



**UNIVERSIDADE FEDERAL DE MINAS GERAIS**

Instituto de Ciências Biológicas

Departamento de Biologia Geral

Programa de Pós-Graduação em Genética

Leonardo Gomes de Lima

EVOLUÇÃO MOLECULAR DO DNA SATÉLITE 1.688 EM ESPÉCIES DE *Drosophila* COM  
GENOMAS SEQUENCIADOS

Belo Horizonte

2013

**Leonardo Gomes de Lima**

**EVOLUÇÃO MOLECULAR DO DNA SATÉLITE 1.688 EM ESPÉCIES DE *Drosophila* COM  
GENOMAS SEQUENCIADOS**

Dissertação apresentada ao Programa de Pós-Graduação em Genética da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do título de Mestre em Genética

Orientador: Prof. Dr. Gustavo Campos e Silva Kuhn

Coorientadora: Profa. Dra. Marta Svartman

Belo Horizonte, 2013

## FICHA CATALOGRÁFICA

- 043 Lima, Leonardo Gomes de.  
Evolução molecular do DNA satélite 1.688 em espécies de *Drosophila* com genomas sequenciados [manuscrito] / Leonardo Gomes de Lima. – 2013.  
105 f. : il. ; 29,5 cm.
- Orientador: Prof. Dr. Gustavo Campos e Silva Kuhn. Coorientadora: Profa. Dra. Marta Svartman.  
Dissertação (mestrado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa de Pós-Graduação em Genética.
1. Genética. 2. DNA Satélite. 3. Evolução Molecular. 4. *Drosophila*. 5. Elementos de DNA Transponíveis. I. Kuhn, Gustavo Campos e Silva. II. Svartman, Marta. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 575



**ATA DA DEFESA DE DISSERTAÇÃO**  
**Leonardo Gomes de Lima**

196/2013  
entrada  
2º/2011  
CPF:  
033.371.015-00

Às oito horas e trinta minutos do dia **25 de setembro de 2013**, reuniu-se, no Instituto de Ciências Biológicas da UFMG, a Comissão Examinadora de Dissertação, indicada pelo Colegiado de Programa, para julgar, em exame final, o trabalho intitulado: **"EVOLUÇÃO MOLECULAR DO DNA SATÉLITE 1.688 EM ESPÉCIES DE Drosophila COM GENOMAS SEQUENCIADOS"**, requisito para obtenção do grau de Mestre em Genética. Abrindo a sessão, o Presidente da Comissão, **Dr. Gustavo Campos e Silva Kuhn**, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra ao candidato, para apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa do candidato. Logo após, a Comissão se reuniu, sem a presença do candidato e do público, para julgamento e expedição de resultado final. Foram atribuídas as seguintes indicações:

Prof./Pesq.	Instituição	Indicação
Dr. Gustavo Campos e Silva Kuhn	UFMG	APROVADO
Dra. Marta Svartman	UFMG	APROVADO
Dr. Jerônimo da Conceição Ruiz	FIOCRUZ	APROVADO
Dra. Gisele Pires de Mendonça Dantas	UFMG	Aprovado

Pelas indicações, o candidato foi considerado: APROVADO  
O resultado final foi comunicado publicamente ao candidato pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora.

Belo Horizonte, 25 de setembro de 2013.

Dr. Gustavo Campos e Silva Kuhn - Orientador

*Gustavo Kuhn*

Dra. Marta Svartman - Coorientadora

*Marta Svartman*

Dr. Jerônimo da Conceição Ruiz

*Jerônimo Ruiz*

Dra. Gisele Pires de Mendonça Dantas

*Gisele Pires de Mendonça Dantas*

***A meus pais e à Renata Vaz, pelo incentivo e carinho.***

Agradeço,

À Universidade Federal de Minas Gerais por possibilitar a realização deste trabalho.

Ao Programa de Pós-Graduação em Genética.

Aos professores Dr. Fabrício Rodrigues dos Santos, Dra. Mônica Bucciarelli Rodriguez e Dra. Andrea Maria Amaral Nascimento, por auxiliarem a realização deste trabalho ao fornecerem acesso a equipamentos.

À CAPES, pelo auxílio financeiro no formato de bolsa demanda social cedida a mim durante o período do mestrado.

Aos colegas Guilherme, Naiara, Radarane, Carolina e Pedro do Laboratório de Citogenômica Evolutiva que me auxiliaram no decorrer do período de mestrado.

À professora Dra. Cláudia Carareto, por ter cedido as linhagens de *Drosophila* utilizadas neste trabalho.

Ao professor Alfredo Ruiz, pelas sugestões dadas para o aprimoramento deste trabalho.

À minha coorientadora, professora Dra. Marta Svartman, pelo auxílio constante durante estes dois anos, por toda a sua disponibilidade, além da contribuição para o desenvolvimento intelectual oriundo das constantes conversas e discussões.

Ao meu orientador professor Dr. Gustavo Kuhn pela confiança em mim depositada ao me aceitar como aluno, pela paciência, sabedoria, disponibilidade e por auxiliar no meu crescimento pessoal e profissional. Sei que ainda tenho muito a aprender e que ainda colaboraremos muito nos anos futuros.

Em especial,

Aos meus pais e às minhas irmãs, por terem sido os principais responsáveis pela realização desta etapa. Sem a ajuda de vocês eu nunca teria conseguido.

À minha namorada Renata Valença Vaz, por toda sua ajuda, carinho, compreensão e acima de tudo por ser minha companheira para todas as horas.

## RESUMO

A família de DNA satélite (DNAsat) 1.688, presente em espécies de *Drosophila* do subgrupo *melanogaster*, vem sendo utilizada como um importante modelo para diversos estudos sobre a organização, função e evolução dos DNAs satélites. No entanto, a maioria dos estudos anteriores relacionados a este DNA satélite envolveu a análise de apenas um pequeno número de cópias (<10 cópias), em contraste com sua grande abundância no genoma (p.ex. 15.000 cópias no genoma de *D. melanogaster*). Diante da disponibilidade de genomas sequenciados de várias espécies de *Drosophila*, incluindo os de cinco espécies do subgrupo *melanogaster*, o presente trabalho teve como objetivo estudar a variação, distribuição genômica e evolução do DNAsat 1.688 pela análise deste vasto reservatório de sequências. Através da mineração de sequências 1.688 nos bancos de dados genômicos de *Drosophila*, foram identificadas e caracterizadas 6.849 cópias deste DNA satélite em cinco espécies de *Drosophila* do subgrupo *melanogaster*. Análises filogenéticas permitiram classificar as cópias de todas as espécies em eucromáticas ou heterocromáticas. Relatamos pela primeira vez a presença de várias cadeias do DNAsat 1.688 na eucromatina de *D. simulans*, *D. sechellia*, *D. yakuba* e *D. erecta*. A localização eucromática de cada cadeia foi confirmada através do estudo de sequências flanqueadoras. Nossos dados mostraram que o DNAsat 1.688 estava presente na espécie ancestral do subgrupo *melanogaster* e em ambos domínios da cromatina. As análises das cópias heterocromáticas acrescentaram informações detalhadas sobre a variação das cópias em cada espécie. Nas espécies do subgrupo *melanogaster*, as cópias heterocromáticas 1.688 apresentaram uma variação intraespecífica menor que a interespecífica, evidenciando um padrão de evolução combinada. No entanto, as filogenias obtidas com as cópias 1.688 não foram totalmente congruentes com as relações filogenéticas das espécies, indicando diferentes taxas de evolução. Surpreendentemente, também encontramos quatro cópias 1.688 no genoma de *D. willistoni*, uma espécie distante do subgrupo *melanogaster*. A distribuição filogenética descontínua e a alta similaridade nucleotídica entre cópias 1.688 de *D. melanogaster* e *D. willistoni* sugerem que as cópias de *D. melanogaster* foram introduzidas em *D. willistoni* através de um evento de transferência horizontal (HT). Não existem na literatura relatos anteriores de HT envolvendo DNAs satélites. Também encontramos várias evidências de que esta HT foi mediada por um retrotransposon não-LTR característico de *D. melanogaster* chamado *Doc*.

**Palavras-chave:** DNA satélite; Evolução Molecular; *Drosophila*; DNAs repetitivos; Elementos transponíveis

## ABSTRACT

The 1.688 satellite DNA (satDNA) family, present in species of *Drosophila* from the *melanogaster* subgroup, has been used as an important model in several studies about satDNA organization, function and evolution. Nevertheless, most previous studies of this satDNA were restricted to the analysis of just a few copies (less than ten), contrasting with its great abundance in the genome (for instance, 15,000 copies in *D. melanogaster*). In view of the availability of sequenced genomes of several *Drosophila* species, including five from the *melanogaster* subgroup, we aimed at studying the variation, genomic distribution and evolution of the 1.688 satDNA through the analysis of this vast sequence repertoire. Data mining of 1.688 sequences in the *Drosophila* genomes databases led to the identification and characterization of 6,849 copies of this satDNA in five species of *Drosophila* of the *melanogaster* subgroup. Phylogenetic analyses allowed the classification of the copies as euchromatic or heterochromatic for all the species. We report for the first time the presence of several arrays of the 1.688 satDNA in the euchromatin of *D. simulans*, *D. sechellia*, *D. yakuba* and *D. erecta*. The euchromatic location of each array was confirmed by the analysis of their flanking sequences. Our data revealed that the 1.688 satDNA was present in both chromatin domains in the ancestral species of the *melanogaster* subgroup. Analyses of the heterochromatic copies yielded detailed information on the copy variations in each species. The intraspecific variation among 1.688 heterochromatic copies was lower than the interspecific one in the species of the *melanogaster* subgroup, evidencing a pattern of concerted evolution. Nevertheless, the phylogenies obtained with the 1.688 copies were not in total agreement with the species phylogenetic relationships, pointing to different evolution rates. Surprisingly, four 1.688 copies were also found in the genome of *D. willistoni*, a distantly related species of the *melanogaster* subgroup. The discontinuous phylogenetic distribution and high nucleotide similarity among the 1.688 copies of *D. melanogaster* and *D. willistoni* suggest that the copies of *D. melanogaster* were introduced in *D. willistoni* through a horizontal transfer (HT) event. There are no previous reports of HT involving satellite DNAs. Several hints that this event was mediated by a non-LTR retrotransposon called *Doc* characteristic of *D. melanogaster* were also found.

**Keywords:** Satellite DNA; Molecular Evolution; *Drosophila*; Repetitive DNAs; Transposable Elements

## SUMÁRIO

1. INTRODUÇÃO .....	10
1.1 Paradoxo do valor C e DNAs repetitivos .....	10
1.2 DNAs Satélites: características gerais .....	11
1.3 Papel biológico dos DNAs satélites .....	13
1.4 Evolução dos DNAs satélites .....	14
1.5 O gênero Drosophila e o subgrupo melanogaster .....	17
1.6 Genomas sequenciados de Drosophila e o estudo de DNAs satélites.....	18
1.7 O DNA Satélite 1.688 g/cm <sup>3</sup> do subgrupo melanogaster .....	22
2. OBJETIVO .....	26
2.1 Objetivo Geral .....	26
2.2 Objetivos específicos .....	26
3. MATERIAL E MÉTODOS .....	27
4. RESULTADOS .....	36
5. DISCUSSÃO .....	64
6. CONCLUSÕES .....	80
7. PERSPECTIVAS .....	82
8. REFERÊNCIAS BIBLIOGRÁFICAS .....	83
9. ANEXOS .....	96

## 1. INTRODUÇÃO

### 1.1. Paradoxo do valor C e DNAs repetitivos

Estudos realizados durante a primeira metade do século XX mostraram que a quantidade de DNA poderia diferir entre espécies e principalmente entre eucariotos e procariotos. Estes estudos demonstraram que algumas espécies de anfíbios e peixes continham cerca de vinte vezes mais material genético por núcleo haplóide (Valor C) do que espécies de mamíferos, dentre elas o homem. Esta aparente incongruência entre o grau de complexidade do organismo e o tamanho do genoma ficou conhecida na literatura como “Paradoxo do valor C” (Doolittle & Sapienza 1980; Moran & Morrish 2005). Um exemplo clássico deste paradoxo é o protozoário *Amoeba dubia*, que contém aproximadamente 200 vezes mais material genético do que a quantidade de material genético presente em uma célula humana.

O Paradoxo do valor C perdurou por alguns anos e começou a ser elucidado mesmo antes do advento das técnicas de sequenciamento e análise de genomas completamente sequenciados. Warring & Britten (1966), através do estudo da cinética de reassociação do DNA, demonstraram que a taxa de renaturação dos fragmentos de DNA depende do grau de similaridade entre as fitas da dupla-hélice. Assim, sequências de cópia única possuem taxas de reassociação mais lentas em relação a sequências menos complexas e/ou repetitivas. A identificação das sequências altamente repetitivas contribuiu para a inferência de que estas perfazem uma porção muito maior do genoma do que os genes codificantes de proteínas (revisado por Richard e cols. 2008).

Elementos repetitivos estão amplamente distribuídos no genoma dos eucariotos e compõem, na maioria dos casos, cerca de 50% destes genomas. É possível classificar os elementos repetitivos em duas principais categorias: sequências repetidas em tandem e sequências repetidas dispersas, cada uma delas com diversas subdivisões. Entre os DNAs repetitivos dispersos destacam-se os elementos transponíveis, as duplicações segmentares, famílias gênicas e pseudogenes. Os DNAs repetidos em tandem são compostos basicamente por DNAs satélites, DNAs minissatélites, DNAs microssatélites e algumas famílias gênicas (Charlesworth e cols. 1994; Richard e cols. 2008).

## 1.2. DNAs satélites: características gerais

Entre os DNAs repetidos em tandem, destacam-se os DNAs satélites. Esta denominação foi utilizada pela primeira vez por Kit (1961) através da análise do DNA de vários animais após centrifugação em gradiente de cloreto de céσιο (CsCl). Nesta técnica o DNA fragmentado é suspenso em uma solução de cloreto de céσιο e submetido a intensa centrifugação. A posição em que o DNA se estabelece no tubo após a centrifugação está relacionada à sua densidade, a qual depende da composição nucleotídica do fragmento. As sequências de DNA satélite possuem uma proporção nucleotídica CG/AT que diverge consideravelmente, na maioria dos casos, da proporção encontrada na maior parte do DNA genômico (revisado por Palomeque e Lorite, 2008). Conseqüentemente, após centrifugação em gradiente de cloreto de céσιο, sequências de DNA satélite podem ser visualizadas como “bandas” ou “picos” de DNA com densidade diferente da banda principal contendo a maior parte dos fragmentos do DNA genômico. Tais bandas foram denominadas DNAs satélites (do grego “aquele que acompanha uma entidade maior”) (Kit 1961).

