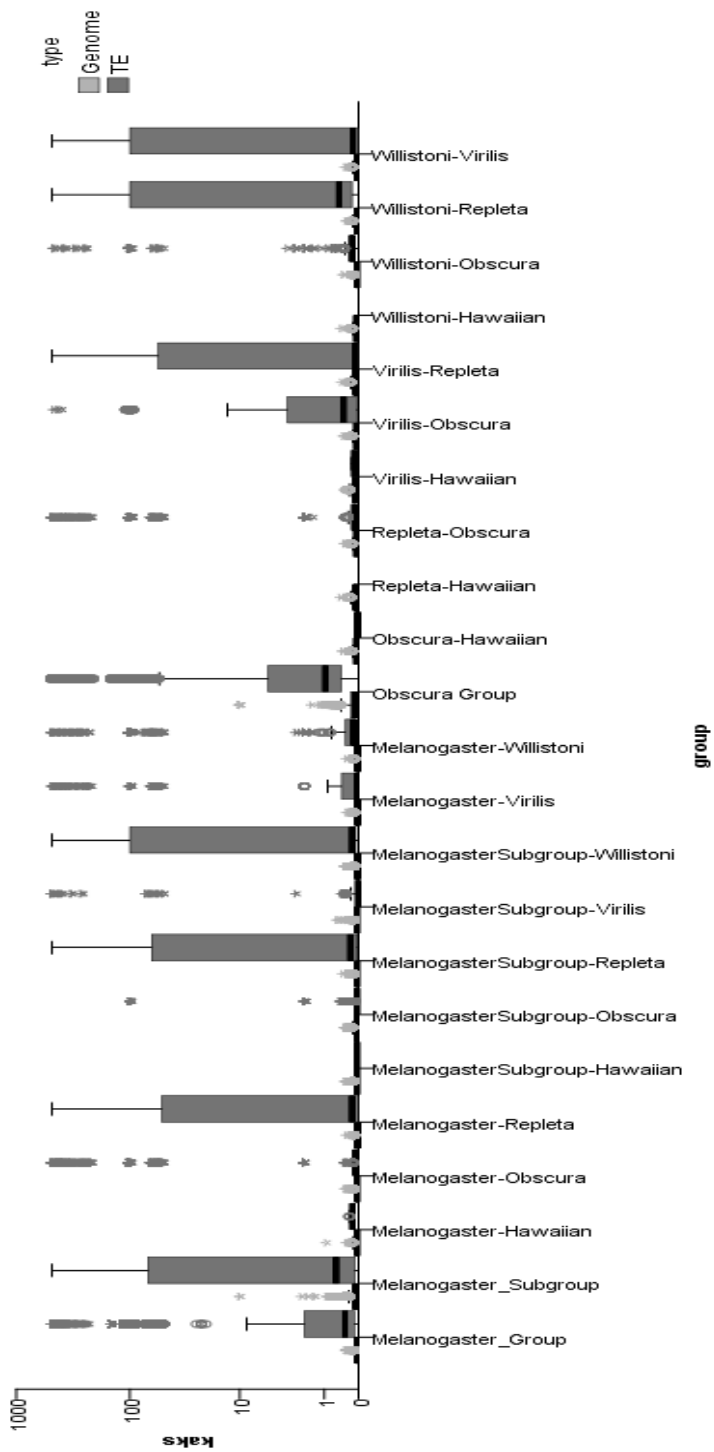


Figura 9 – Diagrama de caixa para dados de relação de sequências gênicas e elementos de transposição para a variável Ka/Ks.

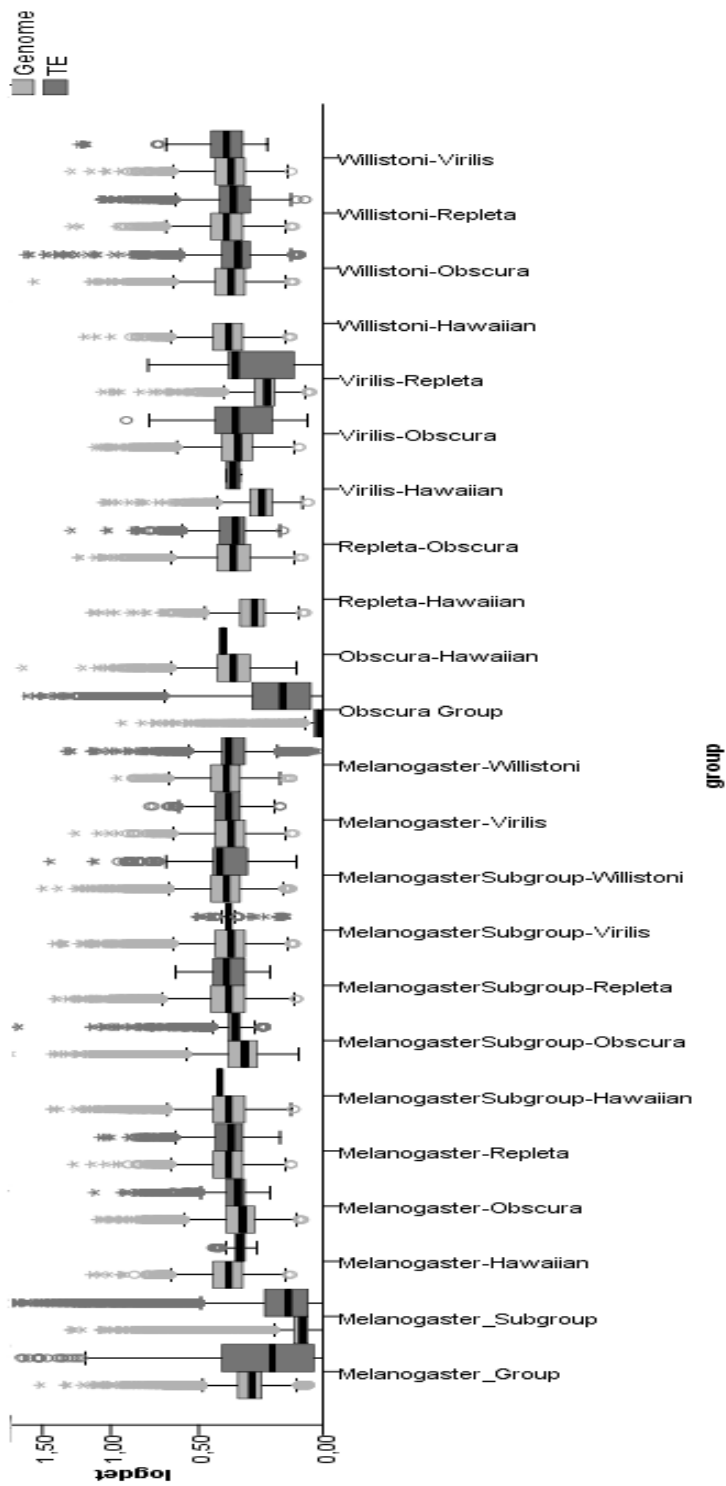


Figura 10 – Diagrama de caixa para dados de relação de sequências gênicas e elementos de transposição para a variável Logdet.

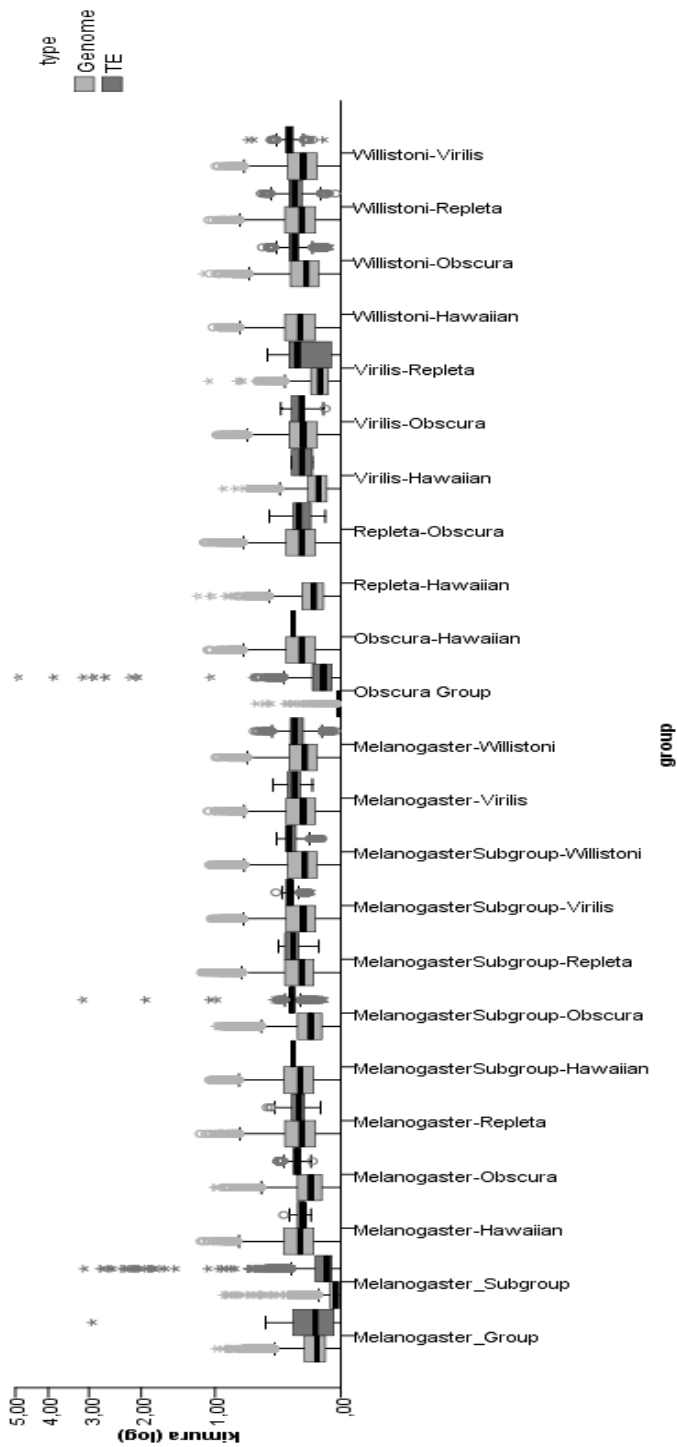


Figura 11 – Diagrama de caixa para dados de relação de sequências gênicas e elementos de transposição para a variável Kimura.

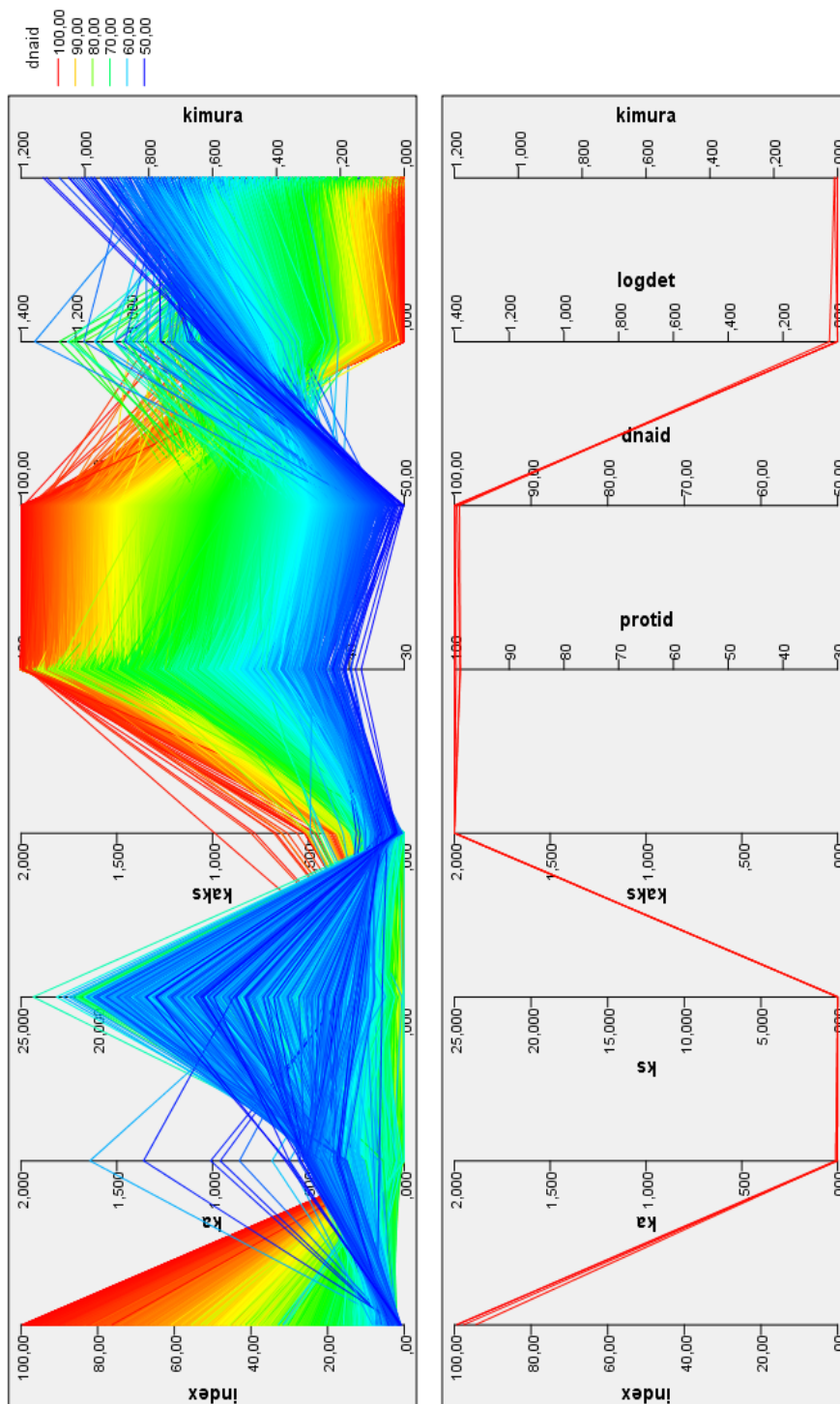


Figura 12 – Gráfico de coordenadas paralelas para comparações entre seqüências gênicas sujeitas a seleção negativa (painel superior) ou seleção positiva (painel inferior).

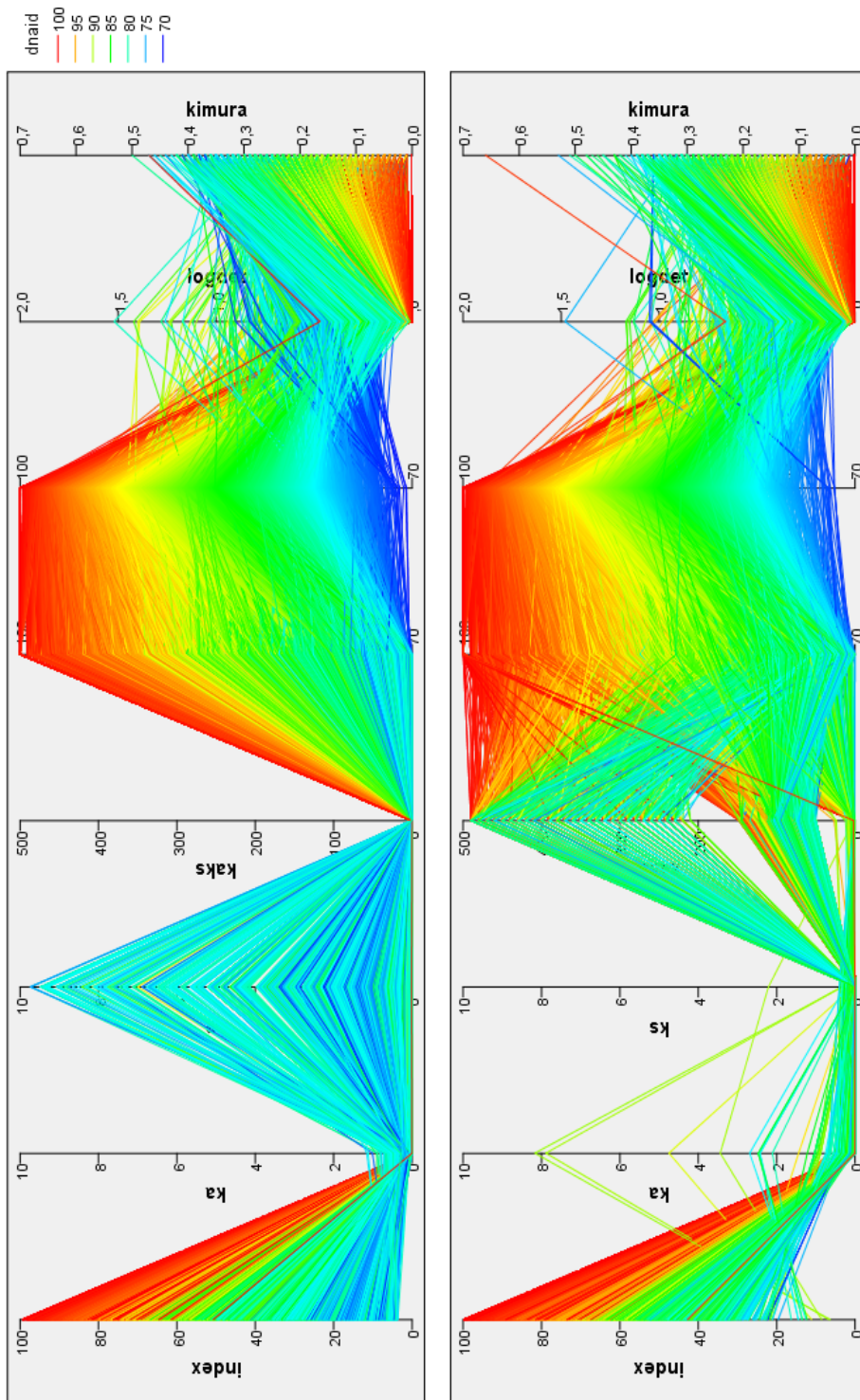


Figura 13 – Gráfico de coordenadas paralelas para comparações entre seqüências de TEs sujeitas a seleção negativa (painel superior) ou seleção positiva (painel inferior).

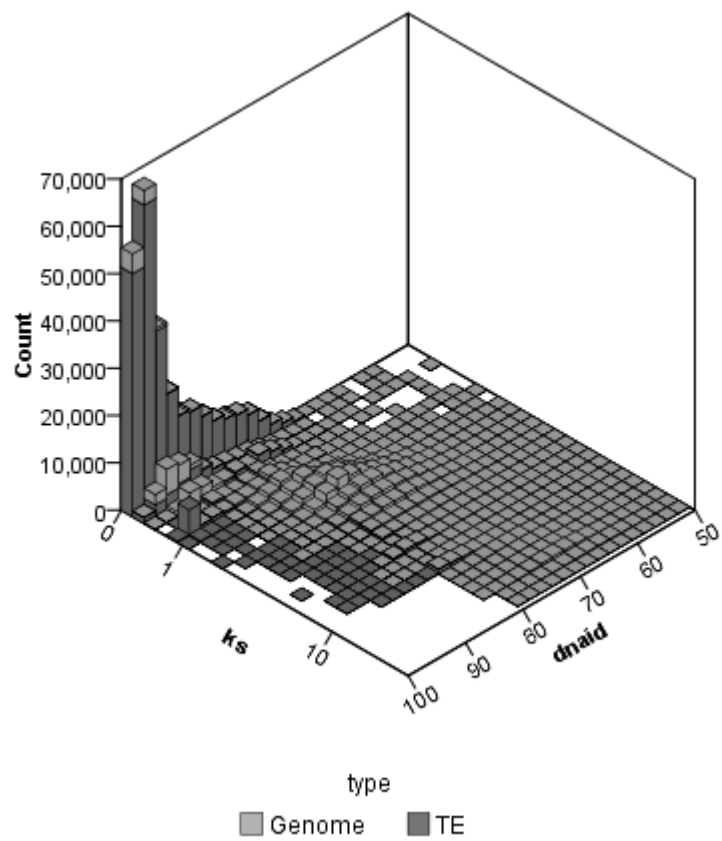


Figura 14 – Histograma 3D para dados de relação de seqüências gênicas e elementos de transposição para as variáveis DNAId e Ks (escala log).

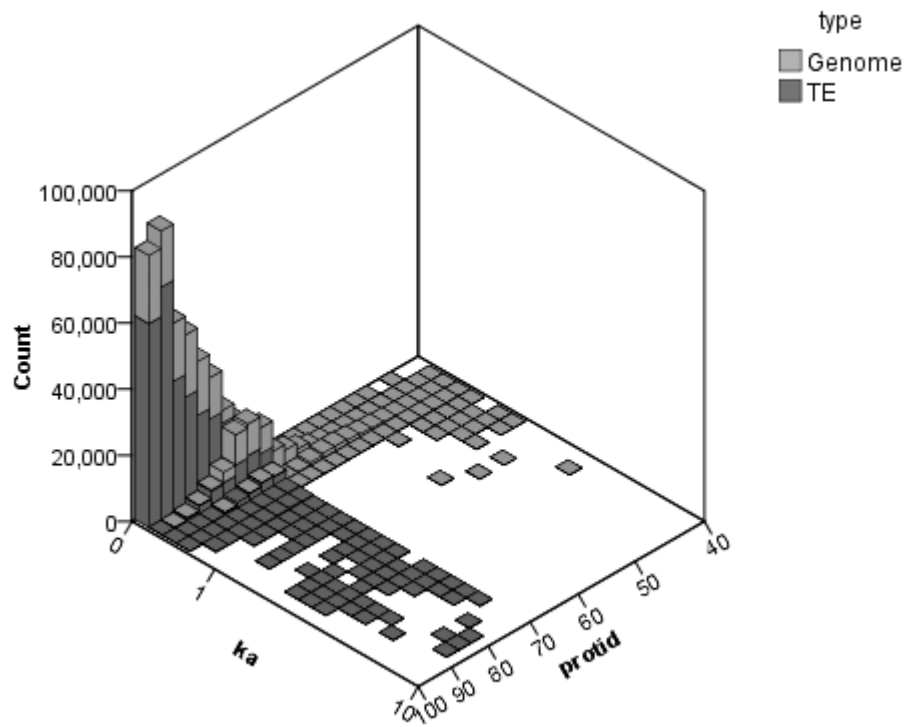


Figura 15 – Histograma 3D para dados de relação de seqüências gênicas e elementos de transposição para as variáveis Ka (escala log) e ProtId.

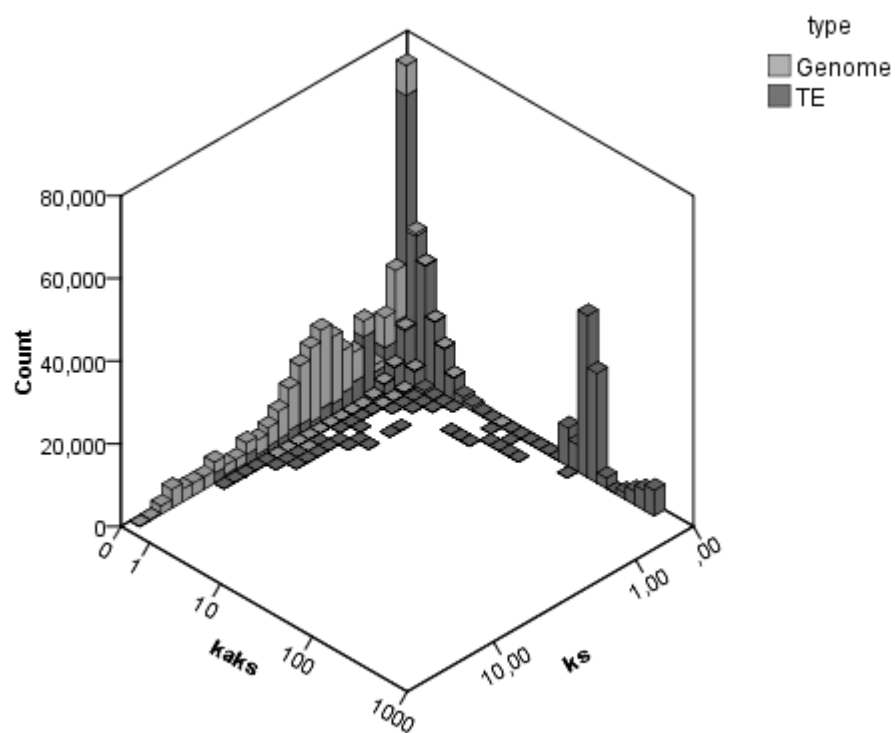


Figura 16 – Histograma 3D para dados de relação de seqüências gênicas e elementos de transposição para as variáveis Ka/Ks (escala log) e Ks (escala log).

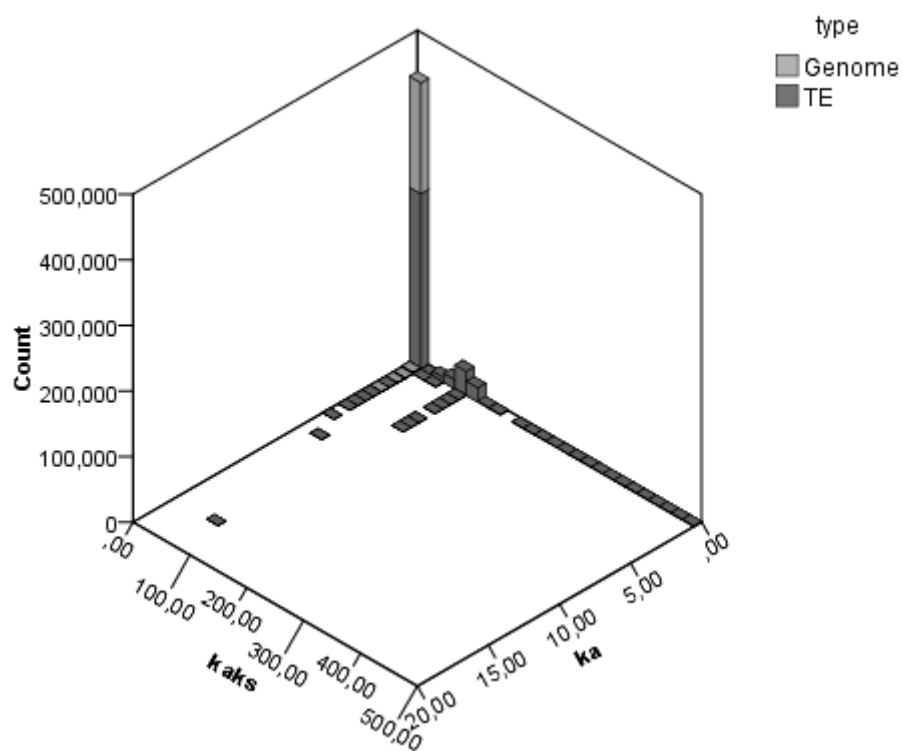


Figura 17 – Histograma 3D para dados de relação de seqüências gênicas e elementos de transposição para as variáveis Ka/Ks e Ka .

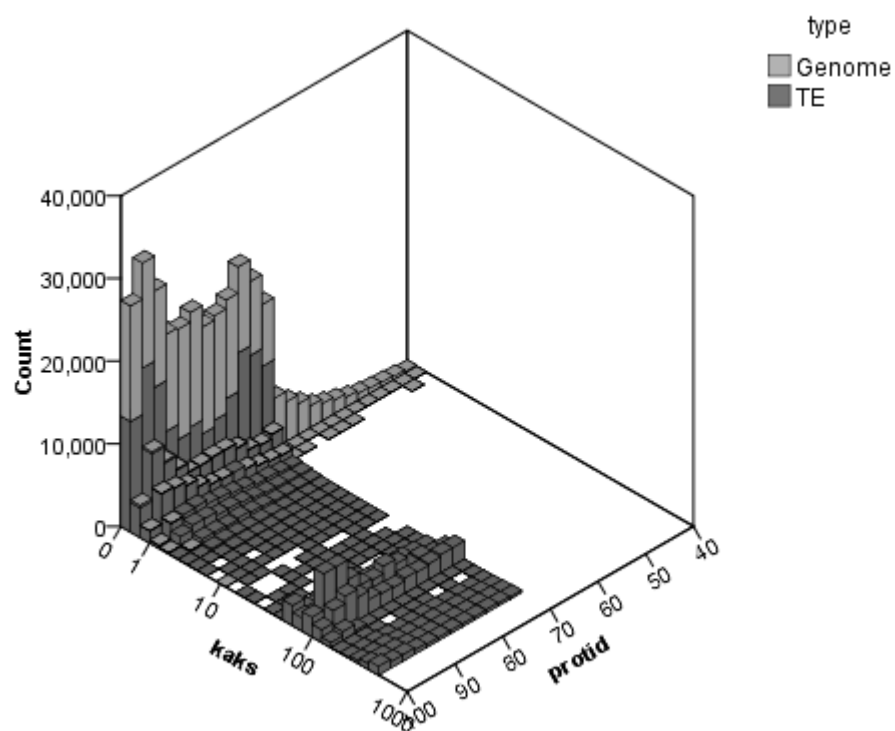


Figura 18 – Histograma 3D para dados de relação de seqüências gênicas e elementos de transposição para as variáveis Ka/Ks (escala log) e ProtId.

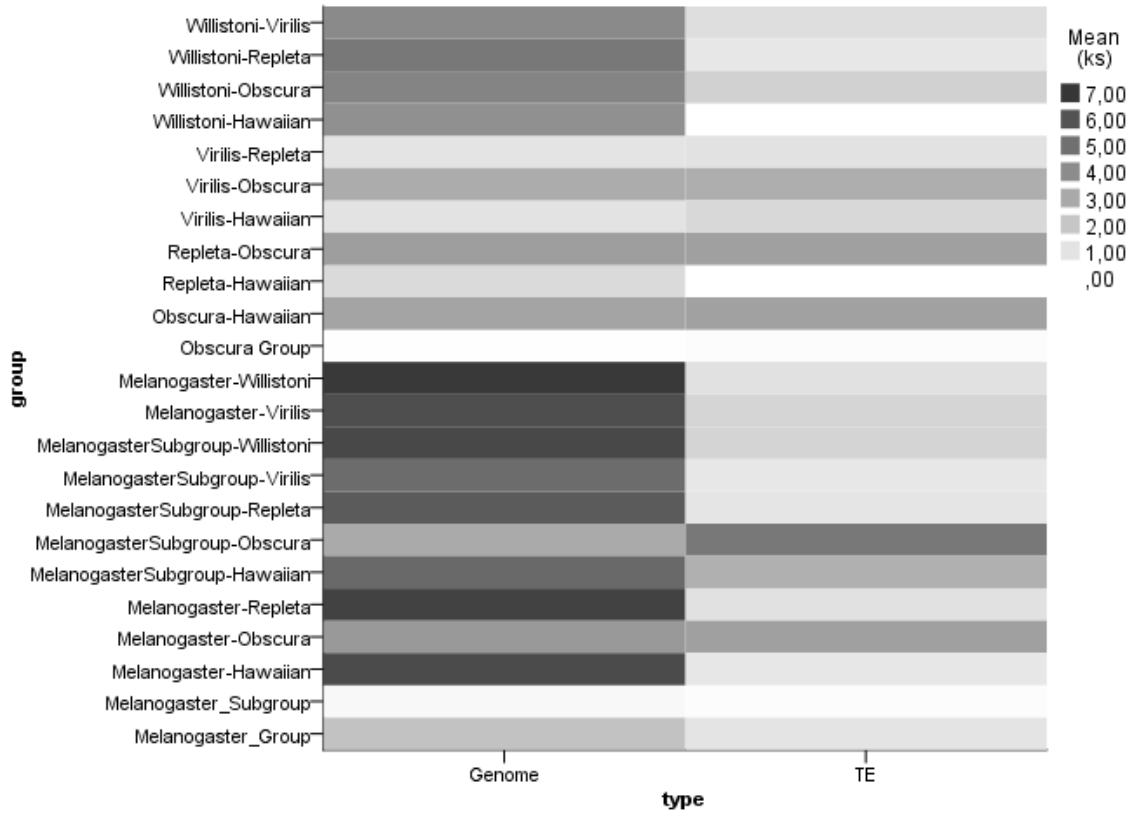


Figura 19 – Mapa de calor dos valores da variável Ks de acordo com cada grupo de comparações genômicas e separado para sequências de gênicas e de elementos de transposição.

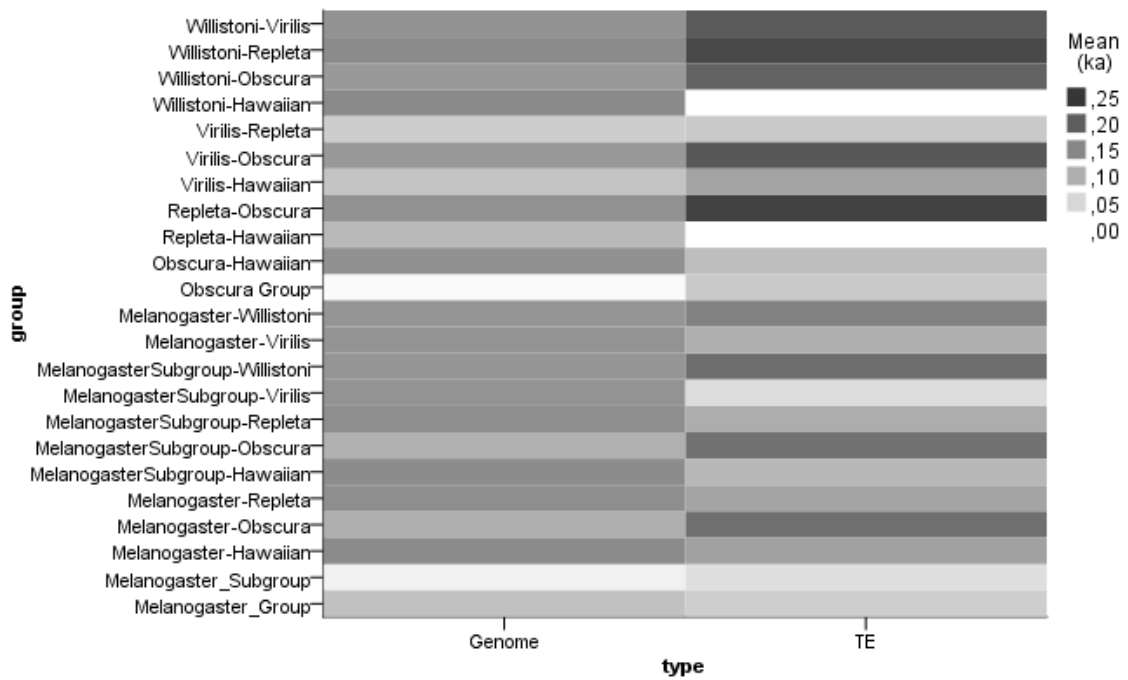


Figura 20 – Mapa de calor dos valores da variável Ka de acordo com cada grupo de comparações genômicas e separado para seqüências de gênicas e de elementos de transposição.

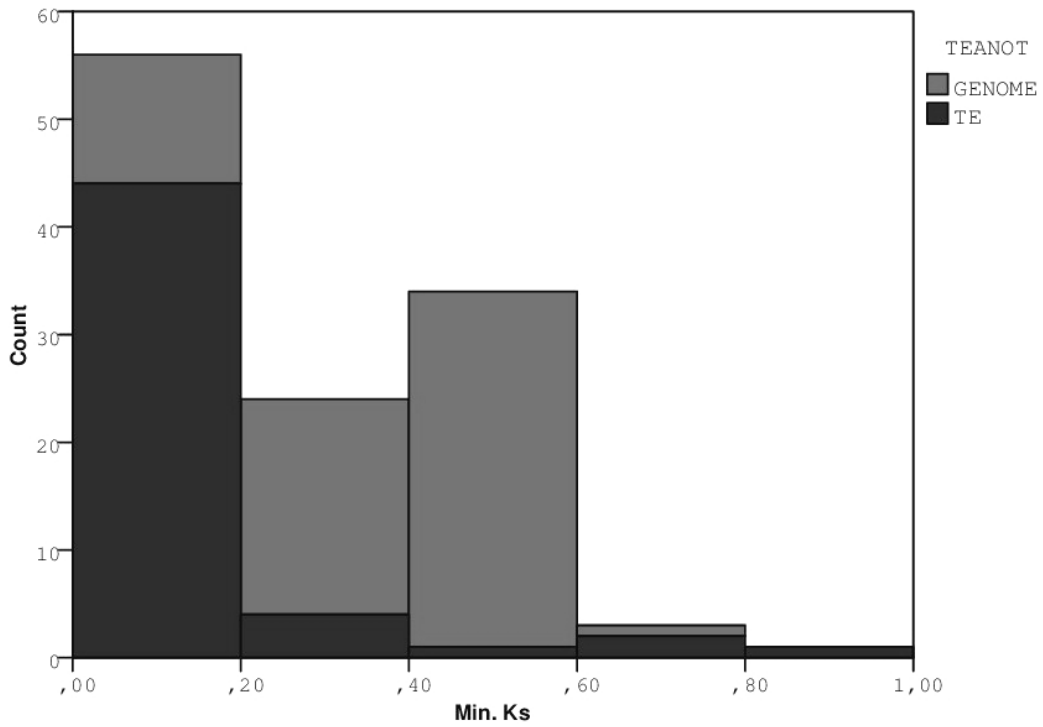


Figura 21 – Histograma dos valores mínimos de Ks para cada comparação entre genomas separados por cor para comparações de sequências gênicas e sequências de elementos de transposição.

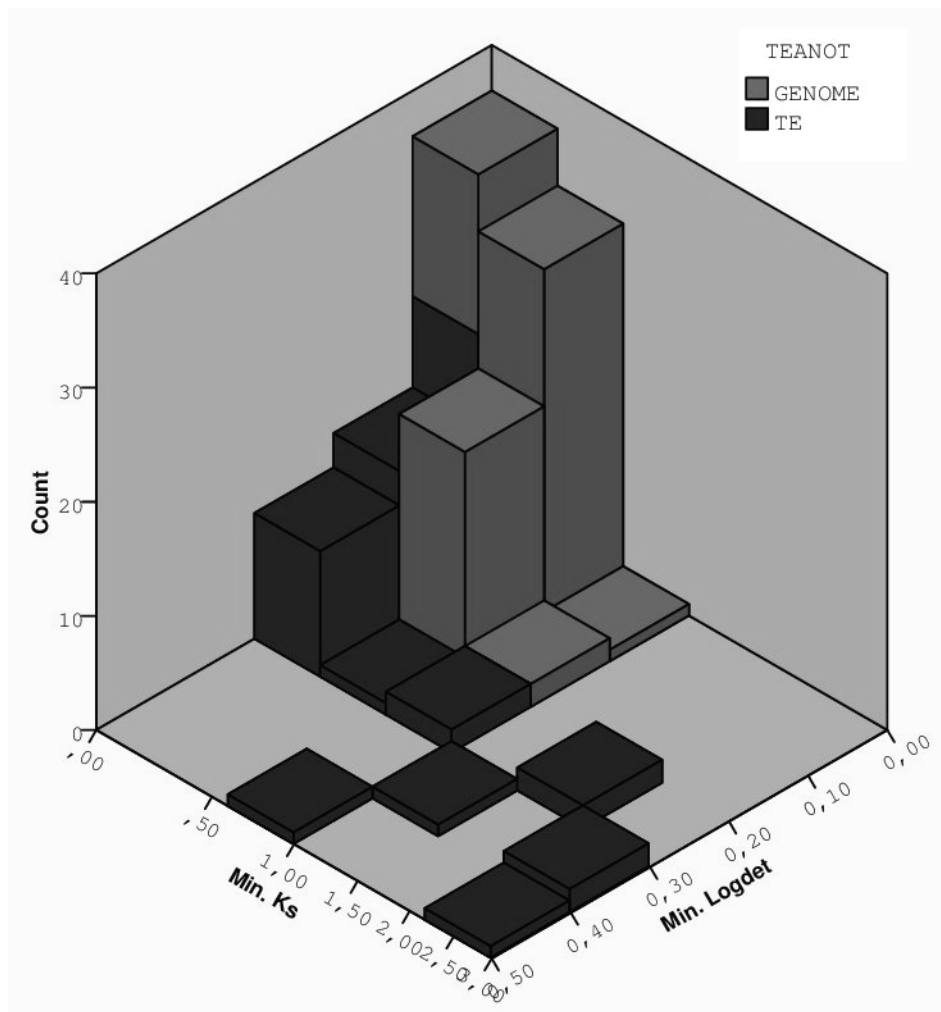


Figura 22 – Histograma 3D dos valores mínimos de Ks (escala log) e Logdet para cada comparação entre genomas separados por cor de acordo com comparações de seqüências gênicas ou de elementos de transposição.

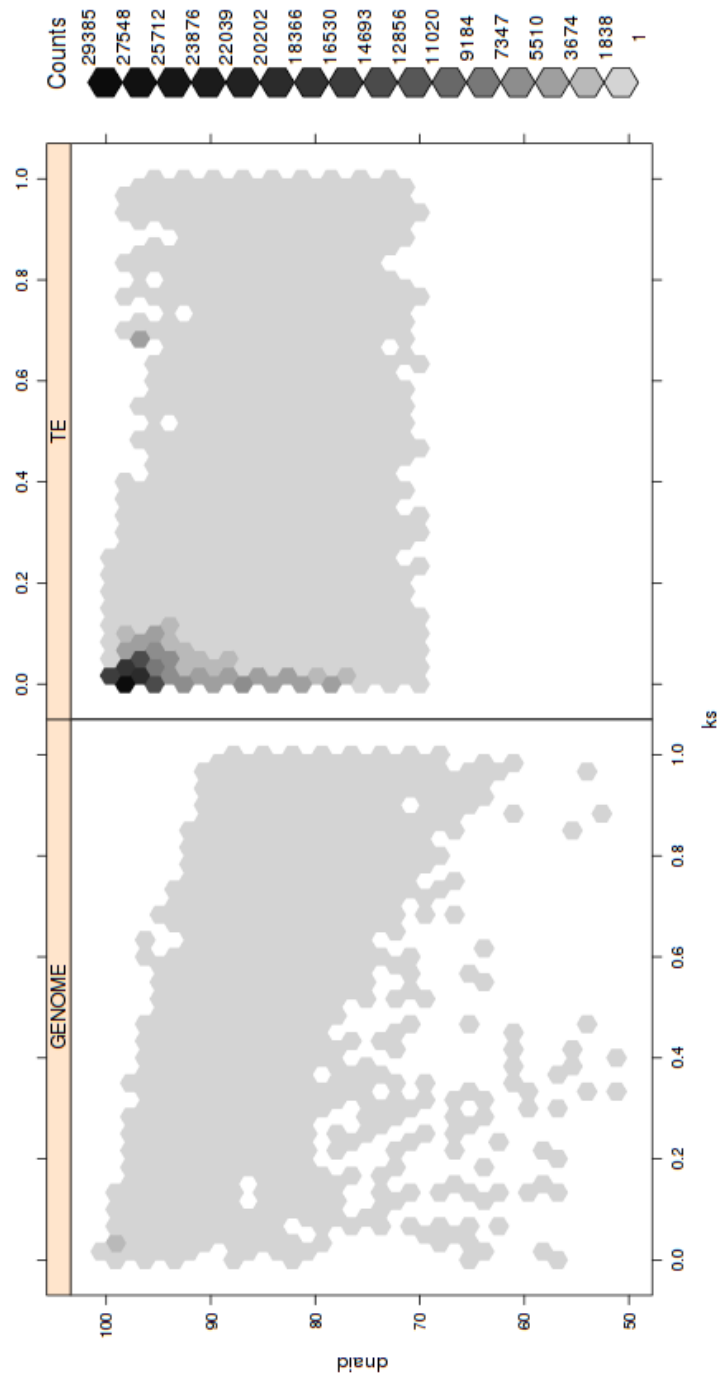


Figura 23 – Gráfico de dispersão para dados de relação de seqüências gênicas e elementos de transposição para as variáveis DNAid e Ks agrupados de acordo com a densidade de pontos de dados no plano cartesiano.

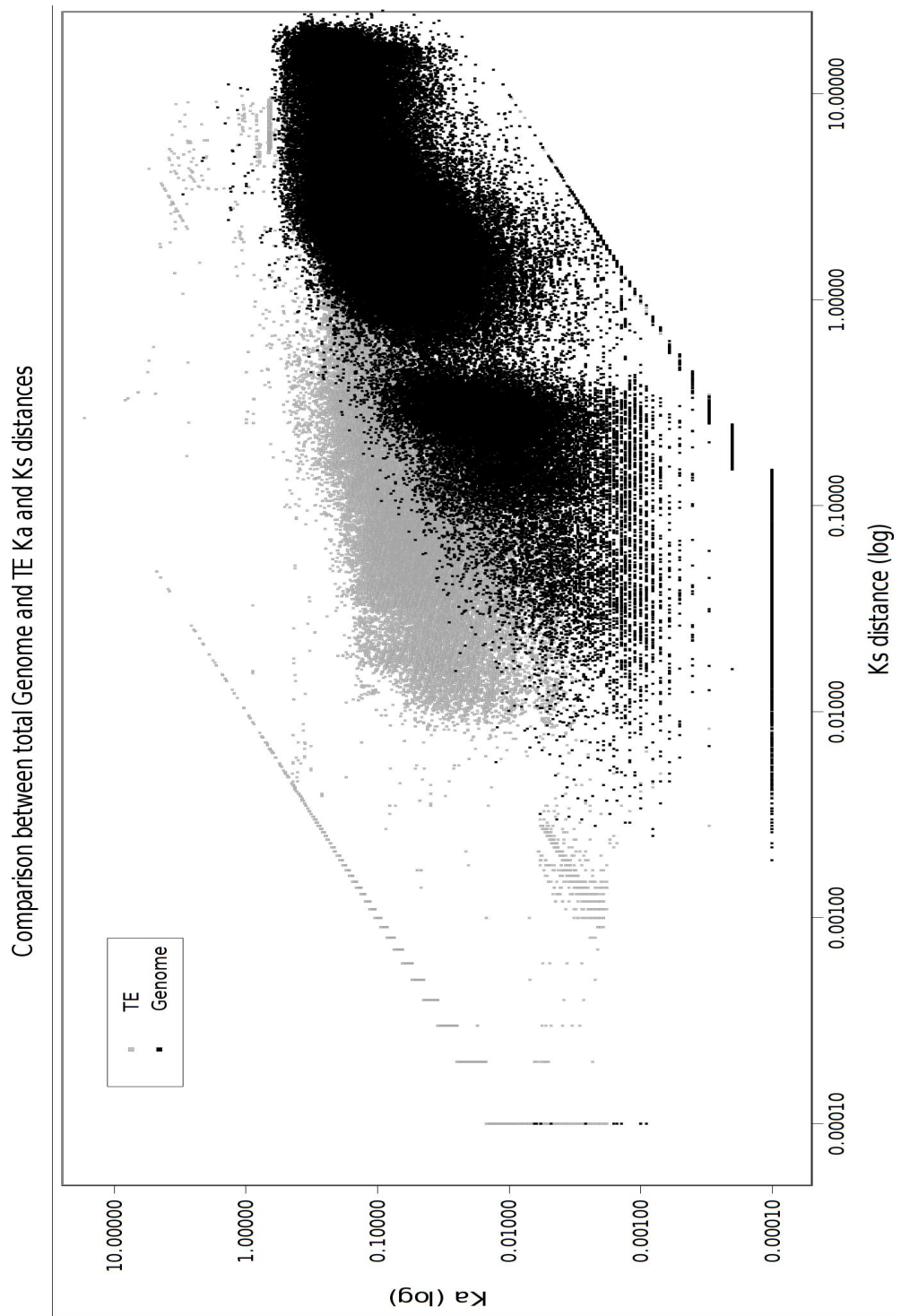


Figura 24 – Gráfico de dispersão para dados de relação de seqüências gênicas e elementos de transposição para as variáveis Ka e Ks.

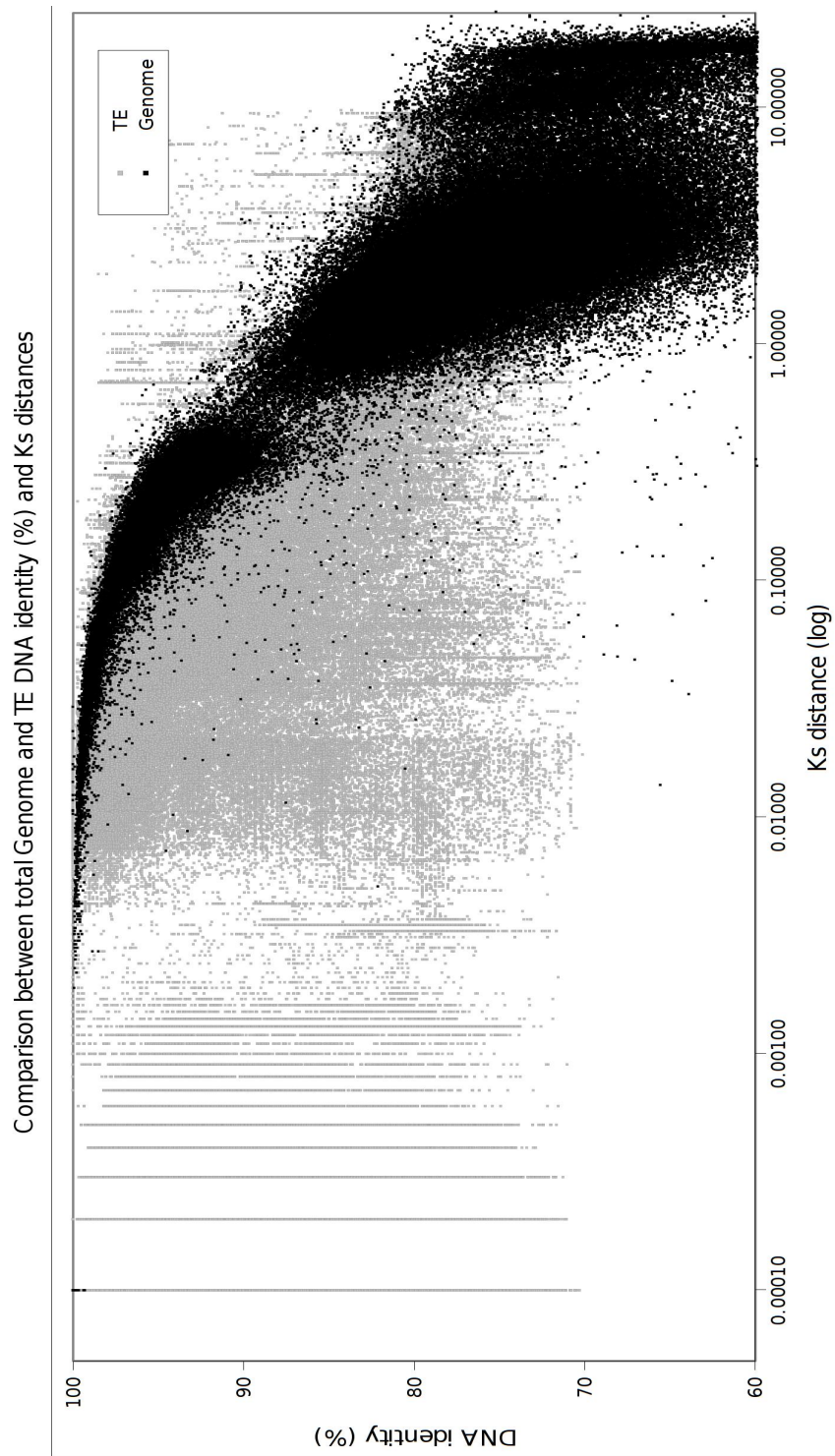


Figura 25 – Gráfico de dispersão para dados de relação de seqüências gênicas e elementos de transposição para as variáveis DNAId e Ks.

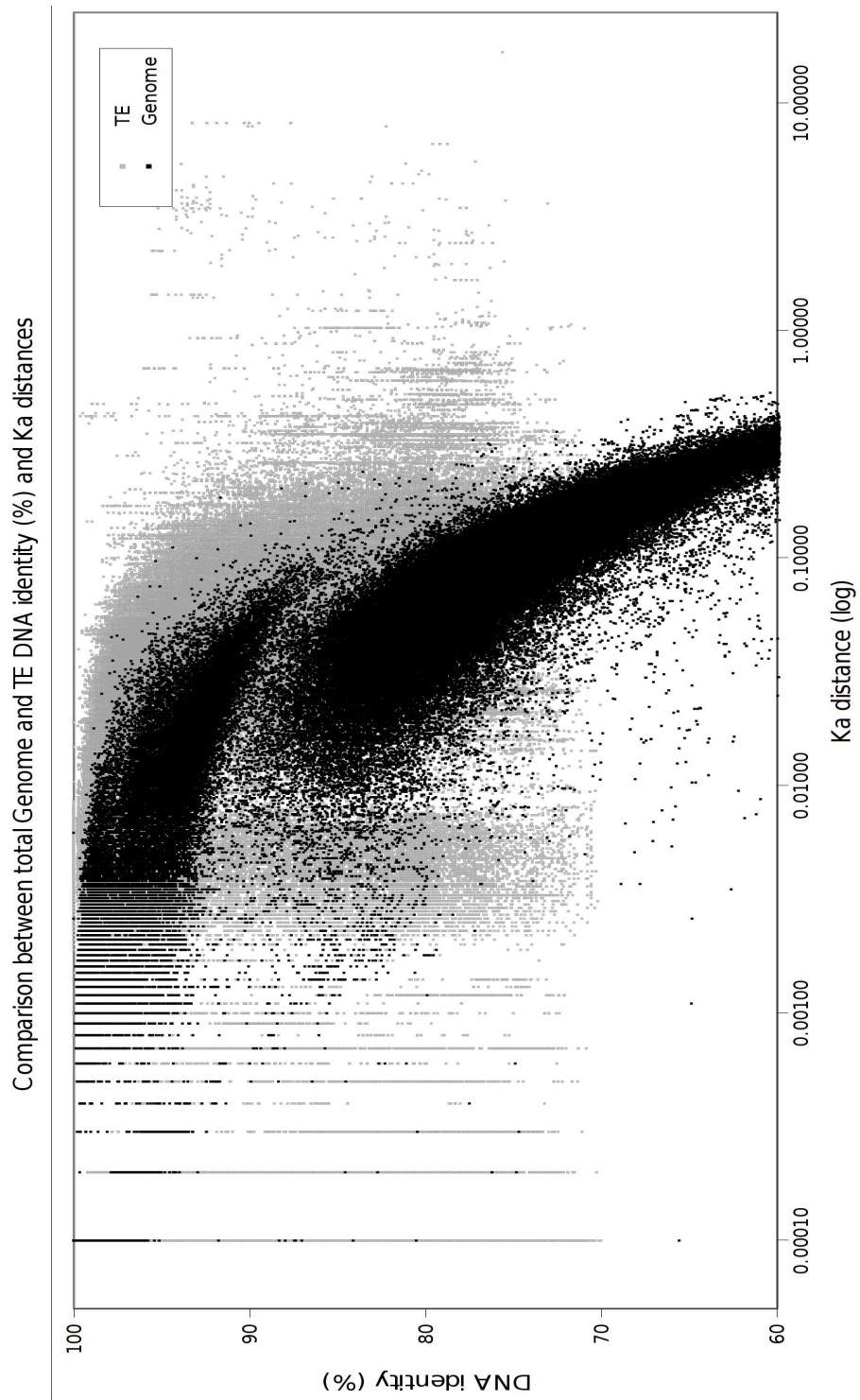


Figura 26 – Gráfico de dispersão para dados de relação de sequências gênicas e elementos de transposição para as variáveis DNAId e Ka.

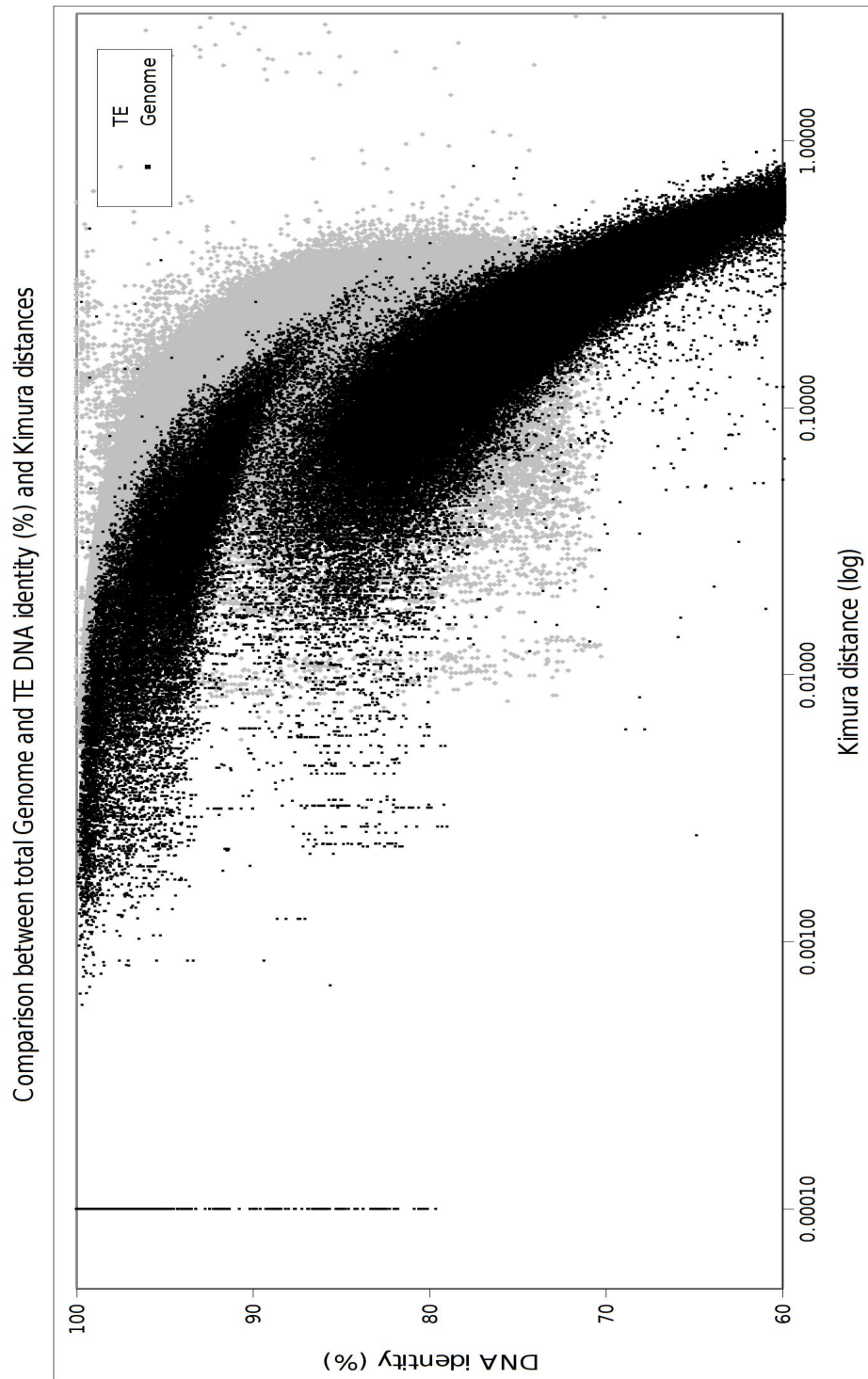


Figura 27 – Gráfico de dispersão para dados de relação de seqüências gênicas e elementos de transposição para as variáveis DNAId e distância Kimura.