Atualmente, os DNAs satélites podem ser definidos como DNAs altamente repetitivos ( $>10^4$  cópias) organizados em tandem dentro de extensas cadeias. Estas cadeias são tipicamente encontradas em regiões heterocromáticas e suas cópias (ou monômeros) não possuem função codificante de proteínas (Beridze 1986; Charlesworth e cols. 1994).

DNAs satélites são componentes característicos da grande maioria das espécies de eucariotos. No entanto, diferem em sequência, complexidade, tamanho da unidade de repetição e abundância, possuindo apenas duas características compartilhadas entre todos: a organização das repetições em longos arranjos em tandem e a localização heterocromática (Plohl 2012).

Quando milhares de cópias (unidades de repetição) parálogas são encontradas arranjadas em tandem ao longo de diferentes posições genômicas, estas cópias são consideradas uma “família” de DNA satélite. Este termo também pode ser utilizado para DNAs satélites homólogos de diferentes espécies. Vários DNAs satélites não-homólogos (diferentes famílias) podem ser encontrados no genoma de uma espécie (Lohe & Roberts 1993). Um único DNA satélite, ou vários DNAs satélites

somados, tipicamente representam mais do que 30% do genoma de espécies de eucariotos (Bachmann e cols. 1996; Pons e cols. 2004).

Cópias de uma mesma família de DNA satélite podem ser encontradas em apenas um cromossomo (De la Herrán e cols. 2001a), em vários (Bachmann e cols. 1989; Kuhn e cols. 2008), ou em todos os cromossomos de uma espécie (Mravinac e cols. 2004, Tyler-Smith & Willard 1993).

Mais de uma cadeia contendo cópias de uma mesma família de DNA satélite pode estar presente em um único cromossomo e mais de uma família de DNA satélite podem estar presentes no mesmo cromossomo (Lohe e cols. 1993; Shiels e cols. 1997; Kuhn e cols. 2009). Por exemplo, em humanos, a região centromérica do cromossomo Y é composta por sete cadeias contendo quatro DNAs satélites diferentes (Cooper e cols. 1993).

Assim como as porções heterocromáticas do genoma, DNAs satélites são mais frequentemente encontrados em regiões centroméricas ou pericentroméricas (Mravinac e cols. 2004; Kuhn e cols. 2008), menos frequentemente em regiões subteloméricas (Ugarkovic & Plohl 2002; Heslop-Harrison e cols. 2003) e, mais raramente, em regiões intercalares (Nagaki e cols. 1999; Heslop-Harrison e cols. 2003; Kuhn e cols. 2012). Há vários exemplos mostrando que a distribuição de DNAs satélites coincide com o padrão de banda C dos cromossomos, indicando que estão entre os elementos mais abundantes da heterocromatina (Miklos 1985; Heslop-Harrison e cols. 2003; Plohl e cols. 2008; Kuhn e cols. 2009).

Embora a grande maioria das repetições de DNAs satélites esteja localizada em regiões heterocromáticas pericentroméricas e subteloméricas, há alguns casos em que algumas poucas repetições foram encontradas em regiões eucromáticas. Em *D. melanogaster* são encontradas oito repetições em tandem do DNA satélite AATAC antecedendo o gene *s38* na eucromatina do cromossomo X (Spralding e cols. 1987). No besouro *Tribolium castaneum*, repetições do DNA satélite 360 pb, abundante em regiões centroméricas e pericentroméricas, encontram-se também dispersas em regiões eucromáticas (Pezer e cols. 2011).

O número de cópias de uma família de DNA satélite é bastante variável entre espécies. Por exemplo, existem 10.000 cópias do DNA satélite STR120 na soja

*Glycine max* (Morgante e cols. 1997) e 6.000.000 cópias no roedor *Ctenomys haigi* (Slamovits e cols. 2001). Eventos de amplificação e deleção de sequências de DNAs satélites ocorreram com frequência em espécies de roedores do gênero *Ctenomys*, resultando em uma variação entre o número de cópias (Slamovits e cols. 2001). A variação também pode ocorrer entre linhagens da mesma espécie, como é o caso da variação de sequências de DNAs satélites entre linhagens aloploidoides de *Nicotiana*, planta da família *Solanaceae* (Koukalova e cols. 2010).

O tamanho das unidades de repetição também pode variar consideravelmente. Foram descritos DNAs satélites com unidades de repetição curtas de apenas 2 pb nos caranguejos *Cancer irroratus* ou *C. borealis*, de 5 ou 10 pb em *Drosophila melanogaster* ou de 5 pb em humanos (Gray & Skiner 1974; Suekoka 1961; Lohe & Roberts 1988). Por outro lado, foram descritos DNAs satélites compostos por monômeros com tamanho superior a 1 kb, como é o caso do DNA satélite de cetáceos (Grétarsdóttir & Árnason 1992), com cópias de 1.740 pb, ou do DNA satélite *PstI* do besouro *Misolampus goudoti* com cópias de 1.061 pb (Pons e cols. 2004). Apesar desta enorme variação no tamanho de repetições, a maioria dos DNAs satélites possui monômeros com sequências que variam entre 100 e 400 pb.

### **1.3. Papel biológico dos DNAs satélites**

Estudos realizados nas décadas de 70 e 80 e confirmados até o presente revelaram que os DNAs satélites não codificam proteínas e que se concentram em regiões caracterizadas por baixos níveis de recombinação e alta compactação da cromatina, isto é, regiões heterocromáticas. Com base nestes dados, DNAs satélites foram historicamente rotulados como “DNA egoísta” e sua presença no genoma foi explicada por sua habilidade de amplificação (Orgel & Crick 1980).

Embora os monômeros de DNAs satélites apresentem uma baixa variabilidade entre si, algumas regiões dos monômeros apresentam uma maior conservação de sequência, enquanto outras possuem uma maior variabilidade (Mravinac e cols. 2004). Esta distribuição não uniforme da variabilidade na sequência da unidade de repetição indica a presença de pressões seletivas para a conservação de determinadas regiões possivelmente funcionais (Heslop-Harrison e cols. 1999; Pezer e cols 2011). Pequenos motivos conservados foram descritos em DNAs satélites centroméricos de arroz

(*Oryza sativa*) e milho (*Zea mays*), sugerindo a manutenção de um mesmo segmento funcional entre espécies separadas há pelo menos 50 milhões de anos (Cheng e cols. 2002).

Os estudos que melhor demonstram a funcionalidade de DNAs satélites estão relacionados com a associação entre as sequências de DNAs satélites e proteínas centroméricas, como CENP-A e CENP-B (proteínas centroméricas A e B, respectivamente) em camundongo e em primatas hominídeos (Marshall & Clarke 1995; Sugimoto 1994). Sítios conservados de 17 pb denominados “CENP-B box” estão presentes nos monômeros de 171 pb do DNA satélite  $\alpha$ , que, por sua vez, está localizado no centrômero de todos os cromossomos destas espécies, exceto o Y. Os sítios de 17 pb que compõem o “CENP-B box” facilitam a interação entre as proteínas formadoras do complexo cinetocórico e a proteína centromérica B (Masumoto 1989). Os sítios “CENP-B box” conservados não estão presentes em todas as cópias de DNA satélite  $\alpha$ , porém estão distribuídos em intervalos regulares nos centrômeros humanos e, aparentemente, são essenciais para sua funcionalidade (Ohzeki e cols. 2002).

A transcrição de DNAs satélites já foi descrita em diversos invertebrados, vertebrados e plantas. A transcrição geralmente ocorre durante estágios iniciais do desenvolvimento e ainda pode ocorrer diferencialmente em tecidos diferentes, sugerindo um papel regulatório (revisado por Ugarkovic 2005). Alguns trabalhos sugerem que transcritos de DNA satélite atuam em mecanismos transregulatórios e que os monômeros possuem potenciais regiões promotoras da transcrição para as enzimas RNA polimerase II e III (Renault e cols. 1999; Pezer e cols. 2011).

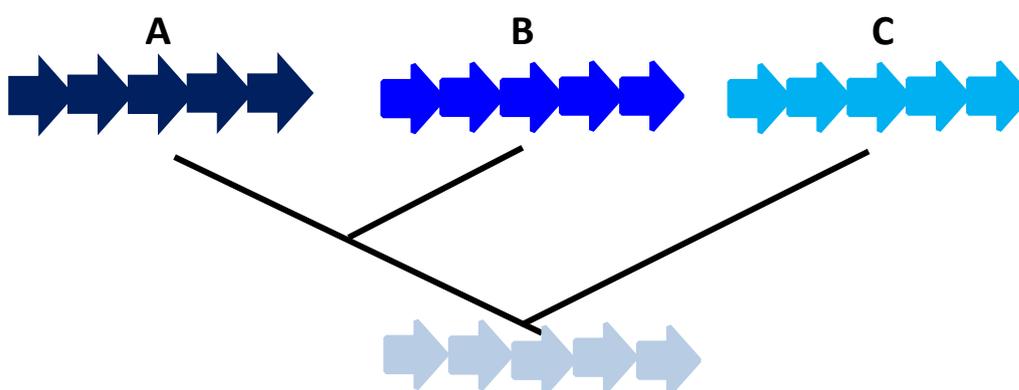
#### **1.4. Evolução dos DNAs satélites**

A dinâmica evolutiva de DNAs satélites é extremamente variável. Em muitos casos, as famílias de DNAs satélites são descritas como sendo espécie-específicas, não sendo detectáveis mesmo entre espécies muito próximas filogeneticamente. Este é o caso do DNA satélite pDoP102 do ortóptera *Dolichopoda schiavazzi* (Bachmann e cols. 1994). Há casos em que famílias de DNAs satélites estão restritas a grupos filogeneticamente próximos, como a família do *pBuM* presente em espécies de *Drosophila* do grupo *buzzatii* (Kuhn e cols. 2008) ou a família ATOC180, restrita ao

genoma de cinco espécies de *Drosophila* do grupo *obscura* (Bachmann & Sperlich, 1993). Existem poucos relatos de DNAs satélites presentes em todas as espécies de um gênero ou família (Grétarsdóttir & Árnason 1992; Mestrovic e cols. 1998; Piras 2010).

De acordo com a teoria de evolução molecular neutra, é esperado que sequências de DNA funcionais, como genes codificadores de proteínas, possuam uma variabilidade menor quando comparadas a sequências com pouco ou nenhum papel funcional (Nei 1987).

DNAs satélites são formados por várias cópias em tandem que não codificam proteínas. Logo, seria de se esperar que estas cópias apresentassem uma grande variabilidade nucleotídica dentro das cadeias. No entanto, há vários trabalhos mostrando uma alta taxa de similaridade entre cópias de um mesmo DNA satélite presentes em uma espécie (revisado em Plohl 2012). Por outro lado, existe uma grande divergência interespecífica entre cópias de uma mesma família de DNA satélite. Isto significa que existe algum processo que leva as cadeias à homogeneização. No caso de isolamento reprodutivo entre populações, este processo pode levar à homogeneização de mutações espécie-específicas, resultando em uma variabilidade intraespecífica menor do que a observada entre cópias de duas ou mais espécies, um fenômeno descrito na literatura como “evolução combinada” ou “evolução em concerto” (Zimmer e cols. 1980) (Figura 1).



**Figura 1.** Esquema dos padrões espécie-específicos de variação entre cópias de um DNA satélite (setas) compartilhado por três espécies (A, B e C), mostrando o fenômeno de

“evolução combinada”. Setas de diferentes cores representam cópias com diferentes perfis mutacionais (variantes).

O fenômeno de evolução combinada também é observado em diferentes famílias de DNAs repetidos em tandem, como famílias gênicas de histonas, globinas, imunoglobulinas, além de diversas sequências não-codificantes (Liebhaber 1981; Scott 1984; Kuhn e cols. 2008; Kuhn e cols. 2012).

O baixo grau de variabilidade entre cópias é explicado pela ação de vários mecanismos moleculares, como a recombinação desigual, a conversão gênica, a derrapagem da polimerase (*replication slippage*) e a amplificação via “círculo-rolante” de DNA extracromossômico circular (Smith 1976; Dover 1986, Cohen & Segal 2009). A influência destes mecanismos na evolução molecular é conhecida como impulso molecular ou *molecular drive* (Dover 1982). Os mecanismos responsáveis pelo *molecular drive* são capazes de alterar a frequência de uma cópia variante dentro de uma família de DNA repetitivo. Ou seja, uma determinada cópia variante pode aumentar sua frequência (homogeneização) ao longo de uma cadeia de cópias e eventualmente se tornar característica das cadeias de todos os indivíduos de uma população ou espécie (fixação) (Dover 1986).

Algumas evidências sugerem que o processo de homogeneização é mais eficiente entre cópias presentes na mesma cadeia, em um cromossomo, do que entre cópias presentes em diferentes cromossomos (Willard e Waye 1987; Hall e cols. 2005; Rudd e cols. 2006). Este fenômeno pode ser explicado pela maior probabilidade de ocorrência de recombinação e, conseqüentemente, de homogeneização, entre cópias de uma mesma cadeia do que entre cópias de cadeias em diferentes cromossomos (Dover 1982). Este tipo de situação pode levar ao surgimento de subfamílias de DNAs satélites cromossomo-específicas, como observado no DNA satélite alfoíde humano (Willard & Waye 1987; Rudd e cols. 2006).

Além dos mecanismos de homogeneização das sequências, há os mecanismos responsáveis pela dispersão de sequências de DNAs satélites por todo o genoma. Sequências de alguns elementos transponíveis apresentam alta similaridade com sequências de DNAs satélites, indicando que elementos móveis podem ser a fonte de origem e/ou dispersão para alguns DNAs satélites, como foi demonstrado em *Drosophila obscura* (Miller e cols. 2000).

Outro importante mecanismo de amplificação e dispersão de DNAs satélites ocorre através de DNAs extracromossômicos circulares (eccDNA). Estes eccDNAs são formados através de eventos de recombinação ectópica entre cópias de uma mesma cromátide, o que acarreta uma deleção de cópias que passam a constituir um DNA circular. Estas moléculas de DNA circular contendo cópias de DNA satélite podem ser amplificadas, caso haja uma região iniciadora de replicação dentro da molécula, e posteriormente inseridas no mesmo ou em outro local do genoma (Cohen & Segal 2009).

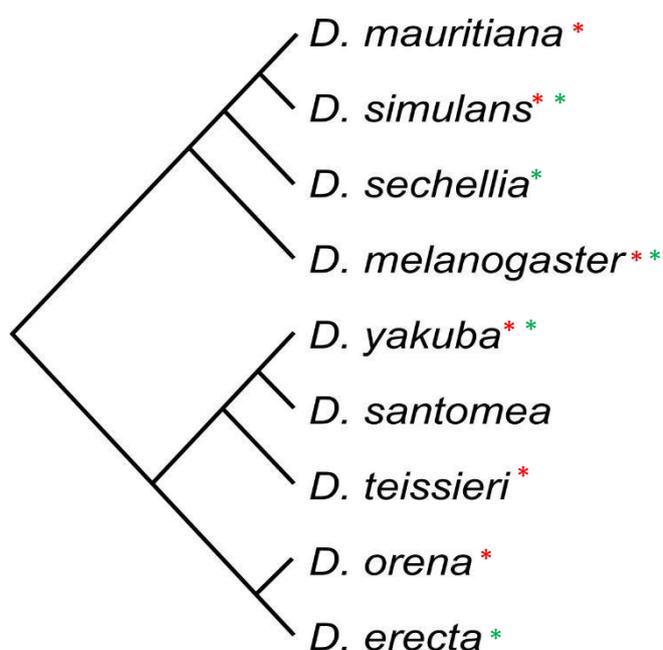
### **1.5. O gênero *Drosophila* e o subgrupo *melanogaster***

O gênero *Drosophila* é composto por mais de 2.000 espécies e é subdividido em dois subgêneros: *Drosophila* e *Sophophora* (Powell, 1997). De acordo com dados biogeográficos e fósseis, a separação entre os dois subgêneros ocorreu entre 40 e 60 milhões de anos atrás (Powell & DeSalle, 1995; Powell 1997).

Espécies do gênero *Drosophila* estão distribuídas por praticamente todo o globo terrestre, em regiões temperadas, tropicais e desertos. A maioria utiliza tecidos de plantas e fungos para a procriação e se alimenta de bactérias e leveduras associadas à decomposição deste material. Devido à grande quantidade de espécies, os drosofilistas criaram um sistema de classificação que reconhece, após o gênero, grupos, complexos e *clusters* de espécies. Embora não reconhecido pelo código internacional de Nomenclatura Zoológica, esta classificação é amplamente utilizada em estudos deste gênero.

O grupo *melanogaster* é formado por 174 espécies (Asburner e cols. 1984), classificadas em doze subgrupos, que se dividem em três subgrupos principais: subgrupo *ananassae*, subgrupo *montium* e subgrupo *melanogaster*. O grupo *melanogaster* teve como centro de diversificação a região sul da Ásia (Throckmorton 1975), enquanto o subgrupo *melanogaster*, formado por *D. melanogaster* e mais oito espécies filogeneticamente próximas, provavelmente se originou na região Afrotropical (Lachise e cols. 1988; Keller 2007) (Figura 2).

*D. melanogaster* (Meigen, 1830) e *D. simulans* (Sturtevant, 1919) passaram por um recente processo de colonização de outros continentes e atualmente são consideradas cosmopolitas. A espécie *D. yakuba* foi identificada no início dos anos 1950 e ocorre na porção ocidental da África, onde hoje é a Costa do Marfim (Burla 1954). Outras cinco espécies foram descobertas nos anos de 1970 e 80 (Tsacas e cols. 1971,1974,1977,1981). Dentre estas, *D. teissieri*, *D. erecta* e *D. orena* ocorrem em porções continentais da África e as outras duas espécies são endêmicas de ilhas da costa oriental africana: *D. sechellia* ocorre no arquipélago de Seychelles e *D. mauritiana* habita as ilhas Maurício. Outra espécie do subgrupo, recentemente identificada na ilha de São Tomé, na porção atlântica do continente africano, recebeu o nome de *D. santomea* (Lachaise e cols. 2000).



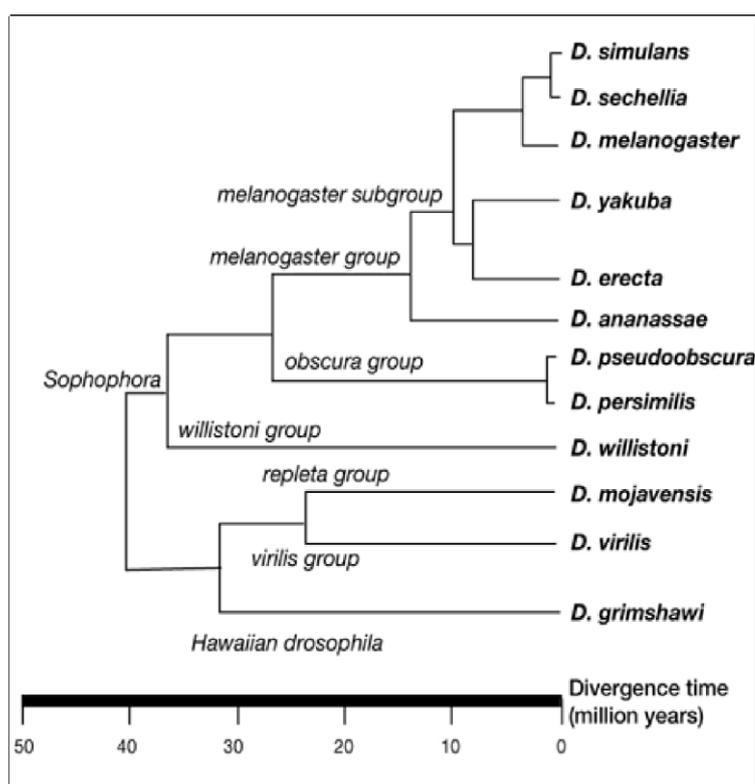
**Figura 2.** Filogenia mostrando as relações evolutivas entre as nove espécies do subgrupo *melanogaster*. Asteriscos em vermelho indicam espécies em que o DNA satélite 1.688 foi estudado (Strachan e cols. 1985). Asteriscos em verde indicam as espécies com genoma completamente sequenciado até 2007. Figura adaptada de Gidaszewski e cols. (2009).

## 1.6. Genomas sequenciados de *Drosophila* e o estudo de DNAs satélites

A publicação do genoma de várias espécies de *Drosophila* fornece um material valioso para biólogos evolutivos, permitindo a caracterização de vários elementos

genéticos e a determinação de sua organização, função e dinâmica evolutiva (*Drosophila* 12 Genomes Consortium 2007). Entre as 12 espécies com genomas completamente sequenciados, há pelo menos uma representante de cada grande grupo taxonômico de *Drosophila* (Figura 3). Portanto, encontra-se disponível um reservatório público imenso de sequências genômicas provenientes de espécies de ampla distribuição filogenética, o que é essencial para estudos evolutivos.

Os DNAs satélites estão entre os principais componentes genômicos de *Drosophila* e a sua contribuição para os genomas das espécies deste gênero pode variar bastante, de 2% em *D. mojavensis* até aproximadamente 50% em *D. virilis* (Bosco e cols. 2007). Apesar desta abundância, poucos destes DNAs satélites foram identificados e estudados em detalhe. De fato, os estudos sobre DNAs satélites no gênero *Drosophila* se concentraram em algumas poucas espécies e foram realizados antes da era genômica de *Drosophila*, resultando em uma análise incompleta e pouco abrangente do papel que tais sequências desempenham na arquitetura e organização do genoma.



**Figura 3.** Filogenia das 12 espécies do gênero *Drosophila* com genomas completamente sequenciados e tempos de divergência estimados (em milhões de anos) (*Drosophila* 12 Genomes Consortium 2007).

Em *Drosophila*, mudanças no conteúdo de DNA satélite durante a evolução estão diretamente relacionadas à expansão e contração do genoma em muitas espécies (Bosco e cols. 2007). Também foi sugerido que estas mudanças podem afetar características de valor seletivo, como o tamanho corporal e o tempo de desenvolvimento (Gregory & Johnston 2008).

Em outro exemplo, Ferree & Barbash (2009) mostraram que o DNA satélite 1.688 do cromossomo X de *D. melanogaster* está diretamente envolvido na letalidade de fêmeas híbridas do cruzamento entre *D. melanogaster* e *D. simulans*. Esta letalidade, observada nos primeiros estágios do desenvolvimento, é causada pela falha na segregação do cromossomo X de *D. melanogaster* no híbrido que, por sua vez, é consequência da falta de heterocromatização de um bloco de DNAs satélite 1.688 presente neste cromossomo. Este fenômeno é provavelmente causado por uma incompatibilidade de ligação deste bloco de DNA satélite de *D. melanogaster* a proteínas de *D. simulans* responsáveis pela modulação da cromatina. Assim, a divergência entre DNAs satélites de espécies próximas pode atuar como um mecanismo de isolamento reprodutivo, contribuindo para o processo de especiação.

Apesar de sua abundância e importância evolutiva, estudos sobre DNAs satélites não foram contemplados no artigo que apresentou a primeira versão “completa” do genoma de *D. melanogaster* (Adams e cols. 2000). Em publicações subsequentes, foi feito progresso na análise dos elementos transponíveis (TEs) e do conteúdo gênico da heterocromatina (Hoskins e cols. 2002, 2007; *Drosophila* 12 Genomes Consortium 2007). No entanto, o estudo dos DNAs satélites vem avançando a uma velocidade bem menor, devido principalmente à associação histórica destas sequências com “DNAs egoístas” ou “DNA entulho” (Orgel & Crick 1980) e também pelas conhecidas dificuldades operacionais e computacionais para montagem de sequências em tandem a partir de genomas sequenciados (Treangen & Salzberg 2011).

Repetições de DNAs satélites representam um desafio técnico para programas de alinhamento e montagem genômica por criarem ambiguidades e vieses na interpretação de dados. A metodologia de sequenciamento por WGS (*Whole Genome Shotgun*), usada no sequenciamento dos genomas de *Drosophila*, gera diversos pequenos fragmentos sequenciados (*reads*) que são montados em *contigs* de acordo

com regiões de similaridade comuns entre elas. Tendo em vista o alto grau de similaridade entre sequências de DNAs satélites, a montagem destas sequências se torna bastante complicada tecnicamente, resultando em uma redução no volume de dados de DNAs repetitivos nos arquivos montados dos genomas (Treangen & Salzberg 2011).

A metodologia do sequenciamento de *WGS* produz em geral grandes *scaffolds* relativamente bem montados que correspondem às regiões eucromáticas do genoma. No entanto, também são produzidos pequenos *scaffolds*, de localização genômica desconhecida, e provavelmente de origem heterocromática. Por exemplo, a montagem inicial do genoma de *D. melanogaster* gerou 178 *scaffolds* grandes (com tamanho médio de 658 kb) pertencentes a regiões de eucromatina e 2.597 *scaffolds* pequenos (com tamanho médio de 8kb) pertencentes às regiões de heterocromatina (Hoskins e cols. 2002).

Segundo Bosco e cols. (2007), estima-se que 25% do genoma de *D. melanogaster* seja composto por DNAs satélites. No entanto, apenas 2% das *reads* do sequenciamento contêm DNAs satélites, indicando também uma maior dificuldade para a clonagem destas sequências (Adams e cols. 2000).

Se o genoma de *D. melanogaster*, espécie modelo e uma das mais amplamente estudadas, foi muito pouco explorado em relação aos DNAs satélites, nas demais espécies sequenciadas de *Drosophila* os estudos são praticamente inexistentes.

Para um melhor entendimento sobre a organização e evolução do genoma, é essencial incluir os DNAs satélites nos estudos genômicos. Entre as espécies de *Drosophila* com genoma sequenciado até 2007, encontram-se cinco espécies do subgrupo *melanogaster*: *D. yakuba*, *D. erecta*, *D. sechellia*, *D. simulans* e *D. melanogaster* (Figura 2). As linhagens selecionadas para o sequenciamento são isogênicas e sequenciadas com altos valores de cobertura, fator que aumenta a qualidade final da montagem dos genomas. Os dados de *D. simulans*, por outro lado, foram obtidos através do sequenciamento de sete linhagens diferentes e com baixos valores de cobertura do genoma a fim de produzir dados diferenciados entre populações (*Drosophila* 12 Genomes Consortium 2007).

Um estudo usando os genomas montados das 12 espécies com genomas completamente sequenciados em 2007 estimou o percentual de DNA satélite presente em cada espécie. As espécies do subgrupo *melanogaster* possuem uma quantidade variável de DNAs satélites, i.e., 9% em *D. erecta*, 17% em *D. simulans*, 23% em *D. yakuba* e 24% em *D. melanogaster* e *D. sechellia* (Bosco e cols. 2007). Embora estes dados indiquem um grande percentual de DNAs satélites na composição destes genomas, nenhum deles foi amplamente estudado, após o sequenciamento dos genomas destas espécies.

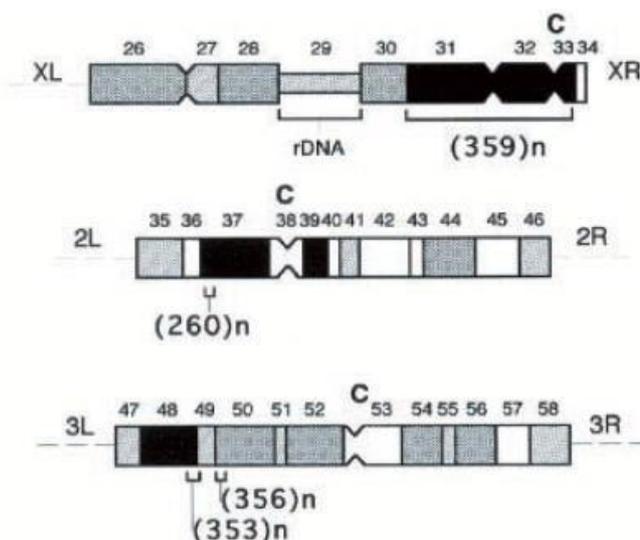
### **1.7. O DNA satélite 1.688 g/cm<sup>3</sup> do subgrupo *melanogaster***

Entre os DNAs satélites que compõem o genoma de *D. melanogaster*, destacase o DNA satélite 1.688 g/cm<sup>3</sup> (Lohe & Roberts 1993), o mais abundante desta espécie, com aproximadamente 15.000 cópias (~5% do genoma) e unidades de repetição de aproximadamente 359 pb (Hsieh & Brutlag 1979; Carlson & Brutlag 1979).

O DNA satélite 1.688 é formado por diferentes subfamílias. A subfamília '359' é a mais frequente e é composta por cópias de 359 pb densamente localizadas na heterocromatina do braço esquerdo do cromossomo X (Brutlag 1979; Lohe e cols. 1986). As subfamílias '353' e '356', compostas por cópias de 353 e 356 pb, respectivamente, estão localizadas na heterocromatina do braço esquerdo do cromossomo 3 (Lohe & Brutlag 1986; Losada & Villasante 1996). A subfamília '260', com cópias de 260 pb, está localizada na heterocromatina do braço esquerdo do cromossomo 2 (Figura 4) (Abad e cols. 2000).