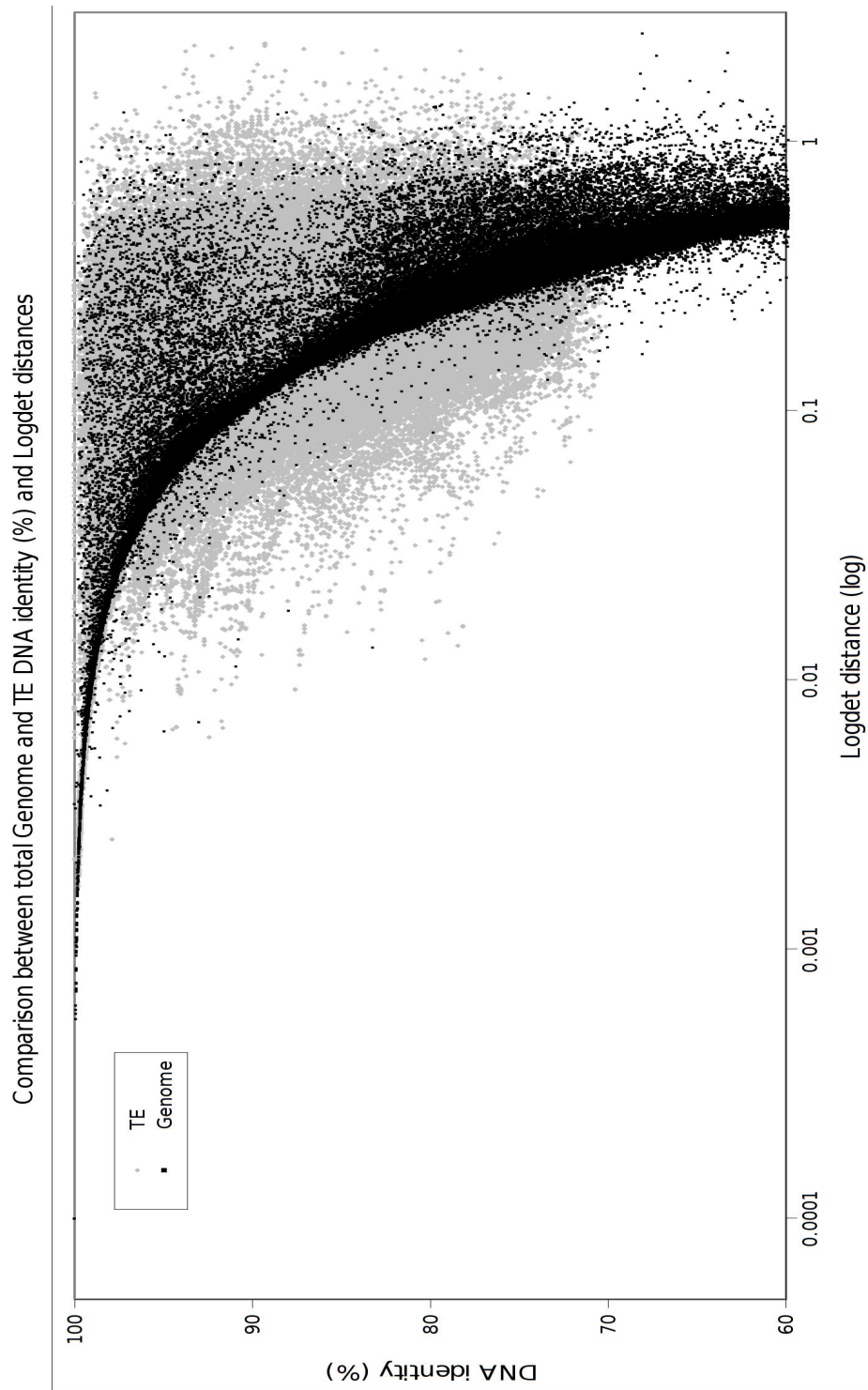


Figura 28 – Gráfico de dispersão para dados de relação de seqüências gênicas e elementos de transposição para as variáveis DNAId e Logdet

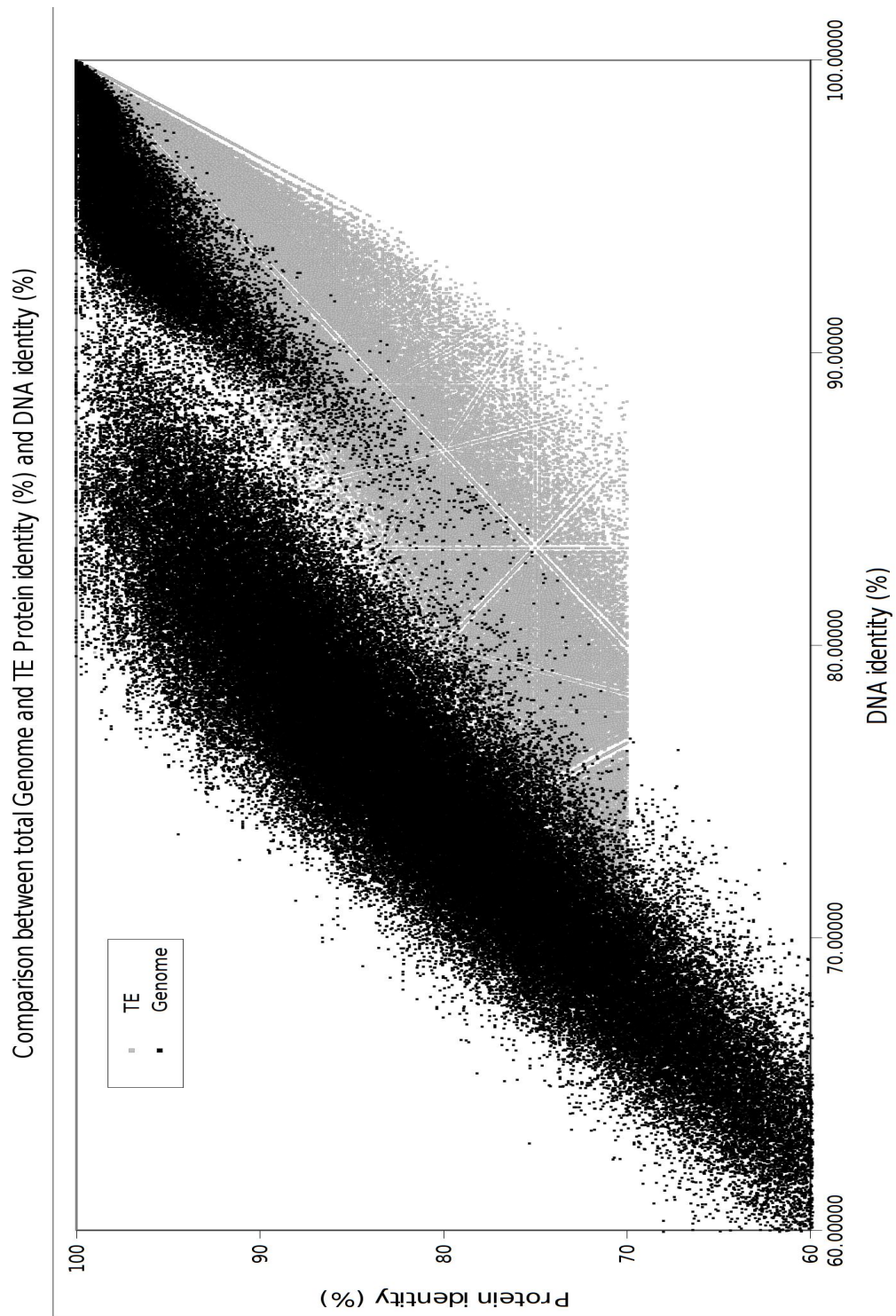


Figura 29 – Gráfico de dispersão para dados de relação de sequências gênicas e elementos de transposição para as variáveis Protid e Logdet.

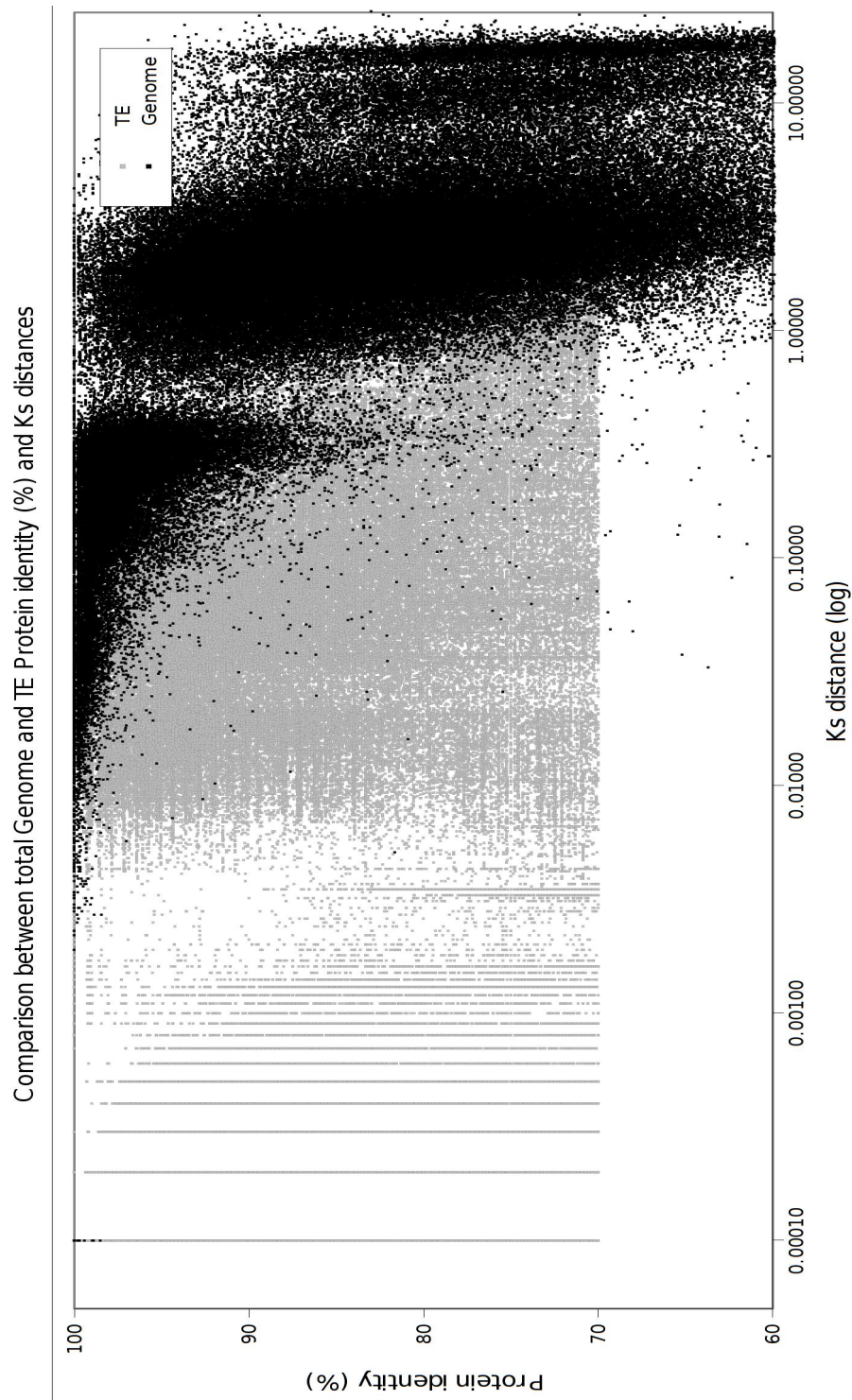


Figura 30 – Gráfico de dispersão para dados de relação de seqüências gênicas e elementos de transposição para as variáveis ProtId e Ks.

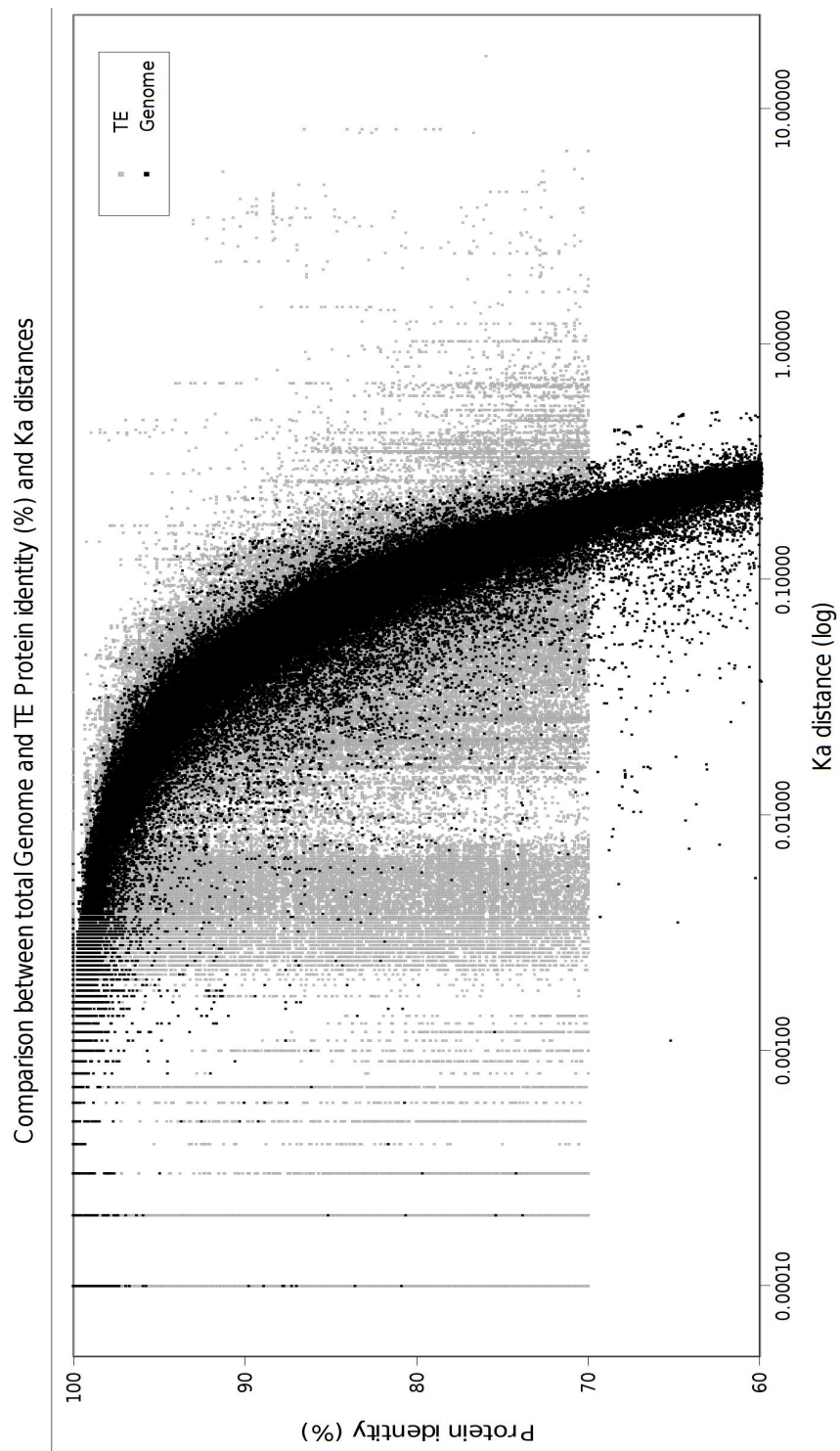


Figura 31 – Gráfico de dispersão para dados de relação de sequências gênicas e elementos de transposição para as variáveis ProtId e Ka.

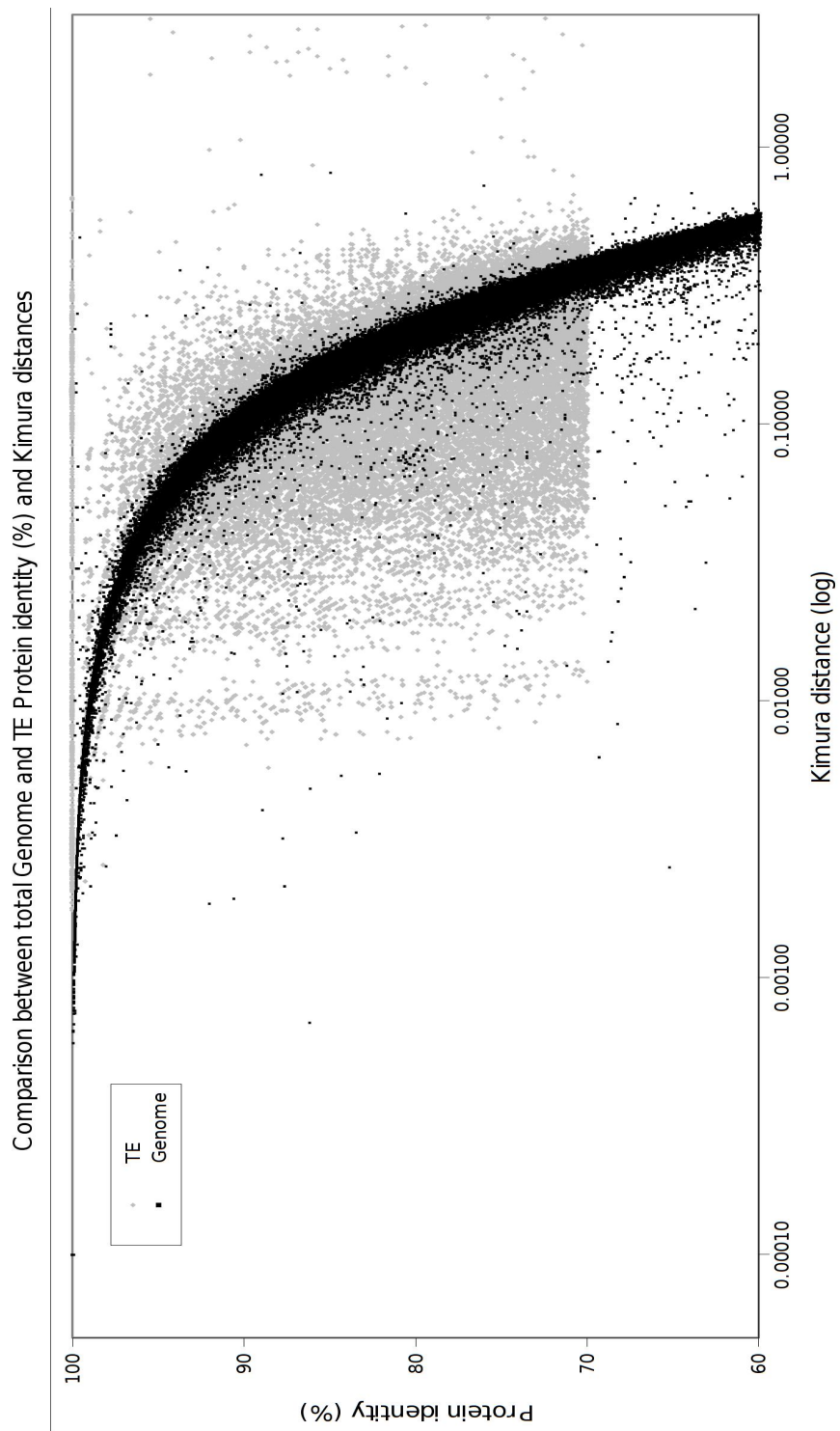


Figura 32 – Gráfico de dispersão para dados de relação de seqüências gênicas e elementos de transposição para as variáveis ProtId e Kimura.

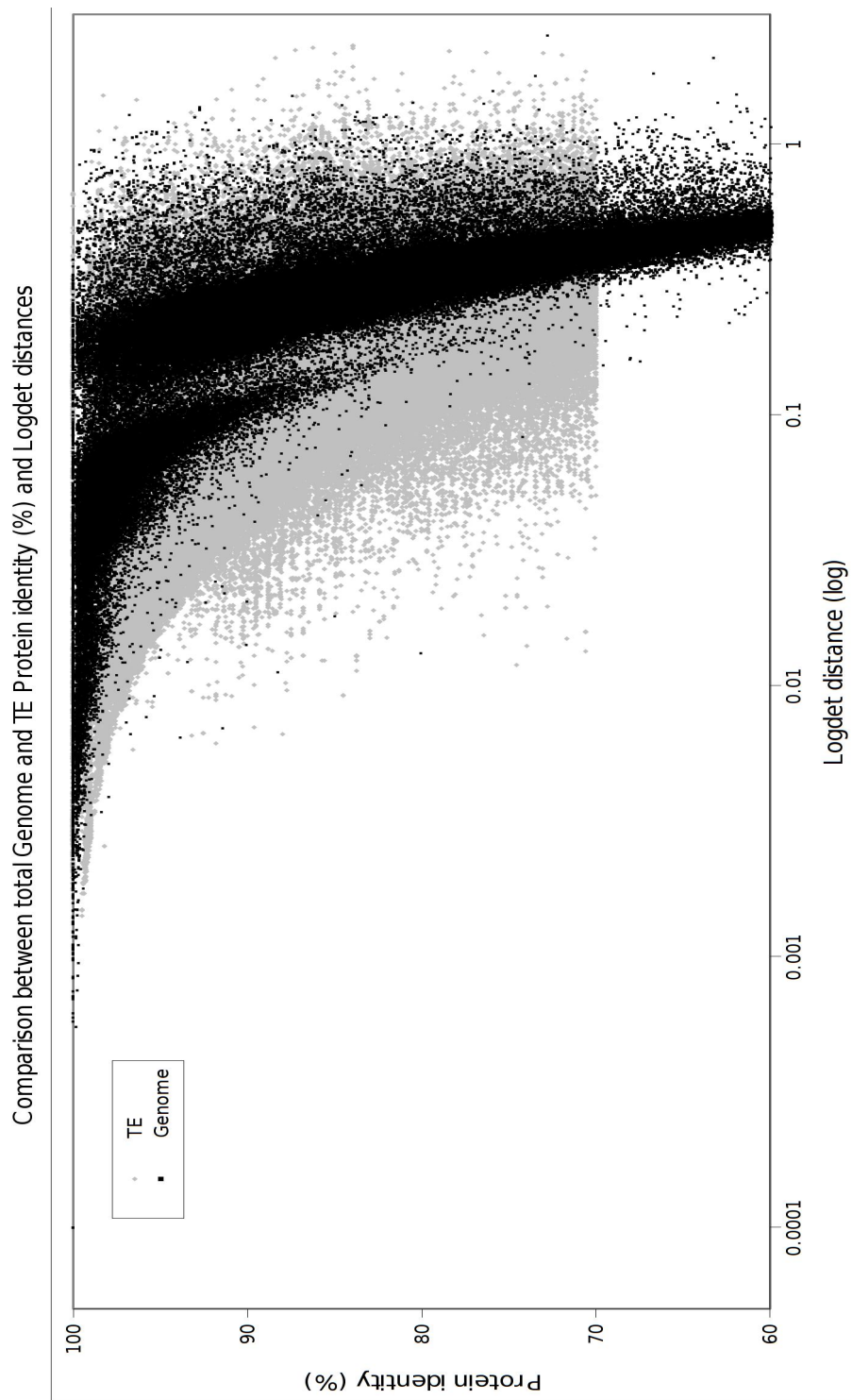


Figura 33 – Gráfico de dispersão para dados de relação de seqüências gênicas e elementos de transposição para as variáveis ProtId e Logdet.

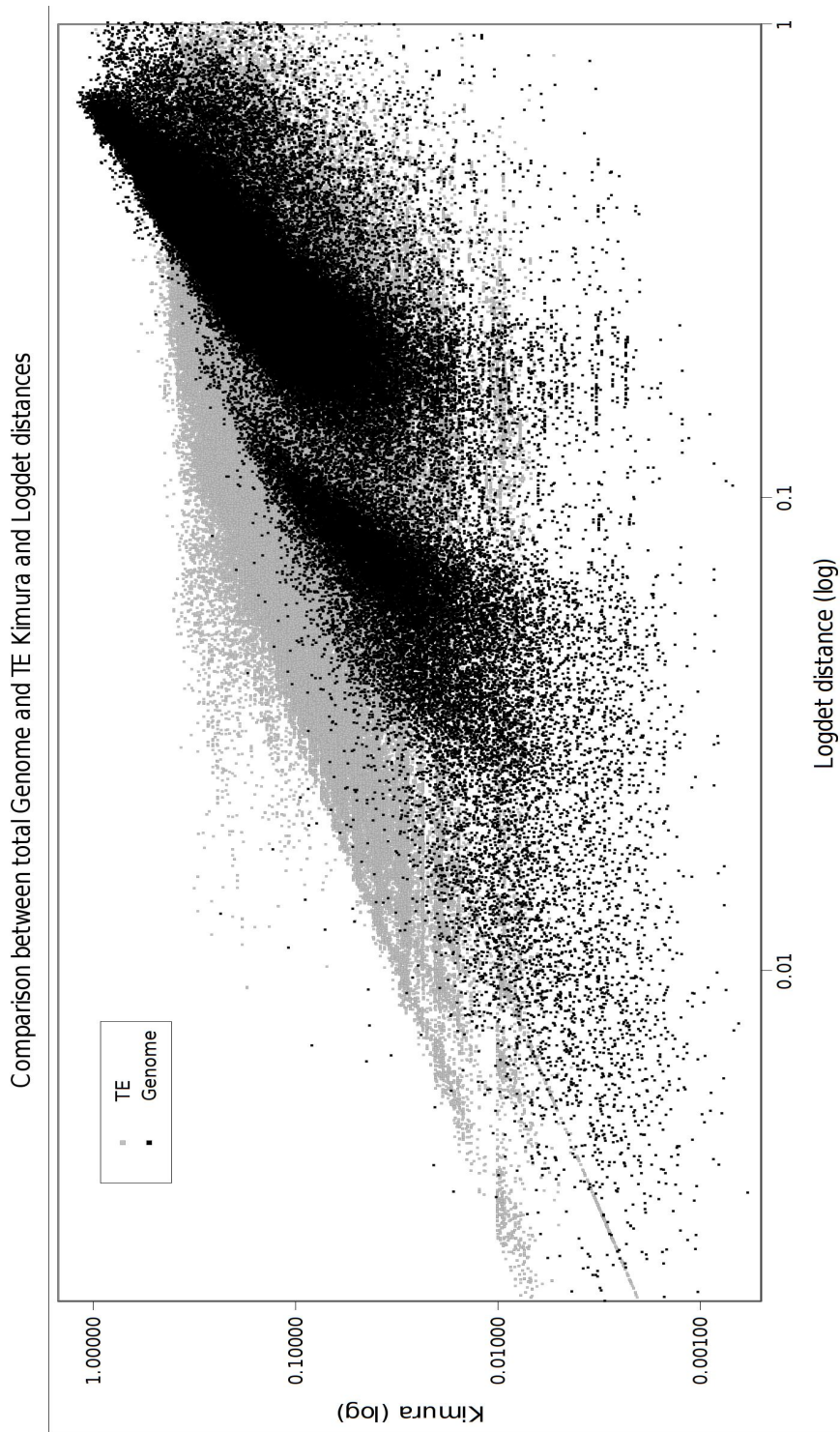


Figura 34 – Gráfico de dispersão para dados de relação de sequências gênicas e elementos de transposição para as variáveis Kimura e Logdet.

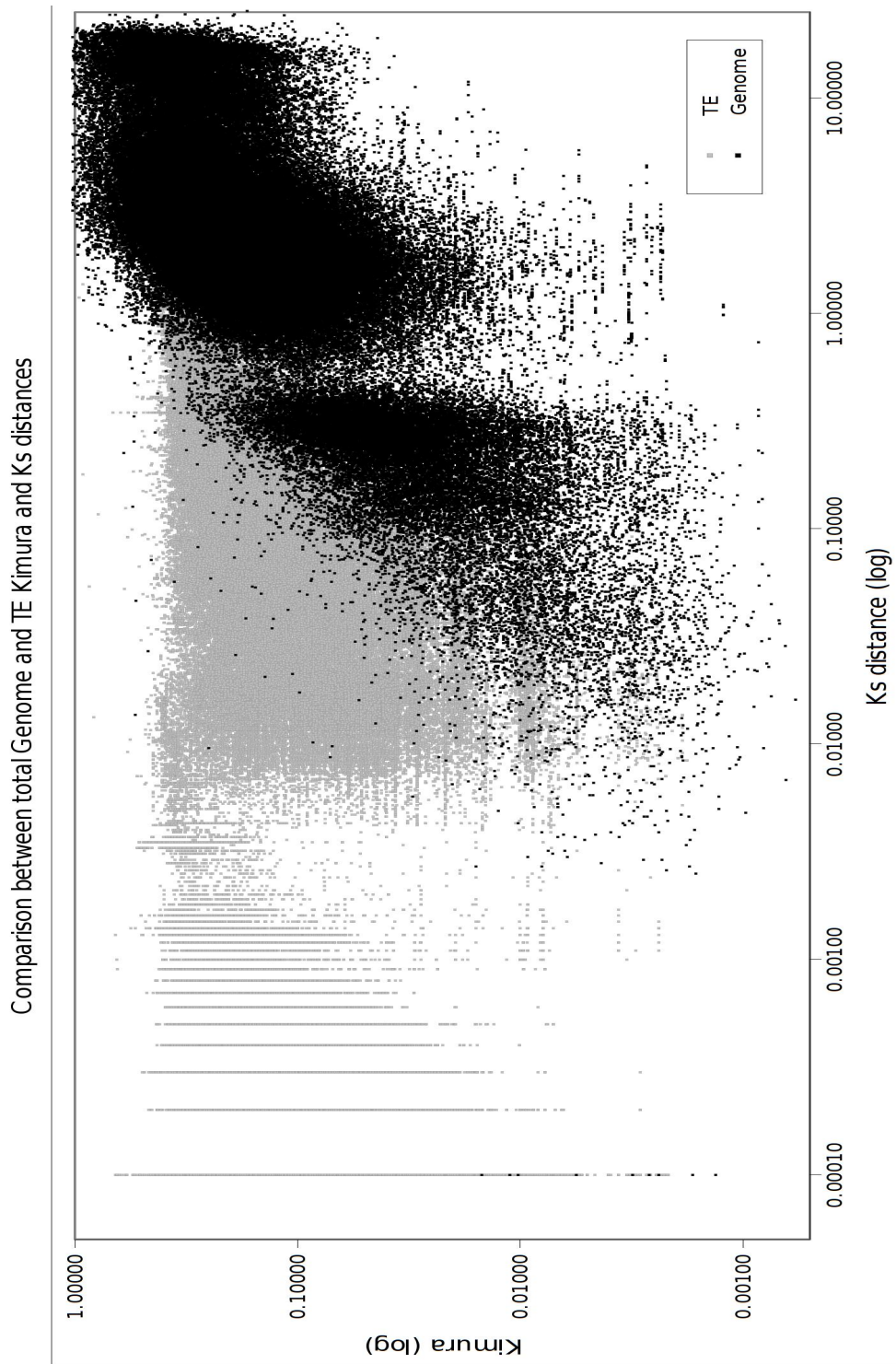


Figura 35 – Gráfico de dispersão para dados de relação de seqüências gênicas e elementos de transposição para as variáveis Kimura e Ks.

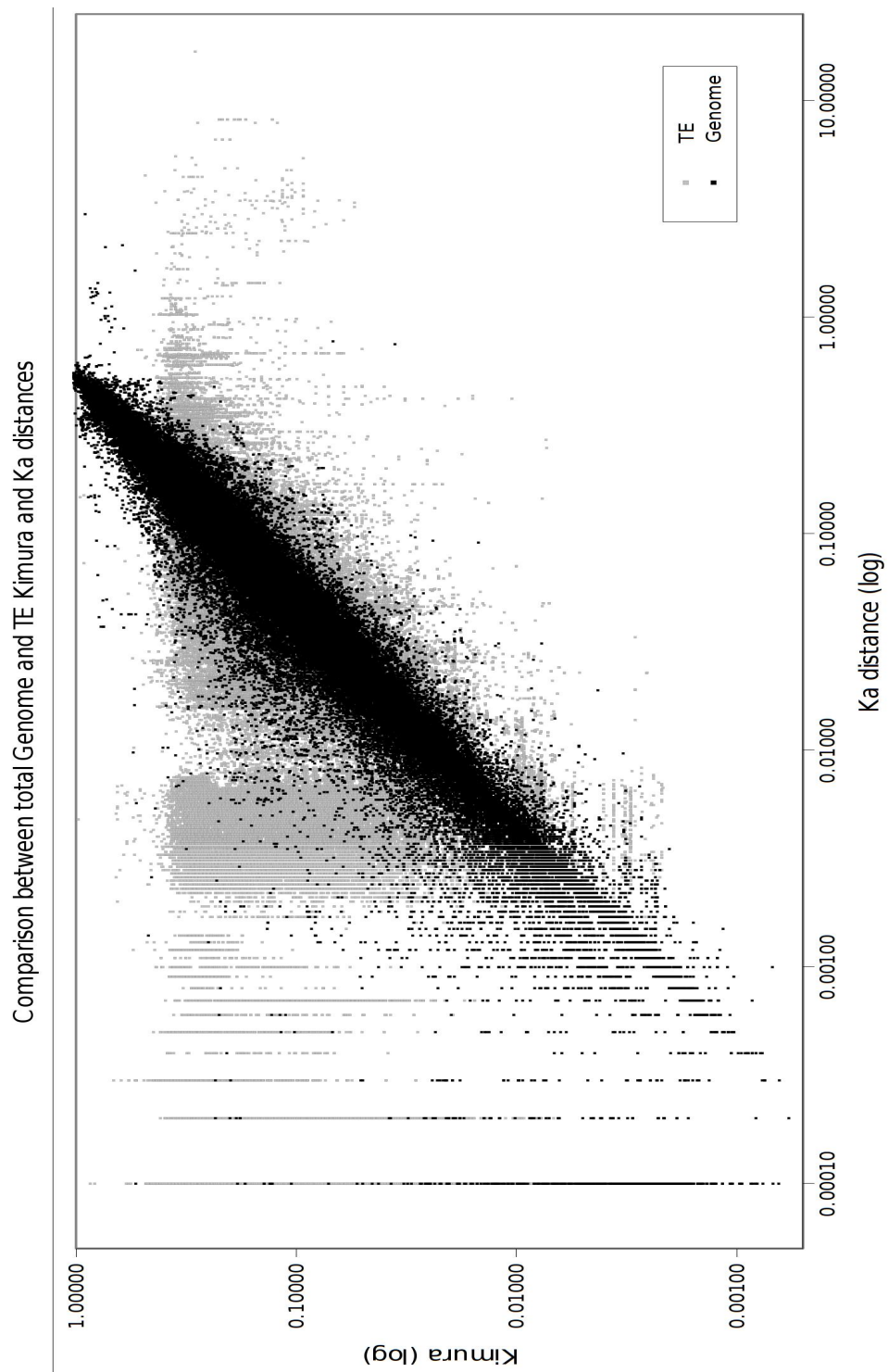


Figura 36 – Gráfico de dispersão para de relação de seqüências gênicas e elementos de transposição para as variáveis Kimura e Ka.

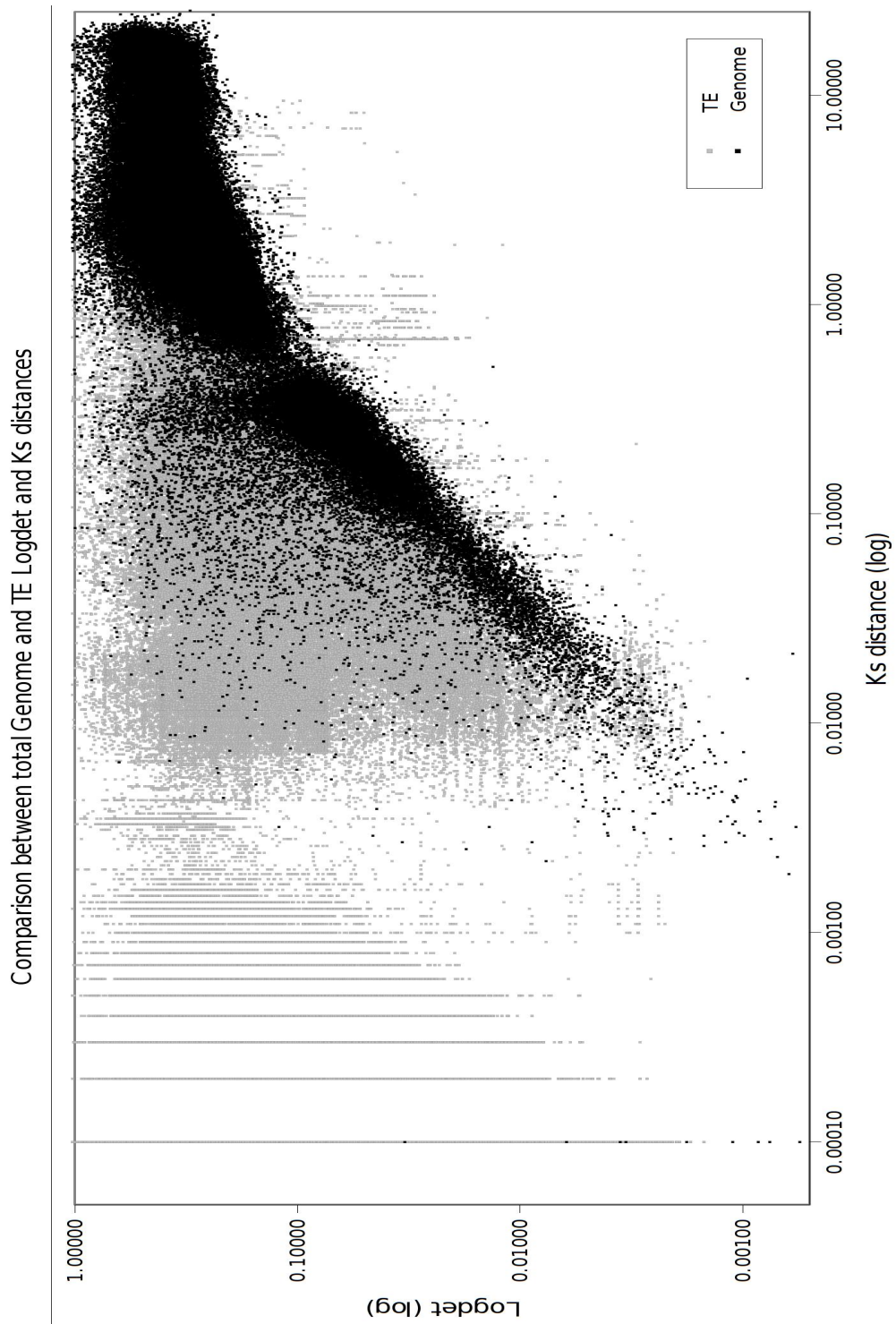


Figura 37 – Gráfico de dispersão para dados de relação de sequências gênicas e elementos de transposição para as variáveis Logdet e Ks.

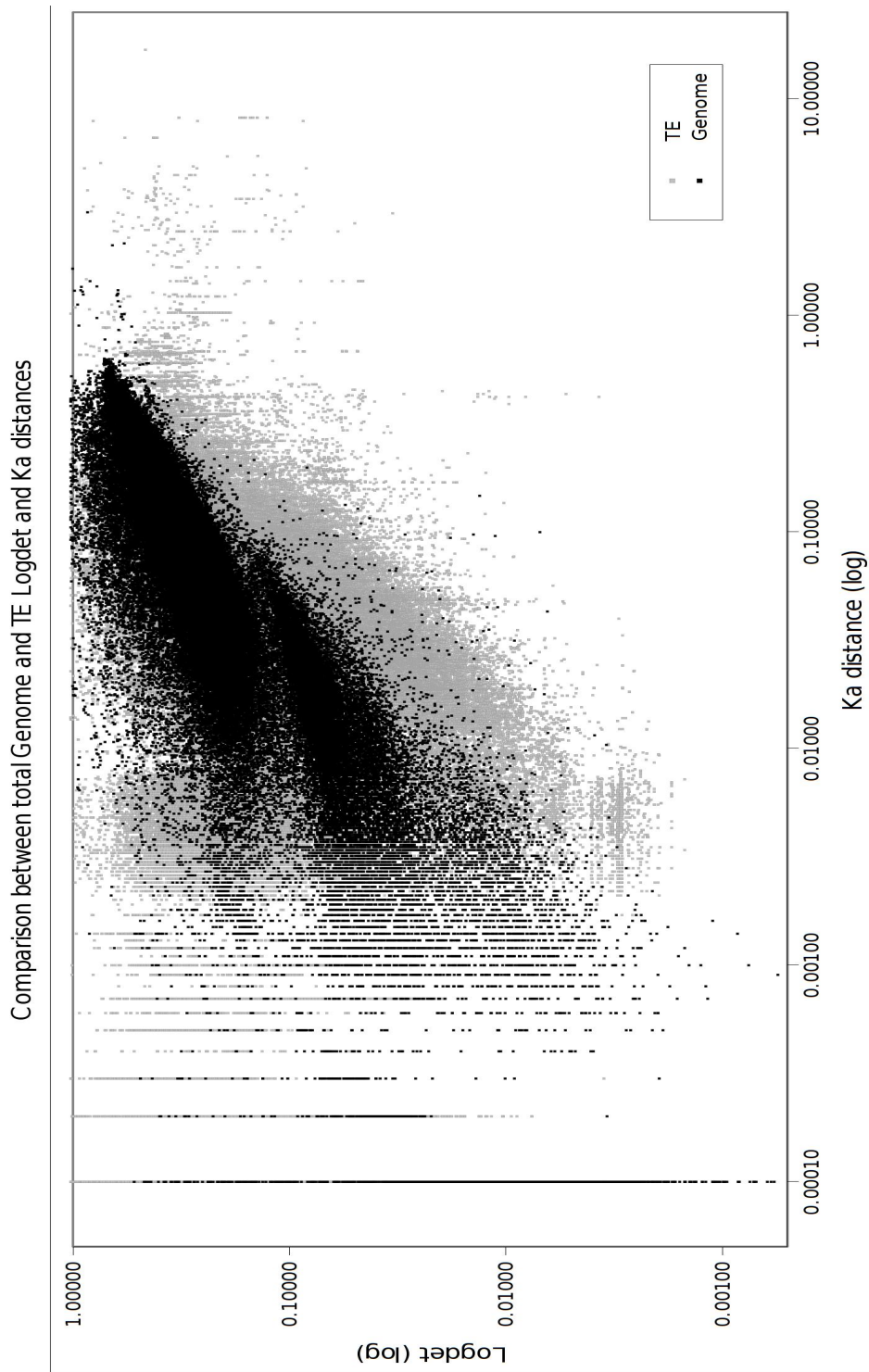


Figura 38 – Gráfico de dispersão para dados de relação de seqüências gênicas e elementos de transposição para as variáveis Logdet e Ka.

Com o intuito de realizar uma comparação direta entre as diferentes variáveis utilizadas para descrever cada par de relação entre sequências de elementos de transposição e sequências gênicas visando a identificação de transferência horizontal foi desenvolvido um index de similaridade. Este index é calculado de forma independente para cada par de comparação de espécies no conjunto de dados de genes ou elementos de transposição. Este index é formado pelos valores das variáveis Ka, Ks, DNAId, ProtID, Logdet e Kimura equacionadas da seguinte forma:

$$\text{Index} = (((\text{DNAId} + \text{ProteinID}) / 2) / (\text{Ka} + \text{Ks} + 1)) / (\text{Logdet} + \text{Kimura} + 1)$$

Esta fórmula gera um valor máximo igual a 100 quando as sequências em um par de comparação são idênticas, apresentando valores proporcionalmente menores a medida que as sequências apresentem maior divergência. Assim, sequências pertencentes a elementos de transposição que apresentarem valores de Index maiores que genes com o maior Index entre duas espécies apresenta-se como candidatas a transferência horizontal pois necessariamente são mais conservadas entre si que o gene mais conservado entre as duas espécies em análise. Um sumário dos valores de Index obtidos neste trabalho pode ser observado na Tabela 9. É importante ressaltar que tal critério é bastante estrigente em relação à identificação de casos de transferência horizontal. Considerando a rápida taxa evolutiva de elementos de transposição, muitos casos mais antigos de transferência horizontal não apresentarão valores de conservação tão elevado e portanto, valores de

index mais baixos, fazendo com que a sua detecção seja dificultada através dessa técnica.

A Figura 39 apresenta uma representação da variação dos valores de Index sumarizados para os diferentes grupos de comparações de genoma, tanto para valores de elementos de transposição quanto para o genoma. É possível observar que em muitos grupos os valores de Index para elementos de transposição são notadamente superiores aos valores médios do genoma, como por exemplo na comparação entre espécies do grupo *Melanogaster* com o grupo *Wilistoni*.

A Figura 40 mostra que de forma geral os elementos de transposição apresentam valores de Index concentrados em torno do valor 80, o que representa uma conservação relativamente alta. Para esta mesma faixa de valor de Index os elementos de transposição também apresentam uma maior quantidade de comparações em relação ao genoma. É importante notar a correlação direta entre os valores DNAid e os valores de Index, corroborando que este pode ser utilizado para comparações diretas entre diferentes pares de genes ou de elementos de transposição.

Tabela 9:- Valores médios, máximo (max), mínimo (min) e desvio padrão do index de similaridade, organizados de acordo com cada comparação entre genomas de *Drosophila* para sequências gênicas e de elementos de transposição.

		Genoma				TE			
		index				index			
		Média	Max	Min	Desvio Padrão	Média	Max	Min	Desvio padrão
Melanogaster_Group	dana_dere	21.67	68.36	1.53	9.10	79.77	97.41	3.29	21.17
	dana_dmel	21.53	67.93	1.33	9.06	26.39	99.53	3.21	12.19
	dana_dsec	21.22	68.84	0.96	9.11	27.51	55.83	4.66	10.12
	dana_dsim	21.13	67.15	0.85	9.10	27.98	57.50	4.66	9.68
	dana_dyak	21.77	67.62	1.41	9.02	22.80	96.89	1.30	21.48
Melanogaster_Subgroup	dmel_dere	63.87	91.89	16.74	8.42	57.67	99.23	5.32	15.97
	dmel_dsec	77.55	94.99	8.32	8.51	67.75	99.78	2.96	16.47
	dmel_dsim	76.62	95.98	6.35	10.04	67.49	99.80	4.37	18.56
	dmel_dyak	62.86	92.32	6.50	8.37	63.62	99.69	5.61	19.61
	dsec_dere	63.46	94.80	7.34	9.47	50.96	96.41	5.95	13.44
	dsec_dsim	84.77	99.74	6.17	12.20	71.62	99.76	3.16	17.63
	dsec_dyak	62.35	94.10	6.60	9.48	66.74	99.23	4.05	17.76
	dsim_dere	62.81	92.80	5.02	10.40	54.60	92.77	3.65	14.96
	dsim_dyak	61.76	92.31	4.97	10.35	65.91	99.33	7.50	18.50
dyak_dere	69.88	93.40	6.61	7.95	59.29	99.70	4.62	20.18	
Melanogaster-Hawaiian	dana_dgri	10.35	49.07	0.82	7.32	26.19	29.40	21.31	2.15
Melanogaster-Obscura	dana_dper	15.91	63.77	0.95	8.73	13.69	52.43	2.40	8.87
	dana_dpse	16.50	63.24	0.92	8.87	27.40	51.74	4.08	15.63
Melanogaster-Repleta	dana_dmoj	9.98	53.35	0.68	7.41	31.44	66.66	3.21	15.15
Melanogaster-Virilis	dana_dvir	10.61	54.03	0.84	7.54	24.44	61.35	4.64	12.53
Melanogaster-Willistoni	dana_dwil	9.48	53.92	0.89	7.02	29.75	89.42	4.09	17.33
MelanogasterSubgroup-Hawaiian	dgri_dere	11.48	44.44	0.93	7.43				
	dgri_dmel	11.85	45.99	0.83	7.64	11.86	11.86	11.86	
	dgri_dsec	11.42	44.07	0.81	7.46	10.83	10.83	10.83	
	dgri_dsim	11.43	45.43	0.77	7.43	11.58	11.58	11.58	
	dgri_dyak	11.73	46.10	0.88	7.57				
MelanogasterSubgroup-Obscura	dmel_dper	17.25	59.10	0.91	8.64	10.82	39.26	4.23	11.31
	dmel_dpse	17.81	58.39	0.95	8.71	20.64	39.19	4.31	12.68
	dper_dere	17.35	60.53	1.08	8.63	16.67	18.31	15.03	2.32
	dper_dsec	17.04	59.20	0.88	8.64	9.63	38.01	4.24	7.12
	dper_dsim	17.04	59.42	0.90	8.63	10.74	38.01	4.25	9.45
	dper_dyak	17.48	58.88	1.13	8.62	10.36	30.86	3.72	4.26
	dpse_dere	17.93	59.81	1.10	8.73	18.16	18.16	18.16	
	dpse_dsec	17.56	58.49	0.90	8.74	18.14	37.86	6.05	11.97
	dpse_dsim	17.55	59.42	0.87	8.73	18.76	37.86	3.94	13.47

	dpse	dyak	18.04	58.18	1.12	8.74	9.29	26.81	2.86	3.85
MelanogasterSubgroup-Repleta	dmel	dmoj	11.13	49.25	0.85	7.58	36.34	51.32	4.94	13.87
	dmoj	dere	10.89	46.05	0.83	7.46				
	dmoj	dsec	10.73	48.82	0.86	7.42	30.32	51.06	4.49	14.77
	dmoj	dsim	10.71	48.25	0.85	7.40	34.61	51.00	5.53	14.33
	dmoj	dyak	11.12	46.98	0.81	7.57	16.85	31.68	4.96	6.76
MelanogasterSubgroup-Virilis	dmel	dvir	12.03	46.71	0.89	7.72	29.75	44.30	9.98	11.18
	dsec	dvir	11.71	46.62	0.91	7.53	30.03	46.33	10.02	11.49
	dsim	dvir	11.64	46.23	1.00	7.51	32.17	44.48	10.02	11.71
	dvir	dere	11.76	46.03	0.93	7.59	44.27	48.12	16.93	5.86
	dvir	dyak	12.01	46.40	0.89	7.65	16.28	24.34	5.54	8.47
MelanogasterSubgroup-Willistoni	dmel	dwil	10.34	48.92	0.99	7.23	29.63	55.04	5.18	16.10
	dsec	dwil	9.84	46.85	1.02	6.99	24.82	64.90	6.58	16.79
	dsim	dwil	9.78	44.74	1.03	6.95	28.40	65.98	6.75	16.75
	dwil	dere	9.89	46.32	0.85	6.94				
	dwil	dyak	10.17	48.11	0.91	7.03	12.12	61.14	5.91	6.78
Obscura Group	dper	dpse	90.22	99.72	18.93	10.88	66.25	99.73	1.72	18.63
Obscura-Hawaiian	dgri	dper	14.79	52.87	0.81	7.90	10.03	10.03	10.03	
	dgri	dpse	15.26	52.92	0.80	7.94	10.01	10.01	10.01	
Repleta-Hawaiian	dgri	dmoj	26.21	72.02	1.34	9.18				
Repleta-Obscura	dmoj	dper	14.27	59.76	0.96	7.95	19.97	52.57	3.47	13.47
	dmoj	dpse	14.75	60.44	0.99	8.05	34.49	51.66	3.49	10.65
Virilis-Hawaiian	dgri	dvir	31.61	74.06	1.48	9.51	19.62	22.15	17.10	3.57
Virilis-Obscura	dper	dvir	15.70	56.28	1.01	8.02	21.07	68.84	3.28	21.04
	dpse	dvir	16.19	56.91	0.93	8.14	49.96	68.84	4.45	13.19
Virilis-Repleta	dmoj	dvir	33.82	77.82	2.48	9.65	46.08	99.76	3.42	31.67
Willistoni-Hawaiian	dgri	dwil	13.63	55.59	1.02	7.91				
Willistoni-Obscura	dper	dwil	12.86	56.93	1.05	7.68	23.75	76.25	3.62	10.21
	dpse	dwil	13.28	56.63	1.04	7.73	26.82	48.59	5.40	6.91
Willistoni-Repleta	dmoj	dwil	12.29	53.22	0.97	7.84	31.45	79.30	3.58	11.52
Willistoni-Virilis	dvir	dwil	13.57	53.39	0.91	8.03	26.85	58.31	3.75	12.01

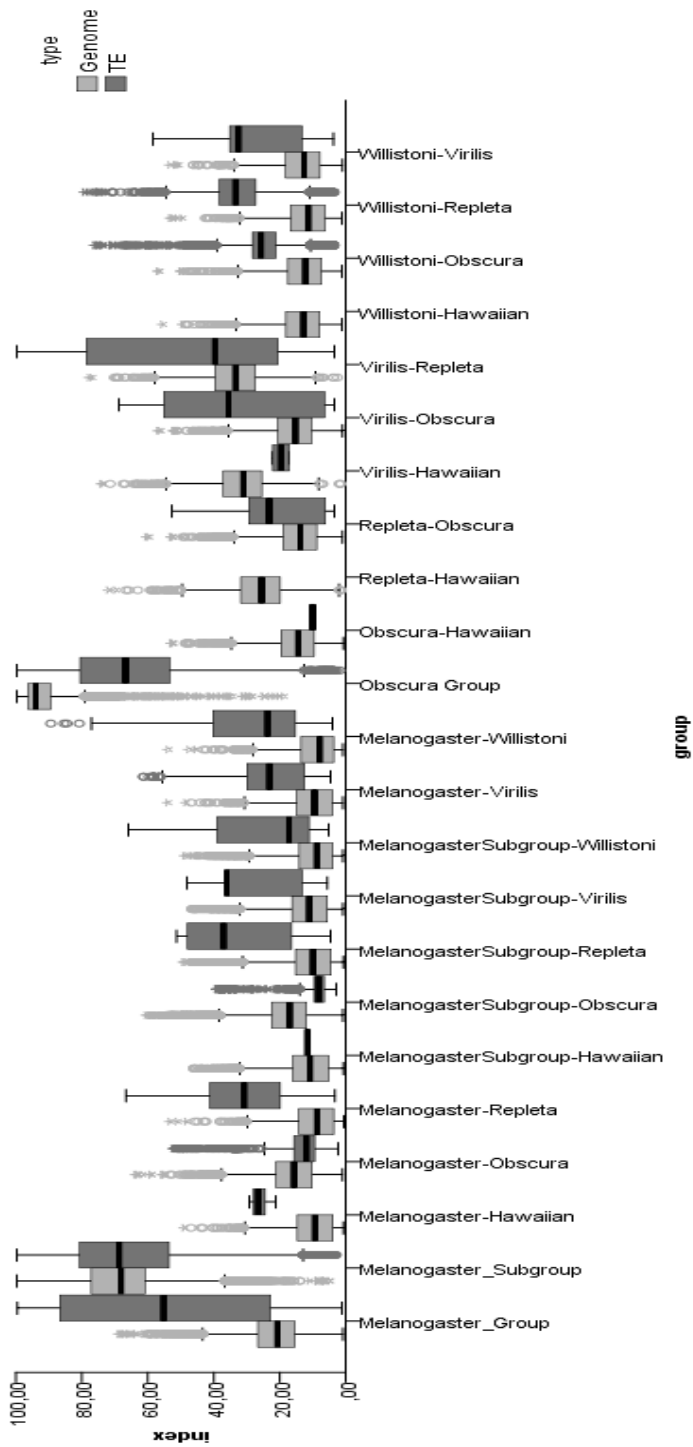


Figura 39 – Diagrama de caixa para dados de relação de sequências gênicas e elementos de transposição para a variável Index.

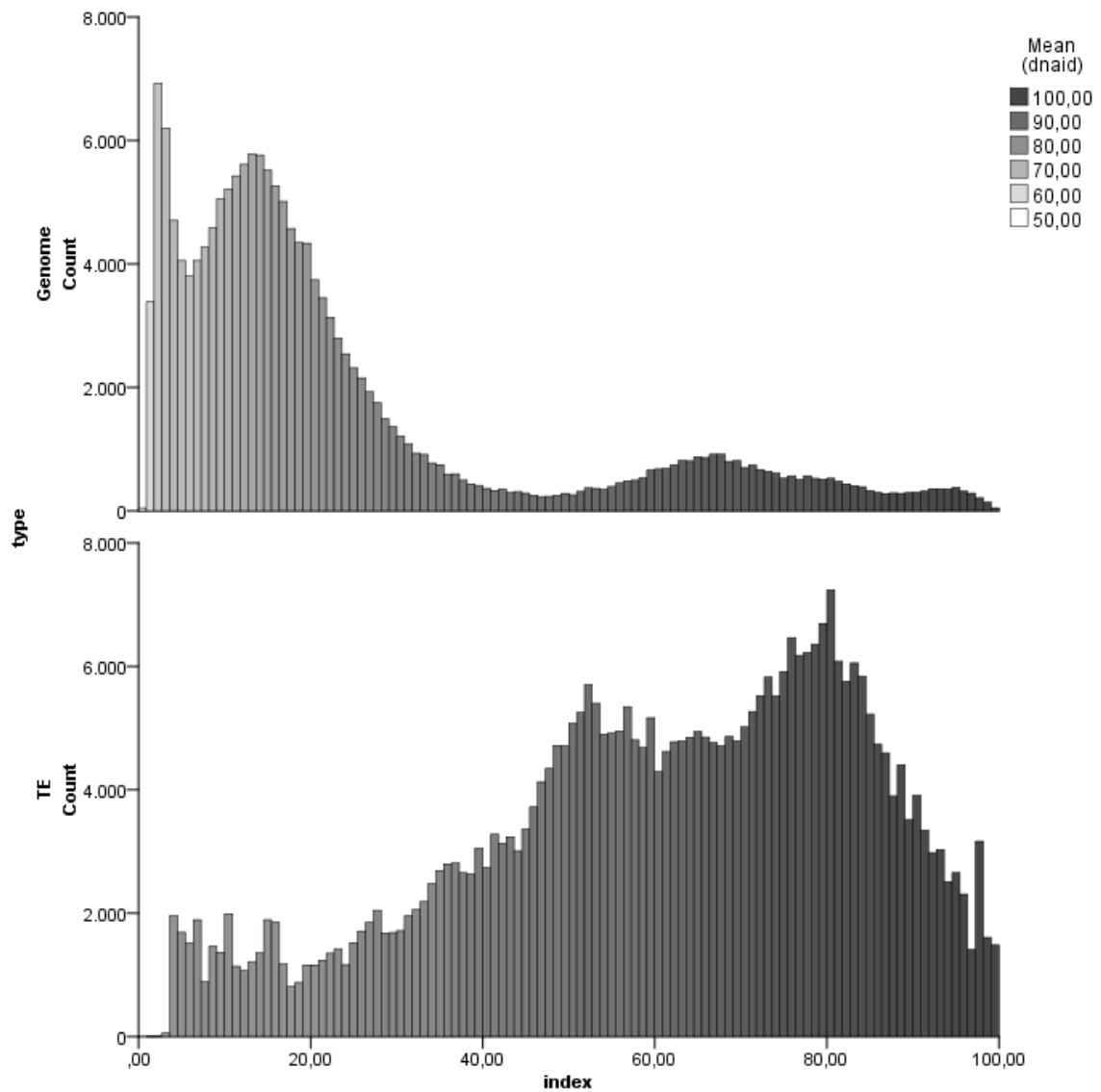


Figura 40 – Histograma dos valores de index separados de acordo com dados derivados de comparações entre sequências gênicas (painel superior) ou dados derivados de comparações entre sequências de elementos de transposição (painel inferior). A cor representa o valor médio da variável DNAId para cada intervalo.

Aplicando-se o cálculo deste Index a todas as comparações entre os 12 genomas de *Drosophila* e seus elementos de transposição foi possível identificar 25 relação entre espécies onde as sequências de transposons apresentavam valores claramente inferiores em relação genes mais conservados. No total 109 casos de transferência horizontal podem ser identificados em 25 relações de espécies, conforme descrito na Tabela 10.

A figura 41 faz uma representação em rede dos eventos de transferência horizontal listados na Tabela 10, também representando os valores de Index como a largura das linhas realizando a conexão entre as espécies onde possivelmente ocorreram eventos de transferência horizontal e o valor de Logdet para comparação como a cor da linha de conexão. As comparações de sequências de elementos de transposição entre as espécies na Figura 41 que não representam eventos de transferência horizontal são representas por uma linha tracejada.

Analisando especificamente as características das comparações identificadas como parte de um possível evento de transferência horizontal verifica-se que estas comparações possuem valores de Ks significativamente menores que as comparações entre genes (Figuras 42 e 45) ao mesmo tempo que apresentam alta similaridade de DNA (Figuras 43 e 47). Contudo, mesmo considerando este subconjunto de dados é possível identificar a repetição do padrão observado quando são analisados todos os elementos de transposição, com valores de variáveis ligadas a substituição não-sinônima em valores mais elevados (Figuras 44, 46 e 48).

Tabela 10 – Listagem os possíveis casos de transferência horizontal conforme inferido pela comparação entre valores do index de similaridade entre genes nucleares e sequencias de elementos de transposição. Os valores das variáveis evolutivas referem-se aos extremos observados dentro dos casos de transferência horizontal.