Atipicamente, foram encontradas diversas cadeias contendo um número reduzido de cópias em tandem (<5 cópias) do DNA satélite 1.688 nas regiões eucromáticas dos cromossomos X, 2 e 3 (Waring & Pollack 1987; DiBartolomeis e cols. 1992; Losada & Villasante 1996; Koryakov e cols. 1999; Kuhn e cols. 2012).

Cópias do DNA satélite 1.688 também foram encontradas em outras espécies de *Drosophila* do subgrupo *melanogaster*, como *D. simulans*, *D. mauritiana*, *D. teissieri*, *D. yakuba* e *D. oreana* (Strachan e cols. 1985) (Figura 2).



**Figura 4.** Regiões heterocromáticas dos cromossomos X, 2 e 3 de *D. melanogaster* contendo a localização das subfamílias '359', '260', '353' e '356' do DNA satélite 1.688 (figura extraída de Abad e cols. 2000). C = centrômero. Números acima dos cromossomos correspondem a diferentes posições citológicas.

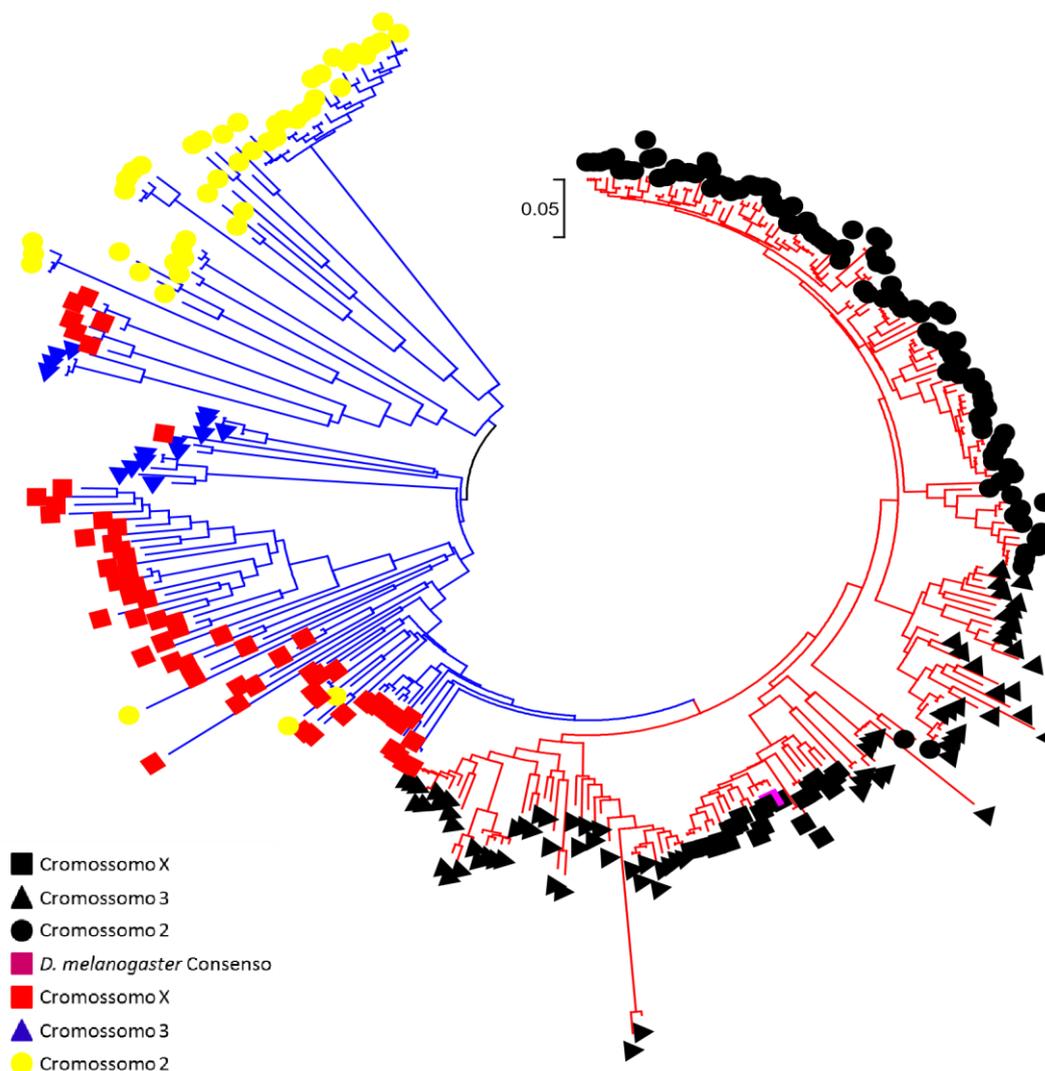
Até o presente, apenas um estudo realizou comparações nucleotídicas entre cópias do DNA satélite 1.688 em diferentes espécies. Strachan e cols (1985) comparam poucas cópias 1.688 (quatro a nove) entre seis espécies do subgrupo *melanogaster* (Figura 2). Análises de similaridade nucleotídica intra- e interespecíficas demonstraram um padrão de evolução combinada para todas as espécies analisadas, exceto para *D. simulans* e *D. mauritiana*, espécies muito próximas filogeneticamente. Os autores ainda observaram que a filogenia das cópias 1.688 refletia a filogenia das espécies, sugerindo que uma evolução gradual. No entanto, nenhuma árvore filogenética foi apresentada.

Existe grande diferença na abundância de cópias do DNA satélite 1.688 entre espécies. Por exemplo, cópias do DNA satélite 1.688 representam apenas 0,11% e 0,24% dos genomas das espécies *D. simulans* e *D. erecta*, respectivamente (Lohe & Brutlag 1986). Informações sobre a presença e abundância do DNA satélites 1.688 em outras espécies são inexistentes e dados a respeito de sua distribuição genômica em outras espécies além de *D. melanogaster*, quando existem, são escassos e imprecisos.

Recentemente, Kuhn e cols. (2012) realizaram o estudo mais extenso sobre o DNA satélite 1.688 em *D. melanogaster* usando principalmente sequências presentes em BACs (cromossomos artificiais de bactérias) com localização cromossômica

determinada. Ao todo, foram analisadas 326 cópias 1.688 eucromáticas e heterocromáticas provenientes de 52 cadeias dos cromossomos X, 2 e 3. Os resultados ampliaram consideravelmente o conhecimento sobre a distribuição cromossômica e variação (tamanho, composição e orientação) deste DNA satélite. Em especial, os dados mostraram que, em geral, cópias 1.688 de uma mesma cadeia ou cromossomo são mais similares entre si do que cópias de diferentes cadeias, e que cópias heterocromáticas são mais similares entre si do que cópias eucromáticas. Conseqüentemente, uma árvore filogenética contendo todas as cópias 1.688 estudadas mostrou agrupamentos cromossomo-específicos e agrupamentos exclusivos de cópias heterocromáticas ou eucromáticas (Figura 5). Foi determinado que a maioria das cadeias eucromáticas está localizada próxima a genes. Os dados sobre a localização genômica das cadeias e a topologia da árvore contendo cópias 1.688 de diferentes regiões cromossômicas obtidos por Kuhn e cols. (2012) fornecem um arcabouço sólido para o estudo de cópias 1.688 em outras espécies de *Drosophila* com genoma sequenciado.

A maior parcela de repetições de DNA satélites presentes em *D. melanogaster* está organizada em grandes *scaffolds* formados por *contigs* concatenados de maneira randômica em um arquivo denominado de “Cromossomo U” (flybase.org). Mais de 2.500 sequências geradas no sequenciamento de *D. melanogaster* não foram mapeadas e foram depositadas no “Cromossomo U”. Até o presente, apenas algumas dezenas de cópias do DNA satélites 1.688 foram anotadas no genoma sequenciado de *D. melanogaster* (revisado por Kuhn e cols. 2012). Portanto, a grande maioria das cópias deve residir no “Cromossomo U” e permanece inexplorada.



**Figura 5.** Árvore filogenética de Máxima Verossimilhança contendo cópias do DNA satélite 1.688 no genoma de *D. melanogaster* estudadas por Kuhn e cols.(2012). Esta árvore ilustra um padrão de evolução combinada intragenômico, com a maioria das cópias provenientes da mesma região genômica agrupadas em ramos específicos. Ramos azuis são compostos por cópias eucromáticas, enquanto os vermelhos são compostos por cópias heterocromáticas. Quadrados lilás: Sequência consenso de *D. melanogaster* descrita por Strachan e cols. (1985). Figura adaptada de Kuhn e cols. (2012).

## 2. OBJETIVO

### 2.1. Objetivo Geral

Este trabalho teve como objetivo estudar a organização e evolução molecular do DNA satélite 1.688 nas espécies do subgrupo *melanogaster* que tiveram seus genomas sequenciados em 2007.

### 2.2. Objetivos específicos

- Isolamento e caracterização das cópias do DNA satélite 1.688 presentes no “Cromossomo U” de *D. melanogaster*;
- Isolamento e caracterização do DNA satélite 1.688 nas espécies *D. simulans*, *D. sechellia*, *D. yakuba* e *D. erecta*;
- Determinação da distribuição genômica das cópias do DNA satélite 1.688 nas espécies *D. simulans*, *D. sechellia*, *D. yakuba* e *D. erecta*;
- Estudo da evolução molecular do DNA satélite 1.688 em *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba* e *D. erecta*;
- Verificação da distribuição do DNA satélite 1.688 nas demais espécies de *Drosophila* com genomas sequenciados não pertencentes ao subgrupo *melanogaster*.

### 3. MATERIAL E MÉTODOS

#### 3.1. Material

##### 3.1.1 Linhagens de *Drosophila*

Foram utilizadas linhagens representando quatro espécies do gênero *Drosophila*: *D. melanogaster*, *D. simulans*, *D. sechellia* e *D. willistoni*. As linhagens das três primeiras espécies foram derivadas de coletas realizadas no estado de São Paulo, em Florianópolis (Santa Catarina) e nas ilhas Seychelles (parte oriental do continente africano). A linhagem de *D. willistoni* estudada foi a mesma utilizada no sequenciamento de seu genoma (estoque 14030-0814.10). Todas as linhagens vêm sendo mantidas no Laboratório de Citogenômica Evolutiva da UFMG através de sucessivas transferências de novas gerações de indivíduos para novos frascos contendo meio de cultura feito à base de fubá e fermento biológico.

##### 3.1.2 Reagentes e produtos químicos

Tampão de Homogeneização	100 mM Tris-base; 160 mM Sacarose; 80 mM EDTA pH ajustado para 8,0 com 10 N NaOH)
Tampão 50 x TAE	0,89 M Tris-Base; 1M ácido acético; 20 mM EDTA
Tampão Tris-EDTA	1 mM EDTA; 10 mM Tris-HCl (pH 7,8)
Tampão de Lavagem	10% SDS; 1 M NaAcPO <sub>4</sub>
Solução de Brometo de Etídeo	10 mg/mL em tampão 1 x TBE
IPTG	100 mM (23 mg/mL em H <sub>2</sub> O destilada)
X-GAL	2 % (20 mg/mL em N,N-dimetilformamida)
LB Agar (1litro)	0,58 g NaCl; 10 g triptona; 5 g extrato de levedura; 0,19 g KCl; 15 g ágar (pH 7,4)
Meio LB (1 litro)	0,58g NaCl; 10g Triptona; 5g extrato de levedura; 0,19g KCl (pH 7,4)
Solução de RNase	10 mg/mL H <sub>2</sub> O
20 x SSC	3 M NaCl; 0,3 M Na <sub>3</sub> citrato (pH 7,0)
Marcadores de Peso Molecular	200 base-pair Ladder: Promega

#### 3.2. Metodologia

### 3.2.1 Identificação e mineração das sequências do DNA satélite 1.688 nos genomas sequenciados de *Drosophila*.

Os genomas sequenciados das espécies de *Drosophila* encontram-se depositados nos bancos de dados “Flybase”, no endereço <http://flybase.org/> ou em <http://rana.lbl.gov/Drosophila>.

A identificação de sequências similares ao DNA satélite 1.688 nos genomas das espécies sequenciadas de *Drosophila* foi realizada através do algoritmo BLAST (Basic Local Alignment Search Tool) com parâmetros padrões, exceto pelo não acionamento da opção “low complexity regions”, e pela seleção por sequências “Somewhat like similars”. Para as buscas foram utilizadas sequências consenso do DNA satélite 1.688 descritas por Strachan e cols. (1985) para as espécies do subgrupo *melanogaster* (Tabela 1).

**Tabela 1.** Sequências consenso utilizadas para a busca de cópias 1.688 no genoma das espécies sequenciadas de *Drosophila*.

Sequência 1.688 consenso utilizada no BLAST (Strachan e cols. 1985)	Espécies analisadas com a sequência consenso
<i>D. melanogaster</i>	<i>D. melanogaster</i> ; <i>D. simulans</i> ; <i>D. sechellia</i> ; <i>D. yakuba</i>
<i>D. simulans</i>	<i>D. simulans</i> ; <i>D. sechellia</i>
<i>D. yakuba</i>	<i>D. yakuba</i> ; <i>D. erecta</i>
<i>D. orena</i>	<i>D. erecta</i>
<i>D. teissieri</i>	<i>D. yakuba</i>

Foram selecionados para análise inicial somente *hits* que apresentaram *e-value* menores que  $10^{-5}$  e com cobertura de homologia mínima de 70% em uma região correspondente a 80% da sequência usada nas buscas. Os *hits* encontrados foram provenientes de três fontes principais: *contigs* cobertos de repetições do DNA satélite 1.688; *scaffolds* formados por um número superior a 2 Mb; cromossomos montados, porém com grande número de *contigs* e *scaffolds* sem a indicação exata da posição genômica, fato que impossibilitou a extração de suas sequências. Para facilitar o isolamento de sequências 1.688 de *scaffolds* ou cromossomos montados foi utilizado o programa *Tandem Repeat Finder* (Benson 1999). Os parâmetros utilizados para o reconhecimento das repetições em tandem foram ajustados para repetições com

tamanho entre 250 pb e 400 pb e composição nucleotídica de aproximadamente 70% de A+T, que consistem em características gerais do DNA satélite 1.688.

Como metodologia alternativa, a mineração de sequências 1.688 presentes em grandes *scaffolds* montados ou arquivos cromossômicos totais foi realizada através do programa GATA 0.7. Este programa utiliza o algoritmo BLAST para realizar um alinhamento comparativo entre sequências e tem como resultado um gráfico *box-linebox* que representa as janelas de *scores* dos alinhamentos locais. Para o BLAST, sequências consenso do DNA satélite 1.688 foram utilizadas, como descrito anteriormente. Foram utilizados os parâmetros “default” do programa, ou seja, 5 para *matches*, -4 para *mismatches*, -10 para geração de *gaps* e -4 para extensão dos *gaps*. A visualização do gráfico foi realizada no programa GATAPlotter, no qual foram utilizados os parâmetros “default”, com *score* mínimo de corte de 70% em uma janela de 24 nucleotídeos. Com a representação gráfica gerada pelo programa GATA 0.7 foi possível visualizar e extrair as sequências 1.688 presentes em grandes arquivos, como os provenientes do cromossomo U, cromossomos montados ou *scaffolds*.

As sequências homólogas ao DNA satélite 1.688 encontradas nos bancos genômicos foram retiradas dos arquivos “.fasta” e comparadas, cadeia por cadeia, com as sequências consenso de *D. melanogaster*, *D. simulans*, *D. yakuba*, *D. orena* e *D. teissieri* pelo método de “dot-plot”, implementado no programa online *Dotlet* do SIB (Swiss institute of Bioinformatics). O “dot-plot” é uma ferramenta de análise gráfica que permite a comparação de duas sequências nucleotídicas e identifica regiões de similaridade entre elas. Através deste método, também foi possível identificar com maior precisão o início e o término das cadeias de sequências 1.688, sendo anotadas como sequências 1.688 aquelas que obtiveram índices de similaridade superiores a 60% em janelas de 15 pb.

### **3.2.3. Determinação *in silico* da distribuição genômica das cópias do DNA satélite 1.688.**

A distribuição genômica das cópias 1.688 foi inferida através dos agrupamentos destas cópias nas árvores filogenéticas contendo cópias 1.688 de *D. melanogaster* com localização citológica (em cromossomos politênicos e metafásicos) já estabelecida (revisado por Kuhn e cols. 2012) (Figura 5). Quando disponíveis, 500 pb

das sequências flanqueadoras das cópias de cada cadeia 1.688 obtida no presente trabalho foram utilizados para a procura de sequências homólogas em *D. melanogaster* através do navegador “GenomeBrowser” (hospedado no portal “<http://flybase.org/>”). Este navegador contém as sequências montadas de cada cromossomo de *D. melanogaster* e um *link* para sua localização citológica em cromossomos politênicos.

### **3.2.4. Alinhamentos nucleotídicos, análises filogenéticas e variabilidade nucleotídica.**

Foram selecionadas para análise apenas cópias completas do DNA satélite 1.688 e apresentando, na maioria dos casos, pelo menos 190 pb de extensão. O tamanho das cópias foi analisado pelo programa *MEGA* versão 5.05 (Tamura e cols. 2011).

As sequências nucleotídicas foram alinhadas com o algoritmo *Muscle* (Edgar 2004), utilizando o padrão “default” do programa, com penalidade de -400 para abertura de *gaps* e sem penalidade para prolongamento do *gap*. Os alinhamentos foram otimizados manualmente quando necessário.

Árvores filogenéticas foram construídas pelo algoritmo *Neighbor Joining* (Saitou & Nei 1987) presente no programa *MEGA 5.05* (Tamura e cols. 2011). As distâncias genéticas entre as sequências foram calculadas através do modelo “Kimura-2Parâmetros” ou “K2P” (Kimura 1980). Um teste estatístico de consistência da topologia foi feito por *bootstrap* (1.000 replicações e 10 turnos de adições aleatórias de sequências).

O programa *MEGA 5.05* também foi utilizado para análises do tamanho, composição nucleotídica, conteúdo A+T e grau de variabilidade entre sequências. Para estas análises foram utilizadas cópias completas, ou seja, cópias que apresentaram as porções de início e fim delimitadas.

### **3.3. Análises Moleculares**

#### **3.3.1. Extração de DNA genômico**

Entre 30-50 moscas de cada espécie foram homogeneizadas com o auxílio de um micropistilo estéril em tubos de microcentrifuga de 1,5 mL contendo 500 µL de Tampão de Homogeneização. Em seguida, foram adicionados 25 µL de 10% SDS e 10 µL de Proteinase K (20mg/mL) e os tubos foram incubados por 1-2 horas a 68°C. Para a remoção das proteínas, 500 µL de uma mistura de Fenol/Clorofórmio (1:1) foram adicionados à solução e os tubos foram centrifugados (3 minutos a 14.000 rpm). Após a centrifugação, a fase aquosa foi transferida para novos tubos, uma nova mistura de Fenol/Clorofórmio foi adicionada e o mesmo procedimento foi repetido. Após a transferência da fase aquosa para novos tubos, 500 µL de clorofórmio foram adicionados às amostras, as quais foram então centrifugadas por 5 minutos a 14.000 rpm. A fase aquosa foi transferida para novos tubos e 45 µL de 8 M Acetato de Potássio foram adicionados à solução. Os tubos permaneceram por 1 hora em gelo para a precipitação das proteínas e RNAs. Após centrifugação por 15 minutos a 14.000 rpm, o sobrenadante foi transferido para novos tubos. O DNA foi então precipitado pela adição de 1 mL de etanol absoluto seguido de centrifugação por 15 minutos a 5.500 rpm. O etanol absoluto foi removido e 100 µL de etanol 70% foram adicionados, seguindo de centrifugação por 5 minutos a 5.500 rpm. Após a remoção do etanol, os tubos permaneceram abertos para secagem. Ao final, 50 µL de Tampão TE foram adicionados e as amostras foram mantidas a -20°C. Após 24 horas da extração de DNA, foi adicionado 1 µL de RNase (10mg/mL) e as amostras foram submetidas à temperatura de 68°C por 10 minutos e armazenadas a -20°C.

### 3.3.2. PCR (Reação em Cadeia da Polimerase)

A reação em cadeia da Polimerase (PCR) foi composta de acordo com a Tabela a seguir:

Reagente	Concentração do estoque	Concentração final	Volume para 1 reação (25µL)
H <sub>2</sub> O estéril	-	-	12µL
Tampão da enzima	5x	1x	5µL
dNTP mix	8mM	0,8mM	2,5µL
MgCl <sub>2</sub>	25mM	1,5mM	1,5µL
Oligo1	10µM	0,4µM	1µL
Oligo2	10µM	0,4µM	1µL
DNA		30 a 50 ng	1µL
Enzima (Taq)	1U/µL	1U	1µL
		<b>Total:</b>	25µL

O programa utilizado consistiu das seguintes configurações:

Denaturação Inicial	95 °C 2min
Ciclo	Denaturação 96 °C 30s
	(30x) Anelamento 50-60 °C 30s
	Extensão 72 °C 1min
Extensão Final	72 °C 5min
Finalização da reação	4 °C

Os oligonucleotídeos iniciadores utilizados nas PCRs foram os seguintes:

### 3.3.3. Eletroforese em géis de agarose

Oligonucleotídeos Iniciadores	Sequência do Oligonucleotídeo
<i>Docsat1</i>	<i>Docsat1: 5' - CCATACAAATCCCCTCGC - 3'</i>

<i>D. mel</i> 359 pb	Foward: 5'- CGTTAGCACTGGTAATTAGCTGC - 3' Reverse: 5'- CGATCCCTATTACTTTTTGAAGG - 3'
<i>D. sim</i> 360 pb	Foward: 5'- GTTTGTTTCTTAAATCCCAATCG - 3' Reverse: 5' - CTCAACGAGGTATGACATTCC -3'

A concentração, pureza e tamanho dos DNAs analisados foram verificados em géis de agarose 1%-1,5% diluída em tampão TAE. As amostras de DNA foram preparadas pela adição de 5 µL do tampão de carregamento Loading Dye 6x (Promega). As eletroforeses foram realizadas a 70mA por 55 a 70 minutos e os géis foram corados com brometo de etídeo e visualizados e fotografados em transluminador UV.

#### 3.3.4. Eluição de amplicons de PCR de géis de agarose

As eluições de fragmentos de DNA amplificados por PCR foram realizadas utilizando o *kit* “Wizard® SV Gel and PCR Clean-Up System” (Promega). Os fragmentos de DNA de interesse foram excisados dos géis de agarose com o auxílio de uma lâmina de bisturi e transferidos para tubos de microcentrífuga de 1,5 mL. Foram adicionados 10 µl da “solução de ligação a membrana” para cada 10 mg de gel de agarose e os tubos foram incubados por 15 minutos a 65°C até a dissolução completa do gel. A solução foi transferida para tubos contendo minicolunas de filtragem, que foram incubados por 1 minuto à temperatura ambiente. Em seguida, os tubos foram centrifugados a 14.000 rpm por 1 minuto. Nesta etapa, espera-se que o DNA permaneça retido na membrana de filtragem das minicolunas. Posteriormente, foi realizada a lavagem da membrana da microcoluna pela adição de “solução de lavagem de membrana” e centrifugação a 14.000 rpm por 1 minuto. A mesma lavagem foi repetida, seguida de centrifugação por 5 minutos. Como etapa final, foram adicionados 50 µl de H<sub>2</sub>O (molecular grade-Sigma) às minicolunas, que foram incubadas à temperatura ambiente. As microcolunas foram então centrifugadas a 14.000 rpm por 1 minuto. O material eluído foi armazenado em freezer a -20°C.

#### 3.3.5. Ligação do DNA amplificado em plasmídios vetores

A ligação dos produtos obtidos através de PCR foi realizada com o *kit* “pGEMT easy vector” (Promega). O plasmídeo pGEM deste kit é linearizado com EcoRV e uma timidina 3' terminal é incorporada em ambas as extremidades. Este procedimento

permite a inserção de produtos de PCR uma vez que a maioria das polimerases disponíveis frequentemente incorpora uma deoxiadenosina nas extremidades 3' dos fragmentos amplificados.

Para o procedimento de ligação foram utilizados tubos de PCR de 0.5 mL devido à sua baixa capacidade de ligação ao DNA em relação a tubos de 1,5mL. A reação de ligação foi composta por:

Reagentes	Amostras	Controle Positivo	Controle Negativo
2x Rapid T4 Ligase Buffer	5 µl	5 µl	5 µl
pGEM vector( 50ng)	1 µl	1 µl	1 µl
PCR product	3 µl	-	-
DNA Controle	-	2 µl	-
T4DNA Ligase (3weisU/unidade)	1 µl	1 µl	1 µl
H2O sigma	-	1 µl	3 µl
Volume Final	10 µl	10 µl	10 µl

Para obter o número máximo de transformantes, a reação foi homogeneizada e incubada por 24h a 4°C.

### 3.3.6. Transformação bacteriana e extração do DNA plasmidial

Os tubos contendo as reações de ligação foram centrifugados e 2 µL de transferidos para tubos de microcentrífuga de 1,5 mL. Os tubos contendo as células competentes (JM109-Promega) foram removidos do freezer – 80°C e colocados em gelo por 5 minutos e 50 µL de células competentes foram adicionados a cada amostra. Para a reação de controle, foram utilizados 100 µL. As amostras foram incubadas em gelo por 20-30 minutos e seguiu-se o choque-térmico em banho-maria a 42°C por 45 segundos. As amostras foram então imediatamente incubadas em gelo por 2 minutos. As células foram transferidas para 950 µL de meio LB líquido sem ampicilina para as ligações e 900 µL para as reações de controle. As amostras foram incubadas por 90 minutos em agitação constante de 230 rpm a 37°C e 100 µL de cada cultura foram plaqueados em placas com meio LB ágar com X-gal (80 µg/mL), IPTG (0.5 mM) e ampicilina (100 µg/mL). As placas foram incubadas em estufa a 37°C por 16 horas.

Os clones recombinantes foram selecionados pelo sistema *blue-white* do gene  $\beta$ -galactosidase. Cada colônia branca das placas contendo meio LB+ampicilina foi transferida individualmente para tubos *Falcon* de 15 mL contendo 5 mL de meio LB líquido com ampicilina. Estes tubos foram incubados a 37°C em agitação contínua de 200-250 rpm por 16 horas. Para a confirmação dos clones recombinantes, foram retirados 0,5  $\mu$ L da cultura dos tubos *Falcon* e que foram submetidos a uma PCR com os parâmetros:

<b>PCR das culturas de bactérias</b>	
<b>Reagentes</b>	<b>Volume</b>
<b>Primer M13 F'</b>	1 $\mu$ L
<b>Primer M13 R'</b>	1 $\mu$ L
<b>dNTP mix</b>	2,5 $\mu$ L
<b>MgCl<sub>2</sub></b>	1,5 $\mu$ L
<b>Taq Buffer (5X)</b>	5 $\mu$ L
<b>Taq DNA Polimerase</b>	1 $\mu$ L
<b>DNA (direto das culturas)</b>	0,5 $\mu$ L
<b>H<sub>2</sub>O</b>	12,5 $\mu$ L
	<b>Volume Final/tubo 25 <math>\mu</math>L</b>

As extrações de DNA plasmidial foram realizadas com o *kit* “Pure Yield MiniPrep” (Promega), de acordo com o protocolo fornecido pelo fabricante.

### **3.3.7. Sequenciamento dos clones**

O sequenciamento das amostras foi realizado no Laboratório de Genética Animal do Departamento de Medicina Veterinária da Universidade Federal de Minas Gerais. O aparelho utilizado para o sequenciamento das amostras foi um sequenciador ABI Prism 3130 (Life Technologies).

Os cromatogramas fornecidos como resultados dos sequenciamentos foram analisados no programa CHROMAS Lite 2.1.

## 4. RESULTADOS

Com a metodologia utilizada foi possível identificar mais de 8.000 sequências do DNA satélite 1.688 nas cinco espécies de *Drosophila* do subgrupo *melanogaster* com genoma sequenciado: *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba* e *D. erecta*. Foram selecionadas para análise apenas cópias completas do DNA satélite 1.688 e com pelo menos 190 pb de extensão. Com isto, o total de sequências analisadas foi reduzido para 6.849 cópias.

As sequências do DNA satélite 1.688 foram encontradas em três diferentes tipos de arquivos genômicos: (i) em *contigs* sem mapeamento cromossômico (ii) em *scaffolds* com ou sem mapeamento cromossômico e (iii) em cromossomos com *scaffolds* montados. A quantidade de sequências distribuídas nestes três tipos de arquivos variou bastante nas diferentes espécies. A principal diferença apresentada era a extensão da montagem genômica disponível para cada espécie. *D. melanogaster*, *D. simulans* e *D. yakuba* apresentaram cópias distribuídas em *contigs*, *scaffolds* e cromossomos montados. Por outro lado, em *D. sechellia* e *D. erecta* cópias 1.688 foram mineradas apenas em *contigs* e *scaffolds*, uma vez que nestas espécies não existem ainda arquivos disponíveis com a montagem do genoma em cromossomos individuais.