<i>Espécie 1</i>	<i>Espécie 2</i>	<i>Min Ka</i>	<i>Min Ks</i>	<i>Max ProtID</i>	<i>Max DNAlD</i>	<i>Min Logdet</i>	<i>Min Kimura</i>	<i>N</i>
Dana	Dere	0.0000	0.0000	100.00	100.00	0.0056	0.0066	1
Dana	Dmel	0.0047	0.0000	100.00	100.00	0.0000	0.0000	1
Dana	Dmoj	0.0000	0.0000	88.57	89.36	0.1477	0.1243	1
Dana	Dwil	0.0000	0.0006	93.09	95.05	0.0527	0.0718	2
Dana	Dyak	0.0010	0.0000	94.83	96.64	0.0776	0.0411	2
Dmel	Dere	0.0000	0.0036	99.68	99.40	0.0074	0.0055	5
Dmel	Dsec	0.0001	0.0000	99.94	99.91	0.0012	0.0016	25
Dmel	Dsim	0.0003	0.0004	100.00	99.98	0.0002	0.0002	26
Dmel	Dwil	0.0003	0.0035	78.63	82.91	0.1703	0.2141	1
Dmel	Dyak	0.0000	0.0012	99.51	99.63	0.0030	0.0062	11
Dmoj	Dper	0.0003	0.0000	89.81	88.33	0.1522	0.0885	1
Dmoj	Dpse	0.0003	0.0000	89.81	87.04	0.2413	0.0885	1
Dmoj	Dvir	0.0000	0.0000	100.00	100.00	0.0000	0.0000	1
Dmoj	Dwil	0.0186	0.0221	90.78	93.35	0.1073	0.0828	6
Dper	Dvir	0.0157	0.0001	90.66	94.30	0.0702	0.0986	2
Dper	Dwil	0.0000	0.0078	91.93	91.42	0.0960	0.0695	2
Dpse	Dvir	0.0104	0.0001	90.31	94.61	0.0560	0.1039	2
Dsec	Dere	0.0003	0.0001	99.19	99.46	0.0056	0.0082	1
Dsec	Dwil	0.0376	0.0621	89.19	93.69	0.0957	0.1170	1
Dsec	Dyak	0.0000	0.0017	99.62	99.46	0.0052	0.0069	6
Dsim	Dere	0.0088	0.0002	98.37	99.05	0.0104	0.0164	1
Dsim	Dwil	0.0335	0.0764	90.09	93.99	0.0901	0.1065	1
Dsim	Dyak	0.0000	0.0014	99.51	99.75	0.0024	0.0075	6
Dwil	Dyak	0.0000	0.0041	84.44	89.88	0.1855	0.1748	1
Dyak	Dere	0.0000	0.0000	100.00	100.00	0.0017	0.0017	2

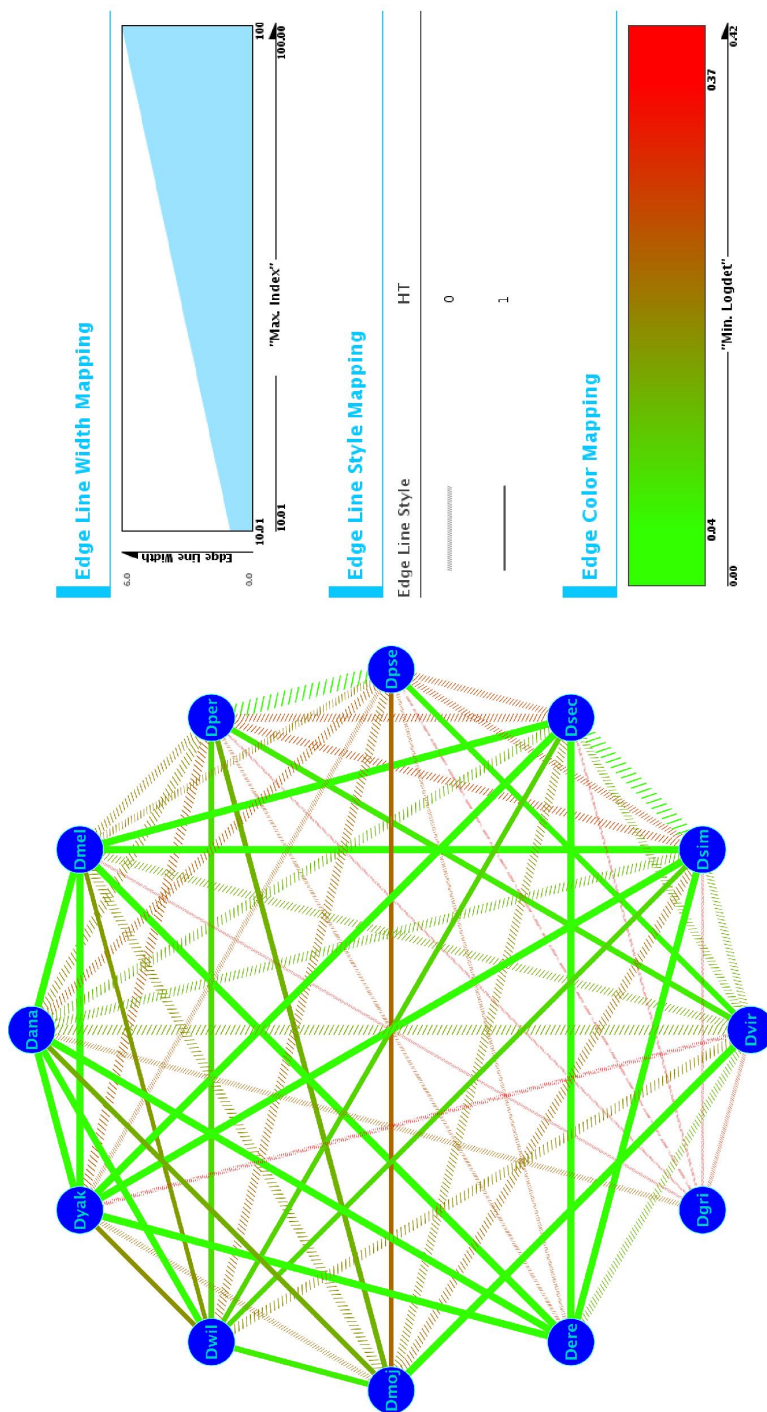


Figura 41- Grafo de conexão entre as 12 espécies de *Drosophila* e os respectivos casos de transferência horizontal conforme inferido pela análise comparativa do index de similaridade.

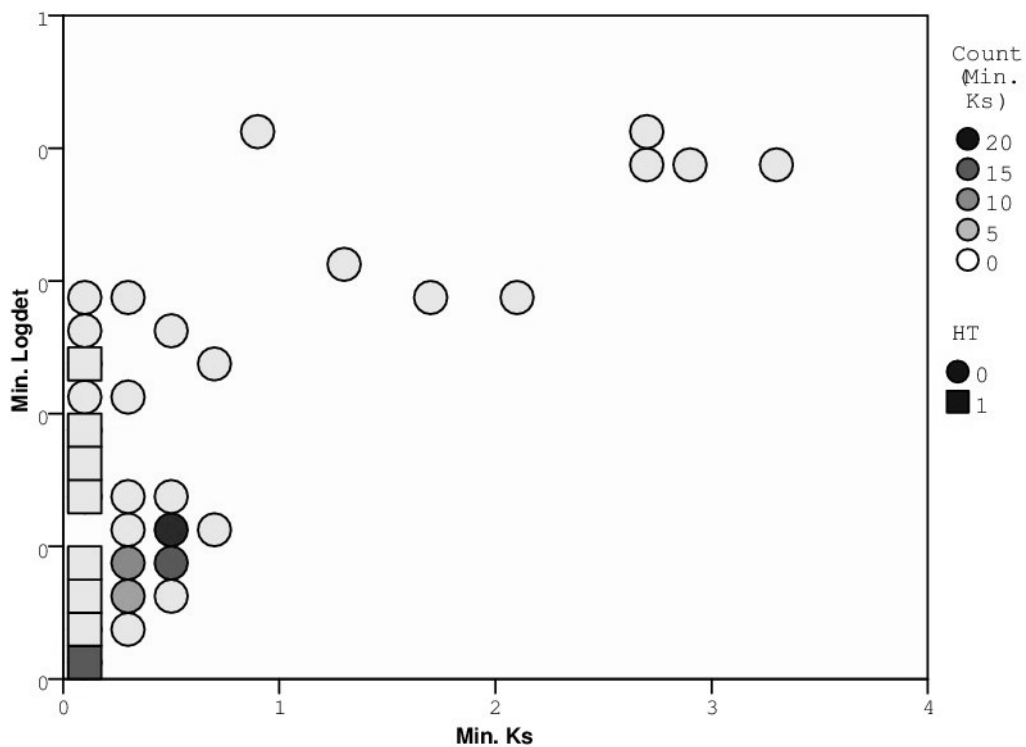


Figura 42 – Gráfico de dispersão para dados de relação de sequências gênicas e elementos de transposição para as variáveis Ks mínimo (entre todas as comparações de genomas) e Logdet mínimo. A intensidade de cor de cada célula representa a quantidade de pontos de dados de comparação agrupados naquela região do plano cartesiano. O formato de cada célula indica a ocorrência de transferência horizontal (quadrado) ou a ausência de transferência horizontal (círculo) para cada grupo de dados representados pela célula.

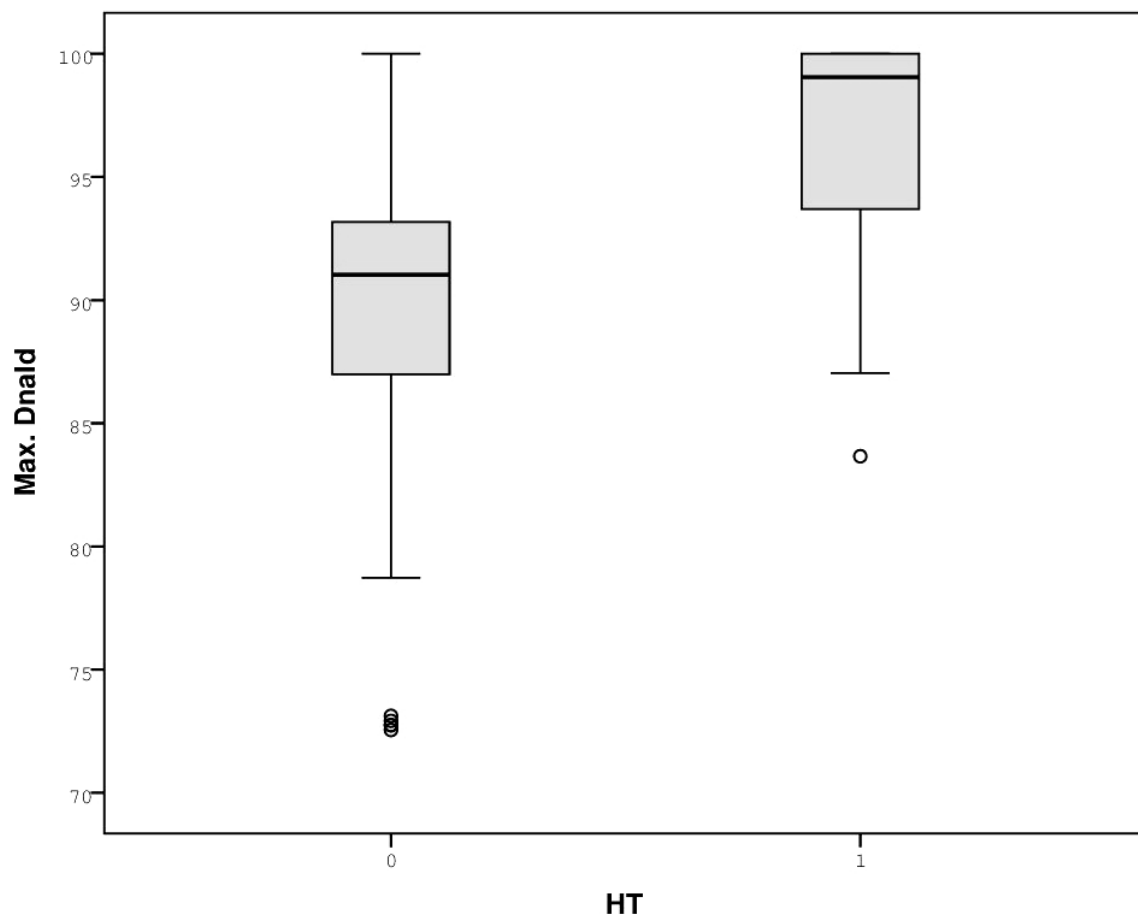


Figura 43 – Diagrama de caixa para dados da variável DNAid de sequências de elementos de transposição separados de acordo com a existência de transferência horizontal (1) ou a ausência de transferência horizontal (0).

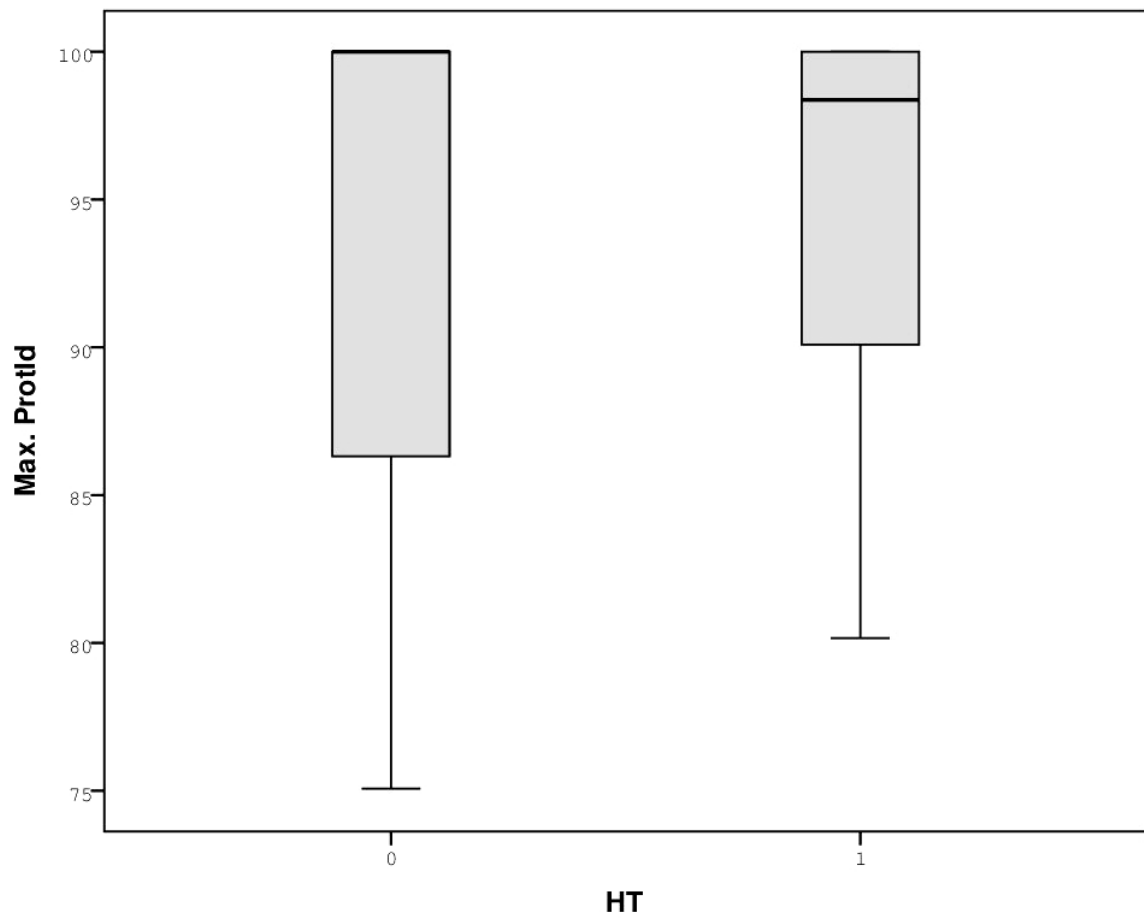


Figura 44 – Diagrama de caixa para dados da variável ProtId de sequências de elementos de transposição separados de acordo com a existência de transferência horizontal (1) ou a ausência de transferência horizontal (0).

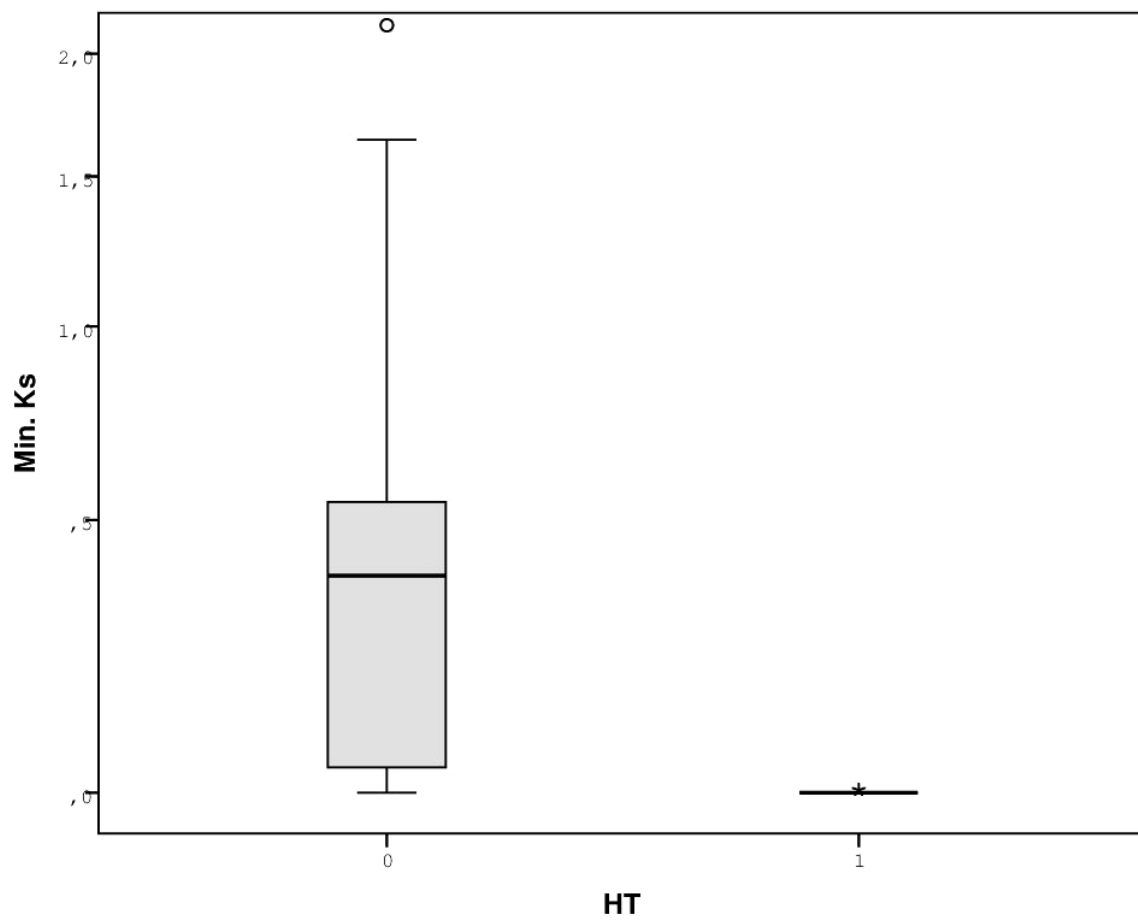


Figura 45 – Diagrama de caixa para dados da variável Ks de sequências de elementos de transposição separados de acordo com a existência de transferência horizontal (1) ou a ausência de transferência horizontal (0).

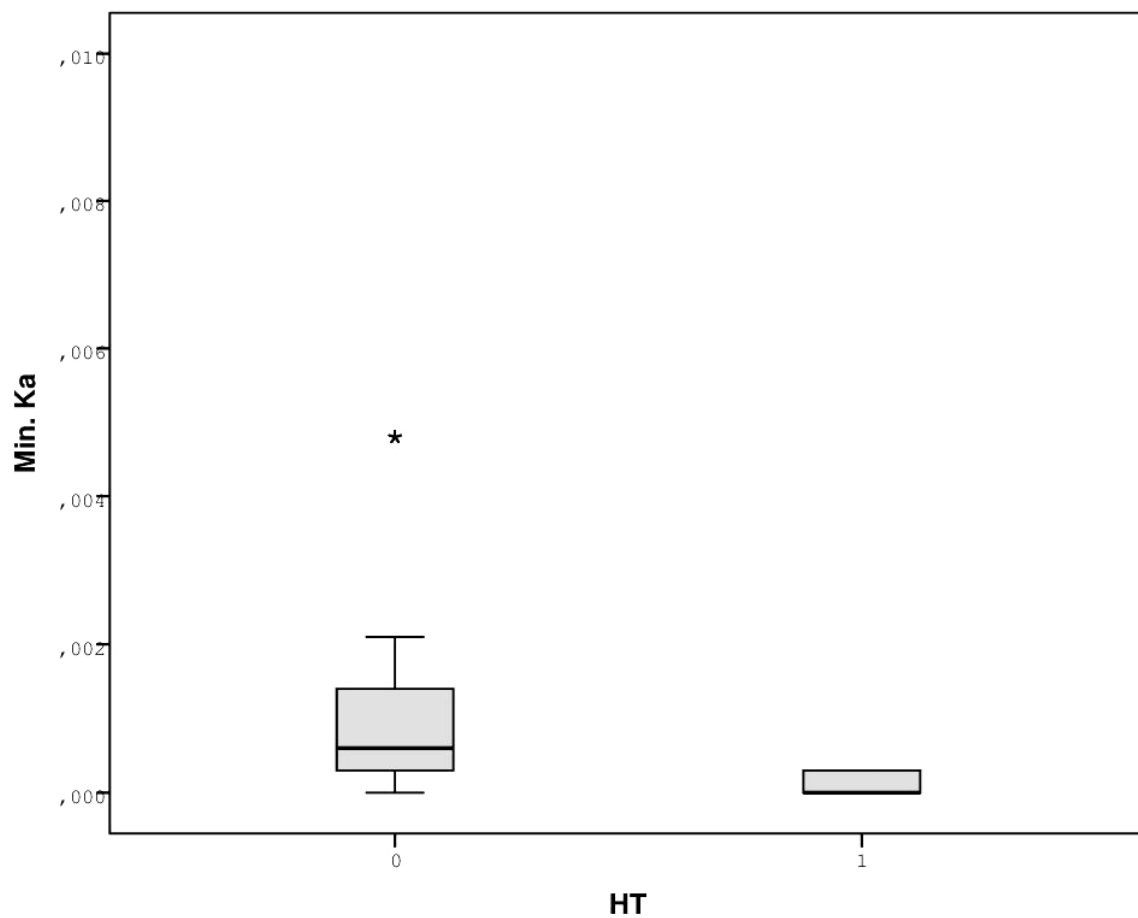


Figura 46 – Diagrama de caixa para dados da variável Ka de seqüências de elementos de transposição separados de acordo com a existência de transferência horizontal (1) ou a ausência de transferência horizontal (0).

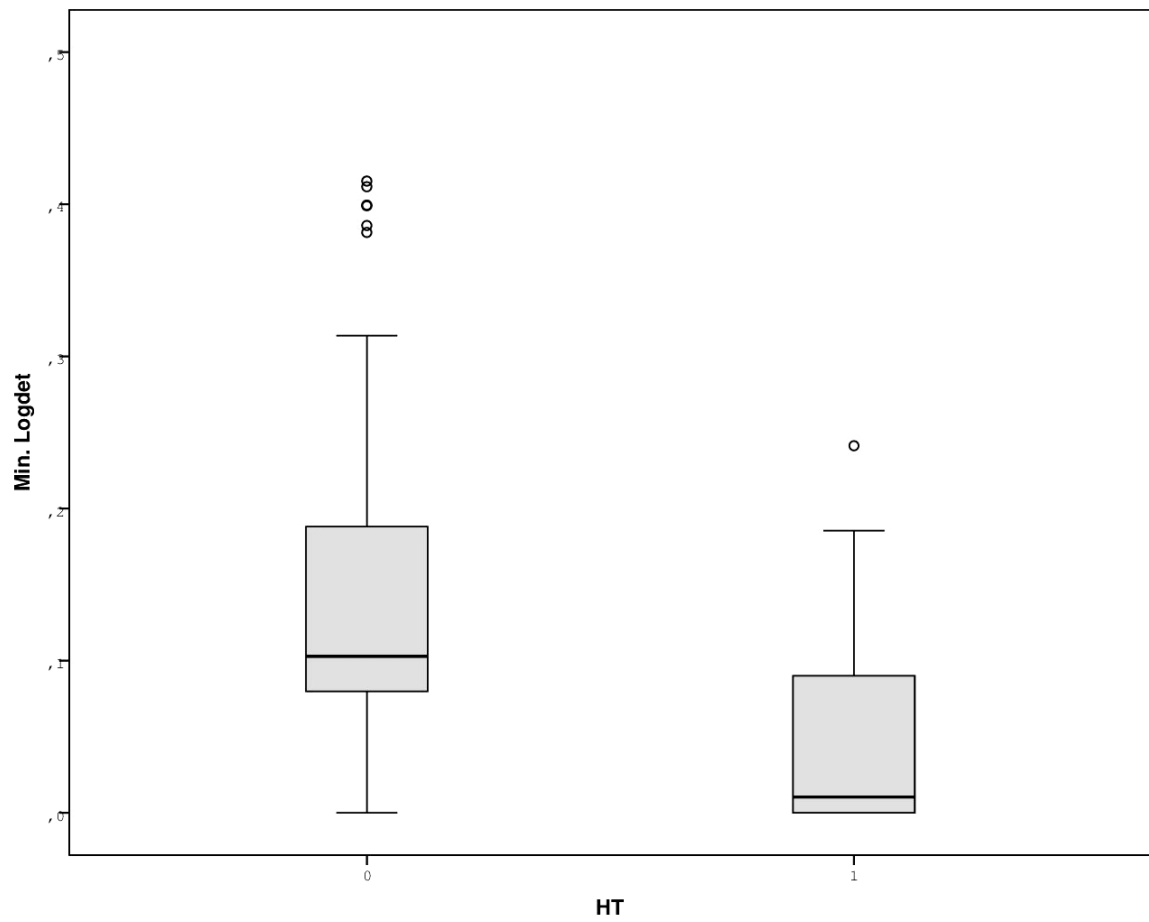


Figura 47 – Diagrama de caixa para dados da variável Logdet de sequências de elementos de transposição separados de acordo com a existência de transferência horizontal (1) ou a ausência de transferência horizontal (0).

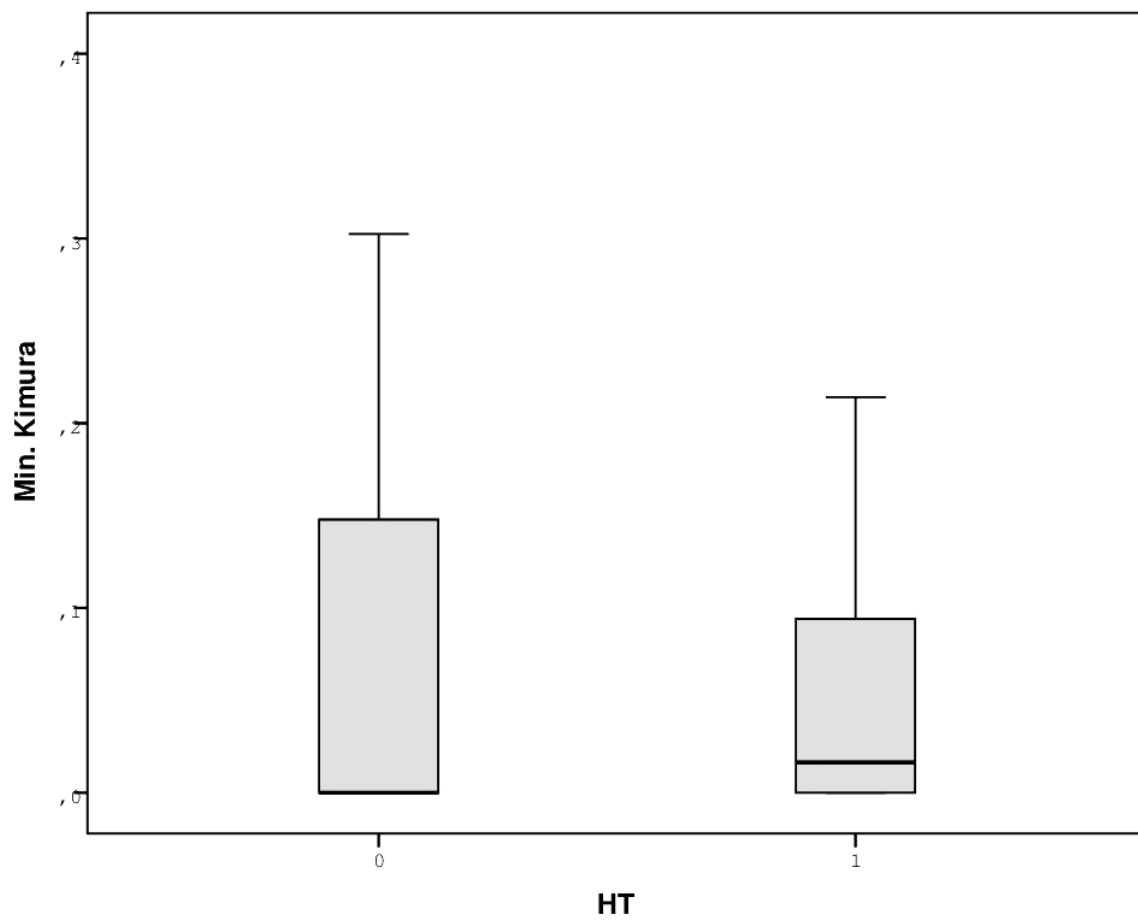


Figura 48 – Diagrama de caixa para dados da variável Kimura de sequências de elementos de transposição separados de acordo com a existência de transferência horizontal (1) ou a ausência de transferência horizontal (0).

Especificamente no caso da relação entre *Drosophila ananassae* e *Drosophila yakuba*, podemos verificar na Figura 49 ao menos uma família de elemento de transposição, o elemento Mariner, com um padrão diferenciado em relação ao genoma e às outras famílias de elementos de transposição.

Estes elementos possuem valores de similaridade de sequência de DNA bem como valores de Ks significativamente diferente do genoma sendo os valores de Ks marcadamente inferiores ao do genoma. Nota-se que existem outras sequências de elemento Mariner que possuem um padrão evolutivo similar às sequências gênicas, com maior valores de Ks e menor similaridade. Pode-se inferir que estas representam sequências ancestrais derivadas de outros processos de colonização, enquanto as sequências de Mariner com alta similaridade são provenientes de um evento recente de transferência horizontal.

Refletindo o padrão mostrado pela Figura 49, a Figura 50 mostra as sequências de elementos Mariner notadamente disjuntas dos valores dos genes mais conservados entre *Drosophila ananassae* e *Drosophila yakuba*, apresentando alta similaridade também mostrada pelos baixos valores de distância Logdet.

É possível também observar elementos Aurora e Helitron com padrões diferenciados em relação ao genoma, contudo não com a conservação observada para elementos Mariner.

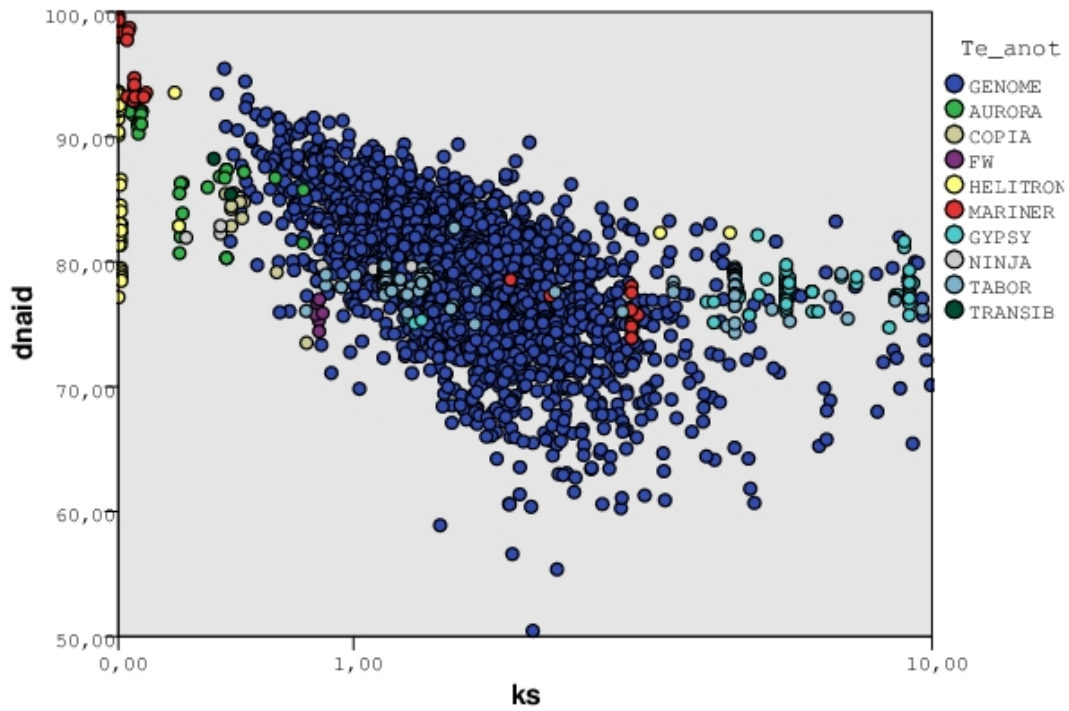


Figura 49 – Gráfico de dispersão para dados de relação de sequências gênicas e elementos de transposição para as variáveis DNAid e Ks entre as espécies *Drosophila ananassae* e *Drosophila yakuba*. A cor de cada ponto de dados indica a família de elemento de transposição respectiva ou sequência gênica. Eixo das abscissa em escala log.

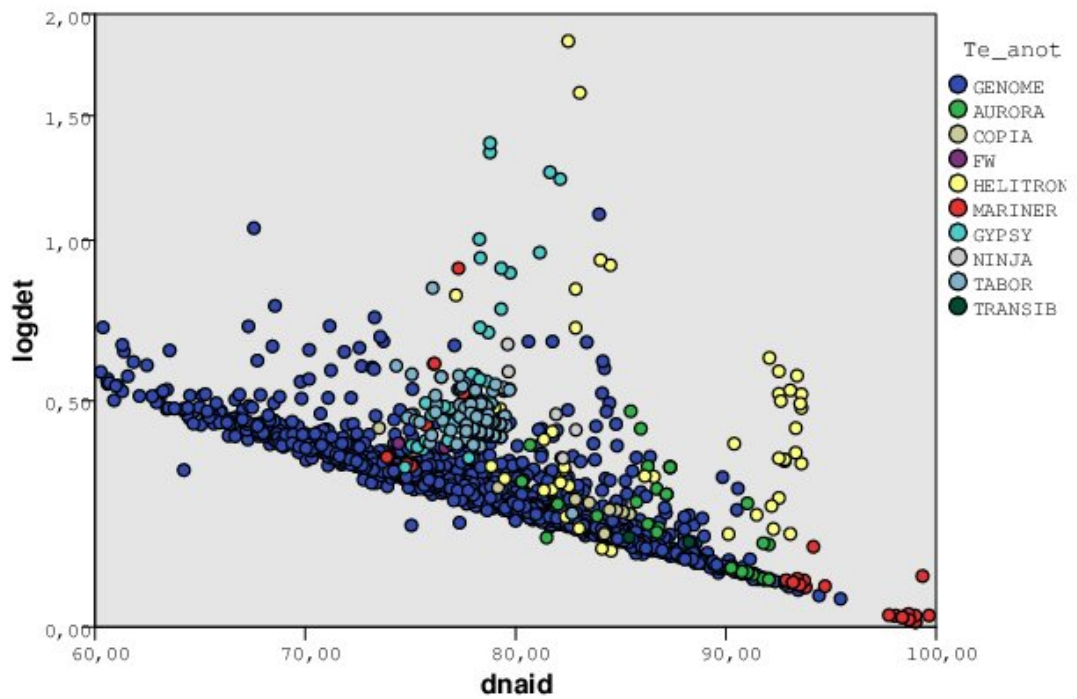


Figura 50 – Gráfico de dispersão para dados de relação de sequências gênicas e elementos de transposição para as variáveis DNAid e Logdet entre as espécies *Drosophila ananassae* e *Drosophila yakuba*. A cor de cada ponto de dados indica a família de elemento de transposição respectiva ou sequência gênica. Eixo das ordenadas em escala log.

Ao utilizar a taxa de substituição sinônima como métrica de evolução neutra é preciso levar em consideração o fenômeno de viés de códon (Hershberg e Petrov, 2008), onde pressupõe-se a existência de seleção sobre sítios sinônimos. Esta seleção é derivada da preferência por códons com maior eficiência de tradução (Vicario *et al.*, 2008), seja por correlação estocástica com a disponibilidade de moléculas de tRNAs (Shah e Gilchrist, 2010) ou acoplamento à maquinaria ribossomal (Fuglsang, 2003). Assim, genes que apresentam maiores taxas de expressão geralmente também possuem maiores índices de viés de códon (Vicario *et al.*, 2008). Altos índices de viés de códon também são associados a genes com maior “essencialidade”(Rocha e Danchin, 2004). Mesmo que estes não apresentem altos níveis de expressão, a necessidade de rápida resposta de expressão por parte desses genes induz à seleção de códons com maior eficiência de tradução.

Esta seleção translacional leva a uma reduzida taxa de substituição sinônima em genes sujeitos a restrições relativas às suas taxas de expressão ou essencialidade (Powell e Moriyama, 1997). Sendo sujeitos à seleção em sítios sinônimas, tais genes podem levar a inferências errôneas caso sejam utilizados como marcadores de evolução neutra. O método de análise via *likelihood* implementado no software PAML (Yang, 2007) introduz o conceito de correção do cálculo de substituições sinônima de acordo com variações no uso de códon, contudo outras evidências indicam que as premissas utilizadas nos modelos de uso de códon não são suficientes para estabelecer um modelo real do processo de evolução neutra (Bauer DuMont *et al.*, 2009).

A existência ampla de viés de códon ligado a seleção translacional de sítio sinônimos

introduz portanto uma possível fonte de erro de cálculo ao realizarmos comparações entre sequências de elementos de transposição e o genoma com o objetivo de identificar transferência horizontal (Vidal, Ludwig e Loreto, 2009) . Uma vez que esteja ocorrendo seleção translacional em sítios sinônimos de elementos de transposição, valores de Ks abaixo do esperado não seriam derivados do fato destas sequências estarem presentes no genoma por um tempo menor do que as sequências do genoma hospedeiro, mas sim à seleção de sítios sinônimos visando à constante otimização translacional (Wallau, Kaminski e Loreto, 2011).

Com o objetivo de testar a extensão do viés de códon encontrado em elementos de transposição e em sequências dos respectivos genomas calculou-se o índice de viés de códon ENC (número efetivo de códons) para cada gene e sequência de TE utilizada neste estudo. O índice ENC é uma medida geral do viés de códon ao longo da sequência, sendo calculado de maneira análoga ao número efetivo de alelos, utilizado em genética de populações (Wright, 1990). Os valores de ENC variam de 20, quando apenas um códon é usado para cada aminoácido na sequência (representando portanto um alto viés de uso de códon) até 61, quando todos os códons são utilizados de forma homogênea para cada aminoácido (representando um baixo viés de códon).

As Tabelas 11 e 12 apresentam um sumário dos dados de ENC calculados como a média dos valores de viés de códon (*avg_nc*) para cada sequência no par de comparação para genes ou elementos de transposição. A Tabela 11 também apresenta a variável *diff_nc*, que representa a relação entre o maior valor de viés de códon em um determinado par de sequência dividido pelo menor valor. Essa relação visa determinar uma relação de mudança

no viés de códon entre duas sequências homólogas (no caso de genes) ou de sequências similares no caso de elementos de transposição. Valores de $diff_{nc}$ próximos de 1 indicam que ambas as sequências do par possuem valores aproximados de viés de códon.

É possível observar que apesar da existência tanto de genes quanto sequências de elementos de transposição com valores de ENC que indicam a existência de viés de códon acentuado, estas apresentam-se em pequena proporção em relação ao total de sequências analisadas, conforme é possível observar nas Figuras 54 e 55. Da mesma forma, apesar de ser possível verificar a existência de uma fraca correlação entre viés de códon e a taxa de substituições sinônimas, esta não é caracterizada por uma tendência direta. Verifica-se na Figura 54 a existência de sequências tanto de transposons quanto de genes com alto viés de códon e baixa taxa de substituição sinônima, sendo o contrário também verdadeiro. Uma pequena correlação entre viés de códon e taxas de substituição sinônima é observada na Figura 55, mas apenas para sequências gênicas, sendo que as sequências de elementos de transposição apresentam uma grande heterogeneidade na relação entre valores de K_s e viés de códon. Esta heterogeneidade pode ser observada pela maior diferença entre viés de códon observada entre os pares de sequências do que em relação às sequências gênicas, como pode-se observar nas Figuras 51, 52 e 53. É possível que essas diferenças em relação à variável ENC nos diferentes pares de sequências de elementos de transposição seja fruto de uma complexa interação entre diferentes forças de seleção e processos de evolução neutra, que no entanto, não se resumem à forças de seleção em sítios sinônimos ou aos mesmos processos evolutivos atuantes sobre as sequências gênicas. Essa hipótese é corroborada pela clara estratificação dos pares de valores de ENC e identidade de DNA

plotada nas Figuras 51 e 52 para as sequências gênicas e o padrão disperso observado para as sequências de elementos de transposição.

Dessa forma, dentro do contexto de análise em larga escala de detecção de transferência horizontal aqui apresentado, onde também são tomadas como referência as variáveis de substituição não-sinônima, dados de similaridade e distância filogenética, pode-se concluir que o fenômeno de viés de códon não é especialmente relevante. Contudo, nesta análise foram estabelecidos parâmetros estridentes para a determinação de ocorrência de transferência horizontal. Ao refinar a análise para a determinação de eventos mais antigos onde exista um processo de melhoria das sequências de elementos de transposição, a influência da seleção sobre substituições sinônimas e outros processos que refletem-se na diversidade de valores de viés de códon em elementos de transposição podem se tornar relevantes.

Tabela 11:- Valores médios de viés de códon medido através da média do índice ENC (*avg_nc*) para cada par de sequência e valores médios da relação entre o maior valor e o menor valor de ENC para cada par de comparação de sequências (*avg_nc*) organizados de acordo com cada comparação entre genomas de *Drosophila* para sequências gênicas e de elementos de transposição

		Genoma		TE	
		<i>avg_nc</i>	<i>diff_nc</i>	<i>avg_nc</i>	<i>diff_nc</i>
		Média		Média	
Melanogaster_Group	dana_dere	48.80	1.10	44.46	1.07
	dana_dmel	49.28	1.10	49.96	1.11
	dana_dsec	48.94	1.10	50.62	1.10
	dana_dsim	48.81	1.10	50.82	1.10
	dana_dyak	48.85	1.10	48.72	1.10
Melanogaster_Subgroup	dmel_dere	48.76	1.04	50.46	1.15
	dmel_dsec	48.91	1.03	53.52	1.10
	dmel_dsim	48.77	1.03	53.36	1.11
	dmel_dyak	48.81	1.05	52.53	1.10
	dsec_dere	48.42	1.04	53.76	1.12
	dsec_dsim	48.43	1.02	54.04	1.10
	dsec_dyak	48.47	1.04	53.37	1.10
	dsim_dere	48.29	1.04	52.88	1.14
	dsim_dyak	48.33	1.04	53.65	1.10
	dyak_dere	48.33	1.03	52.11	1.12
	Melanogaster-Hawaiian	dana_dgri	49.57	1.10	60.82
Melanogaster-Obscura	dana_dper	48.21	1.11	46.50	1.17
	dana_dpse	48.15	1.11	46.38	1.13
Melanogaster-Repleta	dana_dmoj	48.97	1.11	44.02	1.13
Melanogaster-Virilis	dana_dvir	48.62	1.12	44.90	1.14
Melanogaster-Willistoni	dana_dwil	51.22	1.12	47.87	1.18
MelanogasterSubgroup-Hawaiian	dgri_dere	49.05	1.11		
	dgri_dmel	49.53	1.09	44.50	1.05
	dgri_dsec	49.19	1.10	47.26	1.07
	dgri_dsim	49.04	1.10	49.69	1.18
	dgri_dyak	49.09	1.11		
MelanogasterSubgroup-Obscura	dmel_dper	48.17	1.12	48.00	1.11
	dmel_dpse	48.12	1.12	48.29	1.14

	dper	dere	47.69	1.11	50.62	1.23
	dper	dsec	47.83	1.11	49.04	1.10
	dper	dsim	47.69	1.11	47.13	1.15
	dper	dyak	47.74	1.11	48.75	1.16
	dpse	dere	47.63	1.11	39.62	1.06
	dpse	dsec	47.77	1.12	48.37	1.13
	dpse	dsim	47.63	1.12	47.66	1.16
	dpse	dyak	47.68	1.11	49.35	1.12
MelanogasterSubgroup-Repleta	dmel	dmoj	48.95	1.12	44.52	1.13
	dmoj	dere	48.44	1.11		
	dmoj	dsec	48.59	1.12	45.21	1.15
	dmoj	dsim	48.45	1.12	44.73	1.12
	dmoj	dyak	48.50	1.12	46.27	1.14
MelanogasterSubgroup-Virilis	dmel	dvir	48.59	1.11	44.86	1.13
	dsec	dvir	48.24	1.11	44.76	1.13
	dsim	dvir	48.10	1.12	46.34	1.15
	dvir	dere	48.10	1.12	51.57	1.25
	dvir	dyak	48.15	1.11	45.44	1.05
MelanogasterSubgroup-Willistoni	dmel	dwil	51.19	1.11	47.56	1.14
	dsec	dwil	50.85	1.13	48.65	1.15
	dsim	dwil	50.69	1.13	47.51	1.14
	dwil	dere	50.71	1.13		
	dwil	dyak	50.75	1.13	47.73	1.19
Obscura Group	dper	dpse	47.05	1.02	53.10	1.11
Obscura-Hawaiian	dgri	dper	48.46	1.11	49.43	1.17
	dgri	dpse	48.40	1.11	42.03	1.19
Repleta-Hawaiian	dgri	dmoj	49.22	1.09		
Repleta-Obscura	dmoj	dper	47.86	1.10	46.33	1.17
	dmoj	dpse	47.80	1.10	46.30	1.18
Virilis-Hawaiian	dgri	dvir	48.88	1.08	50.72	1.12
Virilis-Obscura	dper	dvir	47.51	1.10	47.00	1.15
	dpse	dvir	47.45	1.10	49.85	1.14
Virilis-Repleta	dmoj	dvir	48.28	1.08	45.37	1.13
Willistoni-Hawaiian	dgri	dwil	51.47	1.09		
Willistoni-Obscura	dper	dwil	50.12	1.15	49.62	1.15
	dpse	dwil	50.05	1.16	50.16	1.15
Willistoni-Repleta	dmoj	dwil	50.87	1.12	46.24	1.18
Willistoni-Virilis	dvir	dwil	50.53	1.13	46.68	1.16

Tabela 12:- Valores médio, máximo (max), mínimo (min) de viés de códon medido através da média do índice ENC (*avg_nc*) para cada par de comparação de sequências organizados de acordo com cada comparação entre genomas de *Drosophila* para sequências gênicas e de elementos de transposição

		Genome			TE		
		<i>avg_nc</i>			<i>avg_nc</i>		
		Média	Max	Min	Média	Max	Min
Melanogaster_Group	dana__dere	48.80	60.27	29.18	44.46	59.96	38.39
	dana_dmel	49.28	59.87	29.19	49.96	61.00	37.81
	dana_dsec	48.94	60.53	28.97	50.62	61.00	37.81
	dana_dsim	48.81	60.24	29.04	50.82	61.00	40.04
	dana_dyak	48.85	59.79	28.28	48.72	59.18	38.80
Melanogaster_Subgroup	dmel_dere	48.76	60.66	29.06	50.46	61.00	34.27
	dmel_dsec	48.91	60.41	28.87	53.52	61.00	30.01
	dmel_dsim	48.77	60.26	28.94	53.36	61.00	28.44
	dmel_dyak	48.81	60.70	28.18	52.53	61.00	36.38
	dsec_dere	48.42	60.60	28.86	53.76	61.00	37.78
	dsec_dsim	48.43	60.63	28.72	54.04	61.00	28.60
	dsec_dyak	48.47	60.60	27.96	53.37	61.00	37.06
	dsim_dere	48.29	61.00	28.93	52.88	61.00	38.51
	dsim_dyak	48.33	60.66	28.03	53.65	61.00	38.19
	dyak_dere	48.33	61.00	28.17	52.11	61.00	36.22
Melanogaster-Hawaiian	dana_dgri	49.57	59.97	32.15	60.82	61.00	59.10
Melanogaster-Obscura	dana_dper	48.21	61.00	28.92	46.50	58.58	37.93
	dana_dpse	48.15	60.49	28.87	46.38	57.51	38.60
Melanogaster-Repleta	dana_dmoj	48.97	60.28	32.02	44.02	58.08	34.99
Melanogaster-Virilis	dana_dvir	48.62	59.35	32.36	44.90	57.57	36.12
Melanogaster-Willistoni	dana_dwil	51.22	60.39	32.71	47.87	61.00	35.18
MelanogasterSubgroup-Hawaiian	dgri_dere	49.05	58.62	32.04			
	dgri_dmel	49.53	59.81	32.05	44.50	44.50	44.50
	dgri_dsec	49.19	59.29	31.83	47.26	47.26	47.26
	dgri_dsim	49.04	59.29	31.90	49.69	49.69	49.69
	dgri_dyak	49.09	60.12	31.14			
MelanogasterSubgroup-Obscura	dmel_dper	48.17	60.46	29.63	48.00	59.52	39.16
	dmel_dpse	48.12	60.45	29.25	48.29	58.45	39.83
	dper_dere	47.69	60.61	29.62	50.62	54.21	47.02
	dper_dsec	47.83	60.35	29.41	49.04	56.77	39.16
	dper_dsim	47.69	60.10	29.48	47.13	53.54	43.10
	dper_dyak	47.74	60.77	28.72	48.75	60.62	41.98
	dpse_dere	47.63	60.61	29.24	39.62	39.62	39.62

	dpse	dsec	47.77	60.35	29.03	48.37	54.76	39.83
	dpse	dsim	47.63	60.10	29.10	47.66	52.47	43.78
	dpse	dyak	47.68	60.65	28.34	49.35	55.75	42.84
MelanogasterSubgroup-Repleta	dmel	dmoj	48.95	60.07	31.60	44.52	56.07	36.22
	dmoj	dere	48.44	58.82	31.19			
	dmoj	dsec	48.59	59.78	31.28	45.21	56.06	36.22
	dmoj	dsim	48.45	59.78	31.13	44.73	52.23	40.16
	dmoj	dyak	48.50	58.96	31.09	46.27	57.66	39.13
MelanogasterSubgroup-Virilis	dmel	dvir	48.59	60.76	32.26	44.86	58.80	37.35
	dsec	dvir	48.24	60.47	32.04	44.76	52.20	37.35
	dsim	dvir	48.10	60.47	32.11	46.34	56.84	41.30
	dvir	dere	48.10	59.49	32.25	51.57	60.40	39.59
	dvir	dyak	48.15	58.94	31.35	45.44	48.90	42.77
MelanogasterSubgroup-Willistoni	dmel	dwil	51.19	60.21	32.25	47.56	61.00	37.56
	dsec	dwil	50.85	60.89	32.32	48.65	61.00	37.56
	dsim	dwil	50.69	60.74	32.38	47.51	60.12	38.59
	dwil	dere	50.71	60.12	32.38			
	dwil	dyak	50.75	59.71	32.25	47.73	58.51	37.36
Obscura Group	dper	dpse	47.05	61.00	28.82	53.10	61.00	25.78
Obscura-Hawaiian	dgri	dper	48.46	59.88	32.59	49.43	49.43	49.43
	dgri	dpse	48.40	59.77	32.21	42.03	42.03	42.03
Repleta-Hawaiian	dgri	dmoj	49.22	60.23	34.74			
Repleta-Obscura	dmoj	dper	47.86	60.02	31.49	46.33	60.81	36.33
	dmoj	dpse	47.80	59.93	31.07	46.30	57.18	37.11
Virilis-Hawaiian	dgri	dvir	48.88	59.83	35.03	50.72	59.43	42.00
Virilis-Obscura	dper	dvir	47.51	60.32	32.79	47.00	60.72	37.47
	dpse	dvir	47.45	60.33	32.41	49.85	60.77	36.32
Virilis-Repleta	dmoj	dvir	48.28	60.40	34.33	45.37	58.01	34.53
Willistoni-Hawaiian	dgri	dwil	51.47	58.90	35.86			
Willistoni-Obscura	dper	dwil	50.12	60.51	33.48	49.62	60.88	37.68
	dpse	dwil	50.05	60.51	32.46	50.16	60.88	39.07
Willistoni-Repleta	dmoj	dwil	50.87	59.74	34.09	46.24	61.00	34.74
Willistoni-Virilis	dvir	dwil	50.53	60.64	35.88	46.68	57.48	37.57

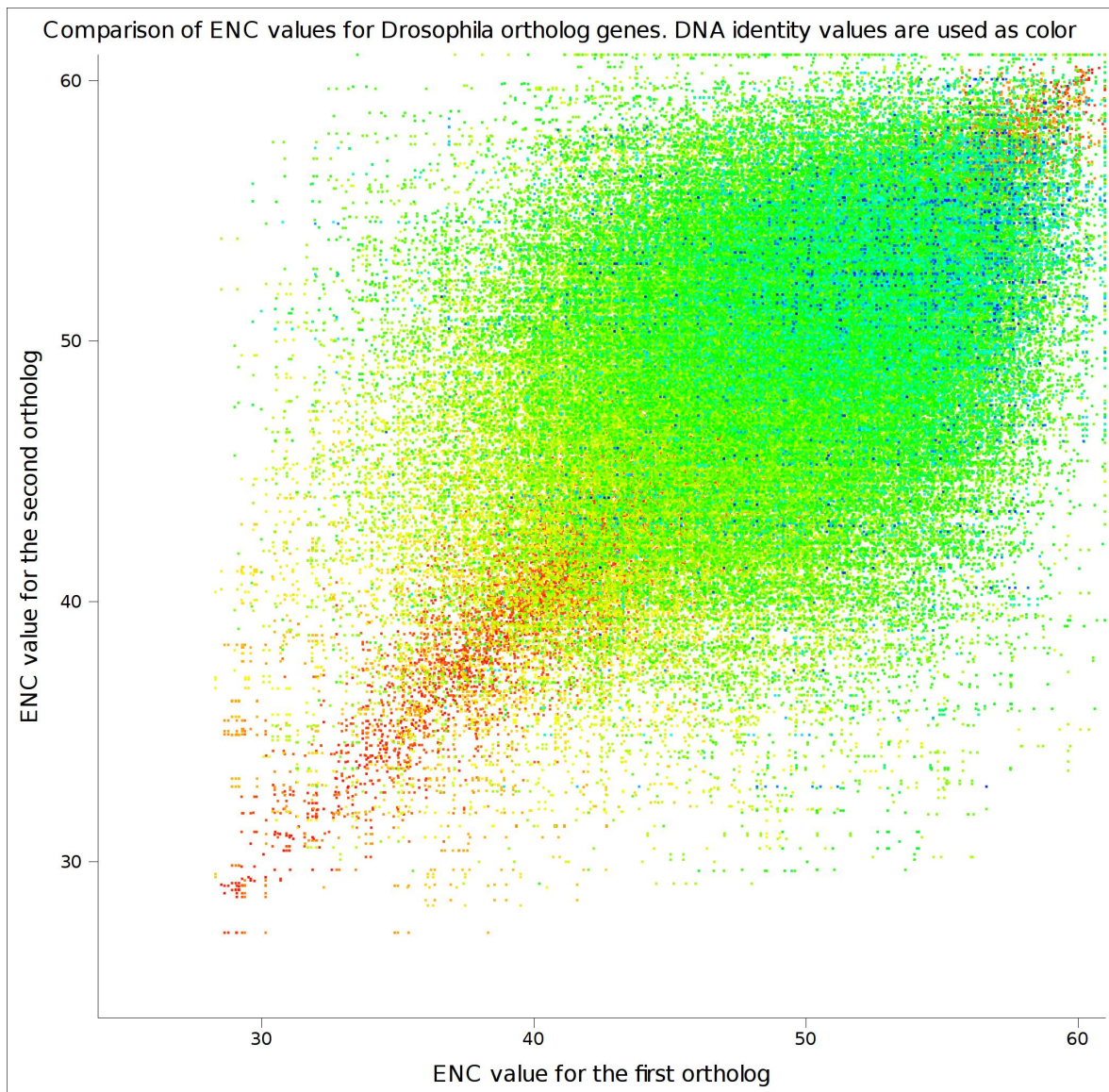


Figura 51 - Gráfico de dispersão entre as variáveis de viés de códon ENC para a primeira e segunda sequência da comparação entre sequências gênicas. A cor de cada ponto de dados representa a intensidade da variável DNAid, sendo vermelho equivalente a 100% e azul ao valor mínimo de 50%.

Comparison of ENC between pairs of TE sequences and DNA identity as the color intensity

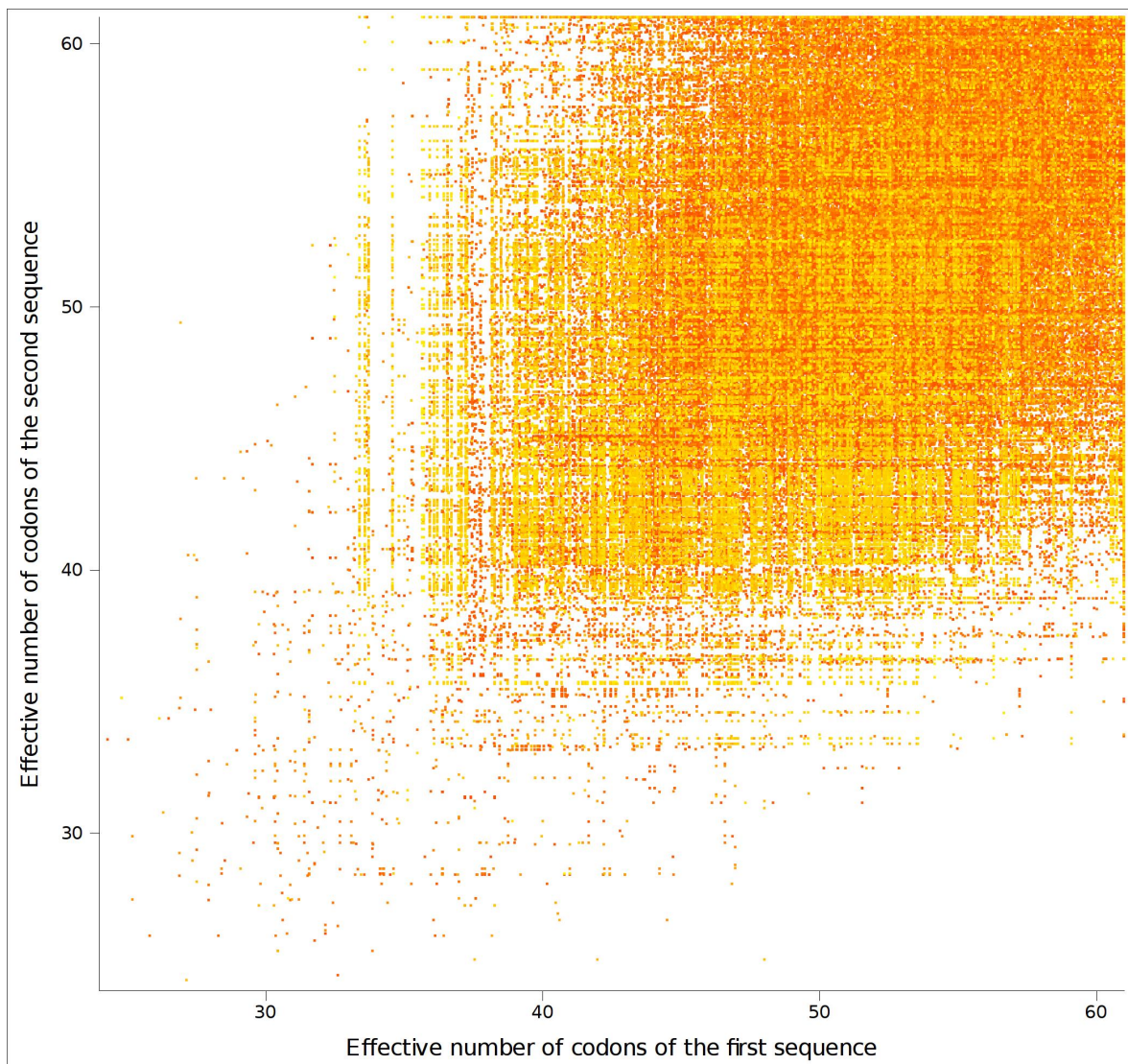


Figura 52 - Gráfico de dispersão entre as variáveis de viés de códon ENC para a primeira e segunda sequência da comparação entre sequências de elementos de transposição. A cor de cada ponto de dados representa a intensidade da variável DNAid, sendo vermelho equivalente a 100% e laranja ao valor mínimo de 70%.

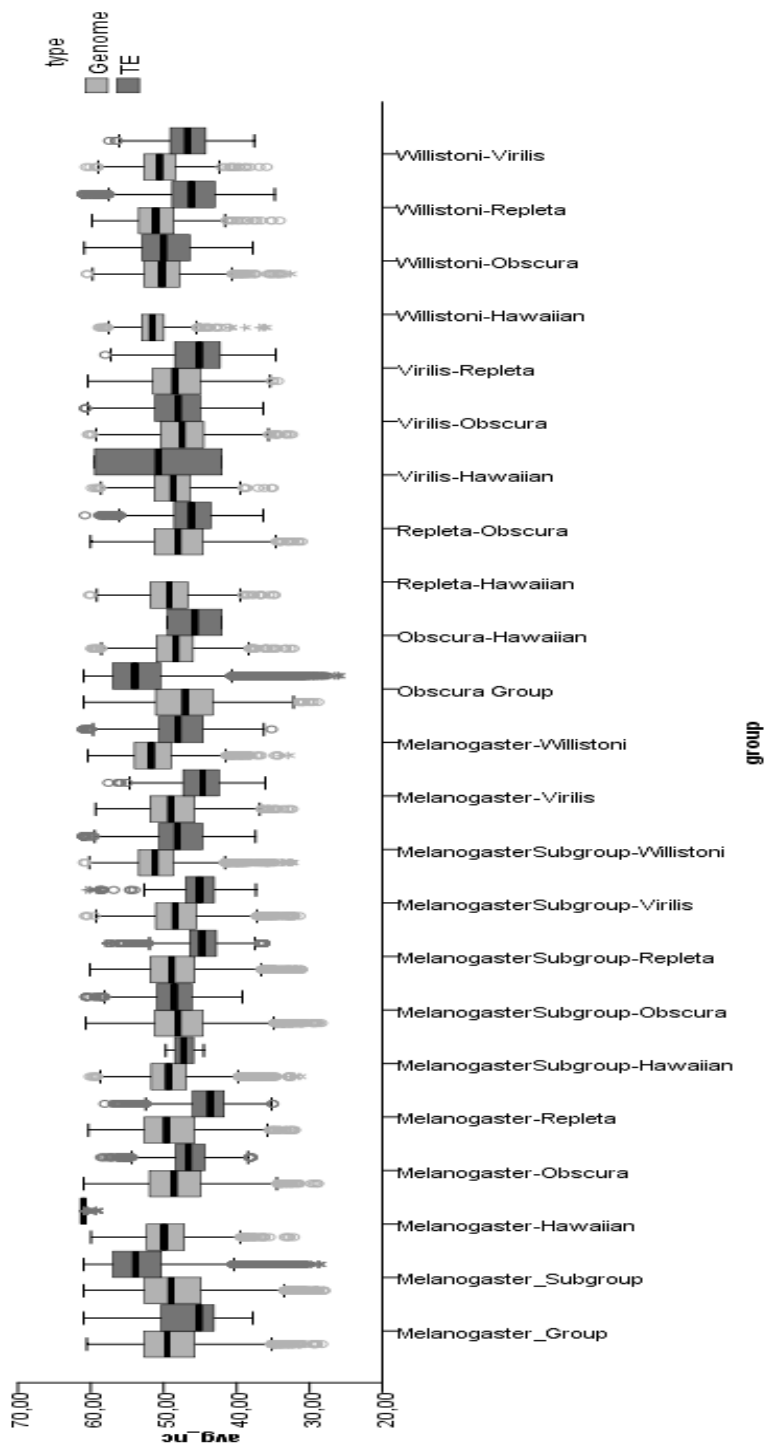


Figura 53 –Diagrama de caixa para dados da variável de viés de códon ENC de seqüências gênica e de elementos de transposição.

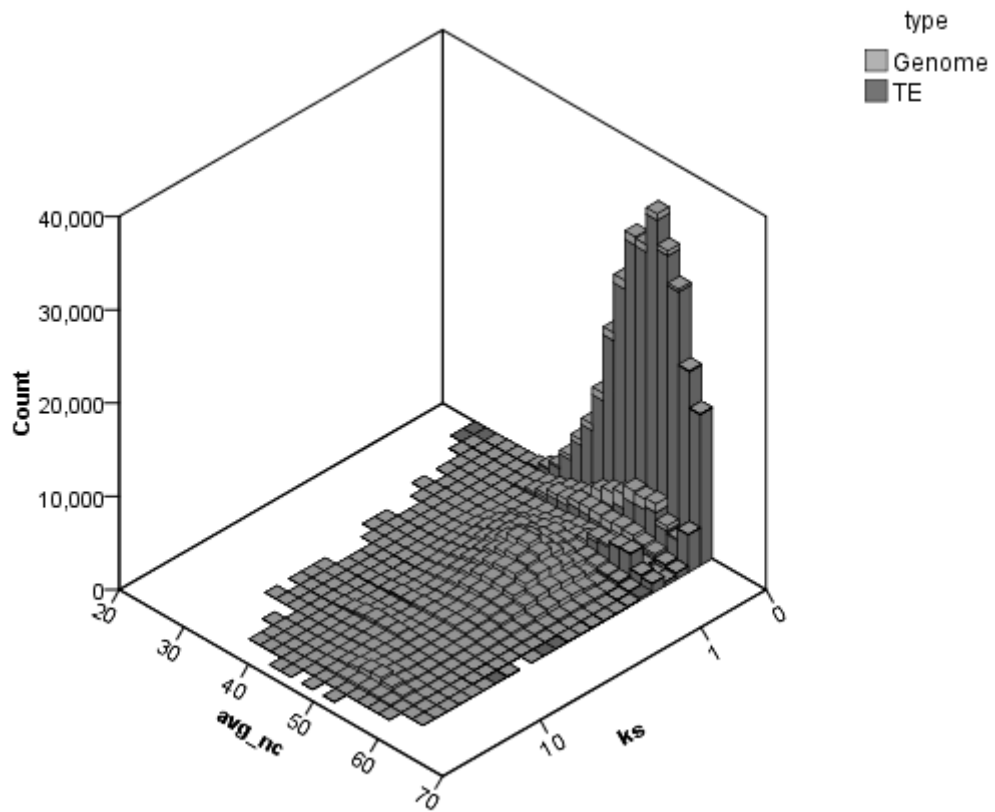


Figura 54 – Histograma 3D das variáveis de viés de códon médio (ENC) para cada par de comparação e valores de Ks (escala log) equivalentes, separados por cor de acordo com a comparação entre sequências gênicas ou de elementos de transposição.

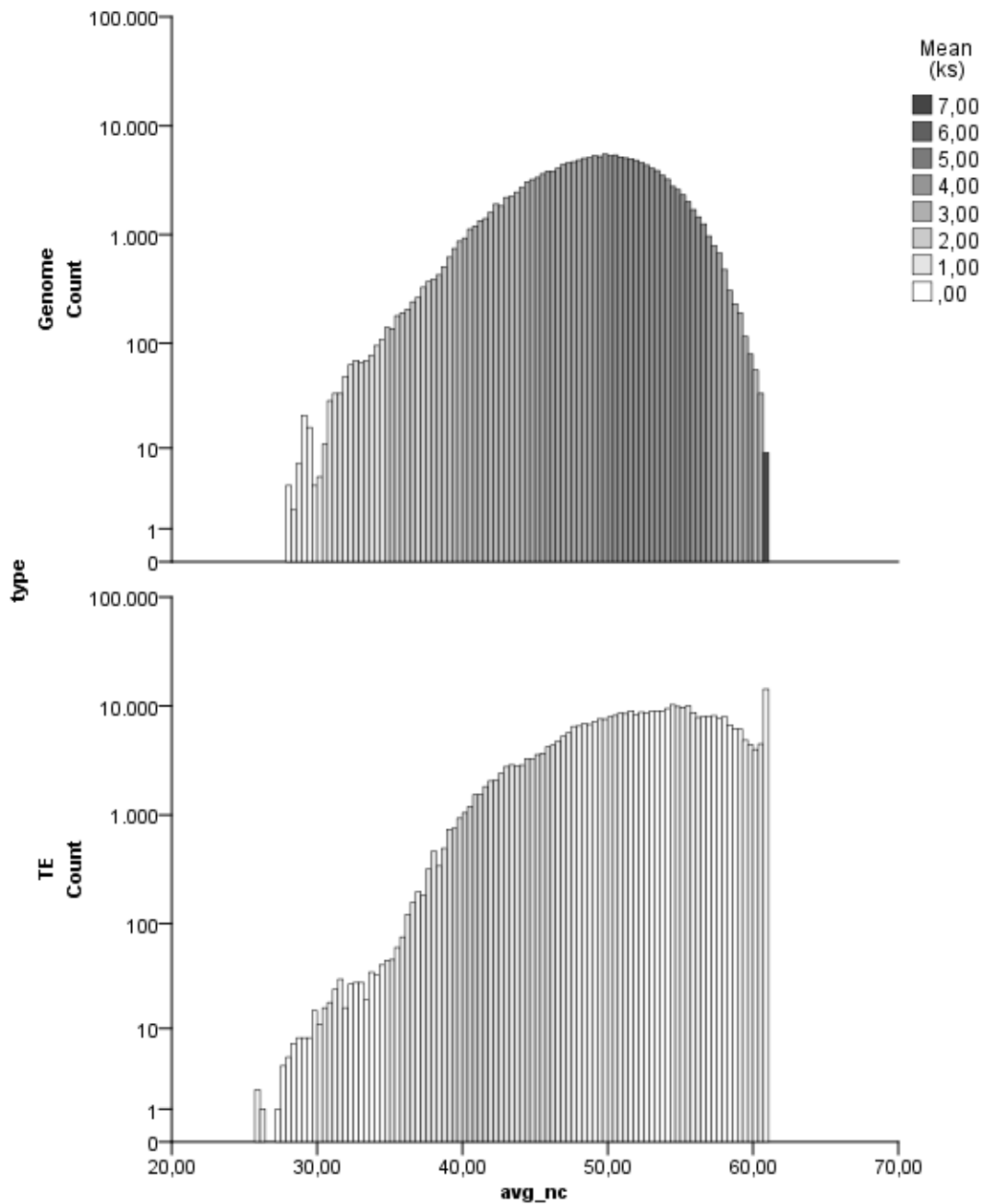


Figura 55 – Histograma dos valores de viés de códon médio (ENC) separados de acordo com dados derivados de comparações entre sequências gênicas (painel superior) ou dados derivados de comparações entre sequências de elementos de transposição (painel inferior). A cor representa o valor médio da variável Ks para cada intervalo. Eixo das ordenadas em escala log.

Considerando os dados obtidos é possível traçar dois cenários relativos a evolução dos elementos de transposição. Primeiramente é preciso considerar a manutenção dos elementos de transposição dentro dos genomas analisados uma vez que estes se caracterizam como sequencias ancestrais, mantendo-se no genoma ao longo de milhões de anos. Em segundo lugar é preciso considerar que o controle da expansão desses elementos é primariamente realizado através do controle da sua expressão, tanto por mecanismos envolvendo a rota de degradação via piRNA (Rozhkov, Hammell e Hannon, 2013) quanto de metilação (Hollister e Gaut, 2009). Esses mecanismos de controle, admitindo-se a sua eficiência, levariam à completa degradação das sequencias de TEs ao longo das linhagens analisadas dada a divergência observada nos genes nucleares e as taxas basais de substituições sinônimas e não-sinônimas. No entanto é possível observar que diversas subpopulações de elementos de transposição são mantidas dentro do genoma com taxas de substituição sinônima muito inferiores aos genes nucleares. Também observam-se sequencias com valores de Ka/Ks muito superiores ao genoma, mas em subpopulações de elementos onde temos baixas taxas de substituições sinônimas. Isso indica uma rápida evolução das sequências de uma subpopulação de elementos de transposição presentes nas espécies de *Drosophila* estudadas. Ao mesmo tempo, uma segunda subpopulação apresenta baixas taxas de substituição sinônima, baixas taxas de substituição não-sinônima, valores de Ka/Ks mais baixos que os observados no genoma bem como uma alta similaridade. Estas são sequências altamente conservadas e que possivelmente representam eventos recentes de transferência horizontal. Ao mesmo tempo observa-se uma terceira

subpopulação de sequências onde aparentemente temos uma total degradação das sequências observada pela menor similaridade das sequências de TE em relação ao genoma bem como maiores taxas de Ka e Ks.

O primeiro cenário evolutivo evidenciados por esses dados em relação aos elementos transponíveis é onde as sequências de elementos de transposição, ao ocorrer um processo inicial de invasão de um genoma, sofrem forte inativação devido aos mecanismos celulares específicos de controle de transposição e seguem uma rota de extinção ao acumular de forma indiscriminada mutações sinônimas e não sinônimas.

O segundo cenário acontece onde ao ocorrer o mecanismo de invasão genômica não temos uma completa inativação da sequência de TE, sendo que este mantém um nível basal de transcrição e tradução, permitindo que o elemento de transposição possa expandir-se no genoma até o ponto de equilíbrio com os efeitos deletérios derivados do processo de transposição. A partir deste equilíbrio é possível especular que mesmo que ainda persista um nível basal de expressão das sequências de elementos de transposição, tais sequências estarão sujeitas aos mesmos mecanismos de degradação existentes no primeiro cenário, uma vez que o processo de colonização do genoma não deverá gerar a pressão de seleção necessária para a manutenção das sequências de TEs.

Apesar dos inúmeros casos que descrevem a cooptação de sequências de elementos de transposição como elementos funcionais do genoma, sejam como sequências envolvidas em regulação gênica (Bourque, 2009) quanto em genes específicos que promovem vantagem adaptativa ao novo hospedeiro (Alzohairy *et al.*, 2013), a taxa com que este fenômeno acontece é muito inferior à taxa de transposição e colonização de genomas

observadas. Caso a taxa de cooptação de elementos de transposição fosse equivalente, observaríamos atualmente um número muito maior de casos de elementos de transposição envolvidos em processos celulares bem como um pequeno número de sequências de TEs sem função celular específica.

Como parasitas intracelulares, os vírus também estão sujeitos a pressões de seleção específicas, manifestadas através de controle pelo mesmo sistema de piRNAs (Wu *et al.*, 2010) também envolvido no controle de transposons e mecanismos de nível mais alto, ocorrendo a nível celular. Como ponto essencial do processo de sobrevivência de sequências virais encontramos a alta taxa de mutação que ocorre durante o processo de replicação (Manrubia *et al.*, 2005). Esta alta taxa de mutação permite que sejam formadas subpopulações virais com uma alta diversidade. Ao serem expostas a grande pressão de seleção exercida pelo hospedeiro, estas subpopulações são reduzidas apenas àquelas sequências que mantêm sua capacidade de completar o ciclo infeccioso até a transferência para um novo hospedeiro ou são extintas, caso a variabilidade gerada não contenha indivíduos com potencial de completar o ciclo após sofrer o processo de seleção (Manrubia *et al.*, 2005). Durante esse processo de geração de variabilidade e seleção é possível observar um grande aumento no número de mutações não-sinônimas sobre mutações sinônimas, evidenciando a extensa seleção positiva a qual os vírus são sujeitos durante o seu ciclo de vida (Duvvuri *et al.*, 2009).

Sugere-se que um processo semelhante ocorra aos elementos de transposição. Contudo, o processo de geração de variabilidade não acontece na mesma escala quantitativa, que no entanto, é compensada na escala temporal.

Elementos de Classe II são copiados pelo mecanismo de replicação celular, e seu incremento no número de cópias é realizado através do mecanismo de reparo de DNA (Liang e Sved, 2009) logo após o processo de transposição.. Logo, a geração de variabilidade de elementos de transposição de Classe II a cada processo de duplicação celular é derivada de mutações sofridas à mesma taxa do genoma. Esta taxa de mutação é notadamente inferior às taxas de mutação observadas em populações virais durante o processo infeccioso, contudo é possível observar uma grande variabilidade em elementos de transposição de Classe II ao longo da árvore evolutiva de *Drosophila*. Mutações estas que foram acumuladas durante o processo de silenciamento e manutenção basal dos elementos de transposição, conforme descrito no cenário 2 acima.

Nos elementos de Classe I, o processo de geração de mutações ocorre tanto durante o processo de transcrição reversa, sendo este um mecanismo que não possui a fidedignidade do processo de replicação celular, quanto pela sua manutenção no genoma hospedeiro, sujeito aos mecanismos celulares de silenciamento.

Conforme a comparação com populações virais, caso a subpopulação viral selecionada durante o processo infeccioso não realize sua transferência para outro hospedeiro, a linhagem selecionada está fadada a extinção, da mesma forma caso a linhagem infecciosa original falhar em dar origem a uma subpopulação resistente capaz de completar o ciclo de infecção.

Podemos seguir o paralelo com relação aos elementos de transposição, uma vez que, considerando o cenário dois, onde estes mantêm um nível basal de atividade, caso um novo genoma hospedeiro não seja colonizado, as sequências originais estão fadadas a

degradação devido aos mecanismos de controle de transposição, que eventualmente levarão à extinção completa das sequências de transposon que originalmente colonizaram o genoma. Assim, compara-se o processo de seleção do sistema imune sobre sequências virais ao processo de seleção dos mecanismos celulares de controle de transposição. Igualmente, compara-se o processo de geração de variabilidade na replicação viral ao processo de acúmulo de mutações ao longo do tempo em que as sequências de transposons permanecem silenciadas ou sob um nível basal de expressão.

Em suma, é imperioso que um novo genoma seja “infectado” para que o ciclo de evolução das sequências de elementos de transposição seja mantido, sendo essa nova infecção mediada por elementos que evoluíram durante o período de silenciamento no genoma. Nesse contexto, esta “infecção” genômica é representado pelo processo de transferência horizontal.

Este modelo é suportado pela evidência de sequências de elementos de transposição com similaridade muito inferior ao genoma (teste de amostras independentes Mann-Whitney U, $p < 0.05$), ao mesmo tempo que apresentam baixos valores de K_a , K_s , distâncias Logdet e Kimura. Considerando-se que o processo de cooptação já citado acontece de forma rara ao longo do processo evolutivo de sequências de elementos de transposição, tal similaridade entre espécies distantes como, por exemplo, *Drosophila willistoni* e *Drosophila mojavensis* é provavelmente explicado pelo processo de transferência horizontal.

Após o processo de transferência horizontal, novas sequências de elementos de transposição não são totalmente silenciadas pelo novo genoma hospedeiro, permitindo que

ocorra expansão do seu número de cópias. Durante o processo de expansão genômica e posterior silenciamento, estas sequências acumulam mutações, gerando uma nova população de elementos de transposição sujeita tanto ao processo de seleção por parte dos mecanismos celulares de contenção de transposição assim como pelo próprio processo de transferência horizontal, uma vez que estas devem se manter funcionais para que ocorra a colonização de um novo genoma.

Pode-se sugerir que a seleção exercida pelos mecanismos celulares de controle de transposição não são suficientemente efetivos a ponto de causar a inativação de elementos de transposição ao longo do processo evolutivo do genoma hospedeiro. Assim, considerando que exista uma taxa basal de transposição que mantenha sempre um número definido de sequências funcionais de transposons, tais sequências não seriam extintas, já que também são passadas verticalmente a novos indivíduos através do processo de reprodução do hospedeiro. Contudo é preciso levar em consideração o conceito definido como “catraca de Muller”(Engelstadter, 2008). Este conceito, desenvolvido inicialmente para explicar padrões evolutivos em espécies de reprodução assexuada, pressupõe que na ausência de recombinação existe o acúmulo direcional de mutações ligeiramente deletérias que por sua vez podem levar à extinção da espécie. Apesar de não ser possível aplicar diretamente esse conceito aos elementos de transposição, é plausível supor que durante o tempo de silenciamento a que os mesmos são sujeitos, o acúmulo de mutações ligeiramente deletérias pode iniciar um processo semelhante à catraca de Muller, onde eventualmente as últimas cópias funcionais de uma determinada família de elemento de transposição são inativadas sendo a transferência horizontal e a colonização de um novo genoma o único

mecanismo de escape possível.

De forma geral é factível propor que a evolução dos elementos de transposição é um caso especial do modelo de *quasispecies* evidenciado em sequencias virais, onde são geradas subpopulações de sequencias com alta variabilidade induzida por processos mutacionais neutros e submetidas às pressões de seleção do hospedeiro a cada processo de colonização de um novo genoma. Assim, da mesma forma que a infecção de novos indivíduos é crucial para a manutenção de espécies virais, o mecanismo de transferência horizontal não é apenas um fenômeno possível pelas características de mobilidade dos elementos de transposição, mas um processo integral necessário à sua manutenção em diferentes linhagens ao longo da evolução.

Da mesma forma, tomando-se o processo de evolução neutro dos elementos de transposição subsequente à invasão de um novo genoma é possível identificar casos de transferência horizontal pela comparação direta entre as taxas de substituição sinônima observadas entre genes do genoma hospedeiro e das sequências de transposons em diferentes espécies.

Perspectivas futuras

Como perspectivas para desenvolvimentos a partir dos resultados obtidos neste trabalho bem como das observações operacionais de trabalho com análise em larga escala de transferência horizontal, visa-se desenvolver:

- Um protótipo de ontologia voltada para a anotação de elementos de transposição e a sua aplicação para a anotação de transposons em genomas de insetos

- A aplicabilidade do uso de métodos de aprendizado de máquina em larga escala para a classificação de elementos de transposição que sofreram transferência horizontal e elementos que são majoritariamente herdados verticalmente. Considera-se que métodos de aprendizado de máquina possuem um potencial de sensibilidade maior que comparações baseadas em índices e também a possibilidade de identificação de transferência horizontal sem a necessidade da comparação direta com o genoma hospedeiro, uma vez que se espera identificar “assinaturas evolutivas de transferência horizontal” dentro das variáveis utilizadas na comparação entre espécies.

- A análise comparativa aprofundada entre os valores obtidos de K_s e K_a para os diferentes genomas e elementos de transposição e valores de viés de códon já obtidos. Estima-se que o viés de códon é fortemente influenciado pelas taxas de mutação sinônima a qual um determinado gene está sujeito. Contudo, as análises realizadas durante o curso deste trabalho evidenciam que esta influência pode ser bem menor do que o esperado.

– Apesar de uma análise global de todas as classes e grupos de elementos de transposição em conjunto fornecer uma figura compreensiva e global sobre os padrões evolutivos aos quais estes elementos estão sujeitos, sabe-se que cada família de elementos de transposição possui particularidades evolutivas que dificilmente serão identificadas em análises conjuntas. Assim, é imperativo que sejam realizados estudos no mesmo molde do que foi aqui apresentado, mas com foco em grupos específicos de elementos de transposição, como por exemplo os Helitrons.

- O projeto “*Drosophila* Population Genomics Project” está realizando o sequenciamento de mais de 350 genomas de populações de *Drosophila melanogaster* e *Drosophila simulans*. Esta grande quantidade de dados de sequencia fornece a oportunidade de análise diferenciada onde será possível testar não só as taxas de substituição sinônima e não-sinônima em uma escala de tempo menor, como também inferir as taxas de eliminação de elementos de transposição por eventos de recombinação.

- Apesar de mostrar-se uma métrica de fácil uso na análise comparativa e determinação de transferência horizontal, o index de similaridade utilizado neste estudo ainda é uma simplificação do processo evolutivo representado pelas variáveis que o compõe. Assim, é importante o seu desenvolvimento junto a uma modelagem matemática mais rigorosa que permita uma análise estatística formal das características de distribuição dos seus valores.

Referências

- ABAD, J. P. *et al.* TAHRE, a novel telomeric retrotransposon from *Drosophila melanogaster*, reveals the origin of *Drosophila* telomeres. **Molecular biology and evolution**, v. 21, n. 9, p. 1620-1624, 1 set. 2004.
- AI, J.; SMITH, B.; DAVID, W. T. Saliva Ontology: an ontology-based framework for a Salivaomics Knowledge Base. **BMC bioinformatics**, v. 11, n. 1, 2010.
- ALZOHAIRY, A. M. *et al.* Transposable elements domesticated and neofunctionalized by eukaryotic genomes. **Plasmid**, v. 69, n. 1, p. 1-15, jan. 2013.
- ANISIMOVA, M.; BIELAWSKI, J. P.; YANG, Z. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. **Molecular biology and evolution**, v. 18, n. 8, p. 1585-92, ago. 2001.
- ASHBURNER, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. **Nature genetics**, v. 25, n. 1, p. 25-29, 1 maio. 2000.
- ASKITIS, N.; SINHA, R. RepMaestro: scalable repeat detection on disk-based genome sequences. **Bioinformatics (Oxford, England)**, v. 26, n. 19, p. 2368-74, 1 out. 2010.
- AZAD, R. K.; LAWRENCE, J. G. Towards more robust methods of alien gene detection. **Nucleic acids research**, v. 39, n. 9, p. e56-e56, 1 maio. 2011.
- BAI, L.; BRUTNELL, T. P. The activator/dissociation transposable elements comprise a two-component gene regulatory switch that controls endogenous gene expression in maize. **Genetics**, v. 187, n. 3, p. 749-59, mar. 2011.
- BAO, Z.; EDDY, S. R. Automated de novo identification of repeat sequence families in sequenced genomes. **Genome research**, v. 12, n. 8, p. 1269-76, ago. 2002.
- BAPTESTE, E. *et al.* The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba. **Proceedings of the National Academy of Sciences of the United States of America**, v. 99, n. 3, p. 1414-9, 5 fev. 2002.
- BAPTESTE, E. *et al.* Alternative Methods for Concatenation of Core Genes Indicate a Lack of Resolution in Deep Nodes of the Prokaryotic Phylogeny. **Molecular Biology and Evolution**, v. 25, n. 1, p. 83-91, 1 jan. 2008.
- BAUER DUMONT, V. L. *et al.* Locus-Specific Decoupling of Base Composition Evolution at Synonymous Sites and Introns along the *Drosophila melanogaster* and *Drosophila sechellia* Lineages. **Genome Biol Evol**, v. 2009, n. 0, p. 67-74, 22 jun. 2009.
- BECHER, V.; DEYMONNAZ, A.; HEIBER, P. Efficient computation of all perfect repeats in genomic sequences of up to half a gigabyte, with a case study on the human genome. **Bioinformatics**, v. 25, n. 14, p. 1746-1753, jul. 2009.
- BIERNE, N.; EYRE-WALKER, A. The problem of counting sites in the estimation of the synonymous and nonsynonymous substitution rates: implications for the correlation

- between the synonymous substitution rate and codon usage bias. **Genetics**, v. 165, n. 3, p. 1587-97, nov. 2003.
- BIRNEY, E.; CLAMP, M.; HUBBARD, T. Databases and tools for browsing genomes. **Annual review of genomics and human genetics**, v. 3, p. 293-310, jan. 2002.
- BLOMBERG, J. *et al.* Classification and nomenclature of endogenous retroviral sequences (ERVs): problems and recommendations. **Gene**, v. 448, n. 2, p. 115-23, dez. 2009.
- BLOT, M. Transposable elements and adaptation of host bacteria. **Genetica**, v. 93, n. 1-3, p. 5-12, jan. 1994.
- BOITARD, S.; SCHLOTTERER, C.; FUTSCHIK, A. Detecting Selective Sweeps: A New Approach Based on Hidden Markov Models. **Genetics**, v. 181, n. 4, p. 1567-1578, 1 abr. 2009.
- BOURQUE, G. Transposable elements in gene regulation and in the evolution of vertebrate genomes. **Current opinion in genetics & development**, v. 19, n. 6, p. 607-612, 13 dez. 2009.
- BULDYREV, S. V. *et al.* Analysis of DNA sequences using methods of statistical physics. **Physica A: Statistical Mechanics and its Applications**, v. 249, n. 1-4, p. 430-438, jan. 1998.
- BURNS, K. H.; BOEKE, J. D. Human transposon tectonics. **Cell**, v. 149, n. 4, p. 740-52, 11 maio. 2012.
- BURTCH, K. O. **Linux shell scripting with Bash**. [s.l.] Sams, 2004. v. 1p. 412
- CAPY, P. *et al.* Relationships between transposable elements based upon the integrase-transposase domains: is there a common ancestor? **Journal of molecular evolution**, v. 42, n. 3, p. 359-68, mar. 1996.
- CAPY, P. *et al.* Stress and transposable elements: co-evolution or useful parasites? **Heredity**, v. 85 (Pt 2), p. 101-106, ago. 2000.
- CARVALHO, M. O. DE; LORETO, E. L. S. Methods for detection of horizontal transfer of transposable elements in complete genomes. **Genetics and molecular biology**, v. 35, n. 4 (suppl), p. 1078-84, dez. 2012.
- CHAISSON, M. J.; BRINZA, D.; PEVZNER, P. A. De novo fragment assembly with short mate-paired reads: Does the read length matter? **Genome research**, v. 19, n. 2, p. 336-46, fev. 2009.
- CHEN, J. *et al.* ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. **Nucleic acids research**, v. 37, n. Web Server issue, 1 jul. 2009.
- CHOTHIA, C. *et al.* Evolution of the protein repertoire. **Science (New York, N.Y.)**, v. 300, n. 5626, p. 1701-3, 13 jun. 2003.
- CLAVERIE, J. M. From bioinformatics to computational biology. **Genome research**, v. 10, n. 9, p. 1277-9, set. 2000.

- COENYE, T. *et al.* Towards a prokaryotic genomic taxonomy. **FEMS microbiology reviews**, v. 29, n. 2, p. 147-67, abr. 2005.
- COHEN, J. E. Mathematics is biology's next microscope, only better; biology is mathematics' next physics, only better. **PLoS biology**, v. 2, n. 12, p. e439, dez. 2004.
- COMFORT, N. Two genes, no enzyme: a second look at Barbara McClintock and the 1951 Cold Spring Harbor Symposium. **Genetics**, v. 1166, n. 140, p. 1161-1155, 1995.
- COMFORT, N. C. From controlling elements to transposons: Barbara McClintock and the Nobel Prize. **Endeavour**, v. 25, n. 3, p. 127-30, set. 2001.
- CROSBY, M. A. *et al.* FlyBase: genomes by the dozen. **Nucleic acids research**, v. 35, n. Database issue, p. D486-91, jan. 2007.
- DANIELS, S. *et al.* Evidence for horizontal transmission of the P transposable element between Drosophila species. **Genetics**, v. 124, n. 2, p. 339-355, fev. 1990.
- DELCHER, A. L. *et al.* Fast algorithms for large-scale genome alignment and comparison. **Nucleic acids research**, v. 30, n. 11, p. 2478-83, 1 jun. 2002.
- DUDLEY, J. T.; BUTTE, A. J. In silico research in the era of cloud computing. **Nature biotechnology**, v. 28, n. 11, p. 1181-1185, nov. 2010.
- DUVVURI, V. R. *et al.* Role of Positive Selection Pressure on the Evolution of H5N1 Hemagglutinin. **Genomics, proteomics & bioinformatics / Beijing Genomics Institute**, v. 7, n. 1-2, p. 47-56, jun. 2009.
- EDGAR, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic acids research**, v. 32, n. 5, p. 1792-7, jan. 2004.
- EISEN, J. A. A phylogenomic study of the MutS family of proteins. **Nucleic acids research**, v. 26, n. 18, p. 4291-300, 15 set. 1998.
- EISEN, J. A.; FRASER, C. M. Phylogenomics: Intersection of Evolution and Genomics. **Science**, v. 300, n. 5626, p. 1706-1707, jun. 2003.
- EMERSON, R. A. Genetical Studies of Variegated Pericarp in Maize. **Genetics**, v. 2, n. 1, p. 1-35, jan. 1917.
- ENGELSTADTER, J. Muller's Ratchet and the Degeneration of Y Chromosomes: A Simulation Study. **Genetics**, v. 180, n. 2, p. 957-967, 1 out. 2008.
- ENRIGHT, A.; DONGEN, S. VAN; OUZOUNIS, C. An efficient algorithm for large-scale detection of protein families. **Nucleic acids research**, v. 30, n. 7, p. 1575-1584, 1 abr. 2002.
- FATAKIA, S. N.; COSTANZI, S.; CHOW, C. C. Computing Highly Correlated Positions Using Mutual Information and Graph Theory for G Protein-Coupled Receptors. **PLoS ONE**, v. 4, n. 3, 5 mar. 2009.
- FEDOROFF, N. V. McClintock's challenge in the 21st century. **Proceedings of the National Academy of Sciences of the United States of America**, v. 109, n. 50, p.

20200-3, 11 dez. 2012.

FESCHOTTE, C. The piggyBac transposon holds promise for human gene therapy. **Proceedings of the National Academy of Sciences of the United States of America**, v. 103, n. 41, p. 14981-14982, 10 out. 2006.

FESCHOTTE, C. *et al.* Exploring repetitive DNA landscapes using REPCLASS, a tool that automates the classification of transposable elements in eukaryotic genomes. **Genome biology and evolution**, v. 1, n. 0, p. 205-220, 1 jan. 2009.

FESCHOTTE, C.; PRITHAM, E. J. DNA transposons and the evolution of eukaryotic genomes. **Annual review of genetics**, v. 41, p. 331-68, jan. 2007.

FINNEGAN, D. Transposable elements in eukaryotes. **International review of cytology**, v. 93, p. 281-326, 1985.

FINNEGAN, D.; FAWCETT, D. Transposable elements in *Drosophila melanogaster*. **Oxford surveys on eukaryotic genes**, v. 3, p. 1-62, 1986.

FLAJNIK, M. F.; KASAHARA, M. Origin and evolution of the adaptive immune system: genetic events and selective pressures. **Nature Reviews Genetics**, v. 11, n. 1, p. 47-59, 8 dez. 2009.

FLEISCHMANN, R. D. *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. **Science (New York, N.Y.)**, v. 269, n. 5223, p. 496-512, jul. 1995.

FROST, L. S. *et al.* Mobile genetic elements: the agents of open source evolution. **Nature Reviews Microbiology**, v. 3, n. 9, p. 722-732, 1 set. 2005.

FU, L. *et al.* CD-HIT: accelerated for clustering the next-generation sequencing data. **Bioinformatics (Oxford, England)**, v. 28, n. 23, p. 3150-2, 1 dez. 2012.

FUGLSANG, A. The effective number of codons for individual amino acids : some codons are more optimal than others. v. 320, p. 185-190, 2003.

FULTON, D. *et al.* Improving the specificity of high-throughput ortholog prediction. **BMC Bioinformatics**, v. 7, n. 1, 28 maio. 2006.

GALAS, D. J.; CHANDLER, M. On the molecular mechanisms of transposition. **Proceedings of the National Academy of Sciences of the United States of America**, v. 78, n. 8, p. 4858-62, ago. 1981.

GEE, H. Evolution: ending incongruence. **Nature**, v. 425, n. 6960, p. 782, 23 out. 2003.

GERSTEIN, M. B. *et al.* What is a gene, post-ENCODE? History and updated definition. **Genome research**, v. 17, n. 6, p. 669-81, jun. 2007.

GOLDSTEIN, R. A.; POLLOCK, D. D. Observations of amino acid gain and loss during protein evolution are explained by statistical bias. **Molecular biology and evolution**, v. 23, n. 7, p. 1444-9, jul. 2006.

GOLLERY, M. *et al.* What makes species unique? The contribution of proteins with

- obscure features. **Genome biology**, v. 7, n. 7, p. R57, jan. 2006.
- GONZÁLEZ, J.; PETROV, D. MITES—The ultimate parasites. **Science**, v. 325, n. September, p. 1352-1353, 2009.
- GREEN, M. M. Transposable Elements in Drosophila and Other Diptera. **Annual Review of Genetics**, v. 14, n. 1, p. 109-120, 1980.
- HAIMOVICH, A. D. *et al.* Wavelet analysis of DNA walks. **Journal of computational biology : a journal of computational molecular cell biology**, v. 13, n. 7, p. 1289-98, set. 2006.
- HERSHBERG, R.; PETROV, D. A. Selection on codon bias. **Annual review of genetics**, v. 42, n. iv, p. 287-99, jan. 2008.
- HOEHNDORF, R. *et al.* BOWiki: an ontology-based wiki for annotation of data and integration of knowledge in biology. **BMC Bioinformatics**, v. 10, n. Suppl 5, 2009.
- HOEHNDORF, R.; KELSO, J.; HERRE, H. The ontology of biological sequences. **BMC bioinformatics**, v. 10, p. 377, jan. 2009.
- HOLLISTER, J. D.; GAUT, B. S. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. **Genome research**, v. 19, n. 8, p. 1419-28, ago. 2009.
- HOSKINS, R. A. *et al.* Sequence Finishing and Mapping of Drosophila melanogaster Heterochromatin. **Science**, v. 316, n. 5831, p. 1625-1628, 15 jun. 2007.
- HUA-VAN, A. *et al.* Abundance, distribution and dynamics of retrotransposable elements and transposons: similarities and differences. **Cytogenetic and genome research**, v. 110, n. 1-4, p. 426-40, jan. 2005.
- HUA-VAN, A. *et al.* The struggle for life of the genome's selfish architects. **Biology Direct**, v. 6, n. 1, p. 1-29, 1 dez. 2011.
- HURST, L. D.; FEIL, E. J.; ROCHA, E. P. C. Protein evolution: causes of trends in amino-acid gain and loss. **Nature**, v. 442, n. 7105, p. E11-2; discussion E12, 24 ago. 2006.
- HUYNH, D. *et al.* Google Refine. **Text**, p. 1-17, 2011.
- IWASHITA, S. *et al.* A Tandem Gene Duplication Followed by Recruitment of a Retrotransposon Created the Paralogous Bucentaur Gene (bcntp97) in the Ancestral Ruminant. **Molecular Biology and Evolution**, v. 23, n. 4, p. 798-806, 1 abr. 2006.
- IZSVÁK, Z. *et al.* Translating Sleeping Beauty transposition into cellular therapies: victories and challenges. **BioEssays : news and reviews in molecular, cellular and developmental biology**, v. 32, n. 9, p. 756-767, 22 set. 2010.
- JENSEN, L. J.; BORK, P. Ontologies in Quantitative Biology: A Basis for Comparison, Integration, and Discovery. **PLoS Biol**, v. 8, n. 5, 25 maio. 2010.
- JOHNSON, L. Transposon silencing: the extraordinary epigenetics of a transposon trap. **Heredity**, v. 100, n. 1, p. 5, jan. 2008.

- JOHNSON, L. J. The Genome Strikes Back: The Evolutionary Importance of Defence Against Mobile Elements. **Evolutionary Biology**, v. 34, n. 3, p. 121-129, 1 dez. 2007.
- JONES, N. McClintock's controlling elements: the full story. **Cytogenet Genome Research**, v. 109, p. 90-103, 2005.
- JORDAN, I. K. *et al.* A universal trend of amino acid gain and loss in protein evolution. 2005.
- JUHAS, M. *et al.* Genomic islands: tools of bacterial horizontal gene transfer and evolution. **FEMS Microbiology Reviews**, v. 33, n. 2, p. 376-393, mar. 2009.
- JURKA, J. *et al.* Repbase Update, a database of eukaryotic repetitive elements. **Cytogenetic and genome research**, v. 110, n. 1-4, p. 462-467, 2005.
- KAER, K.; SPEEK, M. Retroelements in human disease. **Gene**, v. 518, n. 2, p. 231-41, 15 abr. 2013.
- KAPITONOV, V. V.; JURKA, J. A universal classification of eukaryotic transposable elements implemented in Repbase. **Nature reviews. Genetics**, v. 9, n. 5, p. 411-2; author reply 414, maio. 2008.
- KAPITONOV, V. V.; JURKA, J. A universal classification of eukaryotic transposable elements implemented in Repbase. **Nat Rev Genet**, v. 9, n. 5, p. 411-412, 1 maio. 2008.
- KAZAZIAN, H. H. Mobile elements: drivers of genome evolution. **Science**, v. 303, n. 5664, p. 1626-1632, 2004.
- KIDWELL, M. Horizontal transfer of P elements and other short inverted repeat transposons. **Genetica**, v. 86, n. 1-3, p. 275-286, 1992.
- KIDWELL, M. G.; LISCH, D. R. Transposable elements and host genome evolution. v. 15, n. 3, p. 95-99, 2000.
- KIDWELL, M.; LISCH, D. Perspective: transposable elements, parasitic DNA, and genome evolution. **Evolution; international journal of organic evolution**, v. 55, n. 1, p. 1-24, jan. 2001.
- KIMURA, M. The neutral theory of molecular evolution. **Scientific American**, v. 241, n. 5, p. 98-100, 102, 108 passim, nov. 1979.
- KIMURA, M. Rare variant alleles in the light of the neutral theory. **Molecular biology and evolution**, v. 1, n. 1, p. 84-93, dez. 1983.
- KOFLER, R.; BETANCOURT, A. J.; SCHLÖTTERER, C. Sequencing of Pooled DNA Samples (Pool-Seq) Uncovers Complex Dynamics of Transposable Element Insertions in *Drosophila melanogaster*. **PLoS Genet**, v. 8, n. 1, 26 jan. 2012.
- KOFLER, R.; SCHLÖTTERER, C. Gowinda: unbiased analysis of gene set enrichment for genome-wide association studies. **Bioinformatics (Oxford, England)**, v. 28, n. 15, p. 2084-5, 1 ago. 2012.
- KONSTANTINIDIS, K. T.; RAMETTE, A.; TIEDJE, J. M. The bacterial species definition

- in the genomic era. **Philosophical transactions of the Royal Society of London. Series B, Biological sciences**, v. 361, n. 1475, p. 1929-40, 29 nov. 2006.
- KONSTANTINIDIS, K. T.; TIEDJE, J. M. Genomic insights that advance the species definition for prokaryotes. **Proceedings of the National Academy of Sciences of the United States of America**, v. 102, n. 7, p. 2567-72, 15 fev. 2005.
- KOONIN, E. V.; WOLF, Y. I.; KAREV, G. P. The structure of the protein universe and genome evolution. **Nature**, v. 420, n. 6912, p. 218-23, 14 nov. 2002.
- KORBEL, J. O. *et al.* SHOT: a web server for the construction of genome phylogenies. **Trends in genetics : TIG**, v. 18, n. 3, p. 158-162, mar. 2002.
- KOVÁČ, L. Information and knowledge in biology: time for reappraisal. **Plant signaling & behavior**, v. 2, n. 2, p. 65-73, mar. 2007.
- KUMAR, A.; COWEN, L. Augmented training of hidden Markov models to recognize remote homologs via simulated evolution. **Bioinformatics**, v. 25, n. 13, p. 1602-1608, 1 jul. 2009.
- KURTZ, S.; SCHLEIERMACHER, C. REPuter: fast computation of maximal repeats in complete genomes. **Bioinformatics**, v. 15, n. 5, p. 426-427, maio. 1999.
- LABRADOR, M. *et al.* Insulator and Ovo Proteins Determine the Frequency and Specificity of Insertion of the gypsy Retrotransposon in *Drosophila melanogaster*. **Genetics**, v. 180, n. 3, p. 1367-1378, 1 nov. 2008.
- LAMBRIX, P. *et al.* Biological Ontologies. *In: Semantic Web*. [s.l.: s.n.]. p. 85-99.
- LANFEAR, R.; WELCH, J. J.; BROMHAM, L. Watching the clock: studying variation in rates of molecular evolution between species. **Trends in ecology & evolution**, v. 25, n. 9, p. 495-503, set. 2010.
- LIANG, X.; SVED, J. A. Repair of P element ends following hybrid element excision leads to recombination in *Drosophila melanogaster*. **Heredity**, v. 102, n. 2, p. 127-32, fev. 2009.
- LIM, J. K.; SIMMONS, M. J. Gross chromosome rearrangements mediated by transposable elements in *Drosophila melanogaster*. **BioEssays : news and reviews in molecular, cellular and developmental biology**, v. 16, n. 4, p. 269-75, abr. 1994.
- LIN, J.; GERSTEIN, M. Whole-genome trees based on the occurrence of folds and orthologs: implications for comparing genomes on different levels. **Genome research**, v. 10, n. 6, p. 808-18, jun. 2000.
- LIU, H.; HU, Z.; WU, C. DynGO: a tool for visualizing and mining of Gene Ontology and its associations. **BMC bioinformatics**, v. 6, 2005.
- LOCKTON, S.; GAUT, B. S. The evolution of transposable elements in natural populations of self-fertilizing *Arabidopsis thaliana* and its outcrossing relative *Arabidopsis lyrata*. **BMC evolutionary biology**, v. 10, n. 1, 12 jan. 2010.
- LORETO, E.; CARARETO, C.; CAPY, P. Revisiting horizontal transfer of transposable

- elements in *Drosophila*. **Heredity**, v. 100, n. 6, p. 545-554, 23 jun. 2008.
- LORETO, E. L.; CARARETO, C. M.; CAPY, P. Revisiting horizontal transfer of transposable elements in *Drosophila*. **Heredity**, v. 100, n. 6, p. 545-554, jun. 2008.
- LOZOVSKAYA, E. Genomic regulation of transposable elements in *Drosophila*. **Current Opinion in Genetics & Development**, v. 5, n. 6, p. 768-773, dez. 1995.
- LYNCH, V. J. *et al.* Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. **Nature Genetics**, v. 43, n. 11, p. 1154-1159, 25 set. 2011.
- MANRUBIA, S. C. *et al.* High mutation rates, bottlenecks, and robustness of RNA viral quasispecies. **Gene**, v. 347, n. 2, p. 273-82, 14 mar. 2005.
- MARDIS, E. R. Next-Generation DNA Sequencing Methods. **Annual Review of Genomics and Human Genetics**, v. 9, n. 1, p. 387-402, 19 jun. 2008.
- MASZCZYK, T.; DUCH, W. Support Feature Machine for DNA Microarray Data. **Neural Networks**, p. 178-186, 2010.
- MCCLINTOCK, B. A cytological and genetical study of triploid maize. 1929.
- MCCLINTOCK, B. Chromosome organization and genic expression. **Cold Spring Harbor symposia on quantitative biology**, v. 16, p. 13-47, jan. 1951.
- METZKER, M. L. Sequencing technologies — the next generation. **Nature Reviews Genetics**, v. 11, n. 1, p. 31-46, dez. 2009.
- MILBURN, J. Beyond the genome : turning data into knowledge. v. 6, n. 17, p. 881-883, 2001.
- MURPHY, W. J. *et al.* Molecular phylogenetics and the origins of placental mammals. **Nature**, v. 409, n. 6820, p. 614-8, 1 fev. 2001.
- NATH, N.; MITCHELL, J. Is EC class predictable from reaction mechanism? **BMC Bioinformatics**, v. 13, n. 1, 2012.
- NEEDLEMAN, S. B.; WUNSCH, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. **Journal of molecular biology**, v. 48, n. 3, p. 443-53, mar. 1970.
- NEWBERG, L. A. *et al.* A phylogenetic Gibbs sampler that yields centroid solutions for cis-regulatory site prediction. **Bioinformatics**, v. 23, n. 14, p. 1718-1727, jul. 2007.
- OHLFEST, J. R.; IVICS, Z.; IZSVÁK, Z. Transposable elements as plasmid-based vectors for long-term gene transfer into tumors. **Methods in molecular biology (Clifton, N.J.)**, v. 542, p. 105-116, 2009.
- OLIVER, K. R.; GREENE, W. K. Transposable elements and viruses as factors in adaptation and evolution: an expansion and strengthening of the TE-Thrust hypothesis. **Ecology and evolution**, v. 2, n. 11, p. 2912-33, nov. 2012.
- PEDEN, J. F. Analysis of Codon Usage. **Phd thesis**, 1999.

- PELL, J. *et al.* **Scaling metagenome sequence assembly with probabilistic de Bruijn graphs**, 2012. Disponível em: <<http://arxiv.org/abs/1112.4193>>
- PETERLONGO, P.; CHIKHI, R. Mapsembler, targeted and micro assembly of large NGS datasets on a desktop computer. **BMC Bioinformatics**, v. 13, n. 1, 23 mar. 2012.
- PISKUREK, O.; OKADA, N. Poxviruses as possible vectors for horizontal transfer of retroposons from reptiles to mammals. **Proceedings of the National Academy of Sciences**, v. 104, n. 29, p. 12046-12051, 17 jul. 2007.
- PLESSIS, L. DU; SKUNCA, N.; DESSIMOZ, C. The what, where, how and why of gene ontology--a primer for bioinformaticians. **Briefings in bioinformatics**, v. 12, n. 6, p. 723-35, nov. 2011.
- POIREL, L. *et al.* Integron Mobilization Unit as a Source of Mobility of Antibiotic Resistance Genes. **Antimicrob. Agents Chemother.**, v. 53, n. 6, p. 2492-2498, 1 jun. 2009.
- POPTSOVA, M. S.; GOGARTEN, P. P. The power of phylogenetic approaches to detect horizontally transferred genes. **BMC evolutionary biology**, v. 7, n. 1, 2007.
- POWELL, J. R.; MORIYAMA, E. N. Evolution of codon usage bias in *Drosophila*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 94, n. 15, p. 7784-90, 22 jul. 1997.
- QIU, Y.; PALMER, J. Phylogeny of early land plants: insights from genes and genomes. **Trends in plant science**, v. 4, n. 1, p. 26-30, jan. 1999.
- R DEVELOPMENT CORE TEAM, R. **R: A Language and Environment for Statistical Computing** (R. D. C. Team, Ed.) **R Foundation for Statistical Computing** R Foundation for Statistical Computing, 2011. Disponível em: <<http://www.r-project.org>>
- RANEA, J. A. G. *et al.* Evolution of protein superfamilies and bacterial genome size. **Journal of molecular biology**, v. 336, n. 4, p. 871-87, 27 fev. 2004.
- RECHENMANN, F. From data to knowledge. **Bioinformatics**, v. 16, n. 5, p. 411-411, 1 maio. 2000.
- RHOADES, M. M. The early years of maize genetics. **Annual review of genetics**, v. 18, p. 1-29, jan. 1984.
- RICE, P.; LONGDEN, I.; BLEASBY, A. EMBOSS: the European Molecular Biology Open Software Suite. **Trends in genetics : TIG**, v. 16, n. 6, p. 276-7, jun. 2000.
- RINALDI, A. Science wikinomics. Mass networking through the web creates new forms of scientific collaboration. **EMBO reports**, v. 10, n. 5, p. 439-443, maio. 2009.
- ROBERTSON, H. M.; LAMPE, D. J. Distribution of Transposable Elements in Arthropods. **Annual Review of Entomology**, v. 40, n. 1, p. 333-357, 1995.
- ROCHA, E. P. C.; DANCHIN, A. An analysis of determinants of amino acids substitution rates in bacterial proteins. **Molecular biology and evolution**, v. 21, n. 1, p. 108-16, jan. 2004.

- ROKAS, A. *et al.* Genome-scale approaches to resolving incongruence in molecular phylogenies. **Nature**, v. 425, n. 6960, p. 798-804, 23 out. 2003.
- ROMANISH, M. T. *et al.* Repeated recruitment of LTR retrotransposons as promoters by the anti-apoptotic locus NAIP during mammalian evolution. **PLoS genetics**, v. 3, n. 1, 12 jan. 2007.
- ROSSUM, G. VAN; ET AL. The Python programming language. **Python Software Foundation**, 2010.
- ROUAULT, J.-D. *et al.* Automatic classification within families of transposable elements: application to the mariner Family. **Gene**, v. 448, n. 2, p. 227-32, dez. 2009.
- ROUZIC, A. LE; BOUTIN, T. S.; CAPY, P. Long-term evolution of transposable elements. **Proceedings of the National Academy of Sciences of the United States of America**, v. 104, n. 49, p. 19375-80, 4 dez. 2007.
- ROUZIC, A. LE; CAPY, P. The first steps of transposable elements invasion: parasitic strategy vs. genetic drift. **Genetics**, v. 169, n. 2, p. 1033-1043, 2005.
- ROUZIC, A. LE; CAPY, P. The first steps of transposable elements invasion: parasitic strategy vs. genetic drift. **Genetics**, v. 169, n. 2, p. 1033-1043, 1 fev. 2005.
- ROUZIC, A. LE; CAPY, P. Population genetics models of competition between transposable element subfamilies. **Genetics**, v. 174, n. 2, p. 785-793, out. 2006.
- ROUZIC, A. LE; CAPY, P. Theoretical Approaches to the Dynamics of Transposable Elements in Genomes, Populations, and Species Transposons and the Dynamic Genome. *In*: LANKENAU, D.-H.; VOLFF, J.-N. (Eds.). [s.l.] Springer Berlin / Heidelberg, 2009. v. 4p. 1-19.
- ROUZIC, A. LE; DECELIERE, G. Models of the population genetics of transposable elements. **Genetical research**, v. 85, n. 3, p. 171-181, jun. 2005.
- ROUZIC, A. LE; DUPAS, S.; CAPY, P. Genome ecosystem and transposable elements species. **Gene**, v. 390, n. 1-2, p. 214-220, 1 abr. 2007.
- ROZHKOVA, N. V; HAMMELL, M.; HANNON, G. J. Multiple roles for Piwi in silencing Drosophila transposons. **Genes & development**, v. 27, n. 4, p. 400-12, 15 fev. 2013.
- SCHAACK, S.; GILBERT, C.; FESCHOTTE, C. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. **Trends in ecology & evolution**, v. 25, n. 9, p. 537-46, set. 2010.
- SCHEIFELE, L. Z. *et al.* Retrotransposon overdose and genome integrity. **Proceedings of the National Academy of Sciences of the United States of America**, v. 106, n. 33, p. 13927-13932, 18 ago. 2009.
- SCHNEIDER, D.; LENSKI, R. E. Dynamics of insertion sequence elements during experimental evolution of bacteria. v. 155, p. 319-327, 2004.
- SCHULTHEISS, S. J. *et al.* KIRMES: kernel-based identification of regulatory modules in euchromatic sequences. **Bioinformatics**, v. 25, n. 16, p. 2126-2133, ago. 2009.

- SEBERG, O.; PETERSEN, G. A unified classification system for eukaryotic transposable elements should reflect their phylogeny. **Nature reviews. Genetics**, v. 10, n. 4, p. 276, abr. 2009.
- SETTA, N. DE *et al.* Copia Retrotransposon in the Zaprionus Genus: Another Case of Transposable Element Sharing with the Drosophila melanogaster Subgroup. **Journal of molecular evolution**, fev. 2011.
- SHAH, P.; GILCHRIST, M. A. Effect of Correlated tRNA Abundances on Translation Errors and Evolution of Codon Usage Bias. **PLoS Genet**, v. 6, n. 9, 16 set. 2010.
- SHAPIRO, J. A. Mutations caused by the insertion of genetic material into the galactose operon of Escherichia coli. **Journal of molecular biology**, v. 40, n. 1, p. 93-105, 28 fev. 1969.
- SIMPSON, J. T.; DURBIN, R. Efficient construction of an assembly string graph using the FM-index. **Bioinformatics**, v. 26, n. 12, p. i367-i373, 15 jun. 2010.
- SINGH, N. D. *et al.* Comparative Genomics on the Drosophila Phylogenetic Tree. **Annual Review of Ecology, Evolution, and Systematics**, v. 40, n. 1, p. 459-480, 2009.
- SJÖLANDER, K. Phylogenomic inference of protein molecular function: advances and challenges. **Bioinformatics (Oxford, England)**, v. 20, n. 2, p. 170-179, 22 jan. 2004.
- SMIT, AFA; HUBLEY, R; GREEN, P. RepeatMasker Open-3.0. 1996-2010 <<http://www.repeatmasker.org>>.
- SMITH, J.; SMITH, N. Synonymous nucleotide divergence: what is “saturation”? **Genetics**, n. 142, p. 1033-1036, 1996.
- SMITH, T. F.; WATERMAN, M. S. Identification of common molecular subsequences. **Journal of molecular biology**, v. 147, n. 1, p. 195-7, 25 mar. 1981.
- SOLTIS, P. S.; SOLTIS, D. E.; CHASE, M. W. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. **Nature**, v. 402, n. 6760, p. 402-4, 25 nov. 1999.
- SONNHAMMER, E. L. L.; KOONIN, E. V. Orthology, paralogy and proposed classification for paralog subtypes. **Trends in genetics : TIG**, v. 18, n. 12, p. 619-20, dez. 2002.
- STADEN, R. Sequence data handling by computer. **Nucleic acids research**, v. 4, n. 11, p. 4037-51, nov. 1977.
- STUPKA, E. Large-scale open bioinformatics data resources. **Current opinion in molecular therapeutics**, v. 4, n. 3, p. 265-74, jun. 2002.
- TCHAGANG, A. B. *et al.* GOAL: a software tool for assessing biological significance of genes groups. **BMC bioinformatics**, v. 11, p. 229, jan. 2010.
- THAN, C. *et al.* Identifiability Issues in Phylogeny-Based Detection of Horizontal Gene Transfer. In: BOURQUE, G.; EL-MABROUK, N. (Eds.). **Comparative Genomics**. Berlin, Heidelberg: Springer Berlin / Heidelberg, 2006. v. 4205p. 215-229.

- THAN, C. *et al.* Identifiability Issues in Phylogeny-Based Detection of Horizontal Gene Transfer. *In: BOURQUE, G.; EL-MABROUK, N. (Eds.). [s.l.] Springer Berlin / Heidelberg, 2006. v. 4205p. 215-229.*
- THAN, C. *et al.* Confounding factors in HGT detection: statistical error, coalescent effects, and multiple solutions. **Journal of computational biology : a journal of computational molecular cell biology**, v. 14, n. 4, p. 517-535, maio. 2007.
- THURMAN, R.; NOBLE, W.; STAMATOYANNOPOULOS, J. Multi-scale correlations in continuous genomic data. **Pac Symp Biocomput**, p. 201-215, 2008.
- TREANGEN, T. J.; SALZBERG, S. L. Repetitive DNA and next-generation sequencing: computational challenges and solutions. **Nat Rev Genet**, v. 13, n. 1, p. 36-46, jan. 2012.
- VICARIO, S. *et al.* Developmental stage and level of codon usage bias in *Drosophila*. **Molecular biology and evolution**, v. 25, n. 11, p. 2269-2277, 1 nov. 2008.
- VIDAL, N. M.; LUDWIG, A.; LORETO, E. L. S. Evolution of Tom, 297, 17.6 and rover retrotransposons in *Drosophilidae* species. **Molecular genetics and genomics : MGG**, v. 282, n. 4, p. 351-62, out. 2009.
- VILLASANTE, A. *et al.* *Drosophila* telomeric retrotransposons derived from an ancestral element that was recruited to replace telomerase. **Genome Research**, v. 17, n. 12, p. 1909-1918, dez. 2007.
- WALLAU, G. L.; KAMINSKI, V. L.; LORETO, E. L. S. The role of vertical and horizontal transfer in the evolution of Paris-like elements in *drosophilid* species. **Genetica**, v. 139, n. 11-12, p. 1487-97, dez. 2011.
- WALLAU, G. L.; ORTIZ, M. F.; LORETO, E. L. S. Horizontal transposon transfer in eukarya: detection, bias, and perspectives. **Genome biology and evolution**, v. 4, n. 8, p. 689-99, jan. 2012.
- WALLAU, G. L.; ORTIZ, M. F.; LORETO, E. L. S. Horizontal transposon transfer in eukarya: detection, bias, and perspectives. **Genome biology and evolution**, v. 4, n. 8, p. 689-99, 1 jan. 2012.
- WANG, B. Limitations of Compositional Approach to Identifying Horizontally Transferred Genes. p. 244-250, 2001.
- WANG, T.; STORMO, G. D. Identifying the conserved network of cis-regulatory sites of a eukaryotic genome. **Proceedings of the National Academy of Sciences of the United States of America**, v. 102, n. 48, p. 17400-5, 29 nov. 2005.
- WEGRZYN, G. The minimal genome paradox. **Journal of applied genetics**, v. 42, n. 3, p. 385-92, jan. 2001.
- WICKER, T. *et al.* A unified classification system for eukaryotic transposable elements. **Nature reviews. Genetics**, v. 8, n. 12, p. 973-982, 1 dez. 2007.
- WILKINS, A. S. The enemy within: an epigenetic role of retrotransposons in cancer initiation. **BioEssays : news and reviews in molecular, cellular and developmental**

biology, v. 32, n. 10, p. 856-865, out. 2010.

WRIGHT, F. The “effective number of codons” used in a gene. **Gene**, v. 87, n. 1, p. 23-9, 1 mar. 1990.

WU, Q. *et al.* Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs. **Proceedings of the National Academy of Sciences of the United States of America**, v. 107, n. 4, p. 1606-1611, 26 jan. 2010.

YANG, Z. PAML 4: phylogenetic analysis by maximum likelihood. **Molecular biology and evolution**, v. 24, n. 8, p. 1586-91, ago. 2007.

YANG, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. **Molecular Biology and Evolution**, v. 24, n. 8, p. 1586-1591, 1 ago. 2007.

YANG, Z.; NIELSEN, R. Mutation-Selection Models of Codon Substitution and Their Use to Estimate Selective Strengths on Codon Usage. **Molecular Biology and Evolution**, v. 25, n. 3, p. 568-579, 1 mar. 2008.

YU, G. *et al.* PUGSVM: a caBIGTM analytical tool for multiclass gene selection and predictive classification. **Bioinformatics (Oxford, England)**, v. 27, n. 5, p. 736-738, 1 mar. 2011.

ZHANG, Z.; YU, J. Evaluation of six methods for estimating synonymous and nonsynonymous substitution rates. **Genomics, proteomics & bioinformatics**, v. 4, n. 3, p. 173-81, ago. 2006.

ZUCKERKANDL, E. Revisiting junk DNA. **Journal of molecular evolution**, v. 34, n. 3, p. 259-71, mar. 1992.

Anexos

Anexo 1: Methods for detection of horizontal transfer of transposable elements in complete genomes

de Carvalho, MO; Loreto, ELS

Abstract:

The new advances in nucleic acid sequencing technology are opening a diverse landscape for the analysis of horizontal transfer in complete genomes. Before limited to prokaryotes, the availability of complete genomes from close eukaryotic species presents an opportunity to validate hypotheses about the patterns of evolution and mechanisms that drive horizontal transfer. Many of those methods can be ported from previously methods used in prokaryotic genomes, as the assumptions for horizontal transfer can be interpreted as the same. Some methods however require a complete adaptation, while other need refinements over sensibility and specificity to deal with the huge datasets generated from next-generation sequencing technologies. Here we list the types of methods used for horizontal transfer detection, as well as theirs strengths and weakness.

Keywords: transposable element, horizontal transfer, genome, computational analysis, evolution



Methods for detection of horizontal transfer of transposable elements in complete genomes

Marcos Oliveira de Carvalho¹ and Elgion L.S. Loreto^{1,2}

¹Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.

²Departamento de Biologia, Universidade Federal de Santa Maria, Santa Maria, RS, Brazil.

Abstract

Recent advances in nucleic acid sequencing technology are creating a diverse landscape for the analysis of horizontal transfer in complete genomes. Previously limited to prokaryotes, the availability of complete genomes from close eukaryotic species presents an opportunity to validate hypotheses about the patterns of evolution and mechanisms that drive horizontal transfer. Many of those methods can be transported from methods previously used in prokaryotic genomes, as the assumptions for horizontal transfer can be interpreted as the same. Some methods, however, require a complete adaptation, while others need refinements in sensitivity and specificity to deal with the huge datasets generated from next-generation sequencing technologies. Here we list the types of methods used for horizontal transfer detection, as well as their strengths and weakness.

Keywords: transposable element, horizontal transfer, genome, computational analysis, evolution.