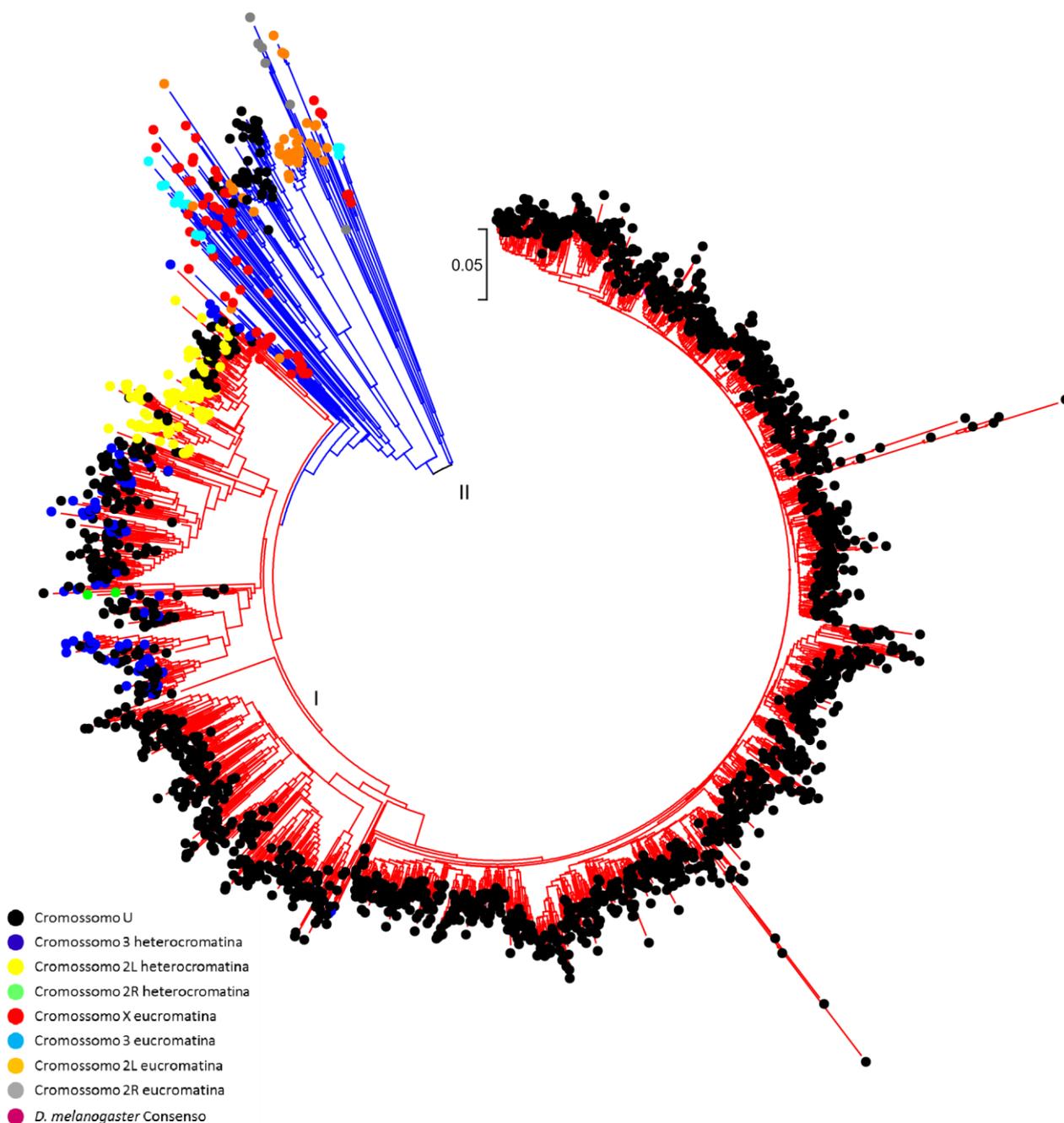
### 4.1. DNA satélite 1.688 em *D. melanogaster*

Em *D. melanogaster* foram identificadas 2.041 sequências homólogas ao DNA satélite 1.688 distribuídas em 395 *contigs* ou *scaffolds* do “Cromossomo U” (arquivo disponível no site [www.flybase.org](http://www.flybase.org)). Deste total, 1.838 cópias foram selecionadas para análise por se encaixarem dentro dos requisitos metodológicos propostos.

As 1.838 cópias do DNA satélite 1.688 obtidas do Cromossomo U foram adicionadas às 323 analisadas anteriormente por Kuhn e cols. (2012) e uma árvore NJ foi construída. De acordo com o padrão de ramificação da árvore NJ resultante, foi possível classificar as 1.838 cópias em eucromáticas e heterocromáticas e inferir a sua localização cromossômica (Figura 6). A árvore NJ separou as cópias 1.688 em dois grandes grupos. O grupo I é composto pela maior parte das sequências 1.793

cópias e contém cópias que se agruparam juntamente com sequências heterocromáticas analisadas por Kuhn e cols (2012). O grupo II corresponde às demais 43 sequências restantes agrupadas com sequências eucromáticas analisadas por Kuhn e cols (2012).

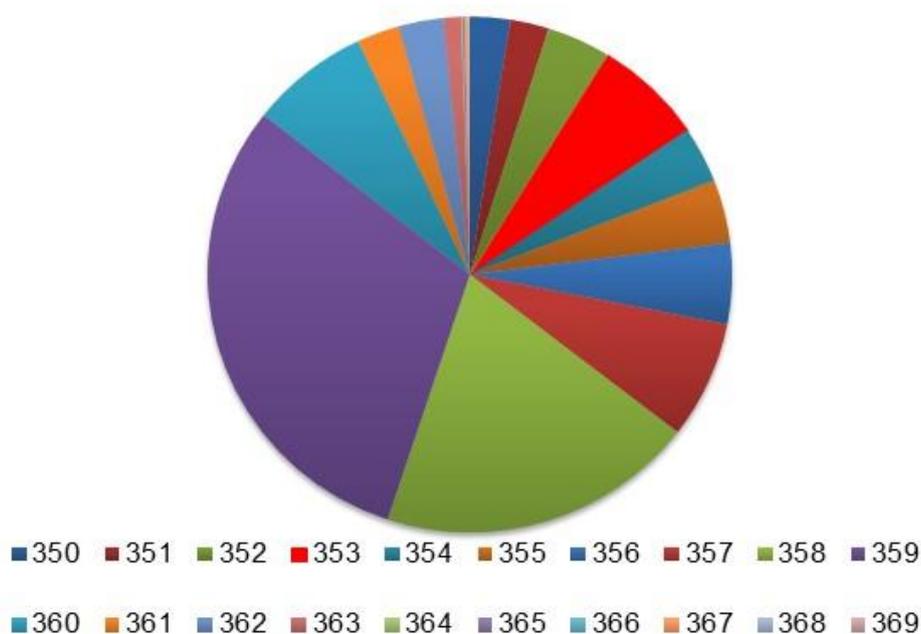
Com relação ao tamanho, as cópias heterocromáticas do 1.688 de *D. melanogaster* estudadas no presente trabalho variam de 200 pb a 386 pb. No entanto, a maioria das cópias possuía um tamanho entre 340 pb e 370 pb, sendo que 59,1% delas (1.097 cópias) estão concentradas no intervalo entre 350 pb a 359 pb (Figura 7). Cópias de 359 pb constituem a variante do DNA satélite 1.688 mais frequentemente encontrada no genoma de *D. melanogaster* (388 cópias ou 20,87% do total).



**Figura 6.** Árvore NJ construída com distâncias genéticas calculadas pelo método K2P e contendo todas as cópias do DNA satélite 1.688 de *D. melanogaster* analisadas no presente trabalho juntamente com as de Kuhn e cols (2012). Ramos vermelhos representam cópias heterocromáticas. Ramos azuis representam cópias eucromáticas. Círculos pretos representam cópias do Cromossomo U analisadas no presente trabalho. O círculo lilás representa a sequência heterocromática consenso de *D. melanogaster* descrita por Strachan e cols. (1985).

As 1.793 cópias heterocromáticas do Cromossomo U foram agrupadas em ramos contendo sequências heterocromáticas características dos cromossomos X, 2

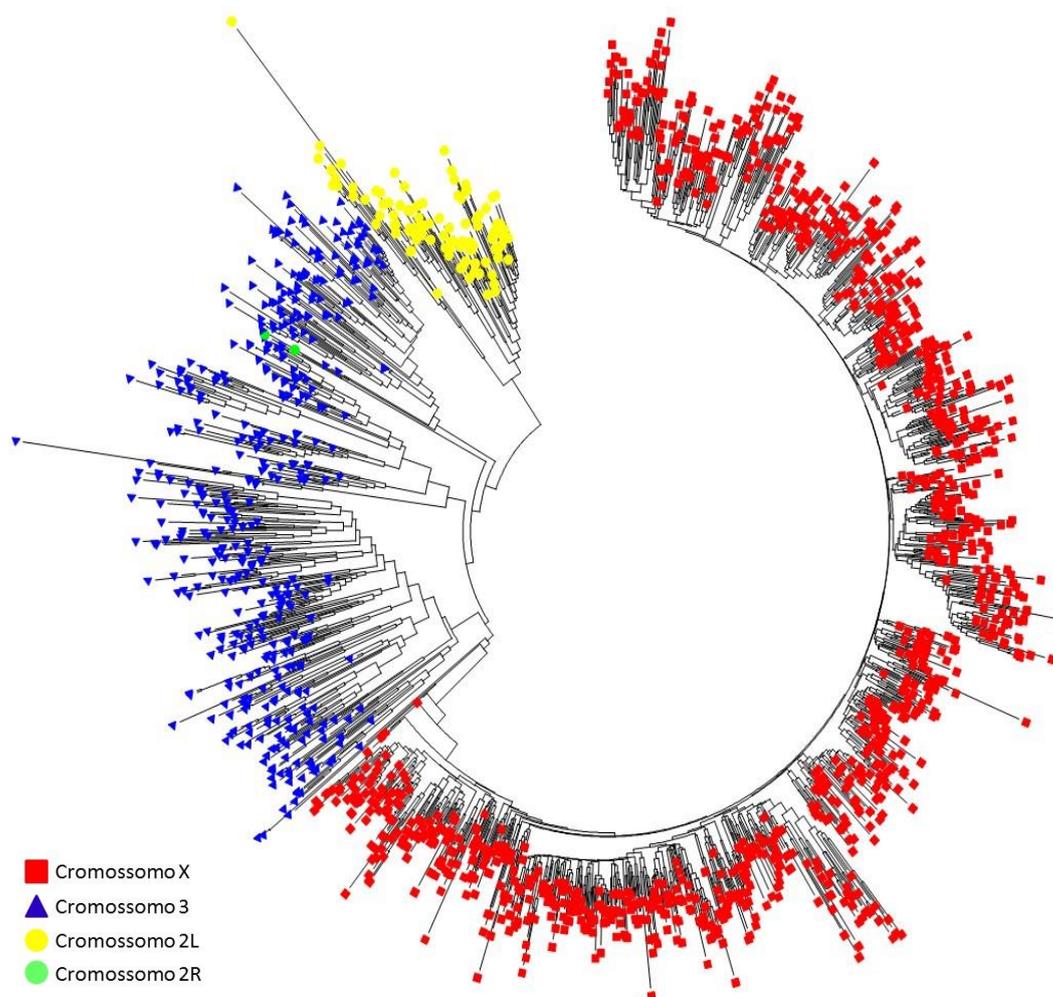
e 3. A maior parte das cópias (1315 sequências ou 72 %) foi agrupada no ramo formado por sequências do cromossomo X. Outras 456 cópias (25%) foram agrupadas nos ramos contendo sequências características do cromossomo 3. Apenas 45 cópias (3%) foram agrupadas juntamente com cópias do cromossomo 2 (Figuras 8 e 9).



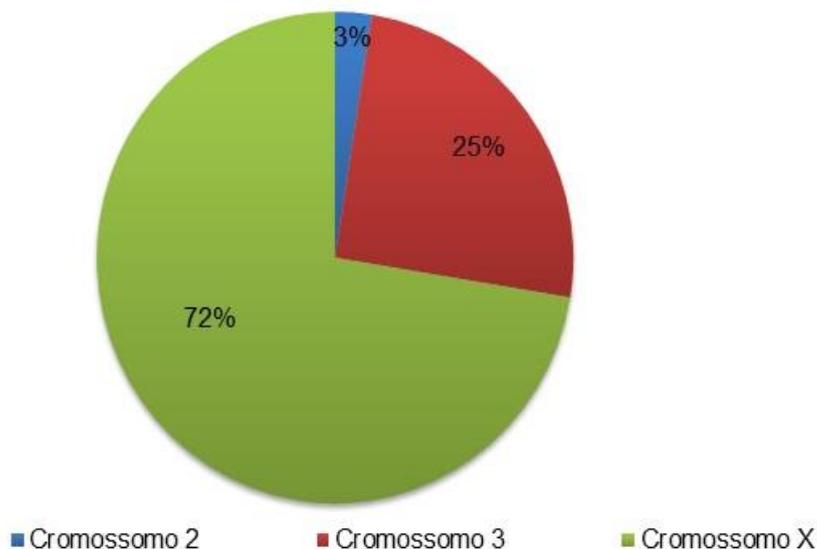
**Figura 7.** Distribuição da abundância de cópias heterocromáticas do DNA satélite 1.688 de *D. melanogaster* por tamanho das unidades de repetição (de 350 a 369 pb).

O cromossomo X teve como principal variante cópias de 359 pb e os cromossomos 2 e 3 tiveram como variantes principais cópias de 257 pb e 353 pb, respectivamente. (Figura 10).