Horizontal Gene Transfer and its Detection

Horizontal transfer can be defined as the exchange of genetic material between species without the aid of any form of sexual mechanism (Gilbert *et al.*, 2009). This phenomenon is widely documented in prokaryote species and plays a special role in prokaryotic and eukaryotic evolution and adaptation (Biémont and Vieira, 2006; Silva *et al.*, 2004). Prokaryotes usually perform horizontal transfer of genetic material through Type IV secretion systems (Juhás *et al.*, 2008), conjugation (Weinert *et al.*, 2009), transformation (Fall *et al.*, 2007) or transduction (Zaneveld *et al.*, 2008), all being biological mechanisms that ease the exchange of DNA. There are many cases of horizontal transfer documented for eukaryotic species as well (Keeling and Palmer, 2008), although with a lower frequency than in prokaryotes, due to the lack of a well defined biological process for the exchange of genetic material between eukaryotic lineages without a sexual mechanism.

However, genetic elements like transposable elements (TEs) are capable of encoding enzymes that permit the integration of their DNA sequences into specific regions of the genome (Schaack *et al.*, 2010). This characteristic enables transposable elements to insert themselves

into different hosts, without the aid of special biological mechanism for DNA transfer. Several processes have been suggested in horizontal transfer of TEs in eukaryotes. For example, direct transfer of episomes (O'Brochta *et al.*, 2009) or some retrotransposons capable of generating viral particles (Kim *et al.*, 1994). Also, transposable elements can putatively explore events like virus infections (Dupuy *et al.*, 2011) and parasite mediated transfer (Gilbert *et al.*, 2010) from one host to another.

Horizontal acquisition of genes is an important force in evolution, with examples of influence in the evolutionary history of many species (Gurudatta and Corces, 2009; Zhou and Wang, 2008). Of special importance is the acquisition of pathogenicity islands in prokaryotic species (Gal-Mor and Finlay, 2006), as those cases of horizontal transfer can promote the development of new pathogenic bacteria strains. Thus, the accessibility of accurate and precise methods to quickly identify horizontally transferred genes is of crucial significance to the study and broad comprehension of the processes that have shaped genomes. However the first challenge a researcher faces when identifying horizontal transfer events in whole genomes, and specifically in the case of transposable element horizontal transfer analysis, is the annotation of TE sequences (Flutre *et al.*, 2011). Currently many methodologies are applied for TE annotation, however, there are no established ontologies like the Gene Ontology (Ashburner *et al.*, 2000) that greatly help

Send correspondence to Marcos Oliveira de Carvalho. Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves 9500, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brazil. E-mail: marcos.carvalho@gmail.com.

the automated curation and annotation of large-scale datasets (Plessis *et al.*, 2011).

Before the large availability of genome sequencing, the support for horizontal transfer detection was sparse, relying on standard molecular biology hybridization protocols or a small number of sequences (Clark *et al.*, 2002; Daniels *et al.*, 1990). As genome sequencing technology progressed, the era of complete genomes brought a great opportunity for large-scale analysis of horizontal transfer events, where a big volume of data could be explored to develop new theories and analysis about the nature of this intricate evolutionary process (Ragan, 2009).

Initially, restrained in its ability to compute and process large quantities of data from full genome projects, indirect methods of detection for horizontal transfer were developed (Ragan, 2001). These methods relied on the detection of differential compositional patterns from nucleotide sequences (Putonti *et al.*, 2006). They were based on the premise that sequences that have been transferred horizontally from a distinct donor to a new host would harbor a different nucleotide composition (Rocha *et al.*, 2006). This would allow one to infer horizontal transfer in a certain region of the genome if its nucleotide composition statistically differs from the genome's average nucleotide composition. These methods, called surrogate methods, proved to be useful in identifying pathogenicity islands in prokaryotic genomes. However, they cannot be widely applied without a careful analysis of the results, since they tend to produce a large number of false positives and false negatives (Azad and Lawrence, 2011). The availability of comparative methods that use phylogenetic inference and tree incongruence greatly expanded the set of tools for horizontal transfer, adding specificity but also increasing the computational costs for each analysis (Lyubetsky and V'yugin, 2003). With the recently developed next-generation sequencing technologies and the low cost for complete genome sequencing genome sequence data is even more readily available. These data can now be used for a data-driven discovery of horizontal transfer, where the phenomenon can be modeled through the use of massive datasets bonded with exploratory and machine learning techniques.

Next, we will present the methods discussed above in more detail, also including a section on large-scale data visualization for horizontal transfer analysis.

Horizontal Transfer Analysis in Whole Genomes

Surrogate methods for detection of horizontal transfer events can be defined as those that do not employ the construction of phylogenetic trees or other direct phylogenetic analysis (Ragan, 2001). Although surrogate methods present advantages in large-scale analysis of horizontal transfer, they have major drawbacks, especially the high rate of false positives that this kind of method can return. Errors of

this type can be traced to the breaking of assumptions of the method. It has been demonstrated that intragenomic variation of codon bias can be large enough to be confounded with true significant variation that otherwise would be attributed to horizontal transfer (Guindon and Perrière, 2001). Moreover, it has been suggested that both codon bias and base compositional indexes, common indexes used in surrogate methods, are poor indicators of horizontal transfer (Koski *et al.*, 2001). Surrogate methods can, however, be used when a quick scan for horizontal transfer is needed, but insufficient genome data are available for comparison. This is the case for a new genome with no closely related species genomes sequenced. In this particular case, comparative methods could not be applied with a high level of confidence, since the lack of paralogous/orthologous data can lead to mistaken results of horizontal transfer (Capy and Gibert, 2004). One remaining weakness of surrogate methods is related to the power of resolution over the time since the occurrence of the horizontal transfer. Genes that were introgressed will slowly acquire the host genome codon usage and compositional values in a phenomenon called amelioration (Marri and Golding, 2008). This will lead to a masking of compositional differences, reducing the sensitivity of surrogate methods (Becq *et al.*, 2010). It also has been suggested that for successful integration and maintenance of genes in the host genome after a horizontal transfer event, both the recipient genome and the transferred genes must have codon bias compatibility (Medrano-Soto *et al.*, 2004). This is a valid assumption for horizontal transfer of bacterial genes; however, its importance for the successful expression of TE genes remains to be accessed.

Specific software implementations exist for different approaches of horizontal transfer identification via surrogate methods. The Alien-Hunter software (Vernikos and Parkhill, 2006) uses Interpolated Variable Order Motifs (IVOMs) to explore compositional biases for detection of horizontal transfer. This method employs variable order motif distributions to capture more reliably the local composition of a sequence compared with fixed-order methods (Vernikos and Parkhill, 2006). It is assumed that the genome has a reasonably constant background sequence composition, derived from uniform mutational pressure over the complete genome. Thus, atypical sequences are inferred as horizontally transferred if the present window of analysis over the genome induces the current HMM to change from the "typical" to "atypical" state. Clustering of proxy variables from sequence composition has also proved to be an effective alternative to discrete modeling through hidden Markov models. The entropic clustering method (Azad and Lawrence, 2007) uses the Jensen-Shannon divergence measure as a variable for posterior clustering of all genes in the genome in analysis, allowing identification of horizontal transferred sequences based on the dissimilarity of the cluster distributions.

However, it is also possible to use more flexible approaches in the identification of horizontal transfer using nucleotide and codon bias data. This can be achieved by the independent calculation of compositional indexes for each gene in the genome (or in a sliding window fashion), be it nucleotide indexes or codon bias values, and run the statistical analysis in a separate framework. The EMBOSS suite of software contains specific software implementations both for nucleotide indexes and codon bias calculations, and can be easily integrated in complex pipelines. The INCA (Supek and Vlahovicek, 2004) system is a user-friendly graphical software program that allows the calculation of many codon bias parameters, also allowing the determination of sequence clusters via the self-organizing maps machine learning method (Wang *et al.*, 2001). The results of the independent codon bias and nucleotide indexes calculations can then be submitted to statistical analysis using the R system or other statistical frameworks.

Comparative methods rely on the existence of evolutionary related sequence data to identify horizontal transfer of genes in a determined species or group of species. This group of methodologies include phylogenetic and tree analysis, phylogenomic approaches and statistical analysis of phylogenetic indexes.

Local alignment similarity searches provide a quick and relatively inexpensive (in computational terms) way to identify related sequences in different databases. This property was explored to develop a method of horizontal transferred genes where simple assumptions are taken into account (Shi *et al.*, 2005). First, for a group of species, all genes are searched against each other in the different species of the dataset. From the results, only hits with an *e*-value of less than $1e-20$ and with the following five hits homologous to the searched taxon are retained. Second, all of its homologs (of the hit selected in the previous step) are from a distant taxon, or the *e*-value of the closest homolog from a distant taxon is significantly lower than the *e*-value of the closest homolog not from the distant taxon (Shi *et al.*, 2005). This method uses the *e*-value as a proxy variable to compare putative horizontally transferred DNA from different genomes, assuming the *e*-value as an indicator of similarity.

The DarkHorse method (Podell and Gaasterland, 2007) also employs local alignment similarity searches as its start point for horizontal transfer detections in genomic scale. However, this methodology uses a specific metric to compare different horizontally transferred gene candidates, instead of the default *e*-value. This metric is called "lineage probability index" or LPI and represents the likelihood that the current gene under search was horizontally transferred taking into consideration the similarity of the gene with distant taxa and also the other genes in the dataset in relation to different taxa.

More recently a method based on phylogenetic distances (Distance Method) was introduced to avoid dealing

with the intrinsic bias introduced in the process of local alignment similarity search (Wei *et al.*, 2008). In this methodology all phylogenetic distances from a gene family are calculated from different species, forming a dataset of all-against-all distance pairs. The distances are then analyzed to identify horizontal transfer through the comparative analysis of distances values between pairs of species according to a pre-defined species phylogeny. The method assumes that between pairs of species from the same branch of the tree, all genes must have smaller distance values between each other than with other species in the tree. By transversing the tree and comparing all distances for the different pairs of species, the Distance Method can identify putatively transferred genes if the distances between taxa from other branches of the tree are smaller than with species within the same branch.

Comparative methods proved to be more sensitive and specific (Poptsova and Gogarten, 2007). However, it is important to note that to work efficiently, comparative phylogenetic methods need a robust phylogenetic tree from the species under analysis as a reference. If this requirement cannot be met, surrogate methods can be applied in conjunction to provide additional support to the results, although a complete phylogenetic validation should be preferentially used if possible (van Passel *et al.*, 2004).

A comprehensive analysis of multiple eukaryotic genomes can generate many megabytes of data, if not gigabytes. This is also true for the analysis of multiple horizontal transfer events, especially if using comparative methodologies. To tackle the problem of identifying data signals that lead to the identification of horizontal transfer, the use of large scale data analysis is imperative. Data clustering is the assignment of data points into subset classes, where the intraclass similarities are statistically more significant than the interclass similarities. Cluster analysis is thus a form of unsupervised learning, where no prior knowledge is used for the determination of the classes. There are two fundamental types of cluster analysis, one that employs a hierarchical approach and another that employs a partition approach to the data classification. Hierarchical methods find successive clusters by seeding the actual cluster with previous classified data. When no cluster is available, like in the initial round of clustering, each data point is considered an initial cluster. The algorithm then progresses by agglomerating similar data points based on metrics derived from distance measures between the data points. This kind of hierarchical clustering is called bottom-up, where each data point starts as a cluster and is fused with other data points as the cycles of clustering progress. It is also possible to use top-bottom approaches, where the whole dataset is initially considered as a single cluster and is progressively separated into small clusters, where the analysis of each data point identifies its separation based on distance measure metrics.

Hierarchical clustering analysis of large datasets is commonplace in many fields of biology today like expression analysis of microarray and RNA-seq data. However, this technique can be successfully exploited in the analysis of whole genome horizontal transfer datasets if the analysis considers each gene as a data point. In this manner, if a gene has a fixed number of variables associated with it, where such variables are indicators of phylogenetic or evolutionary events, those genes can be grouped through the clustering of one or more of those variables. It is important to consider the variance of the variables used for analysis and also the scale of each variable prior to the use of more than one as a group. If there are significant differences between the variables, a normalization step must be considered before the hierarchical clustering. There are many implementations of clustering analysis in software ranging from complete statistical packages like SPSS or SAS to specific library packages for the R platform. However, much of the analysis in whole genome horizontal transfer analysis needs to be integrated in a pipeline fashion, where flexibility is a positive characteristic in a software implementation of hierarchical clustering. In this form, the most useful software platforms to implement this kind of analysis are programming libraries. The Python programming language is considered one of the most flexible and user-friendly modern languages, with multi-paradigm programming capabilities and also clean syntax. Besides the availability of biology-centered programming libraries like Biopython (Cock *et al.*, 2009), PyCogent (Knight *et al.*, 2007) and Corebio, the Python language has many large-scale data analysis libraries, including for hierarchical clustering. One of the most documented and maintained data analysis library comes from the SciPy project, with the *scipy-cluster* plug-in. This library has implemented more than twenty clustering methods, including complete linkage, ward clustering and the centroid/UPGMC algorithm. Those can be quickly implemented in a pipeline alongside other statistical methods allowing for great flexibility in the large-scale analysis of genetic horizontal transfer. The R platform also contains a large number of programming libraries with different clustering methods. One advantage of the R platform is the possibility to prototype specific statistical analysis in integrated graphical interfaces like Rkward and R-Studio, before integration in the pipeline. The R libraries can also be conveniently accessed from the Python environment using the RPy library, making the two technologies very suitable for building large pipelines with many integrated complex statistical analyses. To conduct an exploratory analysis using hierarchical clustering it is also useful to have a graphical interface where it is possible to test different kinds of distance measures and normalizations in a sample of the data that is being analyzed. This kind of analysis can be conveniently performed in the HCE-Explorer software. This tool provides an interactive graphical user interface to explore large datasets before and after the hierarchi-

cal clustering analysis. Also, there are options to compare different runs of clustering, depending on the comparison of the parameters used for the analysis. This software was initially developed for the analysis of gene expression data; however, any dataset can be loaded if it conforms to the standard data format used by HCE-Explorer. This mainly consists of rows describing the main data point with variable values in the respective columns and can be loaded in the CSV data file format. Although only Windows binaries are available, the HCE-Explorer can run on Unix systems under the Wine platform.

Visualization of large-scale datasets can be a powerful tool to help identify patterns of horizontal transfer in genomic data. Although a young subject, many approaches from other fields of science can be applied to the visualization of horizontal transfer, such as graph analysis. Graph visualization is specially suited for the analysis of large-scale horizontal transfer because of the inherently high connection of data points, namely, the genes in a horizontal transfer analysis. These data points are not expected to behave in a tree-like fashion as observed in tree reconstruction phylogenetic analysis, since the variables accounting for genes that were subjected to horizontal transfer should reproduce the characteristics of transfer from one species to another. In a graph, data points that represent genes that have undergone horizontal transfer should connect more distinctly than data points that have not been horizontally transferred. Using variables like phylogenetic distance, one should expect to have connections between data points from distant species as an indicator of putative horizontal transfer. This kind of analysis can provide instant visual information about the patterns of organization of the genes that have undergone horizontal transmission. This kind of visualization was well employed in the analysis of the horizontal transfer events in the genomes of *Mycoplasma synoviae* and *Mycoplasma gallisepticum*, where a cluster of genes that had undergone horizontal transfer was clearly seen in a graph visualization of genes as data points, and the local sequence similarity score as a graph *e.g.* measure (Vasconcelos *et al.*, 2005). Two specific software programs implement useful methods for large-scale graph visualization. The Phylographer software program is a graphical environment that allows for flexible large-scale graph construction, with a simple data file specification where each gene can be considered as a node and any kind of variable can be attributed as an edge. This flexibility allows for the use of either similarity or phylogenetic variables as *e.g.* connection data. A drawback of the Phylographer software is the need for the TK/TCL platform and a less user-friendly interface than most modern packages for graph visualization. However, large graphs (more than 60,000 nodes) can be built in a matter of hours in a high-end workstation, producing a lightweight interactive interface that allows the manipulation of both nodes and edges and identification of dissimilar graph regions. A more modern system for graph visualization is

the Gephi system, used extensively in the graph analysis community and under active development. The Gephi system is an open-source suite built in the Java language and is capable of multiple graph visualization layouts, including the arbitrary mapping of variables over node and *e.g.* properties. This system allows for more information to be added to the graph than the Phylographer software does. One example would be the mapping of gene length over node color, phylogenetic distance over *e.g.* width and GC content over node color. That kind of graph would allow for a complete picture of three different variables and their relationship with putative horizontal transferred genes, providing useful insights in the development of posterior specific analysis. The Python language also provides many libraries that can be effectively used to integrate large-scale graph visualization with analysis pipelines like the NetworkX and the python-graph libraries.

As a complex biological phenomenon, horizontal transfer is modulated and influenced by a number of variables, many unknown and inaccessible with our current set of tools. This lack of knowledge about the specific factors that drive a specific event can be partially overcome by massive datasets derived from analysis of the phenomenon. These datasets can be created by attaching specific variables to each gene in the analysis, in many related genomes and an integrated search for differential patterns of gene evolution carried out over the complete dataset. If evolutionary assumptions for horizontal transfer like small phylogenetic distance between distance species, low dS

rates in genes in relation to the core genes in the complete genome and codon bias (as in surrogate methods) are taken into account, a specific model of horizontal transfer can be derived from an initial exploratory clustering analysis. This initial model can be supplied to supervised machine learning methodologies to identify similar patterns of genes in genomes of related species. This approach has the advantage of the use of real evolutionary signals, represented as the variables attached to each gene in the genome, in a large dataset to build a model that most closely represents horizontal transfer.

Figure 1 presents an organogram representation of methods used for inference of horizontal transfer in complete genomes.

Conclusions

Large-scale horizontal transfer analysis is a recent development, being fueled mainly by recent developments in sequencing technology. With the availability of large datasets of genomic sequences, many hypotheses about horizontal transfer of sequences between distant species could be elucidated, as well as the development and testing of new hypotheses regarding the specific evolutionary patterns of sequences that have undergone horizontal transfer.

The methods employed for such analysis have evolved at the same rate as data has accumulated. However, many of these lack implementation or have implementations that are not user friendly. This means more expense

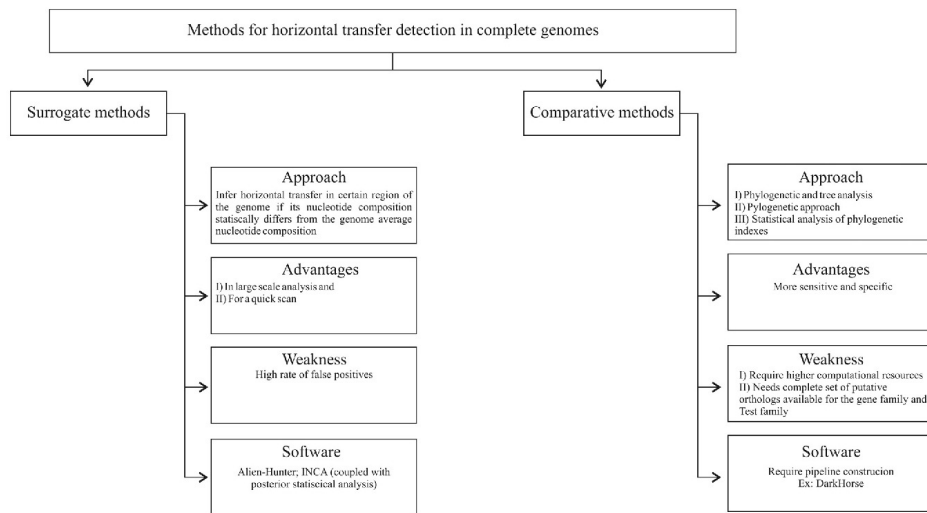


Figure 1 - Organogram representation of methods used for inference of horizontal transfer in complete genomes.

and time spent in the process of analyzing large-scale datasets, a laborious and time consuming endeavor *per se*. Additionally, the lack of a strong community built around the data and methods for analysis of horizontal transfer of transposable elements hinder the development of more advanced tools like common ontologies, essential for an efficient system of data communication. These issues need to be addressed in order to allow effective use of the available data and empower the development of more efficient large-scale analysis of horizontal transfer events.

From the methods presented to identify horizontal transfer, it is clear that surrogate methods should be employed only if no data are suitable for the use of comparative methods. Surrogate methods lag behind comparative methods in terms of resolution and specificity. Comparative methods, however, although more specific and with more power to identify old horizontal transfer events, require higher computational resources for their application on a complete genome scale and increase the complexity of large-scale analysis. Additionally, comparative methods should be applied only when a complete set of putative orthologs are available for the gene family or TE family under analysis.

As a whole, the field of methodologies for large-scale horizontal transfer analysis is in its infancy with many exciting developments under way. Many improvements are needed especially in the automation of TE annotation and in the sensitivity and specificity of the current methods to identify horizontal transfer events. Some of the drawbacks can be addressed with the development of specific machine learning implementations for horizontal transfer detections, as those methodologies can make use of hidden features of evolutionary variables from complete genome datasets to identify the subtle differences between genes inherited vertically or horizontally more accurately.

However, with the growing availability of complete genomes and the growing importance of the understanding of horizontal transfer as a force in the evolution of many species, the methodologies under development should quickly develop into fully established standards. Together with the growing datasets of genomic information, mature methods for horizontal transfer identification could help establish a new way of thinking about the long-term evolution of species.

Acknowledgments

The authors thank CAPES and CNPQ for fellowships. This work was supported by CNPq (grant 473375/2009-5) and PRONEX- FAPERGS (grant 10/0028-7).

References

Ashburner M, Ball C, Blake J, Botstein D, Butler H, Cherry J, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al.* (2000) Gene

ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25:25-29.

Azad RK and Lawrence JG (2007) Detecting laterally transferred genes: Use of entropic clustering methods and genome position. *Nucleic Acids Res* 35:4629-4639.

Azad RK and Lawrence JG (2011) Towards more robust methods of alien gene detection. *Nucleic Acids Res* 39:e56-e56.

Becq J, Churlaud C and Deschavanne P (2010) A benchmark of parametric methods for horizontal transfers detection. *PLoS One* 5:e9989.

Biémont C and Vieira C (2006) Junk DNA as an evolutionary force. *Nature* 443:521-524.

Capy P and Gibert P (2004) *Drosophila melanogaster*, *Drosophila simulans*: So similar yet so different. *Genetica* 120:5-16.

Clark JB, Silva JC and Kidwell MG (2002) Evidence of horizontal transfer of P transposable elements. In: Syvanen M and Kado CI (eds) *Horizontal Gene Transfer*. Academic Press, San Diego, pp 161-171.

Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, *et al.* (2009) Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25:1422-1423.

Daniels S, Peterson K, Strausbaugh L, Kidwell M and Chovnick A (1990) Evidence for horizontal transmission of the P transposable element between *Drosophila* species. *Genetics* 124:339-355.

Dupuy C, Periquet G, Serbielle C, Bézier A, Louis F and Drezén J-MM (2011) Transfer of a chromosomal Maverick to endogenous bracovirus in a parasitoid wasp. *Genetica* 139:489-496.

Fall S, Mercier A, Bertolla F, Calteau A, Gueguen L, Perrière G, Vogel TM and Simonet P (2007) Horizontal gene transfer regulation in Bacteria as a "spandrel" of DNA repair mechanisms. *PLoS One* 2:e1055.

Flutre T, Duprat E, Feuillet C and Quesneville H (2011) Considering transposable element diversification in de novo annotation approaches. *PLoS ONE* 6:e16526.

Gal-Mor O and Finlay BB (2006) Pathogenicity islands: A molecular toolbox for bacterial virulence. *Cell Microbiol* 8:1707-1719.

Gilbert C, Pace JK and Feschotte C (2009) Horizontal SPINning of transposons. *Commun Integr Biol* 2:117-119.

Gilbert C, Schaack S, Pace II JK, Brindley PJ and Feschotte C (2010) A role for host-parasite interactions in the horizontal transfer of transposons across phyla. *Nature* 464:1347-1350.

Guindon S and Perrière G (2001) Intragenomic base content variation is a potential source of biases when searching for horizontally transferred genes. *Mol Biol Evol* 18:1838-1840.

Gurudatta B and Corces VG (2009) Chromatin insulators: Lessons from the fly. *Brief Funct Genomics Proteomics* 8:276-282.

Juhas M, Crook DW and Hood DW (2008) Type IV secretion systems: Tools of bacterial horizontal gene transfer and virulence. *Cell Microbiol* 10:2377-2386.

Keeling PJ and Palmer JD (2008) Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet* 9:605-618.

Kim AC, Terzian P, Santamaria A, Pélisson N, Prud'homme and Bucheton A (1994) Retroviruses in invertebrates: The gypsy retrotransposon is apparently an infectious retrovirus of

- Drosophila melanogaster*. Proc Natl Acad Sci USA 91:1285-1289.
- Knight R, Maxwell P, Birmingham A, Carnes J, Caporaso JG, Easton B, Eaton M, Hamady M, Lindsay H, Liu Z, *et al.* (2007) PyCogent: A toolkit for making sense from sequence. Genome Biol 8:R171.
- Koski LB, Morton RA and Golding GB (2001) Codon bias and base composition are poor indicators of horizontally transferred genes. Mol Biol Evol 18:404-412.
- Lyubetsky VA and V'yugin VV (2003) Methods of horizontal gene transfer determination using phylogenetic data. In silico Biol 3:17-31.
- Marri PR and Golding GB (2008) Gene amelioration demonstrated: The journey of nascent genes in bacteria. Genome 51:164-168.
- Medrano-Soto A, Moreno-Hagelsieb G, Vinuesa P, Christen AA and Collado-Vides J (2004) Successful lateral transfer requires codon usage compatibility between foreign genes and recipient genomes. Mol Biol Evol 21:1884-1894.
- O'Brochta DA, Stosic CD, Pilitt K, Subramanian RA, Hice RH and Atkinson PW (2009) Transpositionally active episomal hAT elements. BMC Mol Biol 10:108-120.
- Passel M, van Bart A, Pannekoek Y and Ende A (2004) Phylogenetic validation of horizontal gene transfer? Nat Genet 36:1028.
- Plessis L, du Skunca N and Dessimoz C (2011) The what, where, how and why of gene ontology - A primer for bioinformaticians. Brief Bioinform 12:723-735.
- Podell S and Gaasterland T (2007) DarkHorse: A method for genome-wide prediction of horizontal gene transfer. Genome Biol 8:R16.
- Poptsova MS and Gogarten JP (2007) The power of phylogenetic approaches to detect horizontally transferred genes. BMC Evol Biol 7:e45.
- Putonti C, Luo Y, Katili C, Chumakov S, Fox GE, Graur D and Fofanov Y (2006) A computational tool for the genomic identification of regions of unusual compositional properties and its utilization in the detection of horizontally transferred sequences. Mol Biol Evol 23:1863-1868.
- Ragan MA (2009) Thinking laterally about genomes. Genome informatics. Int Conf Genome Informat 23:221-222.
- Ragan MA (2001) On surrogate methods for detecting lateral gene transfer. FEMS Microbiol Lett 201:187-191.
- Rocha E, Touchon M and Feil E (2006) Similar compositional biases are caused by very different mutational effects. Genome Res 16:1537-1547.
- Schaack S, Gilbert C and Feschotte C (2010) Promiscuous DNA: Horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol Evol 25:537-546.
- Shi S-Y, Cai X-H and Ding D (2005) Identification and categorization of horizontally transferred genes in prokaryotic genomes. Acta Biochim Biophys Sinica 37:561-566.
- Silva JC, Loreto EL and Clark JB (2004) Factors that affect the horizontal transfer of transposable elements. Curr Issues Mol Biol 6:57-71.
- Supek F and Vlahovicek K (2004) INCA: Synonymous codon usage analysis and clustering by means of self-organizing map. Bioinformatics 20:2329-2330.
- Vasconcelos ATR, Ferreira HB, Bizarro CV, Bonatto SL, Carvalho MO, Pinto PM, Almeida DF, Almeida LG, Almeida R, Alves-Filho L, *et al.* (2005) Swine and poultry pathogens: The complete genome sequences of two strains of *Mycoplasma hyopneumoniae* and a strain of *Mycoplasma synoviae*. J Bacteriol 187:5568-5577.
- Vernikos GS and Parkhill J (2006) Interpolated variable order motifs for identification of horizontally acquired DNA: Revisiting the Salmonella pathogenicity islands. Bioinformatics 22:2196-2203.
- Wang H, Badger J, Kearney P and Li M (2001) Analysis of codon usage patterns of bacterial genomes using the self-organizing map. Mol Biol Evol 18:792-800.
- Wei X, Cowen L, Brodley C, Brady A, Sculley D and Slonim D (2008) A distance-based method for detecting horizontal gene transfer in whole genomes. In: Mandouiu I, Sunderaman R and Zelikovskiy A (eds) Bioinformatics Research and Applications. Springer, Berlin, pp 26-37.
- Weinert LA, Welch JJ and Jiggins FM (2009) Conjugation genes are common throughout the genus Rickettsia and are transmitted horizontally. Proc Biol Sci R Soc Lond 276:3619-3627.
- Zaneveld JR, Nemerget DR and Knight R (2008) Are all horizontal gene transfers created equal? Prospects for mechanism-based studies of HGT patterns. Microbiology 154:1-15.
- Zhou Q and Wang W (2008) On the origin and evolution of new genes - a genomic and experimental perspective. J Genet Genomics 35:639-648.

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Anexo 2 - Dados sumarizados das comparações de Tes e genes nucleares.

		Seleção Negativa					Seleção Positiva					
		<i>Média</i>	<i>Max</i>	<i>Min</i>	<i>Desvio Padrão</i>	<i>N</i>	<i>Média</i>	<i>Max</i>	<i>Min</i>	<i>Desvio Padrão</i>	<i>N</i>	
Melanogaster _Group	ACCORD	<i>ks</i>	0.17	0.18	0.16	0.01	6					
		<i>ka</i>	0.12	0.12	0.11	0.01	6					
		<i>kaks</i>	0.69	0.75	0.64	0.05	6					
		<i>protid</i>	79.09	79.66	78.81	0.44	6					
		<i>dnaid</i>	88.33	88.42	88.14	0.14	6					
		<i>kimura</i>	0.25	0.25	0.24	0.01	6					
		<i>logdet</i>	0.36	0.36	0.36	0.00	6					
		<i>index</i>	40.41	41.02	39.98	0.48	6					
	AURORA	<i>ks</i>	0.39	0.91	0.19	0.19	281	0.08	0.21	0.03	0.06	18
		<i>ka</i>	0.07	0.27	0.00	0.06	281	0.13	0.21	0.09	0.04	18
		<i>kaks</i>	0.25	0.99	0.00	0.23	281	1.79	3.28	1.03	0.59	18
		<i>protid</i>	82.10	90.38	70.90	3.91	281	79.87	83.08	76.69	1.96	18
		<i>dnaid</i>	87.89	92.21	72.89	3.57	281	90.48	92.82	82.96	2.76	18
		<i>kimura</i>	0.22	0.45	0.10	0.07	281	0.23	0.28	0.19	0.03	18
		<i>logdet</i>	0.28	0.61	0.13	0.11	281	0.14	0.46	0.08	0.09	18
		<i>index</i>	39.87	57.50	20.16	7.01	281	51.76	60.58	36.83	7.10	18
	BLASTOPIA	<i>ks</i>						0.00	0.00	0.00		1
		<i>ka</i>						0.00	0.00	0.00		1
		<i>kaks</i>						470.0	470.00	470.00		1
			<i>protid</i>					100.0	100.00	100.00		1
			<i>dnaid</i>					100.0	100.00	100.00		1
			<i>kimura</i>					0.00	0.00	0.00		1
			<i>logdet</i>					0.00	0.00	0.00		1
		<i>index</i>					99.53	99.53	99.53		1	
	CHOUTO	<i>ks</i>	0.47	0.49	0.44	0.01	20					
		<i>ka</i>	0.09	0.15	0.08	0.02	20					
		<i>kaks</i>	0.19	0.31	0.16	0.04	20					
		<i>protid</i>	82.89	84.47	77.14	1.83	20					
		<i>dnaid</i>	85.29	86.08	80.95	1.10	20					
		<i>kimura</i>	0.20	0.29	0.17	0.03	20					
		<i>logdet</i>	0.16	0.24	0.15	0.02	20					
		<i>index</i>	39.79	41.67	31.70	2.19	20					
	COPIA	<i>ks</i>	0.44	0.74	0.37	0.09	18					
		<i>ka</i>	0.11	0.18	0.09	0.03	18					
		<i>kaks</i>	0.25	0.35	0.21	0.04	18					
<i>protid</i>		81.15	83.50	70.54	3.83	18						
	<i>dnaid</i>	83.86	85.44	73.51	2.98	18						

		<i>kimura</i>	0.22	0.39	0.19	0.06	18						
		<i>logdet</i>	0.24	0.43	0.18	0.05	18						
		<i>index</i>	36.88	40.02	20.61	5.13	18						
	FW	<i>ks</i>	0.84	1.60	0.48	0.11	509						
		<i>ka</i>	0.10	0.16	0.06	0.02	509						
		<i>kaks</i>	0.12	0.23	0.05	0.02	509						
		<i>protid</i>	73.35	76.11	70.18	1.28	509						
		<i>dnaid</i>	75.59	78.17	72.70	0.64	509						
		<i>kimura</i>	0.33	0.42	0.28	0.02	509						
		<i>logdet</i>	0.41	0.75	0.26	0.06	509						
		<i>index</i>	22.23	27.23	15.14	1.37	509						
	G5	<i>ks</i>	3.03	4.97	1.04	0.98	37						
		<i>ka</i>	0.24	0.73	0.09	0.15	37						
		<i>kaks</i>	0.10	0.32	0.02	0.08	37						
		<i>protid</i>	72.92	77.19	70.09	2.04	37						
		<i>dnaid</i>	73.92	77.78	71.03	1.61	37						
		<i>kimura</i>	0.34	0.41	0.25	0.03	37						
		<i>logdet</i>	0.61	0.89	0.39	0.15	37						
		<i>index</i>	9.33	15.13	5.85	2.33	37						
	Genome	<i>ks</i>	2.13	24.26	0.28	1.84	12821						
		<i>ka</i>	0.08	0.60	0.00	0.06	12821						
		<i>kaks</i>	0.04	0.30	0.00	0.03	12821						
		<i>protid</i>	85.44	100.00	40.08	8.79	12821						
		<i>dnaid</i>	78.71	95.46	50.28	5.52	12821						
		<i>kimura</i>	0.17	0.99	0.00	0.12	12821						
		<i>logdet</i>	0.28	2.13	0.05	0.11	12821						
		<i>index</i>	21.46	68.84	0.85	9.08	12821						
	GTWIN	<i>ks</i>	0.25	0.28	0.20	0.05	3						
		<i>ka</i>	0.12	0.12	0.11	0.01	3						
		<i>kaks</i>	0.47	0.54	0.44	0.06	3						
		<i>protid</i>	79.31	79.82	79.05	0.44	3						
		<i>dnaid</i>	86.67	88.60	85.71	1.67	3						
		<i>kimura</i>	0.25	0.25	0.24	0.01	3						
		<i>logdet</i>	0.19	0.21	0.15	0.03	3						
		<i>index</i>	42.26	46.27	40.26	3.47	3						
	GYPSY	<i>ks</i>	3.78	9.94	0.22	3.03	1105						
		<i>ka</i>	0.13	0.36	0.05	0.08	1105						
		<i>kaks</i>	0.14	0.99	0.01	0.21	1105						
		<i>protid</i>	74.62	84.91	70.00	2.42	1105						
		<i>dnaid</i>	76.31	88.25	70.00	2.99	1105						
		<i>kimura</i>	0.31	0.51	0.17	0.04	1105						
		<i>logdet</i>	0.49	1.68	0.17	0.18	1105						
		<i>index</i>	13.33	42.00	3.17	8.67	1105						
	HELITRON	<i>ks</i>	0.58	5.06	0.01	1.29	26	0.00	0.01	0.00	0.00	27	
		<i>ka</i>	0.23	2.97	0.00	0.75	26	0.08	0.60	0.00	0.16	27	
		<i>kaks</i>	0.06	0.66	0.00	0.17	26	211.71	490.00	50.00	155.37	27	
		<i>protid</i>	75.41	88.71	70.08	4.93	26	84.85	89.66	72.60	6.24	27	

		<i>dnaid</i>	81.92	93.55	73.22	4.86	26	90.12	93.60	77.17	4.85	27
		<i>kimura</i>	0.26	0.37	0.10	0.08	26	0.14	0.32	0.07	0.09	27
		<i>logdet</i>	0.53	1.86	0.15	0.41	26	0.39	0.81	0.15	0.16	27
		<i>index</i>	38.29	61.41	5.23	16.35	26	54.68	68.80	35.88	10.89	27
	MARINER	<i>ks</i>	0.09	3.63	0.00	0.32	1526	0.01	0.09	0.00	0.01	2066
		<i>ka</i>	0.03	0.26	0.00	0.04	1526	0.02	0.14	0.00	0.02	2066
		<i>kaks</i>	0.41	1.00	0.00	0.27	1526	92.85	490.00	1.00	96.42	2066
		<i>protid</i>	94.08	100.00	70.00	5.16	1526	95.03	100.00	77.12	2.90	2066
		<i>dnaid</i>	96.42	100.00	73.89	3.93	1526	97.57	100.00	83.05	1.78	2066
		<i>kimura</i>	0.06	0.36	0.01	0.06	1526	0.05	0.20	0.01	0.03	2066
		<i>logdet</i>	0.07	0.90	0.01	0.10	1526	0.04	0.43	0.01	0.04	2066
		<i>index</i>	80.22	97.25	8.33	15.23	1526	86.17	97.41	55.64	7.48	2066
	NINJA	<i>ks</i>	0.43	1.37	0.21	0.21	210					
		<i>ka</i>	0.08	0.22	0.04	0.05	210					
		<i>kaks</i>	0.18	0.40	0.07	0.07	210					
		<i>protid</i>	77.32	84.78	70.00	4.73	210					
		<i>dnaid</i>	84.62	89.29	77.44	3.24	210					
		<i>kimura</i>	0.25	0.38	0.15	0.06	210					
		<i>logdet</i>	0.41	1.02	0.17	0.20	210					
		<i>index</i>	34.22	44.65	13.67	7.88	210					
	PAO	<i>ks</i>	0.34	0.34	0.34		1					
		<i>ka</i>	0.19	0.19	0.19		1					
		<i>kaks</i>	0.54	0.54	0.54		1					
		<i>protid</i>	73.50	73.50	73.50		1					
		<i>dnaid</i>	81.20	81.20	81.20		1					
		<i>kimura</i>	0.34	0.34	0.34		1					
		<i>logdet</i>	0.32	0.32	0.32		1					
		<i>index</i>	30.52	30.52	30.52		1					
	POLINTON	<i>ks</i>	5.05	9.80	0.28	1.68	57					
		<i>ka</i>	0.15	1.06	0.00	0.36	57					
		<i>kaks</i>	0.02	0.14	0.00	0.04	57					
		<i>protid</i>	81.85	93.00	71.95	3.97	57					
		<i>dnaid</i>	76.82	83.86	70.96	2.55	57					
		<i>kimura</i>	0.26	2.93	0.07	0.36	57					
		<i>logdet</i>	0.52	1.04	0.19	0.19	57					
		<i>index</i>	9.46	46.64	1.30	8.77	57					
	R1	<i>ks</i>	0.77	0.95	0.66	0.04	25					
		<i>ka</i>	0.17	0.20	0.16	0.01	25					
		<i>kaks</i>	0.22	0.25	0.18	0.01	25					
		<i>protid</i>	78.27	81.05	72.28	2.33	25					
		<i>dnaid</i>	84.75	86.80	78.55	2.13	25					
		<i>kimura</i>	0.25	0.30	0.23	0.02	25					
		<i>logdet</i>	0.48	0.91	0.39	0.09	25					
		<i>index</i>	24.36	26.14	18.54	1.47	25					
	STALKER	<i>ks</i>	0.68	3.34	0.32	0.94	803	0.00	0.00	0.00	0.00	15
		<i>ka</i>	0.05	0.17	0.00	0.08	803	0.01	0.01	0.00	0.01	15
		<i>kaks</i>	0.15	0.54	0.00	0.24	803	177.60	380.00	52.00	143.99	15

		<i>protid</i>	74.85	80.47	70.31	2.88	803	76.56	76.56	76.56	0.00	15	
		<i>dnaid</i>	80.89	82.87	72.66	1.81	803	81.49	81.51	81.25	0.07	15	
		<i>kimura</i>	0.30	0.37	0.20	0.03	803	0.29	0.29	0.29	0.00	15	
		<i>logdet</i>	0.27	0.54	0.17	0.06	803	0.41	0.42	0.41	0.00	15	
		<i>index</i>	32.97	41.83	10.27	8.48	803	46.15	46.28	45.89	0.10	15	
	TABOR	<i>ks</i>	2.30	9.69	0.36	2.06	234						
		<i>ka</i>	0.12	0.29	0.00	0.03	234						
		<i>kaks</i>	0.09	0.81	0.00	0.07	234						
		<i>protid</i>	73.33	81.13	70.00	1.94	234						
		<i>dnaid</i>	78.15	82.70	74.33	1.15	234						
		<i>kimura</i>	0.33	0.46	0.22	0.03	234						
		<i>logdet</i>	0.44	0.84	0.23	0.05	234						
	TRANSIB	<i>index</i>	15.72	28.31	3.67	5.60	234						
		<i>ks</i>	0.37	0.39	0.32	0.04	3						
		<i>ka</i>	0.10	0.11	0.08	0.02	3						
		<i>kaks</i>	0.27	0.29	0.25	0.02	3						
		<i>protid</i>	80.32	83.81	78.10	3.06	3						
		<i>dnaid</i>	86.24	88.25	85.08	1.75	3						
		<i>kimura</i>	0.23	0.26	0.18	0.04	3						
		<i>logdet</i>	0.17	0.18	0.16	0.01	3						
	TRANSPAC	<i>index</i>	40.51	45.37	37.62	4.24	3						
		<i>ks</i>	1.27	1.35	1.13	0.10	4						
		<i>ka</i>	0.03	0.03	0.02	0.01	4						
		<i>kaks</i>	0.02	0.02	0.02	0.00	4						
		<i>protid</i>	91.04	91.69	90.54	0.59	4						
		<i>dnaid</i>	84.17	85.39	83.19	0.91	4						
		<i>kimura</i>	0.10	0.10	0.09	0.01	4						
		<i>logdet</i>	0.23	0.24	0.20	0.02	4						
	Melanogaster _Subgroup	ACCORD	<i>index</i>	28.91	31.89	27.54	2.03	4					
			<i>ks</i>	0.03	0.28	0.01	0.03	890	0.00	0.14	0.00	0.01	1355
			<i>ka</i>	0.01	0.14	0.00	0.02	890	0.01	0.14	0.00	0.01	1355
			<i>kaks</i>	0.19	0.99	0.00	0.30	890	89.43	490.00	1.02	60.34	1355
			<i>protid</i>	96.49	100.00	70.00	5.91	890	95.99	100.00	72.06	4.57	1355
			<i>dnaid</i>	97.37	100.00	71.34	4.80	890	97.73	100.00	77.25	3.81	1355
			<i>kimura</i>	0.03	0.32	0.00	0.06	890	0.04	0.33	0.00	0.04	1355
<i>logdet</i>			0.05	0.45	0.00	0.08	890	0.04	0.72	0.00	0.06	1355	
AURORA		<i>index</i>	88.40	99.38	39.87	15.05	890	89.66	99.73	39.62	10.68	1355	
		<i>ks</i>	0.08	0.40	0.01	0.09	398	0.02	0.20	0.00	0.04	322	
		<i>ka</i>	0.02	0.19	0.00	0.04	398	0.07	0.27	0.01	0.05	322	
		<i>kaks</i>	0.14	0.98	0.00	0.26	398	64.71	108.00	1.02	45.06	322	
		<i>protid</i>	83.62	100.00	70.08	8.15	398	84.05	98.44	70.08	7.41	322	
		<i>dnaid</i>	88.35	100.00	72.04	7.39	398	89.77	99.48	72.70	5.91	322	
		<i>kimura</i>	0.16	0.39	0.00	0.10	398	0.17	0.43	0.02	0.09	322	
	<i>logdet</i>	0.19	0.58	0.00	0.11	398	0.21	1.11	0.01	0.17	322		
BAGGINS	<i>index</i>	60.69	99.02	33.09	15.20	398	60.20	94.18	26.57	13.29	322		
	<i>ks</i>	0.30	9.33	0.01	0.64	2242	0.02	0.61	0.00	0.05	1088		
	<i>ka</i>	0.03	0.35	0.00	0.05	2242	0.09	1.77	0.00	0.09	1088		
	<i>kaks</i>	0.18	0.99	0.00	0.25	2242	77.50	460.00	1.01	48.59	1088		
		<i>protid</i>	82.06	100.00	70.00	6.65	2242	80.65	97.09	70.00	6.35	1088	

		<i>dnaid</i>	89.55	99.51	73.08	4.46	2242	89.42	98.38	77.13	4.56	1088
		<i>kimura</i>	0.20	0.43	0.01	0.08	2242	0.22	0.47	0.03	0.08	1088
		<i>logdet</i>	0.24	1.56	0.01	0.18	2242	0.26	0.94	0.02	0.19	1088
		<i>index</i>	50.94	96.46	4.95	16.13	2242	54.26	90.52	16.67	12.61	1088
	BATUMI	<i>ks</i>	0.08	0.27	0.02	0.06	59	0.03	0.07	0.02	0.01	142
		<i>ka</i>	0.03	0.22	0.00	0.04	59	0.05	0.10	0.03	0.01	142
		<i>kaks</i>	0.45	0.93	0.00	0.34	59	1.57	3.02	1.02	0.37	142
		<i>protid</i>	86.34	97.87	71.43	5.87	59	85.92	92.23	70.69	5.39	142
		<i>dnaid</i>	88.68	98.58	72.67	6.24	59	92.59	96.53	73.79	4.55	142
		<i>kimura</i>	0.10	0.23	0.01	0.06	59	0.14	0.33	0.05	0.05	142
		<i>logdet</i>	0.21	0.46	0.04	0.11	59	0.23	0.65	0.04	0.11	142
		<i>index</i>	60.99	83.76	39.84	8.70	59	60.50	79.66	37.88	7.47	142
	BEL	<i>ks</i>	0.05	0.35	0.01	0.04	1765	0.02	0.34	0.00	0.05	1996
		<i>ka</i>	0.02	0.35	0.00	0.03	1765	0.04	0.38	0.00	0.06	1996
		<i>kaks</i>	0.42	1.00	0.00	0.29	1765	60.53	480.00	1.00	53.72	1996
		<i>protid</i>	94.86	100.00	70.54	5.77	1765	93.03	100.00	70.54	5.81	1996
		<i>dnaid</i>	96.44	100.00	74.65	4.34	1765	96.00	100.00	74.19	4.73	1996
		<i>kimura</i>	0.06	2.72	0.00	0.09	1765	0.07	0.37	0.00	0.06	1996
		<i>logdet</i>	0.06	0.94	0.00	0.09	1765	0.06	0.57	0.00	0.07	1996
		<i>index</i>	82.15	99.40	13.01	13.32	1765	81.21	99.68	29.43	13.83	1996
	BICA	<i>ks</i>	0.04	0.35	0.01	0.05	177	0.02	0.14	0.00	0.03	188
		<i>ka</i>	0.02	0.20	0.00	0.03	177	0.03	0.15	0.00	0.03	188
		<i>kaks</i>	0.43	0.95	0.00	0.30	177	64.81	490.00	1.01	86.09	188
		<i>protid</i>	92.49	99.32	70.37	6.82	177	92.12	99.09	70.63	6.67	188
		<i>dnaid</i>	94.39	99.54	76.54	5.52	177	94.52	99.70	73.33	6.05	188
		<i>kimura</i>	0.07	0.36	0.01	0.08	177	0.06	0.32	0.01	0.06	188
		<i>logdet</i>	0.13	0.64	0.00	0.13	177	0.08	0.33	0.00	0.08	188
		<i>index</i>	76.36	96.38	27.34	15.54	177	80.35	97.78	46.57	13.66	188
	BLASTOPIA	<i>ks</i>	0.03	0.22	0.01	0.05	994	0.00	0.22	0.00	0.02	1482
		<i>ka</i>	0.02	0.22	0.00	0.05	994	0.01	0.52	0.00	0.03	1482
		<i>kaks</i>	0.15	0.99	0.00	0.32	994	126.9 4	490.00	1.03	146.43	1482
		<i>protid</i>	97.80	100.00	72.58	6.15	994	98.76	100.00	72.58	2.36	1482
		<i>dnaid</i>	98.28	100.00	78.76	4.82	994	99.28	100.00	78.76	1.81	1482
		<i>kimura</i>	0.03	0.34	0.00	0.08	994	0.01	0.34	0.00	0.02	1482
		<i>logdet</i>	0.04	0.30	0.00	0.07	994	0.02	0.58	0.00	0.05	1482
		<i>index</i>	90.23	99.40	32.96	16.30	994	95.67	99.78	32.88	8.87	1482
	BLOOD	<i>ks</i>	0.03	0.49	0.00	0.05	2144	0.01	0.17	0.00	0.01	1848
		<i>ka</i>	0.01	0.45	0.00	0.02	2144	0.03	2.79	0.00	0.08	1848
		<i>kaks</i>	0.35	0.98	0.00	0.34	2144	73.46	490.00	1.00	97.35	1848
		<i>protid</i>	90.50	100.00	70.08	10.27	2144	96.03	100.00	70.87	3.80	1848
		<i>dnaid</i>	93.28	100.00	72.18	8.06	2144	98.03	100.00	77.35	2.96	1848
		<i>kimura</i>	0.11	0.44	0.00	0.13	2144	0.04	0.28	0.00	0.03	1848
		<i>logdet</i>	0.07	1.64	0.00	0.10	2144	0.05	0.72	0.00	0.09	1848
		<i>index</i>	77.37	99.13	33.16	17.88	2144	87.73	99.79	16.10	11.06	1848
	BS	<i>ks</i>	0.07	0.09	0.03	0.03	3					
		<i>ka</i>	0.04	0.06	0.01	0.03	3					
		<i>kaks</i>	0.61	0.69	0.44	0.14	3					
		<i>protid</i>	92.34	97.67	89.60	4.62	3					
		<i>dnaid</i>	95.35	98.71	93.60	2.91	3					
		<i>kimura</i>	0.08	0.11	0.02	0.05	3					
		<i>logdet</i>	0.06	0.10	0.01	0.05	3					

	<i>index</i>	75.13	91.41	66.12	14.13	3					
BS2	<i>ks</i>	0.47	9.49	0.00	1.08	2705	0.01	1.52	0.00	0.07	553
	<i>ka</i>	0.01	0.36	0.00	0.05	2705	0.07	4.43	0.00	0.33	553
	<i>kaks</i>	0.03	1.00	0.00	0.12	2705	168.23	490.00	1.07	151.84	553
	<i>protid</i>	87.48	100.00	70.25	8.25	2705	90.74	99.07	70.89	6.96	553
	<i>dnaid</i>	92.92	100.00	70.80	5.43	2705	94.70	99.69	74.70	5.06	553
	<i>kimura</i>	0.13	0.43	0.00	0.10	2705	0.09	0.36	0.01	0.08	553
	<i>logdet</i>	0.14	2.30	0.00	0.17	2705	0.14	2.30	0.00	0.20	553
	<i>index</i>	62.40	99.49	4.77	25.95	2705	74.31	97.82	10.00	17.37	553
BS3	<i>ks</i>	1.26	1.50	0.20	0.33	13	0.05	0.24	0.01	0.05	37
	<i>ka</i>	0.18	0.21	0.10	0.03	13	0.09	0.28	0.05	0.05	37
	<i>kaks</i>	0.17	0.50	0.12	0.10	13	2.54	5.51	1.04	1.33	37
	<i>protid</i>	72.17	85.59	70.40	4.09	13	84.66	91.18	72.41	5.51	37
	<i>dnaid</i>	74.01	88.98	71.24	4.54	13	91.78	95.75	78.07	4.20	37
	<i>kimura</i>	0.35	0.38	0.16	0.06	13	0.17	0.35	0.09	0.08	37
	<i>logdet</i>	0.52	0.88	0.24	0.16	13	0.18	0.35	0.04	0.10	37
	<i>index</i>	17.53	48.14	12.35	9.29	13	59.30	73.88	32.53	11.23	37
BURDOCK	<i>ks</i>	0.11	0.82	0.01	0.16	120	0.01	0.08	0.00	0.02	319
	<i>ka</i>	0.05	0.26	0.00	0.07	120	0.05	0.18	0.01	0.03	319
	<i>kaks</i>	0.34	0.95	0.00	0.31	120	65.96	145.00	1.01	53.79	319
	<i>protid</i>	90.45	100.00	70.59	9.34	120	92.04	99.16	83.33	3.20	319
	<i>dnaid</i>	93.77	99.74	78.15	6.58	120	95.88	99.44	90.28	2.18	319
	<i>kimura</i>	0.09	0.39	0.00	0.10	120	0.07	0.17	0.01	0.03	319
	<i>logdet</i>	0.13	1.13	0.00	0.18	120	0.17	1.23	0.01	0.24	319
	<i>index</i>	71.25	97.60	22.33	22.66	120	73.64	93.74	38.15	11.18	319
CHIMPO	<i>ks</i>	0.08	2.77	0.01	0.28	588	0.01	0.11	0.00	0.02	767
	<i>ka</i>	0.04	1.25	0.00	0.13	588	0.13	1.44	0.00	0.30	767
	<i>kaks</i>	0.45	0.98	0.00	0.30	588	142.91	490.00	1.00	132.31	767
	<i>protid</i>	89.71	100.00	70.34	7.28	588	87.37	99.15	70.27	8.73	767
	<i>dnaid</i>	92.79	99.72	74.18	6.02	588	91.72	99.72	73.47	7.12	767
	<i>kimura</i>	0.10	0.36	0.00	0.08	588	0.12	0.49	0.01	0.10	767
	<i>logdet</i>	0.15	0.44	0.00	0.09	588	0.17	1.08	0.00	0.13	767
	<i>index</i>	68.96	97.98	10.26	15.82	588	66.52	97.53	18.24	19.85	767
CHOUTO	<i>ks</i>	0.06	0.25	0.01	0.07	57	0.02	0.13	0.00	0.03	102
	<i>ka</i>	0.04	0.15	0.00	0.05	57	0.04	0.15	0.00	0.04	102
	<i>kaks</i>	0.49	0.99	0.00	0.33	57	61.67	400.00	1.01	83.50	102
	<i>protid</i>	93.14	100.00	73.33	7.98	57	92.60	100.00	72.27	6.11	102
	<i>dnaid</i>	95.25	100.00	82.49	5.46	57	96.21	100.00	82.91	3.90	102
	<i>kimura</i>	0.08	0.33	0.00	0.10	57	0.08	0.30	0.00	0.07	102
	<i>logdet</i>	0.14	0.83	0.00	0.21	57	0.08	0.44	0.00	0.10	102
	<i>index</i>	76.55	98.88	30.74	23.56	57	79.96	99.69	46.38	16.82	102
CIRC	<i>ks</i>	0.08	0.59	0.01	0.08	344	0.03	0.19	0.00	0.03	553
	<i>ka</i>	0.04	0.27	0.00	0.04	344	0.05	0.30	0.01	0.04	553
	<i>kaks</i>	0.57	1.00	0.00	0.25	344	20.85	133.00	1.01	39.33	553
	<i>protid</i>	91.52	100.00	71.54	6.88	344	88.58	97.00	70.00	6.05	553
	<i>dnaid</i>	94.21	99.00	74.04	5.16	344	93.60	98.97	75.18	5.32	553
	<i>kimura</i>	0.08	0.36	0.01	0.07	344	0.10	0.34	0.03	0.05	553
	<i>logdet</i>	0.09	0.79	0.01	0.10	344	0.10	0.77	0.01	0.08	553
	<i>index</i>	72.90	94.13	32.81	15.26	344	71.90	92.80	32.78	10.62	553
CIRCE	<i>ks</i>	0.24	0.67	0.04	0.15	38	0.13	0.29	0.00	0.10	25

		<i>ka</i>	0.13	0.36	0.02	0.08	38	0.19	0.33	0.08	0.10	25
		<i>kaks</i>	0.55	0.92	0.33	0.17	38	13.09	97.80	1.00	31.93	25
		<i>protid</i>	82.25	97.12	70.65	6.44	38	81.60	92.52	73.28	6.15	25
		<i>dnaid</i>	86.35	97.44	75.58	5.38	38	87.51	94.83	81.61	4.80	25
		<i>kimura</i>	0.19	0.33	0.03	0.08	38	0.19	0.28	0.03	0.07	25
		<i>logdet</i>	0.19	0.61	0.03	0.16	38	0.26	0.60	0.09	0.14	25
		<i>index</i>	47.57	77.95	25.90	14.37	38	46.65	71.71	27.52	12.70	25
	COPIA	<i>ks</i>	0.11	0.81	0.00	0.13	2246	0.01	0.21	0.00	0.02	1296
		<i>ka</i>	0.03	0.21	0.00	0.03	2246	0.03	0.21	0.00	0.03	1296
		<i>kaks</i>	0.38	1.00	0.00	0.26	2246	92.64	490.00	1.01	80.87	1296
		<i>protid</i>	91.47	100.00	70.14	7.24	2246	91.82	100.00	70.34	7.98	1296
		<i>dnaid</i>	93.91	100.00	72.35	5.95	2246	94.66	100.00	71.82	7.19	1296
		<i>kimura</i>	0.08	0.40	0.00	0.08	2246	0.07	0.36	0.01	0.06	1296
		<i>logdet</i>	0.09	0.82	0.00	0.11	2246	0.08	0.85	0.00	0.11	1296
		<i>index</i>	72.36	97.96	27.76	17.94	2246	80.65	98.34	35.98	14.73	1296
	CR1	<i>ks</i>	1.02	9.19	0.00	1.51	1892	0.02	0.47	0.00	0.07	1095
		<i>ka</i>	0.09	0.47	0.00	0.07	1892	0.07	0.50	0.00	0.09	1095
		<i>kaks</i>	0.21	0.99	0.00	0.26	1892	81.60	490.00	1.01	80.89	1095
		<i>protid</i>	80.98	99.06	70.00	6.55	1892	85.48	100.00	70.00	7.13	1095
		<i>dnaid</i>	85.97	99.37	72.42	5.29	1892	90.17	99.06	72.17	5.66	1095
		<i>kimura</i>	0.21	0.48	0.01	0.08	1892	0.16	2.75	0.01	0.14	1095
		<i>logdet</i>	0.31	1.68	0.01	0.20	1892	0.26	1.19	0.01	0.19	1095
		<i>index</i>	35.54	96.21	4.33	18.03	1892	59.03	93.92	14.07	11.97	1095
	DEC1	<i>ks</i>	0.39	0.57	0.22	0.12	6					
		<i>ka</i>	0.10	0.13	0.07	0.02	6					
		<i>kaks</i>	0.25	0.34	0.18	0.06	6					
		<i>protid</i>	84.20	89.32	81.09	3.13	6					
		<i>dnaid</i>	86.13	89.87	82.20	2.72	6					
		<i>kimura</i>	0.17	0.21	0.12	0.03	6					
		<i>logdet</i>	0.24	0.35	0.15	0.07	6					
		<i>index</i>	41.43	51.01	30.91	7.43	6					
	DIVER	<i>ks</i>	0.07	0.50	0.00	0.06	1190	0.01	0.17	0.00	0.02	1765
		<i>ka</i>	0.02	0.30	0.00	0.03	1190	0.07	0.35	0.00	0.05	1765
		<i>kaks</i>	0.27	1.00	0.00	0.29	1190	93.37	490.00	1.00	71.53	1765
		<i>protid</i>	90.38	100.00	70.27	8.75	1190	85.81	100.00	70.09	7.73	1765
		<i>dnaid</i>	93.37	100.00	70.75	6.85	1190	91.60	100.00	71.03	6.07	1765
		<i>kimura</i>	0.10	0.44	0.00	0.09	1190	0.15	0.41	0.00	0.09	1765
		<i>logdet</i>	0.13	0.95	0.00	0.16	1190	0.18	1.45	0.00	0.18	1765
		<i>index</i>	72.06	99.50	28.34	19.29	1190	64.09	99.69	33.72	15.33	1765
	DIVER-BEL	<i>ks</i>	0.06	0.40	0.00	0.04	2664	0.02	0.19	0.00	0.02	3657
		<i>ka</i>	0.03	0.23	0.00	0.02	2664	0.04	0.24	0.00	0.02	3657
		<i>kaks</i>	0.64	1.00	0.00	0.23	2664	33.93	490.00	1.01	50.41	3657
		<i>protid</i>	91.67	100.00	70.59	4.48	2664	91.56	100.00	71.00	4.29	3657
		<i>dnaid</i>	95.21	100.00	75.08	3.18	2664	96.08	100.00	76.61	2.70	3657
		<i>kimura</i>	0.08	0.34	0.00	0.05	2664	0.09	0.36	0.00	0.05	3657
		<i>logdet</i>	0.07	0.48	0.00	0.08	2664	0.05	0.44	0.00	0.06	3657
		<i>index</i>	75.59	98.42	31.78	11.09	2664	78.66	99.65	35.38	10.16	3657
	DM1731	<i>ks</i>	0.06	0.61	0.00	0.08	1395	0.00	0.52	0.00	0.02	1664
		<i>ka</i>	0.02	0.52	0.00	0.07	1395	0.03	0.71	0.00	0.04	1664
		<i>kaks</i>	0.21	1.00	0.00	0.30	1395	139.06	490.00	1.00	124.82	1664
		<i>protid</i>	90.30	100.00	70.00	9.00	1395	90.97	100.00	70.00	8.64	1664

		<i>dnaid</i>	92.14	100.00	70.75	8.21	1395	93.48	100.00	71.26	7.88	1664
		<i>kimura</i>	0.06	0.33	0.00	0.06	1395	0.07	0.43	0.00	0.07	1664
		<i>logdet</i>	0.11	0.68	0.00	0.11	1395	0.13	1.10	0.00	0.13	1664
		<i>index</i>	74.99	99.16	27.51	16.15	1395	77.20	99.52	28.25	15.71	1664
	DM2	<i>ks</i>	0.04	0.26	0.01	0.03	1128	0.00	0.10	0.00	0.01	765
		<i>ka</i>	0.01	0.15	0.00	0.02	1128	0.01	0.11	0.00	0.02	765
		<i>kaks</i>	0.35	1.00	0.00	0.26	1128	115.54	490.00	1.00	127.76	765
		<i>protid</i>	95.99	100.00	70.63	4.96	1128	96.31	100.00	79.81	4.73	765
		<i>dnaid</i>	96.62	100.00	71.16	4.66	1128	97.70	100.00	82.37	4.05	765
		<i>kimura</i>	0.03	0.21	0.00	0.03	1128	0.04	0.20	0.00	0.04	765
		<i>logdet</i>	0.06	0.57	0.00	0.08	1128	0.04	0.27	0.00	0.07	765
		<i>index</i>	85.73	99.50	41.01	11.38	1128	89.56	99.71	53.09	10.52	765
	DM412	<i>ks</i>	0.27	0.90	0.01	0.17	5843	0.05	0.28	0.00	0.05	1817
		<i>ka</i>	0.07	0.28	0.00	0.05	5843	0.09	0.31	0.00	0.06	1817
		<i>kaks</i>	0.32	1.00	0.00	0.19	5843	40.61	490.00	1.01	80.29	1817
		<i>protid</i>	87.16	100.00	70.10	5.94	5843	90.55	100.00	71.76	5.84	1817
		<i>dnaid</i>	89.47	100.00	71.59	5.60	5843	94.08	100.00	73.50	4.35	1817
		<i>kimura</i>	0.14	0.44	0.00	0.07	5843	0.10	0.43	0.00	0.07	1817
		<i>logdet</i>	0.15	1.10	0.00	0.09	5843	0.10	0.49	0.00	0.09	1817
		<i>index</i>	53.68	99.35	21.89	16.14	5843	70.00	99.80	31.14	15.93	1817
	DMRT	<i>ks</i>	0.19	8.44	0.00	0.57	14490	0.03	3.68	0.00	0.16	10728
		<i>ka</i>	0.06	2.15	0.00	0.08	14490	0.08	5.53	0.00	0.22	10728
		<i>kaks</i>	0.41	1.00	0.00	0.28	14490	85.07	490.00	1.00	123.56	10728
		<i>protid</i>	87.00	100.00	70.00	7.30	14490	86.18	100.00	70.00	6.95	10728
		<i>dnaid</i>	90.54	100.00	70.90	5.95	14490	91.01	100.00	71.52	5.74	10728
		<i>kimura</i>	0.13	2.19	0.00	0.09	14490	0.13	2.59	0.00	0.09	10728
		<i>logdet</i>	0.20	2.21	0.00	0.17	14490	0.20	2.07	0.00	0.16	10728
		<i>index</i>	58.78	99.24	4.52	15.49	14490	62.56	99.46	3.36	13.39	10728
	DNAREP1	<i>ks</i>	0.48	6.36	0.01	1.09	354	0.00	0.04	0.00	0.00	817
		<i>ka</i>	0.03	0.60	0.00	0.09	354	0.45	2.47	0.07	0.31	817
		<i>kaks</i>	0.02	0.83	0.00	0.07	354	99.85	102.67	1.59	3.51	817
		<i>protid</i>	73.99	94.50	70.00	3.06	354	74.50	85.00	70.00	2.63	817
		<i>dnaid</i>	84.91	97.25	77.00	2.07	354	83.88	94.17	78.18	1.74	817
		<i>kimura</i>	0.29	0.82	0.06	0.06	354	0.34	0.47	0.17	0.05	817
		<i>logdet</i>	0.44	1.36	0.03	0.14	354	0.29	1.14	0.18	0.10	817
		<i>index</i>	39.36	87.06	6.32	13.28	354	34.48	56.90	12.24	5.62	817
	DOC2	<i>ks</i>	0.06	0.50	0.01	0.05	288	0.03	0.11	0.00	0.02	291
		<i>ka</i>	0.04	0.20	0.00	0.03	288	0.04	0.15	0.01	0.02	291
		<i>kaks</i>	0.59	1.00	0.00	0.24	288	13.10	112.00	1.00	31.49	291
		<i>protid</i>	85.68	98.92	70.00	7.39	288	85.28	98.92	70.18	7.62	291
		<i>dnaid</i>	89.89	99.64	73.04	6.40	288	90.79	99.64	73.62	6.53	291
		<i>kimura</i>	0.13	0.43	0.01	0.08	288	0.13	0.44	0.01	0.08	291
		<i>logdet</i>	0.16	0.45	0.02	0.09	288	0.16	0.45	0.02	0.11	291
		<i>index</i>	63.55	86.43	29.81	11.64	288	65.22	89.58	35.27	12.44	291

DOC3	<i>ks</i>	0.25	2.67	0.01	0.31	674	0.05	0.32	0.00	0.06	308
	<i>ka</i>	0.08	0.40	0.00	0.06	674	0.10	0.46	0.00	0.09	308
	<i>kaks</i>	0.42	1.00	0.00	0.29	674	43.41	480.00	1.00	80.35	308
	<i>protid</i>	80.60	97.41	70.00	6.06	674	79.79	98.25	70.00	6.94	308
	<i>dnaid</i>	87.11	98.85	75.79	4.89	674	87.66	99.42	76.11	5.34	308
	<i>kimura</i>	0.21	3.11	0.03	0.13	674	0.22	0.41	0.02	0.08	308
	<i>logdet</i>	0.34	1.43	0.02	0.22	674	0.31	1.30	0.01	0.20	308
	<i>index</i>	44.09	91.99	7.52	12.73	674	50.19	94.94	25.40	13.85	308
FROGGER	<i>ks</i>	0.26	0.26	0.26		1	0.01	0.02	0.00	0.01	5
	<i>ka</i>	0.14	0.14	0.14		1	0.02	0.03	0.01	0.01	5
	<i>kaks</i>	0.54	0.54	0.54		1	19.80	93.50	1.26	41.20	5
	<i>protid</i>	80.00	80.00	80.00		1	94.45	97.03	92.56	1.99	5
	<i>dnaid</i>	85.28	85.28	85.28		1	97.75	98.68	96.97	0.85	5
	<i>kimura</i>	0.21	0.21	0.21		1	0.06	0.08	0.03	0.02	5
	<i>logdet</i>	0.29	0.29	0.29		1	0.02	0.03	0.01	0.01	5
	<i>index</i>	39.11	39.11	39.11		1	85.22	90.49	80.59	4.94	5
FW	<i>ks</i>	0.24	9.16	0.00	0.35	49337	0.00	0.24	0.00	0.02	11455
	<i>ka</i>	0.01	1.68	0.00	0.03	49337	0.02	0.29	0.00	0.03	11455
	<i>kaks</i>	0.12	1.00	0.00	0.22	49337	135.40	490.00	1.00	126.78	11455
	<i>protid</i>	92.24	100.00	70.00	6.84	49337	92.04	100.00	70.00	7.02	11455
	<i>dnaid</i>	94.20	100.00	70.03	5.98	49337	94.34	100.00	70.31	6.02	11455
	<i>kimura</i>	0.07	2.16	0.00	0.06	49337	0.07	0.49	0.00	0.06	11455
	<i>logdet</i>	0.13	1.66	0.00	0.12	49337	0.12	0.83	0.00	0.09	11455
	<i>index</i>	66.86	99.09	3.64	16.57	49337	77.54	99.64	30.95	11.90	11455
FW2	<i>ks</i>	0.13	0.35	0.02	0.07	50	0.07	0.12	0.00	0.04	15
	<i>ka</i>	0.07	0.20	0.01	0.04	50	0.09	0.16	0.00	0.05	15
	<i>kaks</i>	0.57	0.98	0.25	0.20	50	31.35	450.00	1.03	115.82	15
	<i>protid</i>	84.86	96.75	71.70	7.45	50	85.32	96.52	71.67	7.13	15
	<i>dnaid</i>	90.99	98.46	80.44	4.87	50	92.22	98.55	83.61	4.17	15
	<i>kimura</i>	0.17	0.36	0.03	0.09	50	0.16	0.33	0.04	0.08	15
	<i>logdet</i>	0.38	0.81	0.02	0.21	50	0.16	0.58	0.02	0.17	15
	<i>index</i>	50.00	89.56	30.38	15.95	50	60.81	92.41	35.05	17.06	15
G5	<i>ks</i>	0.13	8.01	0.01	0.37	1986	0.03	0.59	0.00	0.03	2749
	<i>ka</i>	0.06	2.90	0.00	0.08	1986	0.06	5.03	0.00	0.11	2749
	<i>kaks</i>	0.53	1.00	0.00	0.26	1986	33.16	480.00	1.00	48.92	2749
	<i>protid</i>	87.45	100.00	70.00	7.49	1986	88.43	99.17	70.00	7.04	2749
	<i>dnaid</i>	92.02	99.28	71.66	5.50	1986	93.74	99.72	75.28	4.87	2749
	<i>kimura</i>	0.13	0.51	0.01	0.09	1986	0.12	0.39	0.01	0.08	2749
	<i>logdet</i>	0.15	0.99	0.01	0.14	1986	0.11	1.04	0.00	0.12	2749
	<i>index</i>	63.39	95.00	3.65	17.62	1986	70.55	97.77	5.61	15.58	2749
GATE	<i>ks</i>	0.07	0.09	0.06	0.01	7	0.01	0.06	0.00	0.02	15
	<i>ka</i>	0.02	0.05	0.00	0.02	7	0.03	0.12	0.00	0.04	15
	<i>kaks</i>	0.23	0.64	0.00	0.29	7	137.79	470.00	1.44	157.55	15
	<i>protid</i>	75.97	80.00	71.43	2.92	7	77.20	85.33	70.00	5.01	15

		<i>dnaid</i>	82.79	85.49	80.95	1.80	7	84.72	88.89	76.33	3.56	15
		<i>kimura</i>	0.24	0.28	0.22	0.02	7	0.25	0.37	0.15	0.07	15
		<i>logdet</i>	0.34	0.38	0.26	0.05	7	0.50	0.64	0.14	0.14	15
		<i>index</i>	46.46	48.95	42.84	2.76	7	44.94	50.55	37.00	4.13	15
Genome		<i>ks</i>	0.23	2.33	0.00	0.12	25711	0.00	0.01	0.00	0.01	6
		<i>ka</i>	0.02	0.77	0.00	0.03	25711	0.01	0.01	0.00	0.01	6
		<i>kaks</i>	0.08	0.91	0.00	0.08	25711	4.48	10.00	1.41	4.28	6
		<i>protid</i>	95.78	100.00	39.38	4.37	25711	98.59	100.00	97.02	1.10	6
		<i>dnaid</i>	93.97	100.00	51.18	3.50	25711	99.37	100.00	98.81	0.44	6
		<i>kimura</i>	0.04	0.91	0.00	0.05	25711	0.01	0.03	0.00	0.01	6
		<i>logdet</i>	0.09	1.28	0.00	0.09	25711	0.02	0.06	0.00	0.02	6
		<i>index</i>	68.59	99.59	4.97	12.36	25711	94.81	99.74	88.83	3.89	6
GTWIN		<i>ks</i>	0.28	1.04	0.00	0.16	5669	0.02	0.16	0.00	0.04	415
		<i>ka</i>	0.06	0.25	0.00	0.04	5669	0.05	0.18	0.00	0.05	415
		<i>kaks</i>	0.24	1.00	0.00	0.17	5669	69.04	480.00	1.03	85.19	415
		<i>protid</i>	86.06	100.00	70.21	6.72	5669	88.05	100.00	70.00	8.60	415
		<i>dnaid</i>	88.19	100.00	72.60	5.40	5669	92.07	100.00	76.00	6.53	415
		<i>kimura</i>	0.15	0.47	0.00	0.07	5669	0.12	0.36	0.01	0.09	415
		<i>logdet</i>	0.20	1.27	0.00	0.14	5669	0.21	0.96	0.00	0.20	415
		<i>index</i>	51.54	99.44	16.73	16.01	5669	67.48	97.62	31.41	20.11	415
GYPSY		<i>ks</i>	0.26	8.75	0.00	0.46	7450	0.04	0.31	0.00	0.04	3841
		<i>ka</i>	0.06	0.94	0.00	0.06	7450	0.08	0.35	0.00	0.06	3841
		<i>kaks</i>	0.35	1.00	0.00	0.25	7450	35.26	490.00	1.00	60.90	3841
		<i>protid</i>	85.73	100.00	70.00	7.76	7450	84.21	100.00	70.00	8.40	3841
		<i>dnaid</i>	89.23	100.00	70.06	5.86	7450	91.62	100.00	72.41	5.27	3841
		<i>kimura</i>	0.15	2.33	0.00	0.10	7450	0.17	0.50	0.00	0.10	3841
		<i>logdet</i>	0.22	1.55	0.00	0.17	7450	0.16	2.13	0.00	0.14	3841
		<i>index</i>	53.26	99.00	4.22	17.84	7450	61.69	99.70	21.47	17.08	3841
HAT		<i>ks</i>	0.11	0.14	0.08	0.04	2	0.01	0.03	0.00	0.02	2
		<i>ka</i>	0.06	0.06	0.05	0.01	2	0.03	0.04	0.01	0.02	2
		<i>kaks</i>	0.55	0.76	0.34	0.30	2	29.08	57.00	1.16	39.48	2
		<i>protid</i>	80.38	85.51	75.25	7.25	2	89.23	91.03	87.43	2.55	2
		<i>dnaid</i>	88.95	92.75	85.15	5.37	2	92.36	93.10	91.62	1.05	2
		<i>kimura</i>	0.21	0.26	0.16	0.07	2	0.11	0.13	0.08	0.04	2
		<i>logdet</i>	0.46	0.73	0.20	0.37	2	0.20	0.27	0.12	0.11	2
		<i>index</i>	45.20	55.18	35.21	14.12	2	67.36	67.74	66.98	0.54	2
HELENA		<i>ks</i>	0.09	0.35	0.02	0.09	27	0.03	0.11	0.00	0.04	20
		<i>ka</i>	0.06	0.27	0.00	0.07	27	0.05	0.14	0.01	0.04	20
		<i>kaks</i>	0.60	0.96	0.07	0.27	27	20.44	114.00	1.00	38.95	20
		<i>protid</i>	90.32	99.19	70.34	8.82	27	91.78	97.50	72.65	5.77	20
		<i>dnaid</i>	93.83	98.89	78.81	5.96	27	95.60	99.17	80.91	4.38	20
		<i>kimura</i>	0.11	0.39	0.01	0.10	27	0.09	0.34	0.03	0.07	20
		<i>logdet</i>	0.08	0.24	0.01	0.07	27	0.10	0.54	0.02	0.12	20
		<i>index</i>	71.49	91.28	30.23	19.58	27	75.24	88.97	39.78	14.45	20
HELITRON		<i>ks</i>	0.17	0.34	0.02	0.13	29	0.02	0.08	0.00	0.03	11
		<i>ka</i>	0.08	0.20	0.00	0.07	29	0.03	0.09	0.01	0.02	11
		<i>kaks</i>	0.52	0.92	0.22	0.23	29	29.19	97.00	1.03	39.61	11
		<i>protid</i>	84.83	98.36	71.43	8.95	29	90.43	98.36	71.83	8.38	11
		<i>dnaid</i>	90.07	98.91	82.07	5.86	29	93.67	98.91	79.11	6.47	11
		<i>kimura</i>	0.16	0.36	0.02	0.12	29	0.09	0.26	0.02	0.08	11
		<i>logdet</i>	0.15	0.42	0.01	0.11	29	0.09	0.31	0.01	0.10	11
		<i>index</i>	58.02	93.63	29.53	22.31	29	76.11	95.43	47.70	17.69	11

HETA	<i>ks</i>	0.19	0.29	0.06	0.05	61	0.07	0.07	0.07		1
	<i>ka</i>	0.07	0.10	0.01	0.02	61	0.07	0.07	0.07		1
	<i>kaks</i>	0.39	0.85	0.21	0.11	61	1.08	1.08	1.08		1
	<i>protid</i>	86.80	98.68	79.29	4.08	61	85.71	85.71	85.71		1
	<i>dnaid</i>	89.62	98.45	82.62	3.47	61	91.53	91.53	91.53		1
	<i>kimura</i>	0.14	0.21	0.01	0.04	61	0.16	0.16	0.16		1
	<i>logdet</i>	0.16	0.45	0.02	0.09	61	0.19	0.19	0.19		1
	<i>index</i>	54.86	89.02	37.20	9.49	61	57.57	57.57	57.57		1
HMSBEAGLE	<i>ks</i>	0.11	8.93	0.01	0.44	410	0.01	0.19	0.00	0.02	741
	<i>ka</i>	0.05	3.37	0.00	0.17	410	0.05	0.35	0.00	0.06	741
	<i>kaks</i>	0.44	0.99	0.00	0.29	410	138.81	490.00	1.02	133.63	741
	<i>protid</i>	87.77	99.24	70.07	9.31	410	86.63	100.00	70.18	8.82	741
	<i>dnaid</i>	92.94	99.40	74.12	5.13	410	93.04	100.00	74.32	5.17	741
	<i>kimura</i>	0.14	0.43	0.01	0.11	410	0.15	1.80	0.00	0.13	741
	<i>logdet</i>	0.14	0.91	0.01	0.12	410	0.19	1.73	0.00	0.17	741
	<i>index</i>	66.95	96.27	3.16	19.21	410	66.87	99.78	18.34	18.19	741
HOANA	<i>ks</i>	0.03	0.03	0.03		1	0.02	0.07	0.00	0.04	4
	<i>ka</i>	0.00	0.00	0.00		1	0.05	0.09	0.00	0.04	4
	<i>kaks</i>	0.00	0.00	0.00		1	119.96	290.00	1.17	121.57	4
	<i>protid</i>	95.78	95.78	95.78		1	82.41	96.08	76.09	9.20	4
	<i>dnaid</i>	97.79	97.79	97.79		1	89.48	98.69	85.27	6.23	4
	<i>kimura</i>	0.04	0.04	0.04		1	0.21	0.29	0.04	0.11	4
	<i>logdet</i>	0.21	0.21	0.21		1	0.27	0.41	0.05	0.16	4
	<i>index</i>	74.92	74.92	74.92		1	57.23	88.72	44.37	21.07	4
HOBO	<i>ks</i>	0.05	0.27	0.01	0.05	487	0.02	0.08	0.00	0.02	311
	<i>ka</i>	0.03	0.09	0.00	0.02	487	0.05	0.09	0.00	0.02	311
	<i>kaks</i>	0.59	0.98	0.00	0.29	487	21.28	480.00	1.01	56.30	311
	<i>protid</i>	86.05	100.00	70.43	7.18	487	82.13	95.83	70.43	6.23	311
	<i>dnaid</i>	90.72	100.00	71.30	5.57	487	88.82	98.26	74.49	6.06	311
	<i>kimura</i>	0.16	0.43	0.00	0.09	487	0.21	0.42	0.04	0.08	311
	<i>logdet</i>	0.20	0.56	0.00	0.11	487	0.16	0.51	0.07	0.09	311
	<i>index</i>	62.23	99.17	35.17	14.77	487	58.99	81.69	39.58	7.87	311
HOBO-LIKE	<i>ks</i>	0.34	0.41	0.29	0.06	3					
	<i>ka</i>	0.09	0.12	0.05	0.04	3					
	<i>kaks</i>	0.28	0.37	0.13	0.13	3					
	<i>protid</i>	82.27	85.95	79.71	3.27	3					
	<i>dnaid</i>	86.72	88.65	85.51	1.69	3					
	<i>kimura</i>	0.21	0.24	0.16	0.04	3					
	<i>logdet</i>	0.33	0.59	0.16	0.23	3					
	<i>index</i>	39.12	45.06	32.58	6.26	3					
HOYAK	<i>ks</i>	0.07	0.12	0.01	0.08	2					
	<i>ka</i>	0.04	0.07	0.01	0.04	2					
	<i>kaks</i>	0.72	0.80	0.64	0.11	2					
	<i>protid</i>	93.30	98.18	88.41	6.91	2					
	<i>dnaid</i>	95.68	99.09	92.27	4.82	2					
	<i>kimura</i>	0.07	0.13	0.02	0.08	2					
	<i>logdet</i>	0.06	0.09	0.02	0.05	2					
	<i>index</i>	77.71	93.24	62.18	21.96	2					
IDEFIX	<i>ks</i>	0.86	7.92	0.03	0.39	502	0.03	0.03	0.03	0.00	2
	<i>ka</i>	0.08	0.29	0.01	0.05	502	0.03	0.03	0.03	0.00	2

		<i>kaks</i>	0.11	0.80	0.03	0.07	502	1.06	1.06	1.05	0.01	2
		<i>protid</i>	83.20	95.24	70.19	4.19	502	95.24	95.24	95.24	0.00	2
		<i>dnaid</i>	84.97	98.10	70.19	3.85	502	98.10	98.10	98.10	0.00	2
		<i>kimura</i>	0.17	0.34	0.05	0.04	502	0.05	0.05	0.05	0.00	2
		<i>logdet</i>	0.25	0.64	0.02	0.14	502	0.02	0.02	0.02	0.00	2
		<i>index</i>	31.83	85.49	6.05	8.15	502	84.78	84.89	84.67	0.16	2
	INVADER	<i>ks</i>	0.05	0.36	0.01	0.03	3314	0.01	0.22	0.00	0.02	3213
		<i>ka</i>	0.02	0.28	0.00	0.02	3314	0.03	0.27	0.00	0.03	3213
		<i>kaks</i>	0.43	1.00	0.00	0.27	3314	81.85	490.00	1.00	75.87	3213
		<i>protid</i>	91.05	100.00	70.00	7.09	3314	89.74	100.00	70.00	7.60	3213
		<i>dnaid</i>	93.78	100.00	71.03	6.18	3314	93.51	100.00	71.28	6.39	3213
		<i>kimura</i>	0.08	0.39	0.00	0.06	3314	0.09	0.40	0.00	0.06	3213
		<i>logdet</i>	0.10	0.64	0.00	0.08	3314	0.14	1.77	0.00	0.12	3213
		<i>index</i>	74.51	99.23	28.01	12.63	3314	73.12	99.72	28.65	14.07	3213
	IVK	<i>ks</i>	0.05	0.34	0.01	0.07	72	0.02	0.12	0.00	0.02	156
		<i>ka</i>	0.02	0.10	0.00	0.03	72	0.04	0.14	0.00	0.04	156
		<i>kaks</i>	0.39	0.91	0.00	0.28	72	81.54	390.00	1.04	126.20	156
		<i>protid</i>	94.21	100.00	73.68	6.93	72	91.17	99.25	75.45	7.68	156
		<i>dnaid</i>	96.11	100.00	80.70	5.03	72	95.14	99.75	80.45	5.24	156
		<i>kimura</i>	0.06	0.30	0.00	0.07	72	0.08	0.26	0.01	0.08	156
		<i>logdet</i>	0.08	0.53	0.00	0.10	72	0.09	0.53	0.00	0.10	156
		<i>index</i>	80.49	99.21	35.16	17.91	72	77.86	98.14	49.24	17.55	156
	JOCKEY	<i>ks</i>	0.21	0.48	0.02	0.10	73	0.14	0.14	0.14		1
		<i>ka</i>	0.09	0.21	0.01	0.03	73	0.15	0.15	0.15		1
		<i>kaks</i>	0.48	0.99	0.19	0.21	73	1.07	1.07	1.07		1
		<i>protid</i>	80.66	98.04	70.71	6.10	73	78.51	78.51	78.51		1
		<i>dnaid</i>	86.92	98.69	76.43	4.50	73	87.05	87.05	87.05		1
		<i>kimura</i>	0.21	0.37	0.02	0.08	73	0.24	0.24	0.24		1
		<i>logdet</i>	0.33	0.75	0.07	0.17	73	0.13	0.13	0.13		1
		<i>index</i>	43.84	85.44	24.22	11.71	73	46.69	46.69	46.69		1
	JOCKEY-LIKE	<i>ks</i>	0.11	0.20	0.03	0.05	66	0.00	0.00	0.00		1
		<i>ka</i>	0.03	0.10	0.00	0.02	66	0.01	0.01	0.01		1
		<i>kaks</i>	0.36	0.96	0.00	0.22	66	74.00	74.00	74.00		1
		<i>protid</i>	80.41	93.20	70.19	6.40	66	85.95	85.95	85.95		1
		<i>dnaid</i>	87.02	96.94	79.82	4.72	66	89.55	89.55	89.55		1
		<i>kimura</i>	0.21	0.58	0.08	0.09	66	0.13	0.13	0.13		1
		<i>logdet</i>	0.37	1.58	0.15	0.19	66	0.29	0.29	0.29		1
		<i>index</i>	47.01	62.56	30.30	7.16	66	61.23	61.23	61.23		1
	LINEJ1	<i>ks</i>	0.04	0.19	0.01	0.03	92	0.01	0.09	0.00	0.01	347
		<i>ka</i>	0.02	0.09	0.00	0.03	92	0.02	0.10	0.00	0.02	347
		<i>kaks</i>	0.41	0.95	0.00	0.20	92	50.76	410.00	1.01	42.75	347
		<i>protid</i>	94.11	100.00	72.18	8.22	92	97.40	100.00	71.43	3.31	347
		<i>dnaid</i>	95.98	100.00	81.95	5.48	92	98.39	100.00	81.95	2.49	347
		<i>kimura</i>	0.06	0.30	0.00	0.09	92	0.02	0.32	0.00	0.03	347
		<i>logdet</i>	0.05	0.41	0.00	0.08	92	0.02	0.37	0.00	0.04	347
		<i>index</i>	84.08	98.69	41.16	18.19	92	92.05	99.72	40.48	9.49	347
	LINER1	<i>ks</i>	1.45	1.45	1.45	0.00	2					
		<i>ka</i>	0.00	0.00	0.00	0.00	2					
		<i>kaks</i>	0.00	0.00	0.00	0.00	2					
		<i>protid</i>	73.42	74.47	72.37	1.48	2					
		<i>dnaid</i>	77.85	80.14	75.57	3.23	2					
		<i>kimura</i>	0.28	0.28	0.27	0.01	2					

		<i>logdet</i>	0.51	0.62	0.40	0.16	2						
		<i>index</i>	17.32	18.01	16.62	0.98	2						
M4DM		<i>ks</i>	0.08	0.19	0.03	0.04	21						
		<i>ka</i>	0.04	0.12	0.01	0.03	21						
		<i>kaks</i>	0.49	0.78	0.24	0.17	21						
		<i>protid</i>	91.47	98.02	81.19	4.60	21						
		<i>dnaid</i>	95.47	98.02	88.78	2.76	21						
		<i>kimura</i>	0.09	0.19	0.02	0.05	21						
		<i>logdet</i>	0.08	0.34	0.02	0.11	21						
		<i>index</i>	73.07	88.56	43.21	13.54	21						
MARINER		<i>ks</i>	0.24	5.39	0.00	0.73	611	0.02	0.11	0.00	0.03	439	
		<i>ka</i>	0.11	3.81	0.00	0.46	611	0.04	0.70	0.00	0.07	439	
		<i>kaks</i>	0.50	1.00	0.00	0.27	611	95.12	490.00	1.00	118.07	439	
		<i>protid</i>	87.18	100.00	70.09	7.08	611	91.44	100.00	70.14	7.99	439	
		<i>dnaid</i>	93.09	100.00	78.45	4.30	611	95.63	100.00	81.36	4.26	439	
		<i>kimura</i>	0.14	0.39	0.00	0.08	611	0.09	0.36	0.00	0.09	439	
		<i>logdet</i>	0.13	1.28	0.00	0.14	611	0.11	1.28	0.00	0.18	439	
		<i>index</i>	64.14	99.21	2.96	20.44	611	77.62	99.74	19.39	19.33	439	
MAX		<i>ks</i>	0.06	6.29	0.00	0.09	10740	0.01	0.36	0.00	0.03	8445	
		<i>ka</i>	0.02	0.25	0.00	0.03	10740	0.03	0.47	0.00	0.03	8445	
		<i>kaks</i>	0.46	1.00	0.00	0.29	10740	63.65	490.00	1.00	77.88	8445	
		<i>protid</i>	91.76	100.00	70.00	6.64	10740	91.59	100.00	70.00	6.33	8445	
		<i>dnaid</i>	94.28	100.00	70.34	5.36	10740	95.03	100.00	71.25	4.97	8445	
		<i>kimura</i>	0.07	0.58	0.00	0.06	10740	0.08	0.40	0.00	0.06	8445	
		<i>logdet</i>	0.10	1.17	0.00	0.12	10740	0.10	0.93	0.00	0.11	8445	
		<i>index</i>	75.18	99.46	5.58	14.05	10740	77.41	99.75	29.32	13.49	8445	
MDG1		<i>ks</i>	0.29	0.98	0.00	0.25	3407	0.01	0.28	0.00	0.03	1543	
		<i>ka</i>	0.11	0.29	0.00	0.09	3407	0.03	0.30	0.00	0.03	1543	
		<i>kaks</i>	0.38	0.99	0.00	0.21	3407	90.81	430.00	1.00	55.96	1543	
		<i>protid</i>	83.31	100.00	70.09	10.39	3407	91.06	100.00	70.37	6.29	1543	
		<i>dnaid</i>	89.43	100.00	71.68	6.38	3407	94.51	99.42	72.14	5.14	1543	
		<i>kimura</i>	0.19	0.38	0.00	0.13	3407	0.09	0.35	0.01	0.06	1543	
		<i>logdet</i>	0.17	0.58	0.00	0.10	3407	0.10	0.58	0.01	0.10	1543	
		<i>index</i>	52.49	99.47	22.28	23.02	3407	76.43	97.55	31.31	12.67	1543	
MDG3		<i>ks</i>	0.06	0.20	0.01	0.05	518	0.02	0.16	0.00	0.03	639	
		<i>ka</i>	0.02	0.13	0.00	0.03	518	0.04	0.26	0.00	0.05	639	
		<i>kaks</i>	0.32	0.95	0.00	0.30	518	68.95	480.00	1.02	92.81	639	
		<i>protid</i>	91.95	100.00	70.00	6.86	518	90.86	99.06	70.21	7.04	639	
		<i>dnaid</i>	94.33	99.58	71.70	6.25	518	94.88	99.69	76.33	5.06	639	
		<i>kimura</i>	0.06	0.92	0.00	0.05	518	0.08	0.79	0.01	0.06	639	
		<i>logdet</i>	0.11	0.63	0.00	0.11	518	0.10	0.64	0.00	0.12	639	
		<i>index</i>	75.51	96.47	25.08	13.42	518	76.84	97.84	25.16	16.30	639	
MICROPIA		<i>ks</i>	0.05	0.09	0.01	0.02	227	0.03	0.11	0.00	0.02	262	
		<i>ka</i>	0.03	0.05	0.00	0.01	227	0.04	0.12	0.00	0.02	262	
		<i>kaks</i>	0.64	1.00	0.00	0.22	227	17.20	480.00	1.00	49.23	262	
		<i>protid</i>	83.66	96.06	70.08	7.28	227	85.64	99.00	70.69	6.46	262	
		<i>dnaid</i>	88.09	98.16	71.39	6.77	227	90.43	99.33	72.90	5.91	262	
		<i>kimura</i>	0.15	0.44	0.04	0.08	227	0.13	0.34	0.01	0.07	262	
		<i>logdet</i>	0.18	0.41	0.03	0.08	227	0.20	0.43	0.03	0.09	262	
		<i>index</i>	60.91	86.55	42.33	9.94	227	62.73	90.67	40.56	9.22	262	
NINJA		<i>ks</i>	0.10	0.86	0.00	0.12	1643	0.02	0.23	0.00	0.04	602	
		<i>ka</i>	0.03	0.35	0.00	0.04	1643	0.03	0.26	0.00	0.04	602	

		<i>kaks</i>	0.34	1.00	0.00	0.29	1643	63.88	490.00	1.01	81.22	602
		<i>protid</i>	89.51	100.00	70.00	8.19	1643	91.47	99.09	70.08	7.01	602
		<i>dnaid</i>	92.26	100.00	72.80	6.68	1643	94.73	99.70	74.38	5.24	602
		<i>kimura</i>	0.09	0.41	0.00	0.08	1643	0.08	0.50	0.01	0.08	602
		<i>logdet</i>	0.13	0.60	0.00	0.12	1643	0.12	0.56	0.00	0.12	602
		<i>index</i>	69.45	99.41	22.65	16.90	1643	75.93	97.61	26.99	14.95	602
	NOBEL-BEL	<i>ks</i>	0.19	0.27	0.03	0.14	3					
		<i>ka</i>	0.12	0.17	0.03	0.08	3					
		<i>kaks</i>	0.73	1.00	0.58	0.23	3					
		<i>protid</i>	79.44	94.04	70.23	12.79	3					
		<i>dnaid</i>	88.44	97.13	83.46	7.55	3					
		<i>kimura</i>	0.25	0.38	0.06	0.17	3					
		<i>logdet</i>	0.36	0.53	0.03	0.28	3					
		<i>index</i>	47.11	82.42	28.05	30.61	3					
	NOMAD	<i>ks</i>	0.11	0.73	0.00	0.16	709	0.02	0.14	0.00	0.03	332
		<i>ka</i>	0.03	0.36	0.00	0.04	709	0.03	0.17	0.00	0.03	332
		<i>kaks</i>	0.27	0.98	0.00	0.29	709	103.3 3	490.00	1.01	136.80	332
		<i>protid</i>	93.58	100.00	71.86	6.58	709	95.85	100.00	71.93	4.37	332
		<i>dnaid</i>	94.51	100.00	74.45	6.26	709	97.30	100.00	74.27	3.41	332
		<i>kimura</i>	0.07	0.33	0.00	0.07	709	0.04	0.26	0.00	0.04	332
		<i>logdet</i>	0.12	0.86	0.00	0.14	709	0.08	0.52	0.00	0.10	332
		<i>index</i>	75.71	99.11	22.75	20.67	709	84.28	99.74	46.73	14.14	332
	OPUS	<i>ks</i>	0.04	0.04	0.04	0.00	3					
		<i>ka</i>	0.02	0.02	0.02	0.00	3					
		<i>kaks</i>	0.43	0.43	0.43	0.00	3					
		<i>protid</i>	88.66	94.85	84.54	5.46	3					
		<i>dnaid</i>	90.15	96.56	85.91	5.65	3					
		<i>kimura</i>	0.05	0.05	0.04	0.01	3					
		<i>logdet</i>	0.12	0.16	0.10	0.03	3					
		<i>index</i>	72.31	74.78	70.76	2.16	3					
	OSVALDO	<i>ks</i>	0.87	3.07	0.09	1.02	7	0.01	0.01	0.01		1
		<i>ka</i>	0.17	0.31	0.06	0.08	7	0.01	0.01	0.01		1
		<i>kaks</i>	0.47	1.00	0.02	0.36	7	1.06	1.06	1.06		1
		<i>protid</i>	75.63	86.49	70.27	5.71	7	96.99	96.99	96.99		1
		<i>dnaid</i>	82.34	93.92	76.28	6.09	7	98.75	98.75	98.75		1
		<i>kimura</i>	0.29	0.36	0.15	0.08	7	0.03	0.03	0.03		1
		<i>logdet</i>	0.24	0.39	0.05	0.13	7	0.13	0.13	0.13		1
		<i>index</i>	31.59	59.81	10.86	17.48	7	82.03	82.03	82.03		1
	PAO	<i>ks</i>	0.62	4.63	0.03	0.44	135	0.20	0.20	0.20		1
		<i>ka</i>	0.02	0.20	0.00	0.03	135	0.21	0.21	0.21		1
		<i>kaks</i>	0.12	0.77	0.00	0.23	135	1.02	1.02	1.02		1
		<i>protid</i>	86.41	97.97	70.71	5.01	135	72.39	72.39	72.39		1
		<i>dnaid</i>	89.04	98.65	75.48	3.85	135	82.41	82.41	82.41		1
		<i>kimura</i>	0.08	0.31	0.02	0.05	135	0.36	0.36	0.36		1
		<i>logdet</i>	0.50	1.36	0.10	0.11	135	0.28	0.28	0.28		1
		<i>index</i>	35.83	83.57	6.16	8.92	135	33.32	33.32	33.32		1
	PIFO	<i>ks</i>	0.08	0.52	0.01	0.14	27	0.00	0.06	0.00	0.01	23
		<i>ka</i>	0.04	0.25	0.00	0.07	27	0.02	0.06	0.01	0.01	23
		<i>kaks</i>	0.39	0.99	0.00	0.39	27	81.94	97.50	1.05	32.12	23
		<i>protid</i>	88.29	100.00	73.33	9.88	27	92.11	97.14	76.52	5.46	23
		<i>dnaid</i>	90.35	100.00	75.06	8.75	27	95.06	99.05	78.55	5.26	23

		<i>kimura</i>	0.07	0.29	0.00	0.08	27	0.06	0.12	0.03	0.03	23
		<i>logdet</i>	0.20	1.08	0.00	0.24	27	0.16	0.27	0.01	0.09	23
		<i>index</i>	68.33	98.04	24.48	22.86	27	75.61	90.89	60.60	9.30	23
	POLINTON	<i>ks</i>	0.13	0.17	0.06	0.06	3	0.01	0.01	0.00	0.01	3
		<i>ka</i>	0.00	0.00	0.00	0.00	3	0.04	0.05	0.02	0.02	3
		<i>kaks</i>	0.01	0.02	0.00	0.01	3	44.10	113.50	9.11	60.10	3
		<i>protid</i>	97.88	99.36	96.82	1.32	3	92.09	96.11	88.50	3.82	3
		<i>dnaid</i>	94.91	97.03	93.21	1.95	3	96.94	98.70	95.28	1.71	3
		<i>kimura</i>	0.02	0.03	0.01	0.01	3	0.09	0.13	0.04	0.05	3
		<i>logdet</i>	0.05	0.07	0.03	0.02	3	0.27	0.37	0.07	0.17	3
		<i>index</i>	79.30	89.04	74.04	8.44	3	67.30	72.98	61.18	5.91	3
	PROTOP	<i>ks</i>	0.20	0.24	0.18	0.02	32					
		<i>ka</i>	0.05	0.08	0.03	0.01	32					
		<i>kaks</i>	0.22	0.38	0.17	0.06	32					
		<i>protid</i>	90.17	92.31	84.40	2.63	32					
		<i>dnaid</i>	92.20	93.59	88.07	1.56	32					
		<i>kimura</i>	0.10	0.15	0.08	0.03	32					
		<i>logdet</i>	0.19	0.28	0.15	0.04	32					
		<i>index</i>	56.68	62.13	47.81	4.51	32					
	QUASIMOD O	<i>ks</i>	0.09	0.52	0.01	0.05	2951	0.03	0.11	0.00	0.02	292
		<i>ka</i>	0.03	0.15	0.00	0.02	2951	0.05	0.13	0.03	0.02	292
		<i>kaks</i>	0.40	1.00	0.00	0.23	2951	22.09	111.25	1.01	39.48	292
		<i>protid</i>	91.34	99.01	70.48	5.70	2951	85.63	93.89	72.52	6.03	292
		<i>dnaid</i>	94.03	98.47	73.02	4.13	2951	92.35	97.71	79.58	4.99	292
		<i>kimura</i>	0.08	0.37	0.01	0.05	2951	0.15	0.40	0.06	0.08	292
		<i>logdet</i>	0.09	0.63	0.02	0.08	2951	0.11	0.33	0.03	0.06	292
		<i>index</i>	71.81	90.66	33.06	11.71	2951	66.38	82.99	39.36	10.24	292
	R1	<i>ks</i>	0.10	9.63	0.01	0.17	27171	0.00	0.32	0.00	0.02	34590
		<i>ka</i>	0.03	2.84	0.00	0.05	27171	0.04	2.69	0.00	0.10	34590
		<i>kaks</i>	0.18	0.99	0.00	0.25	27171	131.64	490.00	1.00	111.15	34590
		<i>protid</i>	86.15	100.00	70.00	7.51	27171	86.59	100.00	70.00	7.25	34590
		<i>dnaid</i>	90.89	99.46	70.14	6.49	27171	90.70	100.00	70.37	6.60	34590
		<i>kimura</i>	0.13	2.27	0.01	0.09	27171	0.12	2.29	0.01	0.07	34590
		<i>logdet</i>	0.17	2.27	0.01	0.13	27171	0.17	1.72	0.01	0.12	34590
		<i>index</i>	62.94	96.30	4.39	14.61	27171	67.27	94.78	12.35	12.62	34590
	R2	<i>ks</i>	0.10	0.40	0.01	0.08	2108	0.00	0.07	0.00	0.01	1600
		<i>ka</i>	0.01	0.10	0.00	0.02	2108	0.03	0.14	0.00	0.02	1600
		<i>kaks</i>	0.16	0.95	0.00	0.23	2108	76.40	420.00	1.01	38.71	1600
		<i>protid</i>	88.16	99.29	70.00	8.60	2108	89.57	99.10	70.00	8.53	1600
		<i>dnaid</i>	89.87	99.40	70.40	7.80	2108	92.07	99.40	71.17	7.73	1600
		<i>kimura</i>	0.09	0.42	0.01	0.07	2108	0.07	0.43	0.01	0.07	1600
		<i>logdet</i>	0.17	0.96	0.01	0.14	2108	0.09	0.81	0.01	0.10	1600
		<i>index</i>	65.76	97.09	26.08	15.31	2108	78.14	96.89	36.97	15.08	1600
	ROO	<i>ks</i>	0.03	0.05	0.01	0.03	2					
		<i>ka</i>	0.01	0.01	0.00	0.01	2					

		<i>kaks</i>	0.30	0.51	0.09	0.30	2						
		<i>protid</i>	84.17	84.17	84.17	0.00	2						
		<i>dnaid</i>	86.53	86.67	86.39	0.20	2						
		<i>kimura</i>	0.12	0.12	0.12	0.00	2						
		<i>logdet</i>	0.10	0.16	0.04	0.08	2						
		<i>index</i>	67.75	69.82	65.69	2.92	2						
	ROOA	<i>ks</i>	0.11	0.85	0.01	0.07	5837	0.02	0.32	0.00	0.03	3459	
		<i>ka</i>	0.04	0.24	0.00	0.03	5837	0.05	0.36	0.00	0.03	3459	
		<i>kaks</i>	0.43	1.00	0.00	0.26	5837	54.92	460.00	1.00	51.11	3459	
		<i>protid</i>	88.70	98.40	70.00	6.16	5837	85.78	97.30	70.00	6.37	3459	
		<i>dnaid</i>	92.60	98.95	72.67	4.47	5837	91.21	98.69	72.78	4.99	3459	
		<i>kimura</i>	0.11	0.45	0.02	0.06	5837	0.14	0.40	0.03	0.07	3459	
		<i>logdet</i>	0.16	0.95	0.01	0.10	5837	0.19	1.32	0.02	0.15	3459	
	<i>index</i>	63.38	94.42	21.25	11.97	5837	63.33	89.25	27.08	11.24	3459		
	ROVER	<i>ks</i>	0.11	0.14	0.05	0.02	56	0.02	0.09	0.00	0.03	20	
		<i>ka</i>	0.04	0.05	0.02	0.01	56	0.06	0.10	0.03	0.02	20	
		<i>kaks</i>	0.38	0.58	0.23	0.07	56	61.00	106.60	1.14	45.70	20	
		<i>protid</i>	90.21	94.34	87.74	1.47	56	79.83	87.50	70.68	6.34	20	
		<i>dnaid</i>	94.66	96.86	93.08	0.84	56	85.24	92.86	74.68	7.37	20	
		<i>kimura</i>	0.10	0.13	0.06	0.02	56	0.19	0.27	0.10	0.04	20	
		<i>logdet</i>	0.09	0.17	0.05	0.03	56	0.38	0.86	0.18	0.17	20	
	<i>index</i>	67.75	80.27	58.82	4.73	56	49.51	61.45	39.23	7.33	20		
	SPRINGER	<i>ks</i>	0.10	0.11	0.09	0.01	3						
		<i>ka</i>	0.05	0.06	0.05	0.01	3						
		<i>kaks</i>	0.54	0.58	0.47	0.06	3						
		<i>protid</i>	92.59	92.86	92.06	0.46	3						
		<i>dnaid</i>	93.83	94.18	93.65	0.31	3						
		<i>kimura</i>	0.08	0.09	0.08	0.01	3						
		<i>logdet</i>	0.14	0.15	0.13	0.01	3						
	<i>index</i>	65.83	67.13	65.08	1.13	3							
	STALKER	<i>ks</i>	0.19	0.59	0.01	0.16	502	0.01	0.07	0.00	0.02	169	
		<i>ka</i>	0.08	0.17	0.00	0.06	502	0.01	0.11	0.00	0.02	169	
		<i>kaks</i>	0.48	0.97	0.00	0.17	502	143.2	490.00	1.01	144.56	169	
		<i>protid</i>	83.11	100.00	70.00	8.32	502	96.73	100.00	76.00	5.21	169	
		<i>dnaid</i>	87.14	100.00	73.33	6.68	502	97.79	100.00	78.93	3.98	169	
		<i>kimura</i>	0.19	0.35	0.00	0.11	502	0.03	0.25	0.00	0.05	169	
		<i>logdet</i>	0.20	0.43	0.00	0.09	502	0.05	0.33	0.00	0.07	169	
	<i>index</i>	52.76	98.75	28.53	20.55	502	89.98	99.75	46.89	13.08	169		
	TABOR	<i>ks</i>	0.24	1.49	0.01	0.44	236	0.00	0.08	0.00	0.01	223	
		<i>ka</i>	0.03	0.15	0.00	0.03	236	0.01	0.08	0.00	0.01	223	
		<i>kaks</i>	0.19	0.92	0.00	0.21	236	134.1	390.00	1.08	118.23	223	
		<i>protid</i>	92.80	100.00	70.43	8.44	236	96.76	99.32	75.00	4.97	223	
		<i>dnaid</i>	94.08	100.00	78.26	7.48	236	98.28	99.77	80.95	3.85	223	
		<i>kimura</i>	0.08	0.33	0.00	0.09	236	0.03	0.24	0.01	0.05	223	
		<i>logdet</i>	0.12	1.12	0.00	0.16	236	0.05	0.36	0.00	0.07	223	
	<i>index</i>	72.48	99.19	18.86	25.19	236	90.40	98.27	49.68	10.56	223		
	TAHRE	<i>ks</i>	0.25	0.35	0.11	0.08	40						
		<i>ka</i>	0.11	0.14	0.05	0.03	40						
		<i>kaks</i>	0.43	0.63	0.35	0.06	40						
		<i>protid</i>	80.75	90.30	70.97	5.07	40						

		<i>dnaid</i>	86.87	93.53	79.35	3.40	40						
		<i>kimura</i>	0.22	0.37	0.10	0.07	40						
		<i>logdet</i>	0.36	0.47	0.23	0.08	40						
		<i>index</i>	40.06	58.38	31.68	8.91	40						
TART		<i>ks</i>	0.43	2.53	0.01	0.51	326	0.06	0.18	0.00	0.06	26	
		<i>ka</i>	0.07	0.18	0.00	0.05	326	0.08	0.19	0.00	0.07	26	
		<i>kaks</i>	0.25	0.98	0.00	0.17	326	114.85	470.00	1.01	182.49	26	
		<i>protid</i>	85.28	100.00	70.29	7.97	326	82.56	99.36	70.73	8.49	26	
		<i>dnaid</i>	88.56	100.00	72.22	6.34	326	87.71	99.79	73.37	7.45	26	
		<i>kimura</i>	0.17	0.40	0.00	0.10	326	0.15	0.31	0.01	0.10	26	
		<i>logdet</i>	0.23	0.66	0.00	0.14	326	0.17	0.36	0.00	0.12	26	
		<i>index</i>	48.88	99.23	12.30	21.59	326	60.06	98.27	36.73	19.01	26	
TE50134		<i>ks</i>	0.14	0.29	0.09	0.07	11	0.08	0.16	0.00	0.04	86	
		<i>ka</i>	0.12	0.19	0.08	0.04	11	0.13	0.26	0.05	0.04	86	
		<i>kaks</i>	0.87	1.00	0.65	0.13	11	6.75	102.60	1.09	21.14	86	
		<i>protid</i>	79.22	85.71	73.02	5.29	11	75.14	84.51	70.09	3.29	86	
		<i>dnaid</i>	87.40	92.37	81.88	4.52	11	85.70	93.20	81.74	2.35	86	
		<i>kimura</i>	0.25	0.34	0.16	0.07	11	0.29	0.39	0.15	0.05	86	
		<i>logdet</i>	0.47	0.83	0.16	0.29	11	0.25	0.80	0.14	0.14	86	
		<i>index</i>	39.14	48.09	31.41	5.34	11	43.65	53.80	33.82	4.92	86	
TE50157		<i>ks</i>	0.61	2.35	0.09	0.68	61						
		<i>ka</i>	0.11	0.22	0.06	0.04	61						
		<i>kaks</i>	0.31	0.75	0.06	0.18	61						
		<i>protid</i>	76.59	83.91	70.39	3.58	61						
		<i>dnaid</i>	83.22	92.34	73.47	4.88	61						
		<i>kimura</i>	0.27	0.38	0.15	0.06	61						
		<i>logdet</i>	0.43	1.36	0.10	0.32	61						
		<i>index</i>	33.23	56.29	8.23	12.99	61						
TE50164		<i>ks</i>	0.18	0.21	0.14	0.03	5						
		<i>ka</i>	0.09	0.13	0.06	0.03	5						
		<i>kaks</i>	0.51	0.60	0.41	0.09	5						
		<i>protid</i>	84.36	88.00	81.62	3.10	5						
		<i>dnaid</i>	88.01	90.44	86.13	2.22	5						
		<i>kimura</i>	0.16	0.22	0.08	0.07	5						
		<i>logdet</i>	0.38	0.57	0.24	0.17	5						
		<i>index</i>	45.45	54.47	35.16	9.75	5						
TE50200		<i>ks</i>						0.03	0.06	0.01	0.01	12	
		<i>ka</i>						0.05	0.07	0.04	0.01	12	
		<i>kaks</i>						1.85	3.81	1.22	0.86	12	
		<i>protid</i>						85.35	92.16	77.40	5.25	12	
		<i>dnaid</i>						91.85	96.15	85.62	4.29	12	
		<i>kimura</i>						0.16	0.27	0.08	0.07	12	
		<i>logdet</i>						0.17	0.33	0.04	0.10	12	
		<i>index</i>						62.27	76.67	48.35	8.54	12	
TE50218		<i>ks</i>	2.30	3.21	1.39	1.29	2						
		<i>ka</i>	0.28	0.36	0.19	0.12	2						
		<i>kaks</i>	0.15	0.25	0.06	0.13	2						
		<i>protid</i>	71.44	71.54	71.33	0.15	2						
		<i>dnaid</i>	75.09	75.34	74.83	0.36	2						
		<i>kimura</i>	0.40	0.40	0.40	0.00	2						
		<i>logdet</i>	0.57	0.59	0.55	0.03	2						

	<i>index</i>	10.99	13.46	8.52	3.49	2					
TE50219	<i>ks</i>	0.13	0.25	0.04	0.09	18	0.12	0.21	0.02	0.10	7
	<i>ka</i>	0.10	0.20	0.01	0.09	18	0.14	0.23	0.03	0.10	7
	<i>kaks</i>	0.62	0.91	0.33	0.24	18	1.35	1.58	1.09	0.25	7
	<i>protid</i>	86.12	97.14	75.00	9.79	18	80.34	93.33	70.37	11.88	7
	<i>dnaid</i>	90.28	98.10	81.79	7.39	18	87.88	97.14	80.56	8.42	7
	<i>kimura</i>	0.14	0.25	0.03	0.10	18	0.22	0.35	0.07	0.14	7
	<i>logdet</i>	0.13	0.22	0.03	0.08	18	0.16	0.25	0.05	0.09	7
	<i>index</i>	60.87	86.80	36.98	21.88	18	53.31	80.82	32.82	23.84	7
TE50222	<i>ks</i>	0.24	0.31	0.10	0.11	6	0.00	0.00	0.00	0.00	2
	<i>ka</i>	0.14	0.18	0.03	0.07	6	0.12	0.19	0.06	0.09	2
	<i>kaks</i>	0.54	0.59	0.27	0.13	6	97.55	99.26	95.83	2.43	2
	<i>protid</i>	77.45	82.64	72.73	3.64	6	76.06	78.03	74.10	2.78	2
	<i>dnaid</i>	85.36	89.26	82.32	2.63	6	85.25	87.28	83.21	2.88	2
	<i>kimura</i>	0.21	0.28	0.18	0.04	6	0.27	0.32	0.22	0.07	2
	<i>logdet</i>	0.22	0.42	0.15	0.10	6	0.23	0.25	0.22	0.02	2
	<i>index</i>	42.16	52.30	38.86	5.04	6	48.03	53.28	42.78	7.42	2
TE50237	<i>ks</i>	0.13	0.13	0.13		1					
	<i>ka</i>	0.12	0.12	0.12		1					
	<i>kaks</i>	0.94	0.94	0.94		1					
	<i>protid</i>	78.50	78.50	78.50		1					
	<i>dnaid</i>	89.41	89.41	89.41		1					
	<i>kimura</i>	0.25	0.25	0.25		1					
	<i>logdet</i>	0.48	0.48	0.48		1					
	<i>index</i>	38.61	38.61	38.61		1					
TE50238	<i>ks</i>	0.10	0.18	0.05	0.06	6	0.05	0.14	0.03	0.02	24
	<i>ka</i>	0.08	0.14	0.04	0.04	6	0.07	0.14	0.05	0.02	24
	<i>kaks</i>	0.89	0.98	0.77	0.08	6	1.61	2.98	1.01	0.51	24
	<i>protid</i>	86.03	92.66	71.33	8.17	6	81.17	89.09	70.20	5.60	24
	<i>dnaid</i>	92.13	96.33	81.59	5.76	6	88.50	95.15	73.51	5.08	24
	<i>kimura</i>	0.15	0.33	0.08	0.10	6	0.20	0.33	0.11	0.07	24
	<i>logdet</i>	0.28	0.37	0.13	0.09	6	0.27	0.50	0.06	0.13	24
	<i>index</i>	54.50	68.87	36.63	12.48	6	52.37	68.10	38.17	9.16	24
TELEMAC	<i>ks</i>	0.20	0.39	0.10	0.16	3					
	<i>ka</i>	0.13	0.21	0.09	0.07	3					
	<i>kaks</i>	0.79	0.93	0.55	0.21	3					
	<i>protid</i>	80.79	86.11	73.81	6.32	3					
	<i>dnaid</i>	87.81	92.59	85.38	4.14	3					
	<i>kimura</i>	0.22	0.31	0.15	0.08	3					
	<i>logdet</i>	0.41	0.67	0.16	0.26	3					
	<i>index</i>	41.39	51.05	25.08	14.21	3					
TINKER	<i>ks</i>	0.02	0.05	0.00	0.02	128	0.00	0.05	0.00	0.01	42
	<i>ka</i>	0.00	0.04	0.00	0.01	128	0.01	0.05	0.01	0.01	42
	<i>kaks</i>	0.09	0.85	0.00	0.21	128	96.24	101.00	1.11	21.53	42
	<i>protid</i>	95.66	100.00	70.00	7.66	128	95.43	99.26	70.23	7.15	42
	<i>dnaid</i>	96.61	100.00	70.28	7.60	128	96.99	99.75	72.01	7.07	42
	<i>kimura</i>	0.04	0.32	0.00	0.05	128	0.02	0.04	0.01	0.01	42
	<i>logdet</i>	0.06	0.32	0.00	0.07	128	0.04	0.20	0.00	0.06	42
	<i>index</i>	86.59	99.30	45.01	12.71	128	90.06	97.50	52.32	10.74	42
TIRANT	<i>ks</i>	0.07	0.72	0.01	0.05	354	0.08	0.14	0.00	0.05	146
	<i>ka</i>	0.04	0.15	0.00	0.03	354	0.09	0.14	0.00	0.05	146
	<i>kaks</i>	0.55	0.93	0.00	0.20	354	29.89	470.00	1.01	94.18	146

		<i>protid</i>	92.30	99.14	73.67	4.05	354	86.78	98.10	78.74	6.59	146
		<i>dnaid</i>	95.37	99.43	78.90	3.06	354	90.74	98.85	84.78	4.94	146
		<i>kimura</i>	0.08	0.33	0.01	0.04	354	0.15	0.25	0.02	0.08	146
		<i>logdet</i>	0.15	1.10	0.01	0.15	354	0.43	1.05	0.01	0.37	146
		<i>index</i>	71.07	95.65	25.02	12.95	354	53.68	94.84	31.28	22.55	146
	TRANSIB	<i>ks</i>	0.07	0.25	0.01	0.06	50	0.09	0.23	0.01	0.08	15
		<i>ka</i>	0.03	0.15	0.00	0.03	50	0.10	0.27	0.03	0.08	15
		<i>kaks</i>	0.47	0.97	0.00	0.25	50	1.40	2.75	1.01	0.50	15
		<i>protid</i>	93.44	100.00	75.21	5.84	50	86.43	94.16	70.34	8.14	15
		<i>dnaid</i>	95.90	100.00	84.90	3.77	50	90.90	97.51	79.10	6.06	15
		<i>kimura</i>	0.07	0.30	0.00	0.07	50	0.14	0.28	0.06	0.07	15
		<i>logdet</i>	0.08	0.36	0.00	0.10	50	0.19	0.31	0.03	0.11	15
	<i>index</i>	77.49	99.10	38.79	16.80	50	58.77	81.74	32.82	14.57	15	
	TRANSPAC	<i>ks</i>	0.03	0.46	0.00	0.07	97	0.00	0.00	0.00	0.00	88
		<i>ka</i>	0.00	0.01	0.00	0.00	97	0.00	0.01	0.00	0.01	88
		<i>kaks</i>	0.04	0.40	0.00	0.10	97	221.8 1	490.00	2.32	172.23	88
		<i>protid</i>	99.62	100.00	96.88	0.82	97	100.0 0	100.00	100.00	0.00	88
		<i>dnaid</i>	99.46	100.00	92.33	1.29	97	99.92	100.00	99.69	0.14	88
		<i>kimura</i>	0.00	0.03	0.00	0.01	97	0.00	0.01	0.00	0.00	88
		<i>logdet</i>	0.01	0.09	0.00	0.02	97	0.00	0.01	0.00	0.00	88
		<i>index</i>	95.61	99.51	58.18	7.38	97	99.08	99.76	97.79	0.58	88
	ULYSSES	<i>ks</i>	0.07	0.07	0.07		1					
		<i>ka</i>	0.06	0.06	0.06		1					
		<i>kaks</i>	0.86	0.86	0.86		1					
		<i>protid</i>	90.21	90.21	90.21		1					
		<i>dnaid</i>	94.17	94.17	94.17		1					
		<i>kimura</i>	0.11	0.11	0.11		1					
		<i>logdet</i>	0.09	0.09	0.09		1					
		<i>index</i>	68.55	68.55	68.55		1					
	YOYO	<i>ks</i>	0.22	0.57	0.07	0.17	26					
		<i>ka</i>	0.08	0.11	0.03	0.03	26					
		<i>kaks</i>	0.46	0.81	0.16	0.21	26					
		<i>protid</i>	81.28	90.19	70.59	5.39	26					
		<i>dnaid</i>	88.29	94.97	80.29	4.06	26					
		<i>kimura</i>	0.20	0.33	0.11	0.07	26					
		<i>logdet</i>	0.42	0.80	0.16	0.21	26					
		<i>index</i>	42.48	60.31	24.77	11.44	26					
	ZAM	<i>ks</i>	0.22	0.87	0.01	0.13	303	0.02	0.10	0.00	0.03	31
		<i>ka</i>	0.04	0.18	0.00	0.03	303	0.04	0.16	0.00	0.04	31
		<i>kaks</i>	0.21	0.98	0.00	0.19	303	63.64	410.00	1.02	102.65	31
<i>protid</i>		86.87	98.57	70.20	6.84	303	85.91	94.81	72.09	8.20	31	
<i>dnaid</i>		88.93	98.88	70.60	5.71	303	91.04	98.13	76.55	7.70	31	
<i>kimura</i>		0.13	1.49	0.01	0.10	303	0.12	0.18	0.05	0.05	31	
<i>logdet</i>		0.24	1.68	0.01	0.14	303	0.20	0.50	0.02	0.16	31	
<i>index</i>		53.15	93.71	14.22	13.43	303	64.84	86.71	42.07	13.20	31	
Melanogaster-Hawaiian	Genome	<i>ks</i>	6.27	22.68	0.56	5.38	2557					
		<i>ka</i>	0.14	2.99	0.00	0.11	2557					
		<i>kaks</i>	0.03	0.92	0.00	0.03	2557					
		<i>protid</i>	76.93	100.00	37.19	11.49	2557					
		<i>dnaid</i>	71.72	90.93	50.00	6.55	2557					