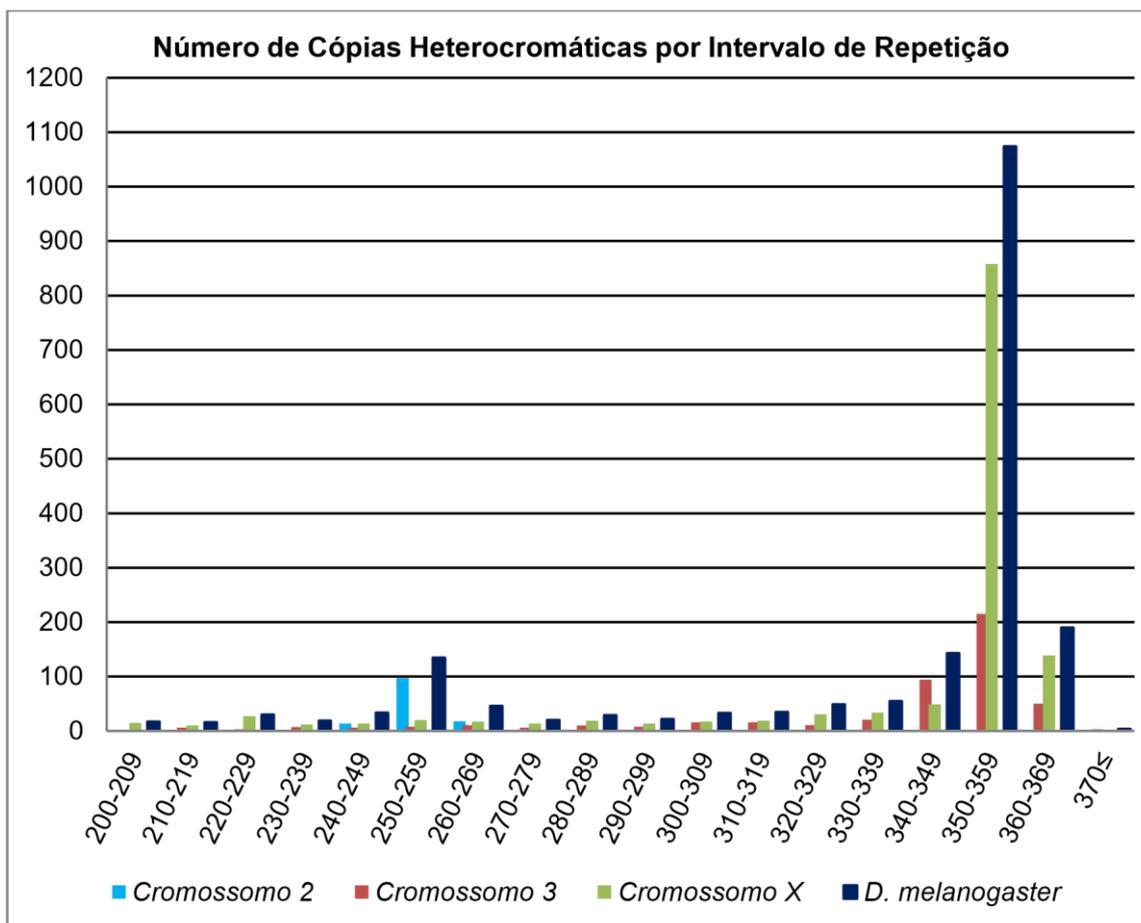
Com relação às 43 cópias do Cromossomo U identificadas como eucromáticas, 39 foram agrupadas em ramos contendo cópias do cromossomo 2, enquanto quatro cópias agruparam em ramos correspondentes ao cromossomo X. As cópias eucromáticas apresentaram tamanho entre 254 pb a 363 pb. Apesar desta variação, as cópias eucromáticas não possuem um tamanho de cópia específico para cada cromossomo.



**Figura 8.** Árvore NJ construída com distâncias genéticas calculadas pelo método K2P com todas as cópias 1.688 heterocromáticas de *D. melanogaster* encontradas no presente trabalho. Círculos verdes representam as cópias do cromossomo 2R descritas por Kuhn e cols. (2012).



**Figura 9.** Percentual das cópias heterocromáticas do DNA satélite 1.688 obtidas do Cromossomo U de *D. melanogaster* e localização cromossômica inferida.



**Figura 10.** Distribuição das cópias heterocromáticas do DNA satélite 1.688 nos cromossomos X, 2 e 3 de *D. melanogaster*.

#### 4.2. DNA satélite 1.688 em *D. simulans*

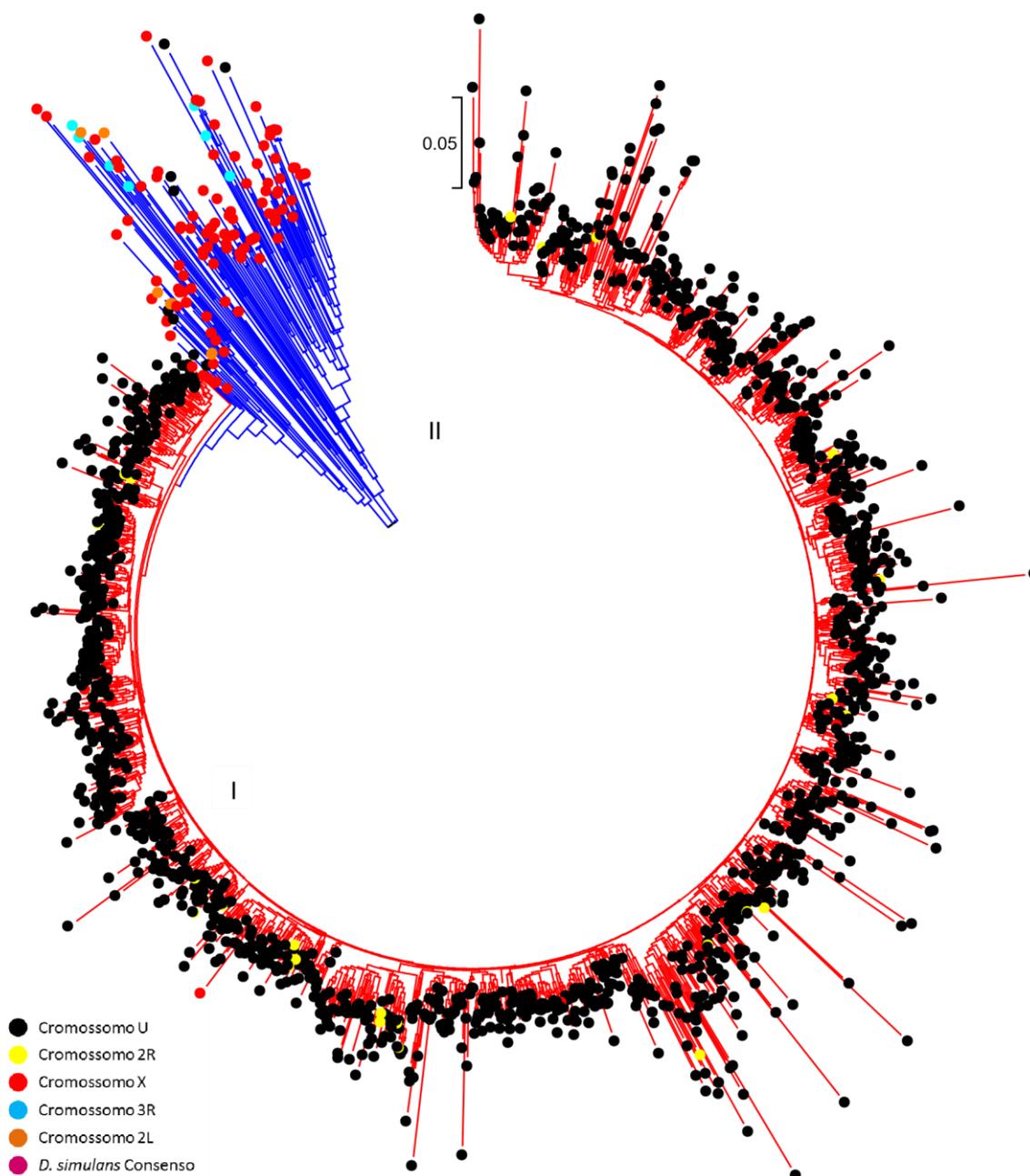
Para uma busca de cópias do DNA satélite 1.688 no genoma de *D. simulans*, foi realizado um BLAST com a sequência 1.688 consenso de *D. melanogaster* e outro BLAST com a sequência 1.688 consenso de *D. simulans*, ambas descritas por Strachan e cols (1985). No total, foram mineradas 1.781 cópias e selecionadas para análise 1.655 cópias nesta espécie.

Através do BLAST com a sequência 1.688 consenso de *D. melanogaster* foi obtido um total de 299 cópias 1.688. Estas cópias foram classificadas como eucromáticas com a seguinte localização cromossômica. 243 foram mapeadas no cromossomo X, 13 no cromossomo 2L, 23 no cromossomo 3R e 14 não possuem localização cromossômica definida. Através do BLAST com a sequência consenso de *D. simulans*, foram obtidas 1.356 cópias do DNA satélite 1.688. Diferentemente do que aconteceu com as sequências extraídas com a consenso de *D. melanogaster*, somente 35 cópias possuem localização genômica determinada, todas no cromossomo X. As demais 1.321 sequências obtidas foram encontradas em *contigs* avulsos e no “Cromossomo U” de *D. simulans*. A redundância dos dados obtidos no BLAST com as duas sequências consenso foi removida manualmente.

No GenBank, foram encontradas oito cópias do DNA satélite 1.688 localizadas nas regiões intrônicas do gene *Distorter of the X* (DOX). Também foram encontradas três sequências do 1.688 no gene *Mother of the Distorter of the X* (MDOX). Ambos os genes são eucromáticos

As 1.655 cópias do DNA satélite 1.688 obtidas de *D. simulans* foram adicionadas às 323 analisadas anteriormente por Kuhn e cols. (2012) e uma árvore NJ foi construída (Figura 11). De acordo com o padrão de ramificação da árvore NJ resultante, foi possível identificar dois grupos. O grupo I contém 1.528 cópias heterocromáticas e o grupo II contém 127 cópias eucromáticas.

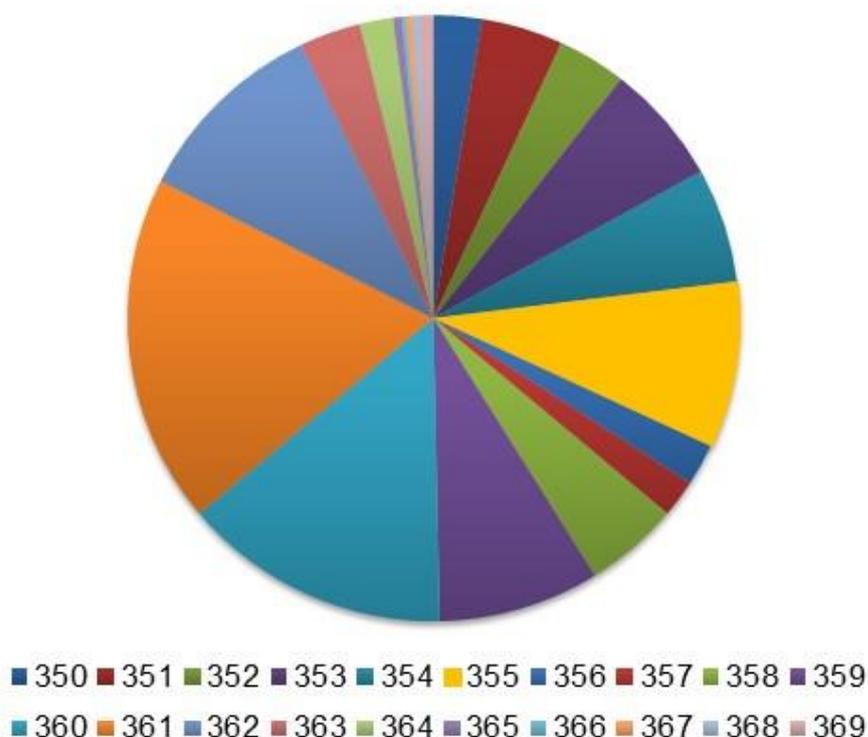
O tamanho das cópias heterocromáticas do DNA satélite 1.688 em *D. simulans* apresentou um padrão similar ao de *D. melanogaster*, com a grande maioria das sequências presentes no intervalo entre 350-359 pb e 360-369 pb, com 521 (31,48%) e 487 (29,42%) cópias para cada intervalo, respectivamente. (Figura 12).



**Figura 11.** Árvore NJ construída com distâncias genéticas calculadas pelo método K2P e contendo todas as cópias do DNA satélite 1.688 de *D. simulans*. Ramos vermelhos representam cópias heterocromáticas e ramos azuis representam cópias eucromáticas. Círculos pretos representam cópias presentes em contigs do cromossomo U; O círculo lilás representa a sequência heterocromática consenso de *D. simulans* descrita por Strachan e cols. (1985).

A cópia mais frequente do DNA satélite 1.688 presente no genoma de *D. simulans* é composta por 361 pb, com 182 repetições encontradas (11%).

Atipicamente, foram encontradas 16 cópias 1.688 completas com apenas 196 pb.



**Figura 12.** Distribuição da abundância de cópias heterocromáticas do DNA satélite 1.688 de *D. simulans* por tamanho das unidades de repetição (de 350 a 369 pb).

O tamanho das cópias eucromáticas de *D. simulans* variou de 200 pb a 365 pb (348 pb na média). A localização cromossômica das sequências flanqueadoras das cadeias 1.688 eucromáticas de *D. simulans* foi inferida através da ferramenta GenomeBrowser, disponível no banco de dados [www.flybase.org](http://www.flybase.org). Desta forma, identificamos cadeias do DNA satélite 1.688 em 37 regiões cromossômicas eucromáticas de *D. simulans* e 16 delas estão localizadas em regiões próximas a genes ortólogos descritos em *D. melanogaster* (Anexo 1).