Melanogaster- Obscura		<i>kimura</i>	0.28	1.16	0.00	0.17	2557						
		<i>logdet</i>	0.37	1.13	0.11	0.11	2557						
		<i>index</i>	10.35	49.07	0.82	7.32	2557						
	GYPSY	<i>ks</i>	0.83	1.03	0.67	0.07	128						
		<i>ka</i>	0.12	0.17	0.09	0.02	128						
		<i>kaks</i>	0.14	0.22	0.11	0.03	128						
		<i>protid</i>	79.05	84.00	71.00	2.75	128						
		<i>dnaid</i>	79.43	81.33	75.00	1.66	128						
		<i>kimura</i>	0.24	0.37	0.18	0.04	128						
		<i>logdet</i>	0.32	0.43	0.24	0.04	128						
	AURORA	<i>index</i>	26.19	29.40	21.31	2.15	128						
		<i>ks</i>	1.52	5.38	0.71	1.42	27						
		<i>ka</i>	0.18	0.27	0.09	0.06	27						
		<i>kaks</i>	0.19	0.33	0.02	0.10	27						
		<i>protid</i>	75.91	79.03	70.97	1.74	27						
<i>dnaid</i>		81.69	82.72	79.30	0.85	27							
<i>kimura</i>		0.29	0.37	0.25	0.03	27							
<i>logdet</i>		0.44	0.64	0.27	0.10	27							
Genome	<i>index</i>	19.72	29.42	6.63	6.27	27							
	<i>ks</i>	3.54	20.29	0.35	3.49	5117							
	<i>ka</i>	0.10	0.69	0.00	0.07	5117							
	<i>kaks</i>	0.04	0.32	0.00	0.03	5117							
	<i>protid</i>	82.06	100.00	37.44	9.91	5117							
	<i>dnaid</i>	76.25	94.04	50.14	5.98	5117							
	<i>kimura</i>	0.21	1.01	0.00	0.14	5117							
	<i>logdet</i>	0.32	1.09	0.07	0.11	5117							
POLINTON	<i>index</i>	16.21	63.77	0.92	8.81	5117							
	<i>ks</i>	1.67	7.64	1.35	0.93	48	0.05	0.27	0.00	0.11	6		
	<i>ka</i>	0.16	1.35	0.00	0.18	48	3.04	16.77	0.30	6.72	6		
	<i>kaks</i>	0.09	0.18	0.00	0.03	48	92.09	97.97	62.68	14.41	6		
	<i>protid</i>	74.14	83.22	71.43	2.80	48	73.84	75.93	71.07	2.18	6		
	<i>dnaid</i>	77.47	78.79	76.24	0.60	48	73.71	75.62	72.18	1.33	6		
	<i>kimura</i>	0.32	0.36	0.18	0.04	48	0.31	0.35	0.29	0.03	6		
	<i>logdet</i>	0.48	1.11	0.39	0.12	48	0.55	0.72	0.38	0.14	6		
TABOR	<i>index</i>	15.76	17.12	3.21	2.27	48	25.48	33.96	2.40	11.71	6		
	<i>ks</i>	3.65	9.89	0.01	2.07	6224	0.00	0.20	0.00	0.01	552		
	<i>ka</i>	0.19	0.76	0.00	0.12	6224	0.08	0.39	0.00	0.12	552		
	<i>kaks</i>	0.05	0.15	0.00	0.02	6224	195.68	490.00	1.94	151.88	552		
	<i>protid</i>	76.02	83.62	70.00	2.85	6224	74.01	79.41	70.00	2.32	552		
	<i>dnaid</i>	80.32	84.98	71.07	1.92	6224	80.16	83.57	75.00	1.48	552		
	<i>kimura</i>	0.27	0.41	0.19	0.04	6224	0.31	0.38	0.24	0.03	552		
	<i>logdet</i>	0.34	0.91	0.19	0.08	6224	0.36	0.93	0.24	0.11	552		
TRANSIB	<i>index</i>	11.75	50.53	3.48	4.46	6224	43.45	52.43	24.50	6.15	552		
	<i>ks</i>	0.88	0.88	0.88		1							
	<i>ka</i>	0.09	0.09	0.09		1							
	<i>kaks</i>	0.10	0.10	0.10		1							
	<i>protid</i>	79.41	79.41	79.41		1							
	<i>dnaid</i>	79.74	79.74	79.74		1							
	<i>kimura</i>	0.24	0.24	0.24		1							
	<i>logdet</i>	0.33	0.33	0.33		1							
<i>index</i>	25.69	25.69	25.69		1								

		<i>ks</i>	1.81	1.85	1.78	0.05	2							
		<i>ka</i>	0.04	0.04	0.04	0.00	2							
		<i>kaks</i>	0.02	0.02	0.02	0.00	2							
		<i>protid</i>	82.77	83.24	82.29	0.67	2							
		<i>dnaid</i>	78.25	78.39	78.11	0.20	2							
		<i>kimura</i>	0.18	0.19	0.17	0.01	2							
		<i>logdet</i>	0.26	0.26	0.25	0.01	2							
		<i>index</i>	19.67	20.14	19.20	0.66	2							
Melanogaster-Repleta	Genome	<i>ks</i>	6.58	23.56	0.47	5.32	2560							
		<i>ka</i>	0.14	1.36	0.00	0.10	2560							
		<i>kaks</i>	0.03	0.28	0.00	0.03	2560							
		<i>protid</i>	77.21	100.00	36.19	11.48	2560							
		<i>dnaid</i>	71.92	92.05	50.02	6.61	2560							
		<i>kimura</i>	0.28	1.17	0.00	0.17	2560							
		<i>logdet</i>	0.37	1.26	0.11	0.11	2560							
		<i>index</i>	9.98	53.35	0.68	7.41	2560							
	GYPSY	<i>ks</i>	1.36	1.43	1.33	0.03	19							
		<i>ka</i>	0.20	0.21	0.19	0.01	19							
		<i>kaks</i>	0.14	0.15	0.14	0.00	19							
		<i>protid</i>	71.18	72.65	70.09	0.83	19							
		<i>dnaid</i>	72.96	73.50	72.22	0.34	19							
		<i>kimura</i>	0.36	0.38	0.34	0.01	19							
		<i>logdet</i>	0.46	0.51	0.44	0.02	19							
		<i>index</i>	15.47	16.31	14.49	0.47	19							
	TABOR	<i>ks</i>	1.45	9.42	0.01	1.90	3970	0.00	0.01	0.00	0.00	1490		
		<i>ka</i>	0.12	0.53	0.00	0.12	3970	0.10	1.04	0.00	0.13	1490		
		<i>kaks</i>	0.13	0.64	0.00	0.17	3970	160.89	490.00	1.48	132.46	1490		
		<i>protid</i>	76.51	87.83	70.00	3.59	3970	79.94	88.57	70.00	3.86	1490		
		<i>dnaid</i>	79.67	88.89	71.62	2.47	3970	84.09	89.36	70.80	2.82	1490		
		<i>kimura</i>	0.27	0.47	0.13	0.05	3970	0.23	0.50	0.12	0.05	1490		
		<i>logdet</i>	0.38	1.07	0.15	0.09	3970	0.29	1.04	0.15	0.11	1490		
		<i>index</i>	24.44	65.48	3.21	10.18	3970	50.29	66.66	19.09	8.87	1490		

Melanogaster-Virilis	Genome	<i>ks</i>	6.17	25.36	0.46	5.18	2560							
		<i>ka</i>	0.13	1.43	0.00	0.09	2560							
		<i>kaks</i>	0.03	0.27	0.00	0.03	2560							
		<i>protid</i>	78.03	100.00	39.68	11.19	2560							
		<i>dnaid</i>	72.42	92.33	50.51	6.41	2560							
		<i>kimura</i>	0.26	1.08	0.00	0.17	2560							
		<i>logdet</i>	0.36	1.25	0.10	0.11	2560							
		<i>index</i>	10.61	54.03	0.84	7.54	2560							
	GYPSY	<i>ks</i>	0.42	0.43	0.41	0.01	2							
		<i>ka</i>	0.11	0.11	0.11	0.00	2							
		<i>kaks</i>	0.27	0.28	0.26	0.01	2							
		<i>protid</i>	70.73	70.73	70.73	0.00	2							
		<i>dnaid</i>	78.19	78.32	78.05	0.19	2							
		<i>kimura</i>	0.34	0.35	0.34	0.01	2							
		<i>logdet</i>	0.30	0.30	0.30	0.00	2							
		<i>index</i>	29.61	29.86	29.35	0.36	2							
	TABOR	<i>ks</i>	1.72	7.19	0.01	1.33	1299	0.00	0.00	0.00	0.00	243		
		<i>ka</i>	0.10	0.56	0.00	0.11	1299	0.07	0.17	0.00	0.08	243		

Melanogaster-Willistoni		<i>kaks</i>	0.10	0.87	0.00	0.16	1299	151.83	490.00	2.00	133.41	243
		<i>protid</i>	75.74	85.29	70.00	3.35	1299	75.68	83.84	70.00	3.92	243
		<i>dnaid</i>	79.35	87.88	72.12	2.24	1299	82.40	88.55	76.83	3.30	243
		<i>kimura</i>	0.29	0.45	0.18	0.05	1299	0.29	0.44	0.17	0.06	243
		<i>logdet</i>	0.37	0.66	0.17	0.08	1299	0.36	0.76	0.15	0.10	243
		<i>index</i>	20.55	58.30	4.64	9.15	1299	45.26	61.35	29.91	5.87	243
		TRANSPA C	<i>ks</i>	2.30	2.30	2.30		1				
	<i>ka</i>		0.09	0.09	0.09		1					
	<i>kaks</i>		0.04	0.04	0.04		1					
	<i>protid</i>		75.64	75.64	75.64		1					
	<i>dnaid</i>		73.56	73.56	73.56		1					
	<i>kimura</i>		0.28	0.28	0.28		1					
	<i>logdet</i>		0.37	0.37	0.37		1					
	Genome	<i>ks</i>	6.93	22.63	0.56	5.50	2560					
		<i>ka</i>	0.13	0.97	0.00	0.09	2560					
		<i>kaks</i>	0.03	0.28	0.00	0.02	2560					
		<i>protid</i>	78.76	100.00	41.49	11.00	2560					
		<i>dnaid</i>	71.04	89.75	50.14	6.18	2560					
		<i>kimura</i>	0.25	0.99	0.00	0.16	2560					
		<i>logdet</i>	0.38	0.96	0.11	0.11	2560					
	HELITRON	<i>ks</i>	0.09	5.06	0.00	0.16	1273	0.01	0.07	0.00	0.01	428
		<i>ka</i>	0.02	2.97	0.00	0.09	1273	0.53	0.68	0.03	0.24	428
		<i>kaks</i>	0.15	0.82	0.00	0.17	1273	96.27	102.62	1.04	13.16	428
		<i>protid</i>	84.36	96.75	70.00	7.38	1273	86.89	94.31	70.59	8.15	428
		<i>dnaid</i>	88.81	98.10	74.24	5.54	1273	90.55	97.56	75.16	7.14	428
		<i>kimura</i>	0.15	0.46	0.03	0.08	1273	0.15	0.56	0.06	0.11	428
		<i>logdet</i>	0.21	1.10	0.02	0.12	1273	0.31	1.31	0.03	0.16	428
HOYAK	<i>ks</i>	0.63	0.63	0.63	0.00	3						
	<i>ka</i>	0.12	0.12	0.12	0.00	3						
	<i>kaks</i>	0.20	0.20	0.20	0.00	3						
	<i>protid</i>	78.11	78.11	78.11	0.00	3						
	<i>dnaid</i>	82.05	82.05	82.05	0.00	3						
	<i>kimura</i>	0.26	0.26	0.26	0.00	3						
	<i>logdet</i>	0.55	0.55	0.55	0.00	3						
STALKER	<i>ks</i>	1.94	3.98	1.07	0.36	130						
	<i>ka</i>	0.16	0.25	0.12	0.02	130						
	<i>kaks</i>	0.09	0.18	0.05	0.02	130						
	<i>protid</i>	76.89	82.14	70.10	2.45	130						
	<i>dnaid</i>	79.37	81.44	71.68	1.47	130						
	<i>kimura</i>	0.27	0.37	0.19	0.03	130						
	<i>logdet</i>	0.44	0.63	0.23	0.07	130						
TABOR	<i>ks</i>	1.57	9.33	0.01	0.95	4814	0.00	0.31	0.00	0.01	996	
	<i>ka</i>	0.14	1.46	0.00	0.08	4814	0.25	0.55	0.00	0.12	996	
	<i>kaks</i>	0.11	1.00	0.00	0.12	4814	114.11	490.00	1.67	69.88	996	
	<i>protid</i>	74.28	89.43	70.00	2.97	4814	74.10	88.79	70.00	3.96	996	
	<i>dnaid</i>	78.84	91.99	72.38	1.73	4814	81.47	90.62	73.52	2.95	996	
	<i>kimura</i>	0.29	0.61	0.11	0.04	4814	0.31	0.52	0.11	0.07	996	

		<i>logdet</i>	0.38	1.02	0.09	0.08	4814	0.35	1.13	0.09	0.11	996	
		<i>index</i>	19.69	71.81	4.09	8.69	4814	38.47	73.61	22.33	9.26	996	
	TRANSIB	<i>ks</i>	1.72	1.72	1.72		1						
		<i>ka</i>	0.18	0.18	0.18		1						
		<i>kaks</i>	0.10	0.10	0.10		1						
		<i>protid</i>	74.23	74.23	74.23		1						
		<i>dnaid</i>	73.88	73.88	73.88		1						
		<i>kimura</i>	0.30	0.30	0.30		1						
		<i>logdet</i>	0.47	0.47	0.47		1						
		<i>index</i>	14.40	14.40	14.40		1						
	Melanogaster Subgroup-Hawaiiian	Genome	<i>ks</i>	5.22	23.86	0.53	4.84	12795					
			<i>ka</i>	0.14	0.59	0.00	0.09	12795					
			<i>kaks</i>	0.04	0.30	0.00	0.03	12795					
			<i>protid</i>	76.88	100.00	39.45	11.36	12795					
			<i>dnaid</i>	72.13	90.25	50.09	6.58	12795					
<i>kimura</i>			0.28	1.06	0.00	0.17	12795						
<i>logdet</i>			0.38	1.42	0.10	0.12	12795						
<i>index</i>			11.58	46.10	0.77	7.51	12795						
TRANSPA C		<i>ks</i>	2.75	2.95	2.63	0.17	3						
		<i>ka</i>	0.09	0.09	0.09	0.00	3						
		<i>kaks</i>	0.03	0.04	0.03	0.01	3						
		<i>protid</i>	76.03	76.47	75.66	0.41	3						
		<i>dnaid</i>	72.93	73.12	72.75	0.19	3						
		<i>kimura</i>	0.30	0.30	0.29	0.01	3						
		<i>logdet</i>	0.40	0.41	0.40	0.01	3						
	<i>index</i>	11.42	11.86	10.83	0.53	3							
Melanogaster Subgroup-Obscura	AURORA	<i>ks</i>	3.88	9.49	1.23	1.61	140						
		<i>ka</i>	0.13	0.33	0.10	0.05	140						
		<i>kaks</i>	0.04	0.16	0.02	0.03	140						
		<i>protid</i>	74.52	79.84	70.16	2.13	140						
		<i>dnaid</i>	79.98	82.26	77.10	1.30	140						
		<i>kimura</i>	0.31	0.43	0.24	0.03	140						
		<i>logdet</i>	0.37	0.76	0.21	0.09	140						
		<i>index</i>	10.15	20.57	3.94	3.19	140						
	Genome	<i>ks</i>	2.96	21.93	0.36	2.91	25613						
		<i>ka</i>	0.10	0.59	0.00	0.07	25613						
		<i>kaks</i>	0.04	0.32	0.00	0.03	25613						
		<i>protid</i>	82.18	100.00	38.64	9.72	25613						
		<i>dnaid</i>	76.69	93.20	50.28	5.92	25613						
		<i>kimura</i>	0.20	0.99	0.00	0.13	25613						

							3						
		<i>logdet</i>	0.32	2.50	0.08	0.13	2561 3						
		<i>index</i>	17.50	60.53	0.87	8.69	2561 3						
	NOBEL-BEL	<i>ks</i>	1.86	9.89	0.42	2.05	32						
		<i>ka</i>	0.16	0.25	0.06	0.05	32						
		<i>kaks</i>	0.21	0.46	0.01	0.17	32						
		<i>protid</i>	75.84	80.95	70.00	2.97	32						
		<i>dnaid</i>	74.03	80.07	70.50	2.62	32						
		<i>kimura</i>	0.29	0.38	0.19	0.05	32						
		<i>logdet</i>	0.55	0.89	0.31	0.18	32						
		<i>index</i>	17.80	30.86	3.72	7.73	32						
	POLINTON	<i>ks</i>	4.62	4.80	1.79	0.63	145	0.00	0.00	0.00	0.00	10	
		<i>ka</i>	0.02	0.32	0.00	0.06	145	0.37	0.37	0.37	0.00	10	
		<i>kaks</i>	0.01	0.18	0.00	0.03	145	97.71	97.71	97.71	0.00	10	
		<i>protid</i>	82.43	92.00	70.37	5.36	145	80.98	90.20	75.21	4.68	10	
		<i>dnaid</i>	78.35	81.96	71.52	2.84	145	79.18	81.67	76.86	1.39	10	
		<i>kimura</i>	0.24	3.14	0.10	0.30	145	0.30	1.06	0.15	0.27	10	
		<i>logdet</i>	0.53	1.84	0.27	0.25	145	0.80	1.13	0.52	0.21	10	
		<i>index</i>	8.43	13.57	2.86	1.51	145	28.23	30.64	19.89	3.47	10	
	STALKER	<i>ks</i>	5.62	9.16	0.35	4.00	582	0.05	0.20	0.00	0.09	61	
		<i>ka</i>	0.30	0.67	0.00	0.25	582	0.36	0.39	0.35	0.02	61	
		<i>kaks</i>	0.04	0.13	0.00	0.03	582	74.20	99.89	1.94	43.44	61	
		<i>protid</i>	74.78	78.76	71.54	1.95	582	72.18	74.04	70.79	1.18	61	
		<i>dnaid</i>	78.40	80.83	75.88	1.32	582	80.39	81.65	75.32	1.41	61	
		<i>kimura</i>	0.30	0.34	0.24	0.02	582	0.32	0.33	0.31	0.01	61	
		<i>logdet</i>	0.34	0.38	0.21	0.03	582	0.34	0.40	0.30	0.04	61	
		<i>index</i>	15.92	39.26	4.23	15.29	582	32.61	34.90	27.81	2.84	61	
	TABOR	<i>ks</i>	4.73	7.47	0.83	1.08	1543						
		<i>ka</i>	0.14	0.39	0.05	0.03	1543						
		<i>kaks</i>	0.03	0.25	0.01	0.02	1543						
		<i>protid</i>	74.98	89.53	70.00	2.93	1543						
		<i>dnaid</i>	79.73	84.11	71.75	1.38	1543						
		<i>kimura</i>	0.31	0.45	0.13	0.04	1543						
		<i>logdet</i>	0.36	0.95	0.28	0.09	1543						
		<i>index</i>	8.32	24.02	5.22	2.21	1543						
	TRANSPAC	<i>ks</i>	1.95	2.13	1.85	0.11	8						
		<i>ka</i>	0.03	0.03	0.03	0.00	8						
		<i>kaks</i>	0.02	0.02	0.01	0.00	8						
		<i>protid</i>	85.48	86.32	84.81	0.51	8						
		<i>dnaid</i>	80.13	80.71	79.66	0.34	8						
		<i>kimura</i>	0.16	0.17	0.15	0.01	8						
		<i>logdet</i>	0.28	0.29	0.28	0.00	8						
		<i>index</i>	19.27	19.87	18.16	0.67	8						
Melanogaster Subgroup-RepY leta	G3-JOCKEY	<i>ks</i>	1.50	1.50	1.50	0.00	2						
		<i>ka</i>	0.18	0.18	0.18	0.00	2						
		<i>kaks</i>	0.12	0.12	0.12	0.00	2						
		<i>protid</i>	71.57	71.57	71.57	0.00	2						
		<i>dnaid</i>	73.53	73.53	73.53	0.00	2						
		<i>kimura</i>	0.36	0.36	0.36	0.00	2						
		<i>logdet</i>	0.35	0.35	0.35	0.00	2						

		<i>index</i>	15.88	15.88	15.88	0.00	2						
	Genome	<i>ks</i>	5.72	24.53	0.53	5.00	1280 4						
		<i>ka</i>	0.14	1.30	0.00	0.09	1280 4						
		<i>kaks</i>	0.04	0.38	0.00	0.03	1280 4						
		<i>protid</i>	77.14	100.00	37.57	11.35	1280 4						
		<i>dnaid</i>	72.20	91.30	50.00	6.60	1280 4						
		<i>kimura</i>	0.28	1.15	0.00	0.17	1280 4						
		<i>logdet</i>	0.38	1.40	0.09	0.12	1280 4						
		<i>index</i>	10.91	49.25	0.81	7.49	1280 4						
	GYPSY	<i>ks</i>	0.93	0.93	0.93		1						
		<i>ka</i>	0.17	0.17	0.17		1						
		<i>kaks</i>	0.18	0.18	0.18		1						
		<i>protid</i>	70.08	70.08	70.08		1						
		<i>dnaid</i>	76.38	76.38	76.38		1						
		<i>kimura</i>	0.38	0.38	0.38		1						
		<i>logdet</i>	0.36	0.36	0.36		1						
		<i>index</i>	20.07	20.07	20.07		1						
	MARINER	<i>ks</i>	0.73	0.85	0.70	0.05	27						
		<i>ka</i>	0.07	0.08	0.07	0.00	27						
		<i>kaks</i>	0.10	0.11	0.09	0.01	27						
		<i>protid</i>	77.23	84.11	74.24	2.74	27						
		<i>dnaid</i>	81.05	83.80	78.09	1.39	27						
		<i>kimura</i>	0.25	0.31	0.18	0.04	27						
		<i>logdet</i>	0.27	0.40	0.23	0.06	27						
		<i>index</i>	28.97	30.50	26.34	1.57	27						
	STALKER	<i>ks</i>	1.27	6.19	0.34	2.12	481	0.00	0.00	0.00	0.00	526	
		<i>ka</i>	0.10	0.53	0.00	0.19	481	0.01	0.01	0.00	0.00	526	
		<i>kaks</i>	0.06	0.40	0.00	0.12	481	205.01	490.00	2.39	165.22	526	
		<i>protid</i>	75.00	79.81	70.31	2.25	481	77.39	80.19	70.00	1.80	526	
		<i>dnaid</i>	80.12	83.97	72.66	1.74	481	81.26	83.33	77.88	1.22	526	
		<i>kimura</i>	0.30	0.37	0.21	0.03	481	0.27	0.37	0.23	0.03	526	
		<i>logdet</i>	0.28	0.53	0.19	0.07	481	0.38	0.61	0.22	0.06	526	
		<i>index</i>	31.15	41.46	5.51	11.16	481	48.12	51.32	37.17	2.78	526	
	TABOR	<i>ks</i>	1.62	9.66	0.74	1.05	448						
		<i>ka</i>	0.22	0.61	0.04	0.05	448						
		<i>kaks</i>	0.16	0.24	0.01	0.05	448						
		<i>protid</i>	72.80	87.50	70.09	2.47	448						
		<i>dnaid</i>	75.93	82.69	71.79	1.21	448						
		<i>kimura</i>	0.36	0.41	0.13	0.05	448						
		<i>logdet</i>	0.44	0.62	0.19	0.04	448						
		<i>index</i>	15.47	31.68	4.49	2.69	448						
Melanogaster Subgroup-Viri lis	Genome	<i>ks</i>	5.15	23.78	0.55	4.72	1279 9						
		<i>ka</i>	0.13	1.35	0.00	0.09	1279 9						

		<i>kaks</i>	0.04	0.55	0.00	0.03	1279	9						
		<i>protid</i>	77.94	100.00	40.41	11.03	1279	9						
		<i>dnaid</i>	72.77	90.57	50.06	6.41	1279	9						
		<i>kimura</i>	0.26	1.04	0.00	0.16	1279	9						
		<i>logdet</i>	0.37	1.41	0.10	0.12	1279	9						
		<i>index</i>	11.83	46.71	0.89	7.60	1279	9						
	GYPSY	<i>ks</i>	0.29	0.34	0.26	0.02	38							
		<i>ka</i>	0.09	0.12	0.08	0.01	38							
		<i>kaks</i>	0.32	0.41	0.27	0.03	38							
		<i>protid</i>	81.72	83.74	78.86	1.33	38							
		<i>dnaid</i>	87.44	88.35	86.18	0.60	38							
		<i>kimura</i>	0.21	0.25	0.18	0.02	38							
		<i>logdet</i>	0.15	0.26	0.13	0.03	38							
		<i>index</i>	44.98	48.12	41.12	1.97	38							
	STALKER	<i>ks</i>	0.94	3.26	0.27	0.99	286	0.00	0.00	0.00	0.00	0.00	31	
		<i>ka</i>	0.04	0.14	0.00	0.06	286	0.00	0.01	0.00	0.01	0.01	31	
		<i>kaks</i>	0.02	0.07	0.00	0.03	286	240.02	490.00	2.54	188.44	188.44	31	
		<i>protid</i>	73.69	75.00	71.09	1.53	286	72.91	73.33	70.48	0.56	0.56	31	
		<i>dnaid</i>	78.05	81.65	75.73	1.08	286	77.33	77.60	75.24	0.40	0.40	31	
		<i>kimura</i>	0.33	0.36	0.31	0.02	286	0.36	0.37	0.35	0.01	0.01	31	
		<i>logdet</i>	0.37	0.51	0.21	0.04	286	0.38	0.39	0.27	0.02	0.02	31	
	<i>index</i>	28.20	37.68	9.98	10.76	286	43.09	44.30	42.69	0.33	0.33	31		
	TABOR	<i>ks</i>	2.16	5.82	0.89	1.89	6							
		<i>ka</i>	0.23	0.50	0.06	0.19	6							
		<i>kaks</i>	0.10	0.16	0.06	0.04	6							
		<i>protid</i>	73.14	77.36	70.37	3.35	6							
		<i>dnaid</i>	79.36	83.33	76.32	3.14	6							
		<i>kimura</i>	0.34	0.43	0.27	0.06	6							
		<i>logdet</i>	0.42	0.45	0.36	0.03	6							
	<i>index</i>	16.21	24.34	5.54	7.58	6								
	TRANSPAC	<i>ks</i>	1.71	1.91	1.59	0.15	4							
		<i>ka</i>	0.08	0.08	0.08	0.00	4							
		<i>kaks</i>	0.04	0.05	0.04	0.01	4							
		<i>protid</i>	76.90	77.36	76.50	0.36	4							
		<i>dnaid</i>	75.94	76.31	75.36	0.43	4							
		<i>kimura</i>	0.27	0.27	0.26	0.01	4							
		<i>logdet</i>	0.33	0.34	0.33	0.01	4							
	<i>index</i>	17.18	17.92	16.03	0.88	4								
Melanogaster Subgroup-Wil listoni	G3-JOCKEY	<i>ks</i>	0.67	0.67	0.67		1							
		<i>ka</i>	0.21	0.21	0.21		1							
		<i>kaks</i>	0.31	0.31	0.31		1							
		<i>protid</i>	73.72	73.72	73.72		1							
		<i>dnaid</i>	78.21	78.21	78.21		1							
		<i>kimura</i>	0.32	0.32	0.32		1							
		<i>logdet</i>	0.67	0.67	0.67		1							
		<i>index</i>	20.33	20.33	20.33		1							

Genome	<i>ks</i>	6.39	25.89	0.49	5.29	1281	1						
	<i>ka</i>	0.13	1.01	0.00	0.09	1281	1						
	<i>kaks</i>	0.03	0.32	0.00	0.02	1281	1						
	<i>protid</i>	78.62	100.00	40.31	10.93	1281	1						
	<i>dnaid</i>	71.22	89.10	50.02	6.13	1281	1						
	<i>kimura</i>	0.25	1.06	0.00	0.16	1281	1						
	<i>logdet</i>	0.39	1.50	0.11	0.12	1281	1						
	<i>index</i>	10.00	48.92	0.85	7.03	1281	1						
GYPSY	<i>ks</i>	1.46	2.98	1.01	0.72	23							
	<i>ka</i>	0.15	0.18	0.12	0.02	23							
	<i>kaks</i>	0.12	0.18	0.05	0.04	23							
	<i>protid</i>	73.00	79.69	70.15	3.01	23							
	<i>dnaid</i>	75.06	77.60	71.58	2.14	23							
	<i>kimura</i>	0.33	0.39	0.24	0.04	23							
	<i>logdet</i>	0.49	0.86	0.32	0.13	23							
	<i>index</i>	16.45	20.06	7.90	3.47	23							
HELITRON	<i>ks</i>	0.58	1.91	0.00	0.75	88							
	<i>ka</i>	0.08	0.17	0.00	0.05	88							
	<i>kaks</i>	0.37	0.88	0.00	0.26	88							
	<i>protid</i>	81.32	90.09	70.19	6.55	88							
	<i>dnaid</i>	85.96	93.99	71.47	9.10	88							
	<i>kimura</i>	0.22	0.38	0.11	0.09	88							
	<i>logdet</i>	0.26	0.82	0.09	0.17	88							
	<i>index</i>	45.24	65.98	12.77	21.07	88							
MARINER	<i>ks</i>	1.02	1.02	1.02	0.00	3							
	<i>ka</i>	0.11	0.11	0.11	0.00	3							
	<i>kaks</i>	0.11	0.11	0.11	0.00	3							
	<i>protid</i>	80.38	80.43	80.29	0.08	3							
	<i>dnaid</i>	81.60	81.64	81.51	0.08	3							
	<i>kimura</i>	0.23	0.23	0.22	0.01	3							
	<i>logdet</i>	0.43	0.44	0.43	0.01	3							
	<i>index</i>	22.89	22.90	22.89	0.01	3							
R2	<i>ks</i>	0.27	0.38	0.20	0.05	75							
	<i>ka</i>	0.14	0.19	0.12	0.02	75							
	<i>kaks</i>	0.56	0.91	0.34	0.14	75							
	<i>protid</i>	71.82	76.19	70.00	1.02	75							
	<i>dnaid</i>	74.59	79.37	71.43	1.58	75							
	<i>kimura</i>	0.35	0.38	0.31	0.01	75							
	<i>logdet</i>	0.74	1.44	0.31	0.16	75							
	<i>index</i>	24.97	31.68	17.87	1.86	75							
STALKER	<i>ks</i>	2.18	3.90	0.35	1.04	210	0.00	0.00	0.00	0.00	452		
	<i>ka</i>	0.25	0.38	0.00	0.16	210	0.35	0.35	0.35	0.00	452		
	<i>kaks</i>	0.09	0.14	0.00	0.06	210	99.89	99.89	99.89	0.00	452		
	<i>protid</i>	73.72	78.65	70.15	1.53	210	74.81	78.63	70.10	2.89	452		
	<i>dnaid</i>	79.13	80.91	71.14	1.49	210	80.85	82.91	76.63	1.77	452		

Obscura Group	TABOR	<i>kimura</i>	0.31	0.36	0.25	0.03	210	0.29	0.41	0.21	0.05	452
		<i>logdet</i>	0.33	1.11	0.19	0.11	210	0.34	0.91	0.17	0.13	452
		<i>index</i>	17.04	37.57	8.84	10.02	210	35.52	41.62	23.71	3.77	452
		<i>ks</i>	2.83	6.50	0.35	1.17	947	0.00	0.00	0.00	0.00	378
		<i>ka</i>	0.16	0.23	0.00	0.03	947	0.02	0.17	0.00	0.04	378
		<i>kaks</i>	0.07	0.18	0.00	0.03	947	201.09	490.00	2.46	157.20	378
		<i>protid</i>	72.03	77.78	70.00	1.56	947	77.18	80.16	72.63	2.09	378
		<i>dnaid</i>	76.78	81.75	72.43	0.99	947	81.83	83.66	78.95	1.22	378
		<i>kimura</i>	0.35	0.42	0.26	0.03	947	0.29	0.34	0.24	0.02	378
	<i>logdet</i>	0.43	0.81	0.26	0.05	947	0.30	0.50	0.20	0.08	378	
	<i>index</i>	11.51	38.81	5.18	4.48	947	49.58	55.04	40.94	3.70	378	
	AURORA	<i>ks</i>	0.03	0.07	0.01	0.03	4	0.01	0.03	0.00	0.02	3
		<i>ka</i>	0.01	0.02	0.00	0.01	4	0.03	0.04	0.02	0.01	3
		<i>kaks</i>	0.15	0.33	0.00	0.17	4	63.86	103.50	1.08	54.99	3
		<i>protid</i>	96.95	98.39	95.33	1.53	4	95.95	97.30	94.59	1.36	3
		<i>dnaid</i>	98.29	99.46	96.88	1.06	4	97.98	98.80	97.30	0.76	3
		<i>kimura</i>	0.03	0.04	0.01	0.02	4	0.04	0.06	0.03	0.02	3
		<i>logdet</i>	0.03	0.05	0.01	0.02	4	0.08	0.20	0.01	0.10	3
	<i>index</i>	89.34	95.52	81.49	5.82	4	83.75	92.62	74.94	8.84	3	
	BATUMI	<i>ks</i>	0.09	0.21	0.02	0.07	14	0.01	0.04	0.00	0.02	5
		<i>ka</i>	0.04	0.08	0.01	0.03	14	0.03	0.07	0.01	0.02	5
		<i>kaks</i>	0.55	0.94	0.28	0.19	14	55.04	100.00	1.43	49.12	5
		<i>protid</i>	87.54	96.00	70.63	7.54	14	88.74	95.65	75.81	8.11	5
		<i>dnaid</i>	92.31	98.22	78.09	6.22	14	94.20	98.19	85.48	5.21	5
		<i>kimura</i>	0.14	0.39	0.04	0.10	14	0.12	0.29	0.04	0.10	5
		<i>logdet</i>	0.20	0.56	0.02	0.13	14	0.14	0.40	0.02	0.16	5
	<i>index</i>	61.66	88.02	40.32	15.56	14	72.70	90.34	42.78	19.66	5	
	BEL	<i>ks</i>	0.20	9.71	0.00	0.76	1672	0.03	2.21	0.00	0.18	1651
<i>ka</i>		0.08	3.17	0.00	0.24	1672	0.20	8.17	0.00	1.02	1651	
<i>kaks</i>		0.43	1.00	0.00	0.28	1672	88.73	490.00	1.00	90.52	1651	
<i>protid</i>		90.76	100.00	70.00	8.23	1672	90.79	100.00	70.00	7.12	1651	
<i>dnaid</i>		93.56	100.00	73.71	6.47	1672	94.71	100.00	74.71	5.04	1651	
<i>kimura</i>		0.10	0.52	0.00	0.10	1672	0.09	0.49	0.00	0.08	1651	
<i>logdet</i>		0.17	2.16	0.00	0.19	1672	0.15	1.21	0.00	0.19	1651	
<i>index</i>	67.63	99.30	3.31	21.68	1672	72.65	99.72	4.13	20.29	1651		
BILBO	<i>ks</i>	0.19	0.76	0.01	0.16	278	0.02	0.13	0.00	0.04	38	
	<i>ka</i>	0.07	0.23	0.00	0.05	278	0.10	0.50	0.00	0.16	38	
	<i>kaks</i>	0.39	0.97	0.00	0.22	278	64.06	400.00	1.00	83.87	38	
	<i>protid</i>	85.87	100.00	71.35	7.17	278	89.15	99.06	70.48	9.60	38	
	<i>dnaid</i>	88.94	100.00	76.52	6.08	278	93.33	99.37	75.87	6.73	38	
	<i>kimura</i>	0.16	0.46	0.01	0.09	278	0.12	0.34	0.01	0.11	38	
	<i>logdet</i>	0.22	1.12	0.01	0.15	278	0.20	0.88	0.01	0.20	38	
<i>index</i>	54.04	95.29	19.01	17.15	278	68.49	97.25	22.19	23.49	38		
BLOOD	<i>ks</i>	0.19	0.27	0.04	0.13	3						
	<i>ka</i>	0.15	0.22	0.01	0.12	3						
	<i>kaks</i>	0.59	0.82	0.12	0.40	3						
	<i>protid</i>	84.68	96.05	78.99	9.85	3						
	<i>dnaid</i>	88.97	96.05	85.43	6.13	3						
	<i>kimura</i>	0.22	0.31	0.04	0.16	3						
	<i>logdet</i>	0.34	0.59	0.09	0.25	3						
<i>index</i>	47.93	81.15	29.16	28.85	3							
BS	<i>ks</i>	0.09	0.26	0.04	0.07	9	0.09	0.25	0.00	0.11	4	

		<i>ka</i>	0.07	0.25	0.02	0.07	9	0.10	0.25	0.01	0.11	4
		<i>kaks</i>	0.74	0.98	0.42	0.22	9	17.88	68.00	1.01	33.41	4
		<i>protid</i>	82.91	95.51	70.92	7.55	9	84.97	98.57	71.63	13.11	4
		<i>dnaid</i>	87.83	97.75	82.05	5.39	9	90.94	99.52	83.45	8.49	4
		<i>kimura</i>	0.20	0.36	0.05	0.10	9	0.18	0.35	0.01	0.16	4
		<i>logdet</i>	0.29	0.47	0.03	0.13	9	0.19	0.30	0.00	0.14	4
		<i>index</i>	51.85	85.43	30.83	14.89	9	61.12	96.52	31.23	29.10	4
	BURDOCK	<i>ks</i>	0.14	0.14	0.14		1	0.01	0.03	0.00	0.01	8
		<i>ka</i>	0.10	0.10	0.10		1	0.05	0.09	0.03	0.02	8
		<i>kaks</i>	0.69	0.69	0.69		1	38.73	102.83	1.97	50.28	8
		<i>protid</i>	81.90	81.90	81.90		1	87.04	95.80	75.34	7.98	8
		<i>dnaid</i>	87.36	87.36	87.36		1	92.25	98.37	82.19	6.20	8
		<i>kimura</i>	0.20	0.20	0.20		1	0.13	0.23	0.04	0.07	8
		<i>logdet</i>	0.08	0.08	0.08		1	0.26	0.42	0.04	0.15	8
		<i>index</i>	53.45	53.45	53.45		1	61.04	72.83	49.54	7.27	8
	BUT2	<i>ks</i>	0.12	0.29	0.02	0.10	10	0.00	0.00	0.00	0.00	2
		<i>ka</i>	0.07	0.22	0.01	0.08	10	0.03	0.03	0.02	0.01	2
		<i>kaks</i>	0.61	0.85	0.21	0.20	10	105.41	106.33	104.50	1.29	2
		<i>protid</i>	84.75	95.70	71.30	7.38	10	87.97	93.01	82.94	7.12	2
		<i>dnaid</i>	90.77	98.21	81.16	5.14	10	93.44	97.67	89.22	5.98	2
		<i>kimura</i>	0.16	0.30	0.04	0.08	10	0.13	0.19	0.07	0.08	2
		<i>logdet</i>	0.23	0.55	0.02	0.14	10	0.12	0.20	0.03	0.12	2
		<i>index</i>	56.18	87.84	27.54	17.69	10	72.26	83.90	60.62	16.46	2
	CHIMPO	<i>ks</i>	0.27	0.27	0.27		1					
		<i>ka</i>	0.12	0.12	0.12		1					
		<i>kaks</i>	0.44	0.44	0.44		1					
		<i>protid</i>	77.31	77.31	77.31		1					
		<i>dnaid</i>	84.59	84.59	84.59		1					
		<i>kimura</i>	0.26	0.26	0.26		1					
		<i>logdet</i>	0.34	0.34	0.34		1					
		<i>index</i>	36.59	36.59	36.59		1					
	CHOUTO	<i>ks</i>	0.10	0.17	0.04	0.05	6	0.00	0.00	0.00	0.00	2
		<i>ka</i>	0.06	0.12	0.01	0.04	6	0.01	0.02	0.01	0.01	2
		<i>kaks</i>	0.49	0.82	0.31	0.21	6	113.75	126.00	101.50	17.32	2
		<i>protid</i>	91.70	97.14	86.84	4.27	6	96.69	97.22	96.15	0.76	2
		<i>dnaid</i>	93.81	98.33	88.59	3.79	6	98.73	99.07	98.40	0.47	2
		<i>kimura</i>	0.09	0.15	0.03	0.05	6	0.04	0.04	0.03	0.01	2
		<i>logdet</i>	0.18	0.36	0.02	0.11	6	0.03	0.04	0.01	0.02	2
		<i>index</i>	64.95	88.89	51.82	14.81	6	90.84	93.37	88.31	3.58	2
	COPIA	<i>ks</i>	0.11	1.44	0.01	0.14	369	0.01	0.14	0.00	0.02	259
		<i>ka</i>	0.05	0.49	0.00	0.07	369	0.02	0.19	0.00	0.02	259
		<i>kaks</i>	0.52	1.00	0.00	0.26	369	83.70	470.00	1.00	101.17	259
		<i>protid</i>	90.67	100.00	71.21	7.38	369	95.68	99.23	77.04	3.22	259
		<i>dnaid</i>	93.64	100.00	76.02	5.44	369	97.84	99.74	84.20	2.31	259
		<i>kimura</i>	0.09	0.36	0.00	0.07	369	0.04	0.23	0.01	0.03	259
		<i>logdet</i>	0.13	0.50	0.00	0.12	369	0.05	0.43	0.00	0.09	259
		<i>index</i>	69.38	98.12	19.71	18.81	369	86.95	98.06	43.30	10.25	259
	CR1	<i>ks</i>	0.09	0.31	0.00	0.06	429	0.02	0.09	0.00	0.02	125
		<i>ka</i>	0.04	0.16	0.00	0.02	429	0.04	0.10	0.00	0.02	125
		<i>kaks</i>	0.50	1.00	0.00	0.23	429	28.65	410.00	1.01	62.43	125
		<i>protid</i>	87.21	99.33	70.43	5.57	429	90.25	99.16	70.54	5.10	125
		<i>dnaid</i>	91.18	99.78	74.21	4.57	429	94.16	99.72	75.89	3.80	125

		<i>kimura</i>	0.12	0.32	0.01	0.06	429	0.10	0.26	0.01	0.05	125
		<i>logdet</i>	0.18	0.84	0.00	0.13	429	0.14	0.46	0.00	0.09	125
		<i>index</i>	61.59	98.01	30.33	11.54	429	70.80	97.96	45.44	11.42	125
DD37E		<i>ks</i>	4.25	5.23	3.53	0.50	14	0.29	4.33	0.00	1.08	16
		<i>ka</i>	3.10	3.84	2.25	0.58	14	0.39	5.41	0.02	1.34	16
		<i>kaks</i>	0.73	0.89	0.51	0.12	14	30.30	98.71	1.16	43.49	16
		<i>protid</i>	87.95	93.04	73.68	5.01	14	87.85	93.60	74.46	4.41	16
		<i>dnaid</i>	90.73	93.62	75.44	4.69	14	92.77	97.60	83.15	3.08	16
		<i>kimura</i>	0.16	1.05	0.05	0.26	14	0.12	0.31	0.06	0.06	16
		<i>logdet</i>	0.54	1.32	0.35	0.31	14	0.17	0.40	0.04	0.10	16
		<i>index</i>	6.69	8.05	2.70	1.50	14	61.94	78.99	6.70	16.17	16
DIVER		<i>ks</i>	0.17	0.39	0.04	0.11	7	0.06	0.06	0.06		1
		<i>ka</i>	0.11	0.30	0.04	0.09	7	0.06	0.06	0.06		1
		<i>kaks</i>	0.67	0.94	0.57	0.14	7	1.07	1.07	1.07		1
		<i>protid</i>	83.30	89.83	75.57	4.54	7	91.75	91.75	91.75		1
		<i>dnaid</i>	87.73	93.79	80.92	3.89	7	94.50	94.50	94.50		1
		<i>kimura</i>	0.17	0.26	0.10	0.05	7	0.09	0.09	0.09		1
		<i>logdet</i>	0.24	0.27	0.16	0.04	7	0.24	0.24	0.24		1
		<i>index</i>	49.13	67.61	30.78	10.76	7	62.94	62.94	62.94		1
DIVER-BE L		<i>ks</i>	0.14	0.55	0.01	0.13	17	0.05	0.16	0.00	0.06	17
		<i>ka</i>	0.08	0.22	0.00	0.06	17	0.10	0.24	0.00	0.08	17
		<i>kaks</i>	0.60	0.93	0.00	0.27	17	84.73	490.00	1.08	126.12	17
		<i>protid</i>	78.21	96.30	70.59	8.14	17	79.34	95.37	70.52	7.73	17
		<i>dnaid</i>	82.76	97.53	75.26	7.42	17	85.51	97.22	77.78	6.18	17
		<i>kimura</i>	0.24	0.40	0.04	0.12	17	0.21	0.35	0.05	0.10	17
		<i>logdet</i>	0.59	1.84	0.16	0.44	17	0.38	1.01	0.10	0.22	17
		<i>index</i>	39.24	79.46	20.87	14.80	17	46.49	62.18	32.64	9.16	17
DM1731		<i>ks</i>	0.57	1.02	0.19	0.28	15					
		<i>ka</i>	0.32	0.48	0.11	0.17	15					
		<i>kaks</i>	0.56	0.70	0.39	0.11	15					
		<i>protid</i>	86.66	90.62	83.00	3.02	15					
		<i>dnaid</i>	89.51	93.12	86.67	2.64	15					
		<i>kimura</i>	0.13	0.18	0.10	0.03	15					
		<i>logdet</i>	0.22	0.32	0.04	0.12	15					
		<i>index</i>	38.37	61.58	24.51	15.16	15					
DMR_DV		<i>ks</i>	0.22	0.33	0.09	0.08	19	0.12	0.24	0.00	0.10	10
		<i>ka</i>	0.16	0.24	0.03	0.05	19	0.16	0.28	0.05	0.09	10
		<i>kaks</i>	0.74	0.92	0.30	0.15	19	11.80	102.17	1.01	31.77	10
		<i>protid</i>	78.00	93.26	70.47	5.78	19	78.50	90.71	70.05	6.42	10
		<i>dnaid</i>	83.96	96.20	76.34	5.34	19	85.79	95.45	78.47	5.80	10
		<i>kimura</i>	0.22	0.43	0.07	0.09	19	0.25	0.49	0.10	0.13	10
		<i>logdet</i>	0.35	0.54	0.09	0.12	19	0.28	0.41	0.17	0.08	10
		<i>index</i>	38.93	73.36	26.27	11.06	19	43.98	69.07	26.84	12.81	10
DMRT		<i>ks</i>	0.14	0.41	0.01	0.12	43	0.01	0.03	0.00	0.01	4
		<i>ka</i>	0.06	0.18	0.00	0.05	43	0.03	0.05	0.02	0.02	4
		<i>kaks</i>	0.41	0.98	0.00	0.23	43	23.05	86.00	1.35	41.97	4
		<i>protid</i>	87.55	95.07	73.91	7.15	43	96.00	96.69	94.17	1.23	4
		<i>dnaid</i>	90.66	96.48	79.47	5.26	43	97.56	98.82	95.83	1.25	4
		<i>kimura</i>	0.14	0.32	0.05	0.09	43	0.04	0.06	0.04	0.01	4
		<i>logdet</i>	0.25	0.68	0.03	0.19	43	0.04	0.07	0.01	0.03	4
		<i>index</i>	56.67	79.96	28.45	15.06	43	85.23	91.51	80.19	4.68	4
G5		<i>ks</i>	0.19	0.36	0.02	0.10	14	0.02	0.03	0.01	0.01	2

		<i>ka</i>	0.13	0.23	0.00	0.07	14	0.06	0.06	0.05	0.01	2
		<i>kaks</i>	0.67	0.84	0.27	0.16	14	2.99	3.51	2.47	0.74	2
		<i>protid</i>	79.47	93.79	70.59	7.25	14	85.91	89.84	81.99	5.55	2
		<i>dnaid</i>	86.20	94.94	79.66	5.14	14	90.34	93.94	86.75	5.08	2
		<i>kimura</i>	0.24	0.36	0.06	0.10	14	0.16	0.21	0.11	0.07	2
		<i>logdet</i>	0.27	0.47	0.10	0.12	14	0.27	0.32	0.21	0.08	2
		<i>index</i>	44.12	75.38	27.36	14.65	14	58.29	65.62	50.96	10.37	2
	GALILEO	<i>ks</i>	2.89	6.34	0.01	2.93	6	0.01	0.05	0.00	0.02	5
		<i>ka</i>	0.98	3.62	0.00	1.51	6	0.09	0.17	0.00	0.08	5
		<i>kaks</i>	0.30	0.57	0.00	0.25	6	107.54	340.00	1.29	138.44	5
		<i>protid</i>	85.52	98.77	70.30	11.54	6	83.46	94.78	70.69	11.80	5
		<i>dnaid</i>	88.42	99.09	73.06	11.22	6	88.16	96.80	77.39	9.90	5
		<i>kimura</i>	0.09	0.14	0.01	0.05	6	0.16	0.29	0.05	0.11	5
		<i>logdet</i>	0.22	0.50	0.01	0.20	6	0.18	0.44	0.04	0.16	5
		<i>index</i>	34.94	96.03	4.38	35.08	6	61.13	87.56	38.61	20.76	5
	Genome	<i>ks</i>	0.04	0.70	0.00	0.05	2545	0.01	0.03	0.00	0.01	16
		<i>ka</i>	0.01	0.41	0.00	0.02	2545	0.01	0.03	0.00	0.01	16
		<i>kaks</i>	0.12	0.99	0.00	0.15	2545	6.13	10.00	1.01	4.53	16
		<i>protid</i>	98.57	100.00	56.57	3.51	2545	98.62	100.00	95.50	1.50	16
		<i>dnaid</i>	98.44	100.00	57.14	2.98	2545	99.20	100.00	96.87	1.04	16
		<i>kimura</i>	0.01	0.60	0.00	0.03	2545	0.01	0.05	0.00	0.02	16
		<i>logdet</i>	0.05	0.93	0.00	0.10	2545	0.05	0.69	0.00	0.17	16
		<i>index</i>	90.21	99.72	18.93	10.87	2545	93.11	99.57	53.33	11.81	16
	GTWIN	<i>ks</i>						0.06	0.09	0.01	0.04	3
		<i>ka</i>						0.07	0.10	0.02	0.04	3
		<i>kaks</i>						1.38	1.86	1.11	0.41	3
		<i>protid</i>						77.89	85.81	70.80	7.54	3
		<i>dnaid</i>						87.49	89.89	82.89	3.98	3
		<i>kimura</i>						0.26	0.37	0.13	0.12	3
		<i>logdet</i>						0.31	0.40	0.19	0.11	3
		<i>index</i>					48.45	64.14	37.99	13.84	3	
	GYPSY	<i>ks</i>	0.13	8.01	0.00	0.38	4759	0.03	0.26	0.00	0.03	3131
		<i>ka</i>	0.05	0.76	0.00	0.06	4759	0.05	1.51	0.00	0.05	3131
		<i>kaks</i>	0.44	1.00	0.00	0.29	4759	60.85	490.00	1.00	105.75	3131
		<i>protid</i>	89.73	100.00	70.00	7.73	4759	89.73	100.00	70.07	7.35	3131
		<i>dnaid</i>	92.78	100.00	70.96	5.82	4759	94.02	100.00	74.39	5.14	3131
		<i>kimura</i>	0.11	3.86	0.00	0.12	4759	0.11	2.66	0.00	0.11	3131
		<i>logdet</i>	0.18	1.42	0.00	0.17	4759	0.16	1.51	0.00	0.17	3131
		<i>index</i>	66.01	99.46	4.17	19.22	4759	70.71	99.71	16.73	17.32	3131
	HARBINGER	<i>ks</i>	0.02	0.03	0.01	0.01	4	0.01	0.01	0.00	0.01	4
		<i>ka</i>	0.00	0.01	0.00	0.01	4	0.01	0.01	0.00	0.01	4
		<i>kaks</i>	0.24	0.36	0.00	0.17	4	107.30	320.00	1.09	150.34	4
		<i>protid</i>	98.74	99.68	98.39	0.63	4	96.89	97.74	96.05	0.69	4
		<i>dnaid</i>	99.25	99.46	98.93	0.23	4	98.88	99.25	98.68	0.25	4
		<i>kimura</i>	0.02	0.02	0.00	0.01	4	0.03	0.04	0.02	0.01	4
		<i>logdet</i>	0.19	0.34	0.04	0.17	4	0.36	0.38	0.34	0.02	4
		<i>index</i>	81.70	92.79	72.18	11.05	4	69.39	70.33	67.79	1.10	4
	HAT	<i>ks</i>	0.14	0.49	0.01	0.10	56	0.03	0.21	0.00	0.04	30
		<i>ka</i>	0.07	0.17	0.00	0.04	56	0.06	0.22	0.00	0.04	30
		<i>kaks</i>	0.51	0.98	0.00	0.22	56	26.38	400.00	1.03	76.65	30
		<i>protid</i>	85.92	100.00	70.63	7.98	56	87.73	98.40	73.19	5.81	30
		<i>dnaid</i>	91.01	99.72	78.52	5.42	56	93.53	99.43	77.29	4.32	30

	<i>kimura</i>	0.13	0.33	0.00	0.09	56	0.12	0.26	0.02	0.06	30
	<i>logdet</i>	0.19	0.97	0.00	0.19	56	0.14	0.39	0.01	0.09	30
	<i>index</i>	59.28	98.40	30.76	18.53	56	67.69	95.61	39.86	12.14	30
HELENA	<i>ks</i>	0.02	0.02	0.02	0.00	2					
	<i>ka</i>	0.01	0.01	0.01	0.00	2					
	<i>kaks</i>	0.45	0.54	0.35	0.13	2					
	<i>protid</i>	94.61	98.58	90.64	5.61	2					
	<i>dnaid</i>	95.62	99.05	92.20	4.84	2					
	<i>kimura</i>	0.06	0.11	0.01	0.07	2					
	<i>logdet</i>	0.10	0.14	0.05	0.06	2					
	<i>index</i>	79.62	83.02	76.22	4.81	2					
HELITRON	<i>ks</i>	0.17	0.58	0.00	0.12	152	0.04	0.16	0.00	0.04	115
	<i>ka</i>	0.08	0.20	0.00	0.06	152	0.06	0.19	0.00	0.05	115
	<i>kaks</i>	0.51	0.98	0.00	0.29	152	70.73	480.00	1.01	133.06	115
	<i>protid</i>	85.01	100.00	70.37	9.23	152	89.45	100.00	70.00	9.04	115
	<i>dnaid</i>	90.31	100.00	77.96	6.04	152	94.12	100.00	81.14	5.51	115
	<i>kimura</i>	0.15	0.33	0.00	0.10	152	0.11	0.39	0.00	0.10	115
	<i>logdet</i>	0.30	0.93	0.00	0.21	152	0.18	0.86	0.00	0.19	115
	<i>index</i>	53.17	99.26	25.04	22.59	152	68.92	99.70	33.46	20.37	115
HOANA	<i>ks</i>	0.02	0.02	0.01	0.01	4					
	<i>ka</i>	0.00	0.00	0.00	0.00	4					
	<i>kaks</i>	0.00	0.00	0.00	0.00	4					
	<i>protid</i>	99.02	99.02	99.02	0.00	4					
	<i>dnaid</i>	99.27	99.35	99.02	0.16	4					
	<i>kimura</i>	0.01	0.01	0.01	0.00	4					
	<i>logdet</i>	0.01	0.01	0.01	0.00	4					
	<i>index</i>	96.09	96.47	95.41	0.50	4					
HOYAK	<i>ks</i>	0.03	0.05	0.01	0.02	8	0.03	0.04	0.01	0.01	10
	<i>ka</i>	0.03	0.05	0.01	0.01	8	0.03	0.05	0.01	0.02	10
	<i>kaks</i>	0.77	0.91	0.70	0.09	8	1.07	1.17	1.02	0.06	10
	<i>protid</i>	89.93	98.27	86.67	4.58	8	88.44	97.33	84.42	4.41	10
	<i>dnaid</i>	93.17	99.23	90.60	3.29	8	92.86	98.89	90.26	2.81	10
	<i>kimura</i>	0.08	0.13	0.02	0.05	8	0.14	0.18	0.03	0.05	10
	<i>logdet</i>	0.16	0.23	0.01	0.09	8	0.26	0.41	0.09	0.13	10
	<i>index</i>	70.99	94.58	61.38	11.66	8	62.35	85.87	54.26	11.72	10
ISIS	<i>ks</i>	0.02	0.02	0.02		1					
	<i>ka</i>	0.00	0.00	0.00		1					
	<i>kaks</i>	0.00	0.00	0.00		1					
	<i>protid</i>	99.18	99.18	99.18		1					
	<i>dnaid</i>	99.73	99.73	99.73		1					
	<i>kimura</i>	0.01	0.01	0.01		1					
	<i>logdet</i>	0.00	0.00	0.00		1					
	<i>index</i>	96.45	96.45	96.45		1					
IVK	<i>ks</i>	0.02	0.03	0.01	0.01	3	0.14	0.18	0.07	0.06	3
	<i>ka</i>	0.01	0.02	0.00	0.01	3	0.16	0.19	0.11	0.05	3
	<i>kaks</i>	0.43	0.64	0.00	0.37	3	1.22	1.52	1.06	0.26	3
	<i>protid</i>	96.49	97.60	95.93	0.96	3	78.53	82.61	76.16	3.55	3
	<i>dnaid</i>	98.20	98.93	97.83	0.64	3	86.43	91.30	83.89	4.22	3
	<i>kimura</i>	0.03	0.04	0.02	0.01	3	0.29	0.32	0.24	0.04	3
	<i>logdet</i>	0.18	0.26	0.01	0.14	3	0.32	0.33	0.30	0.02	3
	<i>index</i>	78.29	94.01	70.43	13.61	3	39.95	48.01	35.66	6.98	3
JOCKEY	<i>ks</i>	0.26	3.26	0.01	0.64	25	0.04	0.08	0.00	0.03	14

		<i>ka</i>	0.08	0.25	0.00	0.06	25	0.05	0.10	0.00	0.03	14
		<i>kaks</i>	0.64	0.99	0.01	0.26	25	53.81	440.00	1.03	119.03	14
		<i>protid</i>	84.89	98.76	70.68	8.72	25	85.35	96.53	75.23	8.88	14
		<i>dnaid</i>	90.03	99.38	75.66	6.80	25	92.30	98.65	82.49	5.17	14
		<i>kimura</i>	0.16	0.33	0.01	0.10	25	0.16	0.30	0.04	0.10	14
		<i>logdet</i>	0.33	0.82	0.01	0.21	25	0.18	0.38	0.01	0.11	14
		<i>index</i>	51.99	95.98	10.89	21.28	25	63.64	92.53	40.88	17.86	14
	JOCKEY-LIKE	<i>ks</i>	0.13	0.61	0.01	0.15	24	0.01	0.15	0.00	0.03	41
		<i>ka</i>	0.07	0.43	0.00	0.12	24	0.05	0.17	0.00	0.03	41
		<i>kaks</i>	0.31	0.82	0.00	0.30	24	99.55	410.00	1.10	56.00	41
		<i>protid</i>	88.40	97.30	75.25	7.38	24	88.19	99.03	73.42	5.53	41
		<i>dnaid</i>	91.41	97.75	79.29	5.85	24	92.70	99.68	79.32	4.36	41
		<i>kimura</i>	0.12	0.24	0.03	0.07	24	0.12	0.24	0.01	0.06	41
		<i>logdet</i>	0.28	0.54	0.08	0.12	24	0.32	0.99	0.09	0.20	41
	<i>index</i>	56.80	82.90	25.15	14.26	24	61.02	89.70	32.55	11.32	41	
	LIAN-AAL-LIKE	<i>ks</i>	0.10	0.23	0.01	0.08	10	0.02	0.04	0.00	0.02	3
		<i>ka</i>	0.05	0.14	0.00	0.05	10	0.03	0.05	0.01	0.02	3
		<i>kaks</i>	0.40	0.73	0.00	0.28	10	40.38	119.00	1.02	68.08	3
		<i>protid</i>	91.75	99.06	74.45	8.20	10	96.08	97.08	94.17	1.66	3
		<i>dnaid</i>	93.86	99.69	83.70	5.58	10	97.59	98.67	95.47	1.84	3
		<i>kimura</i>	0.09	0.30	0.01	0.09	10	0.04	0.06	0.03	0.02	3
		<i>logdet</i>	0.12	0.42	0.00	0.15	10	0.03	0.03	0.02	0.01	3
	<i>index</i>	70.75	96.98	36.27	20.63	10	86.75	91.22	79.48	6.35	3	
	LINEJ1	<i>ks</i>	0.17	0.24	0.11	0.06	7					
		<i>ka</i>	0.07	0.10	0.03	0.03	7					
		<i>kaks</i>	0.41	0.64	0.13	0.19	7					
		<i>protid</i>	85.64	90.10	78.74	4.35	7					
		<i>dnaid</i>	88.27	91.75	84.25	2.65	7					
		<i>kimura</i>	0.15	0.21	0.09	0.04	7					
		<i>logdet</i>	0.18	0.29	0.10	0.07	7					
	<i>index</i>	53.43	59.99	43.61	6.81	7						
	LINER1	<i>ks</i>	0.08	0.37	0.01	0.10	85	0.02	0.05	0.00	0.03	3
		<i>ka</i>	0.03	0.23	0.00	0.04	85	0.04	0.07	0.01	0.03	3
		<i>kaks</i>	0.23	0.90	0.00	0.27	85	77.05	122.00	1.47	65.84	3
		<i>protid</i>	87.18	100.00	70.00	7.97	85	94.32	98.63	89.09	4.84	3
		<i>dnaid</i>	90.19	99.28	77.44	6.02	85	96.98	99.09	94.24	2.49	3
		<i>kimura</i>	0.13	0.32	0.01	0.08	85	0.06	0.12	0.01	0.06	3
		<i>logdet</i>	0.34	1.28	0.01	0.28	85	0.25	0.32	0.19	0.07	3
	<i>index</i>	58.02	96.15	28.07	15.51	85	70.45	78.83	57.12	11.67	3	
	MARINER	<i>ks</i>	0.06	0.17	0.01	0.04	40	0.02	0.06	0.00	0.01	93
		<i>ka</i>	0.04	0.12	0.00	0.03	40	0.05	0.08	0.01	0.01	93
		<i>kaks</i>	0.59	1.00	0.00	0.28	40	8.69	106.50	1.01	22.66	93
		<i>protid</i>	91.37	100.00	73.85	5.97	40	89.30	98.44	70.77	5.09	93
		<i>dnaid</i>	95.13	100.00	80.77	4.28	40	94.59	99.22	73.85	4.77	93
		<i>kimura</i>	0.09	0.22	0.00	0.06	40	0.10	0.27	0.02	0.04	93
		<i>logdet</i>	0.09	0.31	0.00	0.07	40	0.09	0.19	0.02	0.05	93
	<i>index</i>	74.03	98.81	44.14	13.87	40	72.16	92.35	56.79	7.38	93	
	MAX	<i>ks</i>	0.10	0.31	0.06	0.09	8					
		<i>ka</i>	0.02	0.08	0.00	0.03	8					
		<i>kaks</i>	0.25	0.88	0.00	0.38	8					
		<i>protid</i>	94.00	97.87	86.02	4.56	8					
		<i>dnaid</i>	96.37	98.82	90.68	3.27	8					

		<i>kimura</i>	0.07	0.16	0.02	0.05	8						
		<i>logdet</i>	0.11	0.58	0.01	0.20	8						
		<i>index</i>	76.40	89.69	37.61	18.59	8						
MDG1		<i>ks</i>	0.14	0.26	0.02	0.12	4						
		<i>ka</i>	0.12	0.23	0.01	0.10	4						
		<i>kaks</i>	0.83	0.90	0.76	0.08	4						
		<i>protid</i>	85.77	97.09	74.36	10.02	4						
		<i>dnaid</i>	89.34	98.06	80.34	8.36	4						
		<i>kimura</i>	0.15	0.29	0.02	0.12	4						
		<i>logdet</i>	0.14	0.19	0.02	0.08	4						
		<i>index</i>	58.31	90.90	35.38	24.94	4						
MINIME		<i>ks</i>	0.14	0.47	0.01	0.08	208	0.04	0.16	0.00	0.04	103	
		<i>ka</i>	0.08	0.27	0.00	0.05	208	0.06	0.25	0.00	0.05	103	
		<i>kaks</i>	0.58	1.00	0.00	0.24	208	51.08	460.00	1.01	98.33	103	
		<i>protid</i>	83.60	100.00	70.00	7.57	208	86.96	99.15	70.00	9.08	103	
		<i>dnaid</i>	88.98	99.81	73.00	5.48	208	92.55	99.72	72.64	5.90	103	
		<i>kimura</i>	0.18	0.40	0.00	0.09	208	0.14	0.34	0.01	0.10	103	
		<i>logdet</i>	0.21	0.71	0.00	0.14	208	0.14	0.68	0.00	0.13	103	
		<i>index</i>	53.97	97.76	27.05	16.15	208	67.54	97.57	30.82	18.63	103	
NETR1B		<i>ks</i>						0.01	0.01	0.01		1	
		<i>ka</i>						0.01	0.01	0.01		1	
		<i>kaks</i>						1.14	1.14	1.14		1	
		<i>protid</i>						97.93	97.93	97.93		1	
		<i>dnaid</i>						98.85	98.85	98.85		1	
		<i>kimura</i>						0.02	0.02	0.02		1	
		<i>logdet</i>						0.01	0.01	0.01		1	
		<i>index</i>					93.95	93.95	93.95		1		
NINJA		<i>ks</i>	0.07	0.13	0.01	0.03	16	0.04	0.14	0.00	0.05	16	
		<i>ka</i>	0.03	0.04	0.00	0.01	16	0.07	0.19	0.00	0.08	16	
		<i>kaks</i>	0.46	0.69	0.27	0.13	16	52.23	490.00	1.21	124.31	16	
		<i>protid</i>	91.76	95.87	85.00	2.76	16	88.50	99.20	78.51	6.44	16	
		<i>dnaid</i>	95.64	97.87	89.17	1.99	16	92.89	99.47	85.52	4.45	16	
		<i>kimura</i>	0.09	0.13	0.04	0.03	16	0.12	0.23	0.01	0.07	16	
		<i>logdet</i>	0.15	0.29	0.02	0.09	16	0.15	0.39	0.01	0.12	16	
		<i>index</i>	69.40	89.00	58.59	8.98	16	68.08	96.08	38.88	17.80	16	
NOBEL-BEL		<i>ks</i>	0.11	7.37	0.01	0.35	1905	0.01	0.22	0.00	0.03	1908	
		<i>ka</i>	0.04	2.81	0.00	0.08	1905	0.03	0.62	0.00	0.04	1908	
		<i>kaks</i>	0.41	1.00	0.00	0.32	1905	116.46	490.00	1.00	129.25	1908	
		<i>protid</i>	88.56	100.00	70.09	7.64	1905	87.31	100.00	70.00	7.77	1908	
		<i>dnaid</i>	92.08	100.00	70.11	6.11	1905	91.60	100.00	72.49	6.52	1908	
		<i>kimura</i>	0.12	2.91	0.00	0.10	1905	0.11	4.91	0.00	0.14	1908	
		<i>logdet</i>	0.17	1.40	0.00	0.16	1905	0.18	1.57	0.00	0.17	1908	
		<i>index</i>	66.29	99.28	1.72	17.32	1905	68.84	99.61	13.88	13.52	1908	
OSVALDO		<i>ks</i>	0.16	0.30	0.11	0.09	4	0.00	0.00	0.00	0.00	3	
		<i>ka</i>	0.03	0.05	0.02	0.01	4	0.02	0.02	0.02	0.00	3	
		<i>kaks</i>	0.24	0.44	0.09	0.15	4	118.50	118.50	118.50	0.00	3	
		<i>protid</i>	90.41	92.06	86.96	2.33	4	81.91	83.87	77.98	3.40	3	
		<i>dnaid</i>	93.38	95.01	89.23	2.77	4	88.83	91.04	84.40	3.83	3	
		<i>kimura</i>	0.10	0.15	0.08	0.03	4	0.23	0.33	0.18	0.09	3	
		<i>logdet</i>	0.38	0.56	0.22	0.17	4	0.36	0.42	0.25	0.10	3	
		<i>index</i>	52.70	62.93	47.76	6.92	4	52.48	53.76	50.21	1.97	3	
P_ELEME		<i>ks</i>	0.04	0.09	0.01	0.03	7	0.01	0.04	0.00	0.02	7	

NT	<i>ka</i>	0.01	0.03	0.00	0.01	7	0.03	0.05	0.02	0.01	7
	<i>kaks</i>	0.12	0.30	0.00	0.15	7	84.12	107.50	1.05	37.75	7
	<i>protid</i>	91.15	100.00	81.72	6.60	7	89.55	96.45	80.39	5.17	7
	<i>dnaid</i>	93.68	100.00	87.43	4.71	7	94.60	97.16	89.87	2.87	7
	<i>kimura</i>	0.08	0.22	0.00	0.08	7	0.11	0.23	0.04	0.06	7
	<i>logdet</i>	0.18	0.44	0.00	0.19	7	0.17	0.41	0.03	0.17	7
	<i>index</i>	74.31	97.93	48.81	20.35	7	71.50	88.51	50.80	13.88	7
PAO	<i>ks</i>	0.15	0.51	0.02	0.14	29	0.02	0.05	0.00	0.02	17
	<i>ka</i>	0.08	0.24	0.00	0.07	29	0.03	0.07	0.01	0.02	17
	<i>kaks</i>	0.54	0.98	0.00	0.25	29	42.03	134.00	1.02	57.44	17
	<i>protid</i>	85.37	99.07	70.55	7.86	29	87.55	99.00	74.14	6.99	17
	<i>dnaid</i>	90.28	99.69	78.91	5.64	29	92.15	99.67	79.89	5.63	17
	<i>kimura</i>	0.15	0.36	0.01	0.10	29	0.14	0.26	0.01	0.07	17
	<i>logdet</i>	0.41	1.00	0.00	0.23	29	0.30	0.41	0.03	0.12	17
<i>index</i>	50.25	95.92	23.50	18.25	29	61.04	94.48	45.85	13.63	17	
PIF	<i>ks</i>	0.08	0.27	0.01	0.10	7	0.01	0.02	0.00	0.01	7
	<i>ka</i>	0.05	0.19	0.01	0.07	7	0.02	0.04	0.01	0.01	7
	<i>kaks</i>	0.70	0.96	0.29	0.23	7	42.22	111.00	1.04	51.54	7
	<i>protid</i>	89.63	98.49	71.96	9.72	7	92.70	98.14	86.31	4.56	7
	<i>dnaid</i>	93.43	99.16	83.18	6.14	7	95.52	99.17	92.06	2.96	7
	<i>kimura</i>	0.11	0.34	0.02	0.12	7	0.07	0.13	0.01	0.05	7
	<i>logdet</i>	0.44	1.12	0.14	0.34	7	0.21	0.44	0.09	0.13	7
<i>index</i>	58.35	80.59	24.57	20.92	7	71.70	81.04	53.98	8.57	7	
POLINTON	<i>ks</i>	0.21	2.36	0.01	0.32	184	0.00	0.03	0.00	0.01	31
	<i>ka</i>	0.03	0.17	0.00	0.04	184	0.13	0.30	0.00	0.13	31
	<i>kaks</i>	0.21	0.98	0.00	0.28	184	106.18	370.00	1.08	88.75	31
	<i>protid</i>	88.57	100.00	73.75	6.60	184	89.10	100.00	71.62	8.11	31
	<i>dnaid</i>	92.26	100.00	79.58	5.29	184	92.78	100.00	79.17	6.24	31
	<i>kimura</i>	0.14	2.90	0.00	0.22	184	0.12	0.34	0.00	0.09	31
	<i>logdet</i>	0.19	1.35	0.00	0.21	184	0.18	0.74	0.00	0.19	31
<i>index</i>	60.91	98.83	5.72	18.60	184	66.75	99.73	27.96	21.12	31	
PROTOP	<i>ks</i>						0.00	0.00	0.00		1
	<i>ka</i>						0.01	0.01	0.01		1
	<i>kaks</i>						4.12	4.12	4.12		1
	<i>protid</i>						97.89	97.89	97.89		1
	<i>dnaid</i>						98.77	98.77	98.77		1
	<i>kimura</i>						0.02	0.02	0.02		1
	<i>logdet</i>						0.18	0.18	0.18		1
<i>index</i>						80.60	80.60	80.60		1	
QUETZAL	<i>ks</i>						0.15	0.15	0.15		1
	<i>ka</i>						0.17	0.17	0.17		1
	<i>kaks</i>						1.16	1.16	1.16		1
	<i>protid</i>						71.05	71.05	71.05		1
	<i>dnaid</i>						85.38	85.38	85.38		1
	<i>kimura</i>						0.37	0.37	0.37		1
	<i>logdet</i>						0.23	0.23	0.23		1
<i>index</i>						37.09	37.09	37.09		1	
R1	<i>ks</i>	0.08	0.47	0.01	0.09	333	0.02	0.16	0.00	0.03	278
	<i>ka</i>	0.04	0.33	0.00	0.05	333	0.04	0.22	0.00	0.04	278
	<i>kaks</i>	0.53	0.98	0.00	0.29	333	55.88	440.00	1.00	55.82	278
	<i>protid</i>	88.66	99.50	70.25	8.39	333	89.28	98.77	70.20	7.62	278
	<i>dnaid</i>	91.73	99.66	76.55	6.21	333	93.34	99.59	78.48	5.72	278

	<i>kimura</i>	0.12	0.39	0.01	0.10	333	0.11	0.37	0.01	0.09	278
	<i>logdet</i>	0.19	0.68	0.00	0.15	333	0.20	0.70	0.00	0.17	278
	<i>index</i>	65.12	97.44	22.16	18.92	333	69.46	96.50	36.45	18.24	278
ROOA	<i>ks</i>	0.14	0.14	0.14	0.00	2					
	<i>ka</i>	0.07	0.07	0.07	0.00	2					
	<i>kaks</i>	0.50	0.52	0.48	0.03	2					
	<i>protid</i>	89.60	90.10	89.11	0.70	2					
	<i>dnaid</i>	92.57	92.74	92.41	0.23	2					
	<i>kimura</i>	0.11	0.11	0.10	0.01	2					
	<i>logdet</i>	0.10	0.10	0.09	0.01	2					
	<i>index</i>	63.23	64.00	62.46	1.09	2					
SPOCK	<i>ks</i>	0.06	0.19	0.01	0.04	36	0.08	0.20	0.00	0.08	15
	<i>ka</i>	0.03	0.08	0.00	0.02	36	0.10	0.26	0.01	0.10	15
	<i>kaks</i>	0.57	0.99	0.00	0.25	36	16.93	145.00	1.04	42.59	15
	<i>protid</i>	93.38	98.39	87.34	3.12	36	89.27	98.39	71.86	8.12	15
	<i>dnaid</i>	96.15	99.19	91.56	2.02	36	93.91	99.46	81.84	5.30	15
	<i>kimura</i>	0.07	0.12	0.02	0.03	36	0.11	0.24	0.02	0.09	15
	<i>logdet</i>	0.14	1.23	0.01	0.27	36	0.17	0.37	0.00	0.14	15
	<i>index</i>	75.02	94.99	30.10	14.29	36	65.72	95.98	37.83	21.62	15
STALKER	<i>ks</i>	0.07	0.12	0.01	0.04	24	0.01	0.04	0.00	0.02	34
	<i>ka</i>	0.04	0.08	0.00	0.02	24	0.02	0.06	0.00	0.02	34
	<i>kaks</i>	0.67	0.98	0.24	0.17	24	41.11	470.00	1.02	88.11	34
	<i>protid</i>	91.54	99.12	85.32	4.82	24	95.52	99.08	87.16	3.72	34
	<i>dnaid</i>	95.31	99.42	92.35	2.37	24	97.98	99.69	94.80	1.51	34
	<i>kimura</i>	0.09	0.16	0.01	0.05	24	0.05	0.14	0.01	0.04	34
	<i>logdet</i>	0.06	0.09	0.01	0.02	24	0.03	0.06	0.00	0.02	34
	<i>index</i>	74.05	94.67	59.96	12.06	24	87.43	97.71	68.94	8.93	34
TABOR	<i>ks</i>	1.46	9.45	0.01	2.83	506	0.00	0.12	0.00	0.02	135
	<i>ka</i>	0.12	1.13	0.00	0.18	506	0.08	0.75	0.00	0.15	135
	<i>kaks</i>	0.20	0.96	0.00	0.22	506	174.99	470.00	1.02	130.35	135
	<i>protid</i>	84.47	100.00	70.00	10.45	506	89.69	100.00	71.96	8.56	135
	<i>dnaid</i>	89.00	100.00	77.68	7.53	506	93.05	99.73	77.68	6.84	135
	<i>kimura</i>	0.18	0.42	0.00	0.12	506	0.11	0.35	0.01	0.10	135
	<i>logdet</i>	0.19	0.59	0.00	0.14	506	0.12	0.50	0.00	0.12	135
	<i>index</i>	52.74	99.14	3.59	31.18	506	72.78	97.94	23.08	20.35	135
TAHRE	<i>ks</i>	0.12	0.12	0.12		1	0.08	0.08	0.08		1
	<i>ka</i>	0.09	0.09	0.09		1	0.11	0.11	0.11		1
	<i>kaks</i>	0.80	0.80	0.80		1	1.45	1.45	1.45		1
	<i>protid</i>	90.65	90.65	90.65		1	80.83	80.83	80.83		1
	<i>dnaid</i>	91.28	91.28	91.28		1	87.50	87.50	87.50		1
	<i>kimura</i>	0.10	0.10	0.10		1	0.19	0.19	0.19		1
	<i>logdet</i>	0.32	0.32	0.32		1	0.23	0.23	0.23		1
	<i>index</i>	52.95	52.95	52.95		1	50.11	50.11	50.11		1
TART	<i>ks</i>	0.08	0.42	0.01	0.09	55	0.02	0.07	0.00	0.03	26
	<i>ka</i>	0.02	0.17	0.00	0.04	55	0.05	0.11	0.00	0.03	26
	<i>kaks</i>	0.16	0.55	0.00	0.18	55	64.19	99.00	1.06	46.58	26
	<i>protid</i>	83.38	97.45	70.07	7.83	55	84.11	95.41	71.60	7.04	26
	<i>dnaid</i>	86.93	98.47	76.91	6.49	55	89.40	97.86	76.82	6.15	26
	<i>kimura</i>	0.17	0.35	0.03	0.09	55	0.16	0.36	0.04	0.08	26
	<i>logdet</i>	0.50	1.49	0.07	0.27	55	0.49	0.94	0.09	0.23	26
	<i>index</i>	48.80	85.68	26.20	12.73	55	50.30	66.71	34.71	8.77	26
TC1	<i>ks</i>	0.12	0.30	0.01	0.09	7	0.02	0.07	0.00	0.04	3

		<i>ka</i>	0.07	0.20	0.00	0.06	7	0.03	0.08	0.00	0.05	3
		<i>kaks</i>	0.49	0.76	0.00	0.27	7	280.35	420.00	1.05	241.88	3
		<i>protid</i>	85.92	95.08	77.07	7.29	7	92.59	99.03	79.70	11.16	3
		<i>dnaid</i>	89.51	97.27	81.10	6.17	7	96.25	99.68	89.72	5.66	3
		<i>kimura</i>	0.15	0.24	0.05	0.07	7	0.09	0.26	0.01	0.14	3
		<i>logdet</i>	0.25	0.43	0.03	0.15	7	0.12	0.21	0.00	0.11	3
		<i>index</i>	55.59	87.88	35.64	17.16	7	77.39	97.62	50.41	24.32	3
	TELEMAC	<i>ks</i>	0.02	0.04	0.01	0.02	3	0.03	0.10	0.00	0.04	9
		<i>ka</i>	0.01	0.02	0.01	0.01	3	0.06	0.10	0.02	0.03	9
		<i>kaks</i>	0.61	0.63	0.60	0.02	3	44.83	106.00	1.03	51.46	9
		<i>protid</i>	97.12	99.04	94.35	2.46	3	78.76	95.16	70.20	8.68	9
		<i>dnaid</i>	98.48	99.09	97.31	1.01	3	85.24	98.12	79.91	6.50	9
		<i>kimura</i>	0.03	0.06	0.01	0.03	3	0.24	0.35	0.05	0.10	9
		<i>logdet</i>	0.04	0.05	0.03	0.01	3	0.23	0.33	0.02	0.12	9
		<i>index</i>	88.37	91.42	82.85	4.79	3	53.16	87.42	40.48	14.48	9
	TF001012	<i>ks</i>	0.13	0.28	0.02	0.08	13	0.02	0.02	0.02	0.00	2
		<i>ka</i>	0.09	0.17	0.01	0.06	13	0.02	0.02	0.02	0.00	2
		<i>kaks</i>	0.76	1.00	0.36	0.22	13	1.09	1.09	1.09	0.00	2
		<i>protid</i>	81.63	98.01	72.06	8.26	13	96.15	96.50	95.80	0.49	2
		<i>dnaid</i>	89.15	98.90	81.13	5.34	13	98.48	98.60	98.37	0.16	2
		<i>kimura</i>	0.21	0.31	0.02	0.10	13	0.04	0.04	0.04	0.00	2
		<i>logdet</i>	0.23	0.55	0.01	0.18	13	0.02	0.02	0.02	0.00	2
		<i>index</i>	51.86	91.96	30.28	19.55	13	89.02	89.63	88.40	0.87	2
	TIANG	<i>ks</i>	0.02	0.02	0.02	0.00	4	0.01	0.01	0.01	0.00	2
		<i>ka</i>	0.01	0.01	0.00	0.01	4	0.02	0.02	0.02	0.00	2
		<i>kaks</i>	0.52	0.77	0.15	0.26	4	1.18	1.18	1.18	0.00	2
		<i>protid</i>	95.03	97.16	93.75	1.49	4	96.03	97.73	94.32	2.41	2
		<i>dnaid</i>	97.63	98.48	97.16	0.59	4	97.91	98.67	97.16	1.07	2
		<i>kimura</i>	0.05	0.07	0.03	0.02	4	0.04	0.06	0.02	0.03	2
		<i>logdet</i>	0.03	0.03	0.02	0.00	4	0.02	0.03	0.01	0.01	2
		<i>index</i>	86.96	90.73	84.31	2.70	4	88.70	91.98	85.42	4.64	2
	TIRANT	<i>ks</i>	0.12	0.12	0.12		1					
		<i>ka</i>	0.08	0.08	0.08		1					
		<i>kaks</i>	0.62	0.62	0.62		1					
		<i>protid</i>	89.57	89.57	89.57		1					
		<i>dnaid</i>	91.59	91.59	91.59		1					
		<i>kimura</i>	0.11	0.11	0.11		1					
		<i>logdet</i>	0.07	0.07	0.07		1					
		<i>index</i>	63.83	63.83	63.83		1					
	TRAM-BEL	<i>ks</i>	0.08	0.29	0.01	0.05	114	0.02	0.09	0.00	0.02	113
		<i>ka</i>	0.04	0.20	0.00	0.04	114	0.04	0.11	0.00	0.02	113
		<i>kaks</i>	0.55	1.00	0.00	0.22	114	16.80	440.00	1.00	53.04	113
		<i>protid</i>	89.23	100.00	70.43	8.27	114	90.64	99.04	71.97	5.00	113
		<i>dnaid</i>	92.81	99.73	81.45	5.93	114	95.60	99.68	79.80	3.53	113
		<i>kimura</i>	0.11	0.42	0.00	0.09	114	0.10	0.29	0.01	0.05	113
		<i>logdet</i>	0.08	0.30	0.00	0.07	114	0.11	0.35	0.00	0.09	113
		<i>index</i>	70.47	98.29	29.71	16.73	114	73.23	97.66	48.35	10.81	113
	TRANSIB	<i>ks</i>	0.09	0.33	0.01	0.08	31	0.03	0.16	0.00	0.03	45
		<i>ka</i>	0.05	0.13	0.00	0.04	31	0.05	0.22	0.01	0.04	45
		<i>kaks</i>	0.64	0.92	0.00	0.22	31	23.17	108.33	1.01	40.89	45
		<i>protid</i>	84.09	98.41	70.25	9.07	31	87.29	96.55	70.19	7.81	45
		<i>dnaid</i>	89.87	98.27	77.60	6.57	31	93.11	98.47	80.16	5.14	45

		<i>kimura</i>	0.18	0.35	0.02	0.10	31	0.14	0.35	0.04	0.09	45
		<i>logdet</i>	0.21	0.53	0.02	0.14	31	0.23	0.56	0.03	0.12	45
		<i>index</i>	58.16	89.36	32.00	17.30	31	62.83	86.47	38.25	13.96	45
	TRANSPAC	<i>ks</i>	0.03	0.03	0.03		1					
		<i>ka</i>	0.00	0.00	0.00		1					
		<i>kaks</i>	0.10	0.10	0.10		1					
		<i>protid</i>	98.43	98.43	98.43		1					
		<i>dnaid</i>	98.60	98.60	98.60		1					
		<i>kimura</i>	0.02	0.02	0.02		1					
		<i>logdet</i>	0.01	0.01	0.01		1					
		<i>index</i>	92.23	92.23	92.23		1					
	TRIM	<i>ks</i>	0.10	0.59	0.01	0.09	244	0.04	0.29	0.00	0.05	1079
		<i>ka</i>	0.07	0.26	0.00	0.06	244	0.09	0.38	0.00	0.06	1079
		<i>kaks</i>	0.59	1.00	0.00	0.28	244	33.82	480.00	1.01	51.86	1079
		<i>protid</i>	88.01	100.00	70.56	7.87	244	86.99	100.00	70.13	5.91	1079
		<i>dnaid</i>	92.06	100.00	76.61	5.64	244	92.88	100.00	78.86	4.12	1079
		<i>kimura</i>	0.13	0.41	0.00	0.09	244	0.14	0.41	0.00	0.07	1079
		<i>logdet</i>	0.18	0.70	0.00	0.14	244	0.18	1.64	0.00	0.15	1079
		<i>index</i>	62.39	98.97	25.36	18.24	244	62.95	99.60	21.18	14.63	1079
	ULYSSES	<i>ks</i>	0.01	0.01	0.01		1					
		<i>ka</i>	0.00	0.00	0.00		1					
		<i>kaks</i>	0.00	0.00	0.00		1					
		<i>protid</i>	96.54	96.54	96.54		1					
		<i>dnaid</i>	97.84	97.84	97.84		1					
		<i>kimura</i>	0.04	0.04	0.04		1					
		<i>logdet</i>	0.23	0.23	0.23		1					
		<i>index</i>	76.01	76.01	76.01		1					
	WORF	<i>ks</i>	0.09	1.31	0.01	0.11	853	0.02	1.07	0.00	0.03	2484
		<i>ka</i>	0.04	1.16	0.00	0.06	853	0.07	2.99	0.00	0.07	2484
		<i>kaks</i>	0.57	1.00	0.00	0.30	853	36.62	480.00	1.00	53.71	2484
		<i>protid</i>	86.56	100.00	70.08	8.11	853	86.27	100.00	70.00	6.41	2484
		<i>dnaid</i>	91.27	100.00	74.86	6.20	853	92.44	100.00	76.70	4.71	2484
		<i>kimura</i>	0.14	0.44	0.00	0.09	853	0.15	0.47	0.00	0.07	2484
		<i>logdet</i>	0.23	1.30	0.00	0.19	853	0.21	0.94	0.00	0.17	2484
		<i>index</i>	61.10	98.93	13.26	17.48	853	62.84	99.34	9.83	13.56	2484
	YOYO	<i>ks</i>	0.04	0.04	0.04		1					
		<i>ka</i>	0.03	0.03	0.03		1					
		<i>kaks</i>	0.76	0.76	0.76		1					
		<i>protid</i>	96.39	96.39	96.39		1					
		<i>dnaid</i>	96.79	96.79	96.79		1					
		<i>kimura</i>	0.04	0.04	0.04		1					
		<i>logdet</i>	0.10	0.10	0.10		1					
		<i>index</i>	79.23	79.23	79.23		1					
Obscura-Hawaiian	Genome	<i>ks</i>	3.15	20.54	0.43	3.07	5111					
		<i>ka</i>	0.14	0.63	0.00	0.09	5111					
		<i>kaks</i>	0.05	0.38	0.00	0.03	5111					
		<i>protid</i>	77.46	100.00	39.35	11.16	5111					
		<i>dnaid</i>	73.48	92.05	50.94	6.51	5111					
		<i>kimura</i>	0.27	1.08	0.00	0.16	5111					
		<i>logdet</i>	0.35	1.67	0.09	0.12	5111					
		<i>index</i>	15.03	52.92	0.80	7.93	5111					
	TRANSPAC	<i>ks</i>	3.29	3.30	3.29	0.01	2					

	C	<i>ka</i>	0.08	0.08	0.08	0.00	2							
		<i>kaks</i>	0.02	0.02	0.02	0.00	2							
		<i>protid</i>	75.21	75.35	75.07	0.20	2							
		<i>dnaid</i>	72.64	72.74	72.55	0.13	2							
		<i>kimura</i>	0.30	0.30	0.30	0.00	2							
		<i>logdet</i>	0.39	0.39	0.38	0.01	2							
		<i>index</i>	10.02	10.03	10.01	0.01	2							
Repleta-Hawaiian	Genome	<i>ks</i>	1.30	18.83	0.23	0.84	2562							
		<i>ka</i>	0.09	2.14	0.00	0.07	2562							
		<i>kaks</i>	0.07	0.45	0.00	0.04	2562							
		<i>protid</i>	84.10	100.00	36.57	8.98	2562							
		<i>dnaid</i>	79.28	94.42	50.17	5.34	2562							
		<i>kimura</i>	0.18	1.20	0.00	0.12	2562							
		<i>logdet</i>	0.27	1.13	0.06	0.09	2562							
<i>index</i>	26.21	72.02	1.34	9.18	2562									
Repleta-Obscura	BURDOCK	<i>ks</i>	1.17	1.28	1.04	0.12	5							
		<i>ka</i>	0.16	0.16	0.15	0.01	5							
		<i>kaks</i>	0.14	0.15	0.13	0.01	5							
		<i>protid</i>	73.92	75.86	72.41	1.38	5							
		<i>dnaid</i>	77.39	77.87	76.72	0.55	5							
		<i>kimura</i>	0.32	0.34	0.28	0.02	5							
		<i>logdet</i>	0.46	0.76	0.35	0.17	5							
	<i>index</i>	18.40	20.46	15.04	2.19	5								
	BUT2	<i>ks</i>	0.61	0.61	0.61		1							
		<i>ka</i>	0.11	0.11	0.11		1							
		<i>kaks</i>	0.19	0.19	0.19		1							
		<i>protid</i>	79.69	79.69	79.69		1							
		<i>dnaid</i>	82.81	82.81	82.81		1							
		<i>kimura</i>	0.24	0.24	0.24		1							
		<i>logdet</i>	0.37	0.37	0.37		1							
	<i>index</i>	29.46	29.46	29.46		1								
	CHOUTO	<i>ks</i>	1.55	1.76	1.44	0.11	6							
		<i>ka</i>	0.17	0.21	0.14	0.03	6							
		<i>kaks</i>	0.11	0.15	0.10	0.02	6							
		<i>protid</i>	72.28	72.99	70.18	1.09	6							
		<i>dnaid</i>	73.37	73.97	72.81	0.49	6							
		<i>kimura</i>	0.35	0.38	0.33	0.02	6							
		<i>logdet</i>	0.41	0.49	0.35	0.06	6							
	<i>index</i>	15.34	16.45	13.79	1.02	6								
	Genome	<i>ks</i>	3.38	21.92	0.43	3.19	5114							
		<i>ka</i>	0.14	1.10	0.00	0.09	5114							
		<i>kaks</i>	0.05	0.33	0.00	0.03	5114							
		<i>protid</i>	77.59	100.00	37.74	11.18	5114							
<i>dnaid</i>		73.63	93.17	50.11	6.58	5114								
<i>kimura</i>		0.27	1.12	0.00	0.17	5114								
<i>logdet</i>		0.36	1.22	0.07	0.12	5114								
<i>index</i>	14.51	60.44	0.96	8.00	5114									
GYPSY	<i>ks</i>	7.01	7.01	7.01		1								
	<i>ka</i>	0.25	0.25	0.25		1								
	<i>kaks</i>	0.04	0.04	0.04		1								
	<i>protid</i>	71.43	71.43	71.43		1								
	<i>dnaid</i>	70.16	70.16	70.16		1								

	STALKER	<i>kimura</i>	0.36	0.36	0.36		1							
		<i>logdet</i>	0.64	0.64	0.64		1							
		<i>index</i>	4.28	4.28	4.28		1							
		<i>ks</i>	5.20	5.20	5.20	0.00	8							
		<i>ka</i>	0.26	0.26	0.26	0.00	8							
		<i>kaks</i>	0.05	0.05	0.05	0.00	8							
		<i>protid</i>	78.27	84.54	71.54	4.67	8							
		<i>dnaid</i>	83.78	87.63	80.49	2.80	8							
		<i>kimura</i>	0.25	0.33	0.16	0.06	8							
	<i>logdet</i>	0.27	0.38	0.14	0.07	8								
	<i>index</i>	8.32	9.94	7.29	0.84	8								
	TABOR	<i>ks</i>	3.74	9.49	0.35	3.54	4485	0.01	0.20	0.00	0.05	584		
		<i>ka</i>	0.25	0.88	0.00	0.21	4485	0.16	1.71	0.00	0.18	584		
		<i>kaks</i>	0.08	0.31	0.00	0.05	4485	148.37	490.00	1.53	132.43	584		
		<i>protid</i>	79.83	89.81	70.00	4.94	4485	75.15	85.19	70.00	3.04	584		
		<i>dnaid</i>	81.23	87.76	71.29	2.71	4485	81.21	88.33	76.32	1.96	584		
		<i>kimura</i>	0.24	0.46	0.09	0.07	4485	0.30	0.48	0.14	0.05	584		
		<i>logdet</i>	0.35	1.28	0.15	0.09	4485	0.35	0.70	0.23	0.08	584		
		<i>index</i>	18.53	44.28	3.47	12.22	4485	41.51	52.57	15.45	7.21	584		
	Virilis-Hawaiian	Genome	<i>ks</i>	0.98	18.49	0.21	0.59	2560						
			<i>ka</i>	0.07	1.64	0.00	0.06	2560						
<i>kaks</i>			0.07	0.32	0.00	0.05	2560							
<i>protid</i>			86.12	100.00	40.46	8.04	2560							
<i>dnaid</i>			81.68	95.40	52.73	4.84	2560							
<i>kimura</i>			0.16	0.91	0.00	0.10	2560							
<i>logdet</i>			0.23	1.05	0.05	0.09	2560							
<i>index</i>			31.61	74.06	1.48	9.51	2560							
GYPSY		<i>ks</i>	1.30	1.30	1.30		1							
		<i>ka</i>	0.19	0.19	0.19		1							
		<i>kaks</i>	0.15	0.15	0.15		1							
		<i>protid</i>	70.50	70.50	70.50		1							
		<i>dnaid</i>	72.66	72.66	72.66		1							
		<i>kimura</i>	0.31	0.31	0.31		1							
		<i>logdet</i>	0.37	0.37	0.37		1							
		<i>index</i>	17.10	17.10	17.10		1							
TRANSPAC		<i>ks</i>	1.43	1.43	1.43		1							
		<i>ka</i>	0.04	0.04	0.04		1							
		<i>kaks</i>	0.03	0.03	0.03		1							
		<i>protid</i>	83.70	83.70	83.70		1							
		<i>dnaid</i>	78.73	78.73	78.73		1							
	<i>kimura</i>	0.17	0.17	0.17		1								
	<i>logdet</i>	0.31	0.31	0.31		1								
	<i>index</i>	22.15	22.15	22.15		1								
Virilis-Obscura	Genome	<i>ks</i>	2.91	21.19	0.46	2.77	5112							
		<i>ka</i>	0.13	1.12	0.00	0.08	5112							
		<i>kaks</i>	0.05	0.36	0.00	0.03	5112							
		<i>protid</i>	78.54	100.00	40.49	10.80	5112							
		<i>dnaid</i>	74.36	92.43	50.99	6.35	5112							
		<i>kimura</i>	0.25	0.98	0.00	0.16	5112							
		<i>logdet</i>	0.34	1.82	0.08	0.12	5112							
		<i>index</i>	15.94	56.91	0.93	8.08	5112							
	GYPSY	<i>ks</i>	0.16	1.06	0.05	0.26	78	0.03	0.10	0.00	0.02	880		

Virilis-Repleta		<i>ka</i>	0.06	0.19	0.02	0.04	78	0.10	0.17	0.03	0.02	880
		<i>kaks</i>	0.60	0.97	0.15	0.25	78	30.67	109.25	1.00	43.44	880
		<i>protid</i>	83.65	89.47	70.41	5.24	78	80.21	90.20	70.13	3.68	880
		<i>dnaid</i>	88.46	93.20	74.15	5.37	78	89.58	94.74	79.64	3.35	880
		<i>kimura</i>	0.18	0.38	0.11	0.06	78	0.23	0.37	0.10	0.04	880
		<i>logdet</i>	0.23	0.56	0.13	0.11	78	0.20	0.90	0.05	0.12	880
		<i>index</i>	53.42	66.10	16.58	12.51	78	53.43	68.43	30.76	6.64	880
		STALKER	<i>ks</i>	0.42	1.09	0.19	0.31	14	0.15	0.15	0.15	
	<i>ka</i>		0.09	0.15	0.05	0.03	14	0.17	0.17	0.17		1
	<i>kaks</i>		0.29	0.54	0.10	0.17	14	1.10	1.10	1.10		1
	<i>protid</i>		77.22	84.62	72.50	4.09	14	72.64	72.64	72.64		1
	<i>dnaid</i>		84.96	91.35	82.42	2.69	14	85.53	85.53	85.53		1
	<i>kimura</i>		0.28	0.37	0.18	0.07	14	0.33	0.33	0.33		1
	<i>logdet</i>		0.31	0.42	0.16	0.09	14	0.16	0.16	0.16		1
	TABOR	<i>index</i>	35.26	51.48	23.18	7.40	14	40.20	40.20	40.20		1
		<i>ks</i>	5.50	9.86	0.11	3.34	1034	0.03	0.20	0.00	0.08	12
		<i>ka</i>	0.32	1.06	0.00	0.26	1034	0.22	0.39	0.00	0.18	12
		<i>kaks</i>	0.07	0.71	0.00	0.10	1034	194.82	470.00	1.94	179.91	12
		<i>protid</i>	76.56	91.18	70.00	4.61	1034	71.38	78.50	70.00	2.32	12
		<i>dnaid</i>	79.08	94.49	72.46	4.28	1034	79.46	85.05	73.08	2.85	12
		<i>kimura</i>	0.28	0.39	0.09	0.06	1034	0.35	0.38	0.25	0.04	12
		<i>logdet</i>	0.39	0.69	0.07	0.10	1034	0.37	0.54	0.19	0.10	12
	TRANSPA C	<i>index</i>	11.62	68.84	3.28	12.60	1034	36.03	44.77	24.05	7.06	12
		<i>ks</i>	1.95	2.01	1.89	0.08	2					
		<i>ka</i>	0.07	0.07	0.07	0.00	2					
		<i>kaks</i>	0.04	0.04	0.03	0.01	2					
		<i>protid</i>	75.77	75.91	75.63	0.20	2					
		<i>dnaid</i>	76.98	77.03	76.94	0.06	2					
		<i>kimura</i>	0.30	0.30	0.30	0.00	2					
		<i>logdet</i>	0.35	0.35	0.35	0.00	2					
	Genome	<i>index</i>	15.34	15.65	15.03	0.44	2					
		<i>ks</i>	0.93	9.99	0.15	0.38	2563					
		<i>ka</i>	0.06	0.60	0.00	0.05	2563					
<i>kaks</i>		0.07	0.41	0.00	0.04	2563						
<i>protid</i>		87.87	100.00	40.03	7.34	2563						
<i>dnaid</i>		82.94	96.09	52.70	4.49	2563						
<i>kimura</i>		0.14	1.07	0.00	0.09	2563						
<i>logdet</i>		0.22	1.06	0.04	0.08	2563						
<i>index</i>		33.82	77.82	2.48	9.65	2563						
HELITRON		<i>ks</i>	0.24	0.24	0.24	0.00	3	0.16	0.18	0.13	0.03	3
		<i>ka</i>	0.24	0.24	0.24	0.00	3	0.24	0.27	0.21	0.03	3
		<i>kaks</i>	0.99	0.99	0.99	0.00	3	1.60	2.16	1.18	0.51	3
		<i>protid</i>	71.26	72.73	70.20	1.31	3	71.37	72.27	70.59	0.85	3
		<i>dnaid</i>	83.41	84.02	83.00	0.54	3	84.06	84.59	83.66	0.48	3
		<i>kimura</i>	0.44	0.50	0.36	0.07	3	0.42	0.46	0.38	0.04	3
	<i>logdet</i>	0.43	0.50	0.39	0.06	3	0.43	0.66	0.25	0.21	3	
	<i>index</i>	28.08	30.31	26.04	2.14	3	30.29	32.68	26.21	3.55	3	
NETR1B	<i>ks</i>	0.04	0.33	0.00	0.04	246	0.00	0.04	0.00	0.01	372	
	<i>ka</i>	0.01	0.17	0.00	0.02	246	0.01	0.06	0.00	0.01	372	
	<i>kaks</i>	0.16	0.98	0.00	0.26	246	188.71	490.00	1.22	143.91	372	
	<i>protid</i>	97.35	100.00	75.61	4.40	246	98.06	100.00	85.37	2.79	372	
	<i>dnaid</i>	98.06	100.00	75.34	3.75	246	98.81	100.00	86.18	2.00	372	

		<i>kimura</i>	0.03	0.24	0.00	0.04	246	0.02	0.16	0.00	0.03	372	
		<i>logdet</i>	0.03	0.32	0.00	0.05	246	0.03	0.32	0.00	0.05	372	
		<i>index</i>	89.47	99.08	47.54	11.41	246	93.63	99.76	64.80	7.48	372	
	TABOR	<i>ks</i>	1.69	9.42	0.01	1.85	1317	0.01	0.37	0.00	0.04	291	
		<i>ka</i>	0.09	0.81	0.00	0.14	1317	0.06	0.95	0.00	0.16	291	
		<i>kaks</i>	0.04	0.79	0.00	0.07	1317	169.96	490.00	1.15	153.95	291	
		<i>protid</i>	75.20	88.12	70.00	3.68	1317	77.07	92.31	70.09	5.58	291	
		<i>dnaid</i>	79.64	89.44	71.57	2.15	1317	82.27	90.10	74.63	3.97	291	
		<i>kimura</i>	0.31	0.49	0.13	0.06	1317	0.26	0.43	0.08	0.07	291	
		<i>logdet</i>	0.37	0.77	0.16	0.07	1317	0.33	0.59	0.15	0.09	291	
		<i>index</i>	24.08	60.95	3.42	12.88	1317	48.57	64.95	18.63	8.60	291	
	Willistoni-Hawaiian	Genome	<i>ks</i>	3.87	21.90	0.38	4.07	2558					
			<i>ka</i>	0.15	2.10	0.00	0.10	2558					
			<i>kaks</i>	0.05	0.43	0.00	0.04	2558					
<i>protid</i>			77.15	100.00	41.22	11.26	2558						
<i>dnaid</i>			71.78	92.19	50.00	6.49	2558						
<i>kimura</i>			0.28	1.02	0.00	0.17	2558						
<i>logdet</i>			0.37	1.18	0.11	0.11	2558						
	<i>index</i>	13.63	55.59	1.02	7.91	2558							
Willistoni-Obscura	BLOOD	<i>ks</i>	0.70	0.71	0.70	0.01	2						
		<i>ka</i>	0.17	0.18	0.17	0.01	2						
		<i>kaks</i>	0.26	0.26	0.25	0.01	2						
		<i>protid</i>	70.66	71.07	70.25	0.58	2						
		<i>dnaid</i>	79.62	79.89	79.34	0.39	2						
		<i>kimura</i>	0.36	0.37	0.35	0.01	2						
		<i>logdet</i>	0.57	0.60	0.54	0.04	2						
		<i>index</i>	20.70	21.10	20.30	0.57	2						
	Genome	<i>ks</i>	4.27	22.18	0.45	4.15	5115						
		<i>ka</i>	0.13	0.88	0.00	0.08	5115						
		<i>kaks</i>	0.04	0.36	0.00	0.03	5115						
		<i>protid</i>	79.23	100.00	39.12	10.74	5115						
		<i>dnaid</i>	72.34	91.03	50.06	6.26	5115						
		<i>kimura</i>	0.24	1.13	0.00	0.15	5115						
		<i>logdet</i>	0.37	1.57	0.10	0.12	5115						
		<i>index</i>	13.07	56.93	1.04	7.71	5115						
	GYPSY	<i>ks</i>	0.86	1.12	0.66	0.11	212						
		<i>ka</i>	0.24	0.36	0.14	0.03	212						
		<i>kaks</i>	0.28	0.38	0.13	0.06	212						
		<i>protid</i>	75.50	78.26	70.00	2.06	212						
		<i>dnaid</i>	76.44	78.43	71.13	1.31	212						
		<i>kimura</i>	0.29	0.41	0.26	0.03	212						
		<i>logdet</i>	0.37	0.66	0.26	0.11	212						
		<i>index</i>	22.06	25.66	15.99	2.34	212						
	P_ELEMENT	<i>ks</i>	2.94	8.63	0.70	3.26	15						
		<i>ka</i>	0.19	0.22	0.14	0.03	15						
		<i>kaks</i>	0.17	0.29	0.02	0.11	15						
		<i>protid</i>	71.99	74.19	70.10	1.28	15						
<i>dnaid</i>		75.46	77.78	73.20	1.36	15							
<i>kimura</i>		0.36	0.46	0.28	0.05	15							
<i>logdet</i>		0.71	1.12	0.46	0.25	15							
	<i>index</i>	13.32	20.20	4.04	6.60	15							
STALKER	<i>ks</i>	0.90	1.86	0.02	0.31	180							

Willistoni-Rep leta		<i>ka</i>	0.11	0.17	0.00	0.03	180					
		<i>kaks</i>	0.12	0.25	0.00	0.04	180					
		<i>protid</i>	78.58	91.00	70.10	3.13	180					
		<i>dnaid</i>	82.35	90.33	73.20	2.36	180					
		<i>kimura</i>	0.24	0.38	0.07	0.04	180					
		<i>logdet</i>	0.39	0.99	0.11	0.11	180					
		<i>index</i>	26.16	71.27	15.66	9.45	180					
	TABOR	<i>ks</i>	1.80	9.82	0.01	2.42	6579	0.02	1.34	0.00	0.14	669
		<i>ka</i>	0.18	2.78	0.00	0.16	6579	0.39	3.47	0.00	0.36	669
		<i>kaks</i>	0.14	0.87	0.00	0.08	6579	101.48	470.00	1.18	43.34	669
		<i>protid</i>	75.30	92.86	70.00	3.41	6579	74.84	85.71	70.00	3.09	669
		<i>dnaid</i>	81.22	92.52	73.39	2.35	6579	78.96	89.80	72.47	2.01	669
		<i>kimura</i>	0.29	0.54	0.07	0.05	6579	0.31	0.49	0.16	0.05	669
		<i>logdet</i>	0.34	2.17	0.08	0.10	6579	0.45	1.48	0.17	0.19	669
	<i>index</i>	23.13	76.25	3.62	10.07	6579	32.72	56.14	7.15	6.28	669	
	TC1	<i>ks</i>	0.68	0.70	0.66	0.03	2					
		<i>ka</i>	0.16	0.17	0.15	0.01	2					
		<i>kaks</i>	0.24	0.26	0.22	0.03	2					
		<i>protid</i>	72.94	74.58	71.31	2.31	2					
		<i>dnaid</i>	78.62	79.10	78.14	0.68	2					
		<i>kimura</i>	0.33	0.35	0.31	0.03	2					
		<i>logdet</i>	0.32	0.33	0.30	0.02	2					
	<i>index</i>	25.05	26.00	24.09	1.35	2						
	BUT2	<i>ks</i>	0.77	1.45	0.10	0.45	20	0.00	0.00	0.00		1
		<i>ka</i>	0.11	0.16	0.03	0.04	20	0.06	0.06	0.06		1
		<i>kaks</i>	0.22	0.61	0.08	0.17	20	94.50	94.50	94.50		1
		<i>protid</i>	79.32	92.70	71.82	6.04	20	75.52	75.52	75.52		1
		<i>dnaid</i>	81.05	95.13	71.82	6.89	20	85.08	85.08	85.08		1
		<i>kimura</i>	0.25	0.37	0.08	0.08	20	0.28	0.28	0.28		1
		<i>logdet</i>	0.42	0.93	0.14	0.20	20	0.21	0.21	0.21		1
	<i>index</i>	29.42	65.99	15.99	15.01	20	51.06	51.06	51.06		1	
	COPIA	<i>ks</i>	1.04	3.43	0.86	0.44	86					
		<i>ka</i>	0.18	0.28	0.14	0.02	86					
		<i>kaks</i>	0.18	0.30	0.05	0.04	86					
		<i>protid</i>	74.28	79.84	70.16	2.26	86					
<i>dnaid</i>		75.26	77.96	71.78	1.27	86						
<i>kimura</i>		0.32	0.38	0.24	0.03	86						
<i>logdet</i>		0.49	0.63	0.22	0.07	86						
<i>index</i>	19.23	22.93	8.56	2.71	86							
G3-JOCKEY	<i>ks</i>	2.52	4.65	0.33	2.16	3						
	<i>ka</i>	0.12	0.17	0.07	0.05	3						
	<i>kaks</i>	0.10	0.21	0.04	0.10	3						
	<i>protid</i>	71.96	72.55	71.43	0.56	3						
	<i>dnaid</i>	74.81	76.47	73.81	1.45	3						
	<i>kimura</i>	0.34	0.35	0.33	0.01	3						
	<i>logdet</i>	0.80	1.06	0.60	0.24	3						
<i>index</i>	13.25	24.90	6.63	10.12	3							
Genome	<i>ks</i>	4.69	22.19	0.48	4.60	2559						
	<i>ka</i>	0.14	0.86	0.00	0.09	2559						
	<i>kaks</i>	0.04	0.29	0.00	0.03	2559						
	<i>protid</i>	77.37	100.00	39.78	11.24	2559						
	<i>dnaid</i>	71.36	90.67	50.50	6.45	2559						

		<i>kimura</i>	0.27	1.07	0.00	0.17	2559						
		<i>logdet</i>	0.38	1.28	0.10	0.11	2559						
		<i>index</i>	12.29	53.22	0.97	7.84	2559						
	GTWIN	<i>ks</i>	0.70	0.81	0.66	0.07	8						
		<i>ka</i>	0.14	0.17	0.13	0.01	8						
		<i>kaks</i>	0.20	0.21	0.20	0.01	8						
		<i>protid</i>	76.19	78.10	73.33	1.61	8						
		<i>dnaid</i>	79.60	80.63	77.46	1.24	8						
		<i>kimura</i>	0.21	0.22	0.20	0.01	8						
		<i>logdet</i>	0.50	0.54	0.42	0.05	8						
		<i>index</i>	24.78	27.38	22.03	1.84	8						
	GYPSY	<i>ks</i>	0.38	0.80	0.02	0.19	29	0.01	0.05	0.00	0.02	19	
		<i>ka</i>	0.06	0.11	0.02	0.02	29	0.04	0.07	0.03	0.01	19	
		<i>kaks</i>	0.22	0.84	0.09	0.18	29	51.09	111.00	1.02	52.09	19	
		<i>protid</i>	83.25	90.55	72.97	6.40	29	91.27	93.43	86.49	1.65	19	
		<i>dnaid</i>	84.62	89.30	76.88	4.93	29	96.01	97.32	88.29	2.01	19	
		<i>kimura</i>	0.19	0.37	0.10	0.08	29	0.09	0.15	0.07	0.02	19	
		<i>logdet</i>	0.26	0.52	0.11	0.11	29	0.13	0.24	0.12	0.03	19	
		<i>index</i>	42.17	64.90	23.67	10.82	29	72.54	77.23	61.34	4.13	19	
	HAT	<i>ks</i>	0.02	0.02	0.02		1						
		<i>ka</i>	0.01	0.01	0.01		1						
		<i>kaks</i>	0.63	0.63	0.63		1						
		<i>protid</i>	88.06	88.06	88.06		1						
		<i>dnaid</i>	89.05	89.05	89.05		1						
		<i>kimura</i>	0.03	0.03	0.03		1						
		<i>logdet</i>	0.06	0.06	0.06		1						
		<i>index</i>	79.30	79.30	79.30		1						
	HELITRON	<i>ks</i>						0.12	0.13	0.11	0.01	2	
		<i>ka</i>						0.14	0.14	0.14	0.00	2	
		<i>kaks</i>						1.16	1.24	1.08	0.11	2	
		<i>protid</i>						78.02	78.02	78.02	0.00	2	
		<i>dnaid</i>						87.73	87.91	87.55	0.25	2	
		<i>kimura</i>						0.26	0.26	0.26	0.00	2	
		<i>logdet</i>						0.40	0.40	0.40	0.00	2	
		<i>index</i>					39.59	39.93	39.25	0.48	2		
	HOANA	<i>ks</i>	1.57	2.57	0.02	0.98	10						
		<i>ka</i>	0.09	0.13	0.01	0.04	10						
		<i>kaks</i>	0.12	0.42	0.03	0.13	10						
		<i>protid</i>	76.81	93.45	70.59	6.42	10						
		<i>dnaid</i>	78.54	95.49	73.20	6.76	10						
		<i>kimura</i>	0.27	0.36	0.08	0.08	10						
		<i>logdet</i>	0.53	0.83	0.09	0.22	10						
		<i>index</i>	22.30	78.21	9.76	20.81	10						
	IVK	<i>ks</i>	0.32	0.38	0.29	0.05	3						
		<i>ka</i>	0.16	0.16	0.15	0.01	3						
		<i>kaks</i>	0.49	0.53	0.41	0.07	3						
		<i>protid</i>	77.26	78.18	76.11	1.05	3						
		<i>dnaid</i>	83.55	84.85	81.42	1.86	3						
		<i>kimura</i>	0.25	0.29	0.22	0.04	3						
		<i>logdet</i>	0.45	0.56	0.23	0.19	3						
		<i>index</i>	32.29	33.77	31.51	1.29	3						
	MARINER	<i>ks</i>	0.31	0.36	0.29	0.02	9						

		<i>ka</i>	0.07	0.09	0.06	0.01	9						
		<i>kaks</i>	0.24	0.27	0.19	0.02	9						
		<i>protid</i>	80.82	85.15	79.17	1.78	9						
		<i>dnaid</i>	84.54	87.13	82.41	1.41	9						
		<i>kimura</i>	0.19	0.22	0.17	0.02	9						
		<i>logdet</i>	0.37	0.51	0.32	0.07	9						
		<i>index</i>	38.39	40.54	36.18	1.43	9						
	QUETZAL	<i>ks</i>	1.17	3.39	0.70	0.85	9						
		<i>ka</i>	0.03	0.06	0.00	0.01	9						
		<i>kaks</i>	0.03	0.05	0.00	0.02	9						
		<i>protid</i>	86.30	89.47	79.31	3.51	9						
		<i>dnaid</i>	82.89	84.80	78.54	1.84	9						
		<i>kimura</i>	0.17	0.44	0.11	0.10	9						
		<i>logdet</i>	0.56	0.83	0.31	0.17	9						
	STALKER	<i>ks</i>	5.63	9.05	4.56	0.98	98	1.80	1.80	1.80		1	
		<i>ka</i>	0.27	0.90	0.25	0.07	98	4.43	4.43	4.43		1	
		<i>kaks</i>	0.05	0.20	0.03	0.02	98	2.46	2.46	2.46		1	
		<i>protid</i>	80.46	88.89	70.00	4.96	98	74.77	74.77	74.77		1	
		<i>dnaid</i>	83.07	89.12	72.96	4.03	98	81.31	81.31	81.31		1	
		<i>kimura</i>	0.22	0.37	0.11	0.07	98	0.29	0.29	0.29		1	
		<i>logdet</i>	0.30	0.89	0.12	0.17	98	0.33	0.33	0.33		1	
	TABOR	<i>ks</i>	1.49	9.42	0.01	1.85	3240	0.00	1.80	0.00	0.07	3220	
		<i>ka</i>	0.19	1.97	0.00	0.16	3240	0.28	4.43	0.00	0.25	3220	
		<i>kaks</i>	0.20	0.80	0.00	0.16	3240	128.83	490.00	1.16	97.05	3220	
		<i>protid</i>	77.41	88.97	70.00	4.58	3240	76.11	86.55	70.00	3.85	3220	
		<i>dnaid</i>	81.34	89.22	73.33	3.07	3240	80.57	89.34	71.67	3.30	3220	
		<i>kimura</i>	0.28	0.53	0.12	0.06	3240	0.30	0.55	0.17	0.06	3220	
		<i>logdet</i>	0.31	0.84	0.13	0.09	3240	0.39	1.06	0.16	0.14	3220	
	TF001011	<i>ks</i>	0.76	0.77	0.72	0.02	8						
		<i>ka</i>	0.07	0.11	0.06	0.02	8						
		<i>kaks</i>	0.09	0.16	0.08	0.03	8						
		<i>protid</i>	85.65	86.93	80.13	2.25	8						
		<i>dnaid</i>	84.63	84.97	83.66	0.42	8						
<i>kimura</i>		0.16	0.23	0.14	0.03	8							
<i>logdet</i>		0.27	0.32	0.20	0.04	8							
TRANSIB	<i>ks</i>	0.26	0.54	0.09	0.18	35							
	<i>ka</i>	0.09	0.18	0.04	0.05	35							
	<i>kaks</i>	0.41	0.55	0.21	0.12	35							
	<i>protid</i>	77.49	88.06	70.27	6.30	35							
	<i>dnaid</i>	84.36	93.78	76.88	5.88	35							
	<i>kimura</i>	0.27	0.38	0.13	0.09	35							
	<i>logdet</i>	0.33	0.46	0.12	0.15	35							
Willistoni-Virilis	Genome	<i>ks</i>	4.10	20.45	0.44	4.24	2558						
	<i>ka</i>	0.14	0.90	0.00	0.09	2558							
	<i>kaks</i>	0.05	0.31	0.00	0.03	2558							
	<i>protid</i>	78.28	100.00	41.15	10.91	2558							
	<i>dnaid</i>	72.10	91.21	51.45	6.34	2558							

		<i>kimura</i>	0.26	0.99	0.00	0.16	2558						
		<i>logdet</i>	0.37	1.28	0.11	0.11	2558						
		<i>index</i>	13.57	53.39	0.91	8.03	2558						
GYPSY		<i>ks</i>	0.26	0.40	0.12	0.20	2						
		<i>ka</i>	0.11	0.18	0.05	0.09	2						
		<i>kaks</i>	0.44	0.45	0.43	0.01	2						
		<i>protid</i>	84.25	90.83	77.67	9.31	2						
		<i>dnaid</i>	87.25	93.58	80.91	8.96	2						
		<i>kimura</i>	0.19	0.28	0.10	0.13	2						
		<i>logdet</i>	0.28	0.36	0.20	0.11	2						
		<i>index</i>	44.08	54.19	33.96	14.30	2						
R1		<i>ks</i>	0.46	0.60	0.41	0.05	33						
		<i>ka</i>	0.10	0.15	0.05	0.02	33						
		<i>kaks</i>	0.21	0.31	0.13	0.04	33						
		<i>protid</i>	77.10	85.71	71.31	3.71	33						
		<i>dnaid</i>	82.24	85.71	79.20	1.63	33						
		<i>kimura</i>	0.28	0.36	0.17	0.05	33						
		<i>logdet</i>	0.38	0.67	0.20	0.13	33						
		<i>index</i>	31.10	36.56	22.61	2.83	33						
STALKER		<i>ks</i>	2.38	2.79	2.18	0.35	3						
		<i>ka</i>	0.20	0.27	0.17	0.06	3						
		<i>kaks</i>	0.09	0.12	0.06	0.03	3						
		<i>protid</i>	74.23	76.29	71.13	2.73	3						
		<i>dnaid</i>	76.98	78.69	75.26	1.72	3						
		<i>kimura</i>	0.32	0.36	0.29	0.04	3						
		<i>logdet</i>	0.48	0.52	0.42	0.06	3						
		<i>index</i>	11.79	12.79	10.61	1.10	3						
TABOR		<i>ks</i>	1.78	8.80	0.24	1.89	646	0.03	0.40	0.00	0.11	344	
		<i>ka</i>	0.14	0.86	0.00	0.19	646	0.33	0.67	0.00	0.17	344	
		<i>kaks</i>	0.07	0.84	0.00	0.13	646	107.51	490.00	1.59	81.10	344	
		<i>protid</i>	73.57	80.34	70.00	2.62	646	72.48	79.82	70.00	2.20	344	
		<i>dnaid</i>	78.43	82.17	72.64	1.67	646	78.79	85.93	73.24	1.69	344	
		<i>kimura</i>	0.32	0.67	0.22	0.05	646	0.34	0.45	0.21	0.04	344	
		<i>logdet</i>	0.38	0.71	0.27	0.08	646	0.37	1.22	0.20	0.16	344	
		<i>index</i>	23.14	41.64	3.75	12.89	646	33.51	58.31	13.53	6.47	344	
TRANSPAC		<i>ks</i>	4.58	4.58	4.58		1						
		<i>ka</i>	0.10	0.10	0.10		1						
		<i>kaks</i>	0.02	0.02	0.02		1						
		<i>protid</i>	76.14	76.14	76.14		1						
		<i>dnaid</i>	70.45	70.45	70.45		1						
		<i>kimura</i>	0.27	0.27	0.27		1						
		<i>logdet</i>	0.42	0.42	0.42		1						
		<i>index</i>	7.65	7.65	7.65		1						