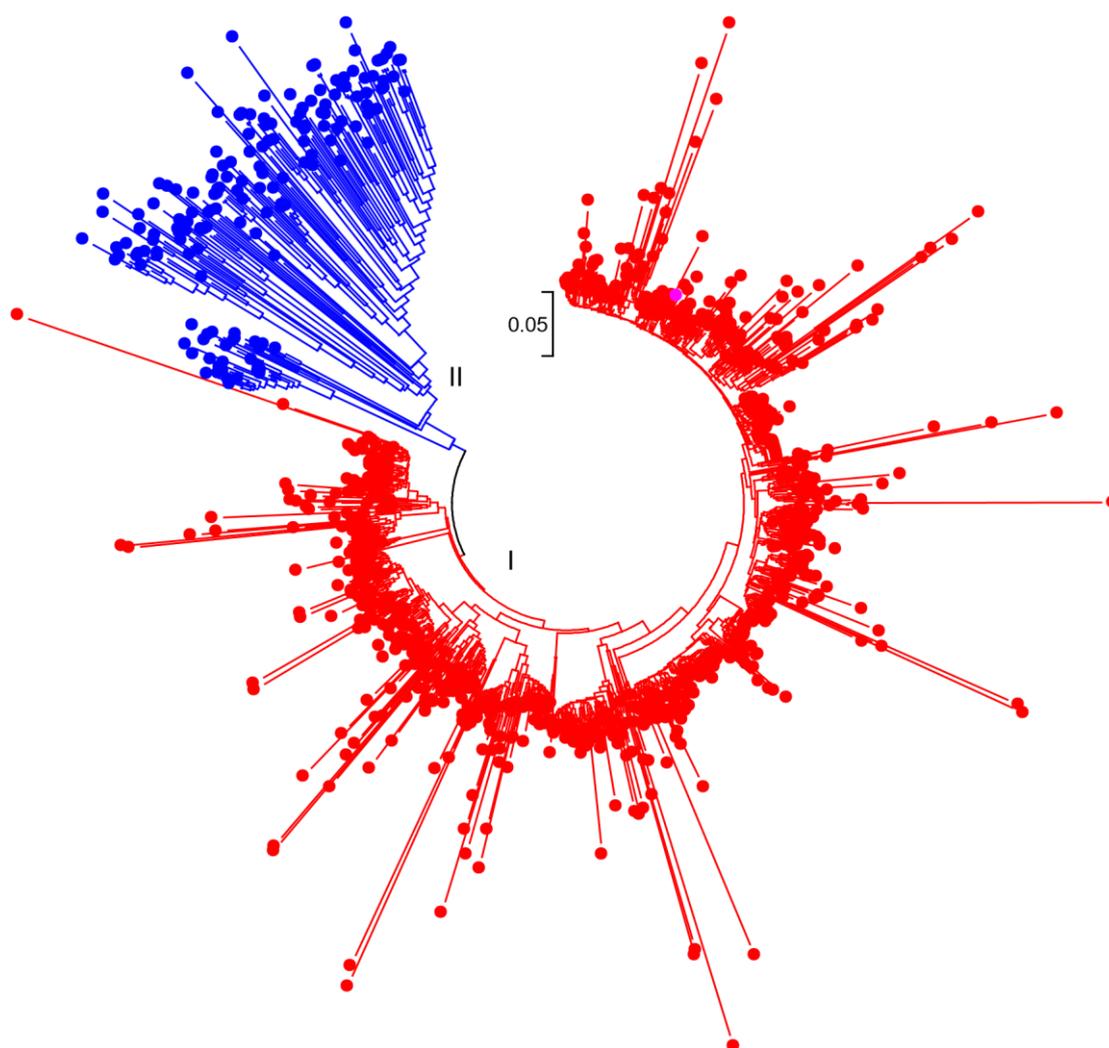
#### 4.3. DNA satélite 1.688 em *D. sechellia*

Para uma busca de cópias do DNA satélite 1.688 no genoma de *D. sechellia*, foi realizado um BLAST com a sequência 1.688 consenso de *D. melanogaster* e outro

BLAST com a sequência 1.688 consenso de *D. simulans*, ambas descritas por Strachan e cols. (1985). No total, foram mineradas 1.201 cópias e selecionadas 1.028 cópias para análise. Devido à falta de montagem cromossômica das sequências de *D. sechellia*, as cópias 1.688 só puderam ser identificadas como pertencentes a *contigs* ou *scaffolds*. Utilizando a sequência consenso de *D. melanogaster* no BLAST, foram obtidas 296 cópias 1.688, sendo 22 extraídas de sete *contigs* e 274 extraídas de sete *scaffolds*.

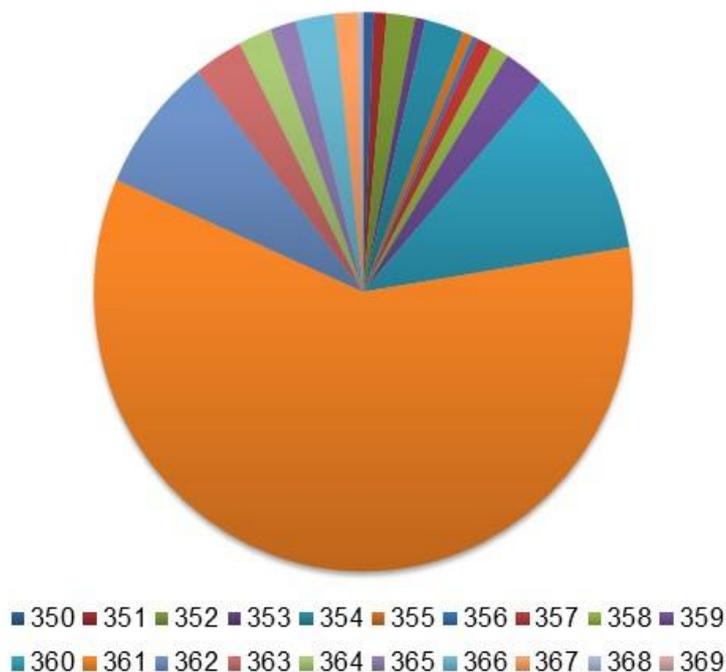
As 1.028 cópias do DNA satélite 1.688 obtidas de *D. sechellia* foram adicionadas às 323 analisadas anteriormente por Kuhn e cols. (2012) e uma árvore NJ foi construída (Figura 13). De acordo com o padrão de ramificação da árvore NJ resultante, foi possível identificar dois grupos. O grupo I contém 883 cópias heterocromáticas e o grupo II contém 145 cópias eucromáticas.

A maioria das cópias heterocromáticas do DNA satélite 1.688 de *D. sechellia* apresentou um tamanho entre 360-369 pb (562 ou 63,64%) e entre 350-359 pb (72 ou 8,16%).(Figura 14). A cópia mais frequente no genoma de *D. sechellia* é composta por 361 pb, sendo encontrada em 382 repetições. Foram encontradas 31 cópias 1.688 com tamanho menor do que 200 pb (entre 180 e 199 pb).



**Figura 13.** Árvore NJ construída com distâncias genéticas calculadas pelo método K2P e contendo todas as cópias estudadas do DNA satélite 1.688 de *D. sechellia*. Ramos e círculos vermelhos representam cópias heterocromáticas. Ramos e círculos azuis representam cópias eucromáticas. O círculo lilás representa a sequência heterocromática consenso de *D. simulans* descrita por Strachan e cols. (1985).

O tamanho das cópias eucromáticas de *D. sechellia* variou entre 233 pb e 374 pb (média de 351,5 pb). As sequências flanqueadoras das cópias eucromáticas foram analisadas com o GenomeBrowser presente no banco de dados Flybase.org. Foram identificadas sequências do DNA satélite 1.688 em 11 regiões cromossômicas eucromáticas de *D. sechellia*, sendo dez localizadas em regiões próximas a genes ortólogos de *D. melanogaster* (Anexo 2).



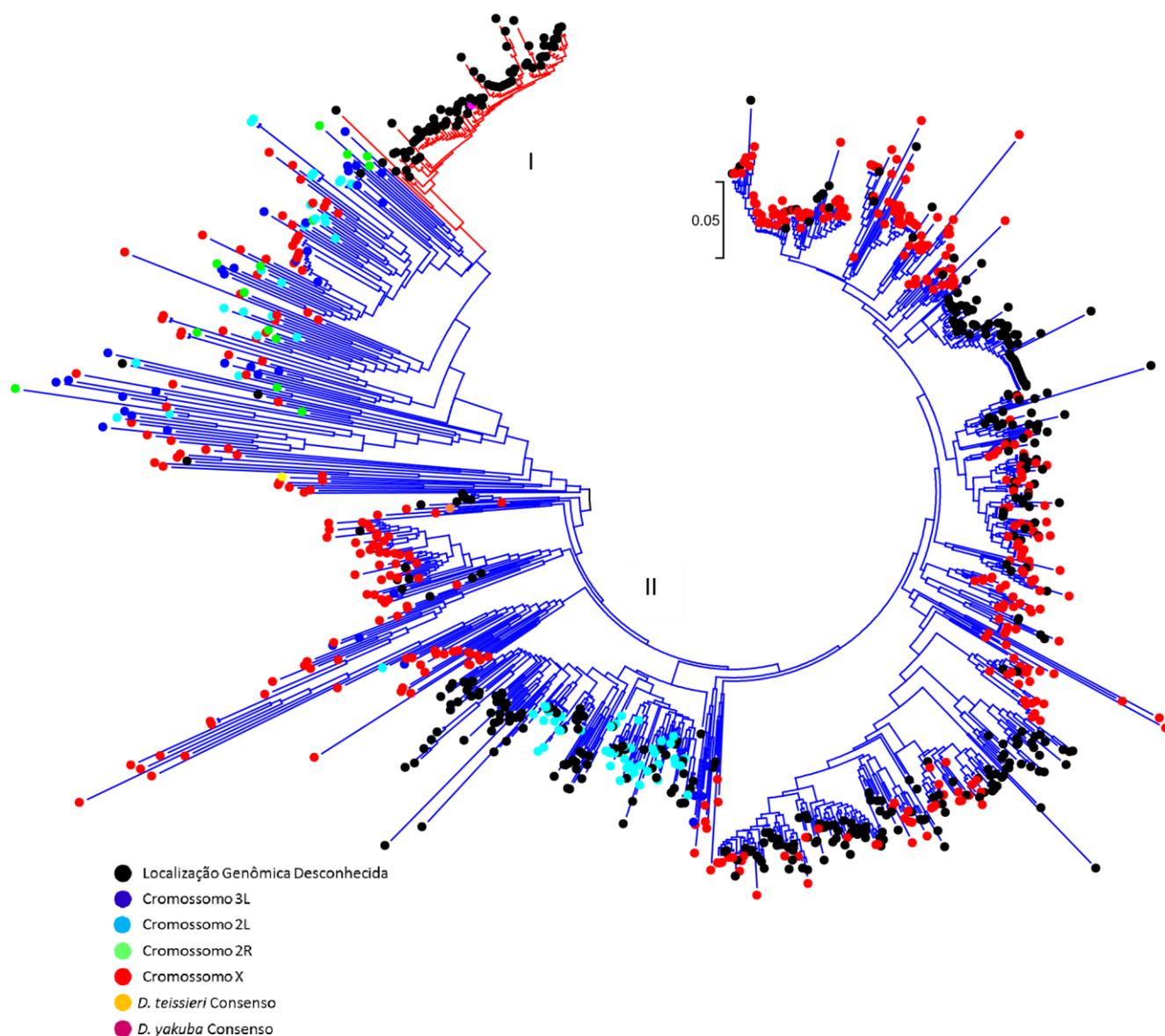
**Figura 14.** Distribuição da abundância de cópias heterocromáticas do DNA satélite 1.688 de *D. sechellia* por tamanho das unidades de repetição (de 350 a 369 pb).

#### 4.4. DNA satélite 1.688 de *D. yakuba*

Para uma busca de cópias do DNA satélite 1.688 no genoma de *D. yakuba*, foi realizado um BLAST com a sequência 1.688 consenso de *D. yakuba* e outro BLAST com a sequência 1.688 consenso de *D. melanogaster*, ambas descritas por Strachan e cols (1985). Também foi realizado um BLAST com a sequência consenso de *D. teissieri*, mas os resultados se mostraram iguais aos obtidos com as sequências consenso das duas outras espécies utilizadas. No total, foram mineradas 1.065 cópias de DNA satélite 1.688 em *D. yakuba*, sendo selecionadas para análise 1.011 cópias.

As 1.011 cópias do DNA satélite 1.688 obtidas de *D. yakuba* foram adicionadas às 323 analisadas anteriormente por Kuhn e cols. (2012) e uma árvore NJ foi construída (Figura 15). De acordo com o padrão de ramificação da árvore NJ resultante, foi possível identificar dois grupos. O grupo I contém 89 cópias heterocromáticas e o grupo II contém 992 cópias eucromáticas. Em relação ao tamanho das cópias, foram observados dois grupos principais: um com 190-199 pb

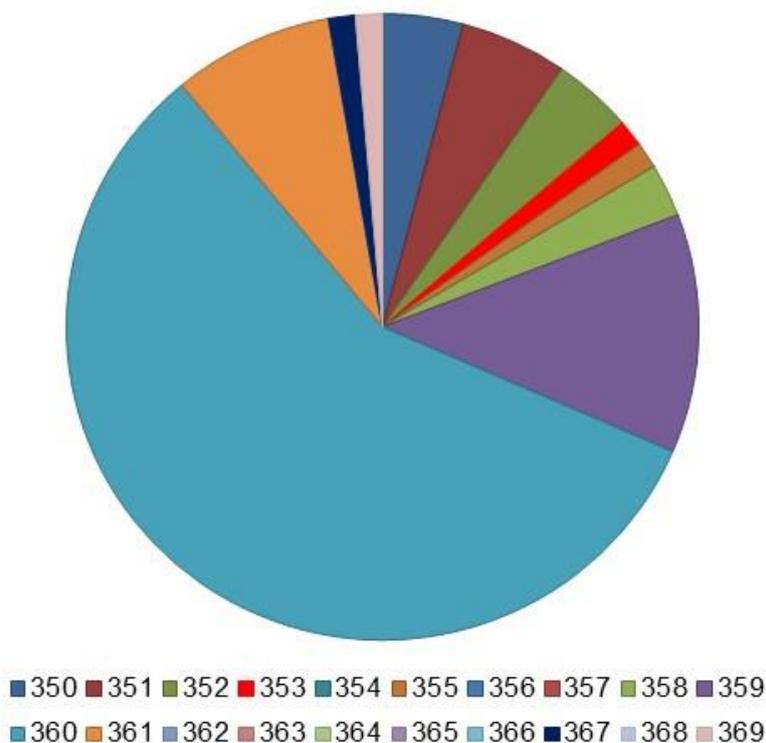
(520 cópias) e outro com 350-369 pb (119), sendo que a maioria das cópias se encontra no primeiro grupo.



**Figura 15.** Árvore NJ construída com distâncias genéticas calculadas pelo método K2P e contendo todas as cópias estudadas do DNA satélite 1.688 de *D. yakuba*. Ramos vermelhos representam cópias heterocromáticas e ramos azuis representam cópias eucromáticas. O círculo lilás representa a sequência consenso de *D. yakuba*, o círculo amarelo representa a sequência consenso de *D. melanogaster*, o círculo laranja representa a sequência consenso do *D. teissieri*, descritas por Strachan e cols. (1985).

Em relação ao tamanho da sequência, a maioria das cópias heterocromáticas do DNA satélite 1.688 de *D. yakuba* apresentou um tamanho entre 360-369 pb (48 cópias ou 53,93%) (Figura 16). A cópia mais frequente presente no genoma de *D.*

*yakuba* é composta por 360 pb, sendo encontrada em 41 repetições. Estas cópias foram extraídas de 30 *contigs* sem localização genômica definida.



**Figura 16.** Distribuição da abundância de cópias heterocromáticas do DNA satélite 1.688 de *D. yakuba* por tamanho das unidades de repetição (de 350 a 369 pb).

As sequências flanqueadoras das cópias 1.688 eucromáticas foram analisadas com o GenomeBrowser presente no banco de dados Flybase.org. Foram encontrados 156 arranjos eucromáticos contendo cópias 1.688 de *D. yakuba* nos cromossomos X (104), 2L (24), 2R (9) e 3L (18) e próximos a diversos genes.

#### 4.5. DNA satélite 1.688 de *D. erecta*

Para uma busca de cópias do DNA satélite 1.688 no genoma de *D. erecta*, foi realizado um BLAST com a sequência 1.688 consenso de *D. yakuba* e outro BLAST

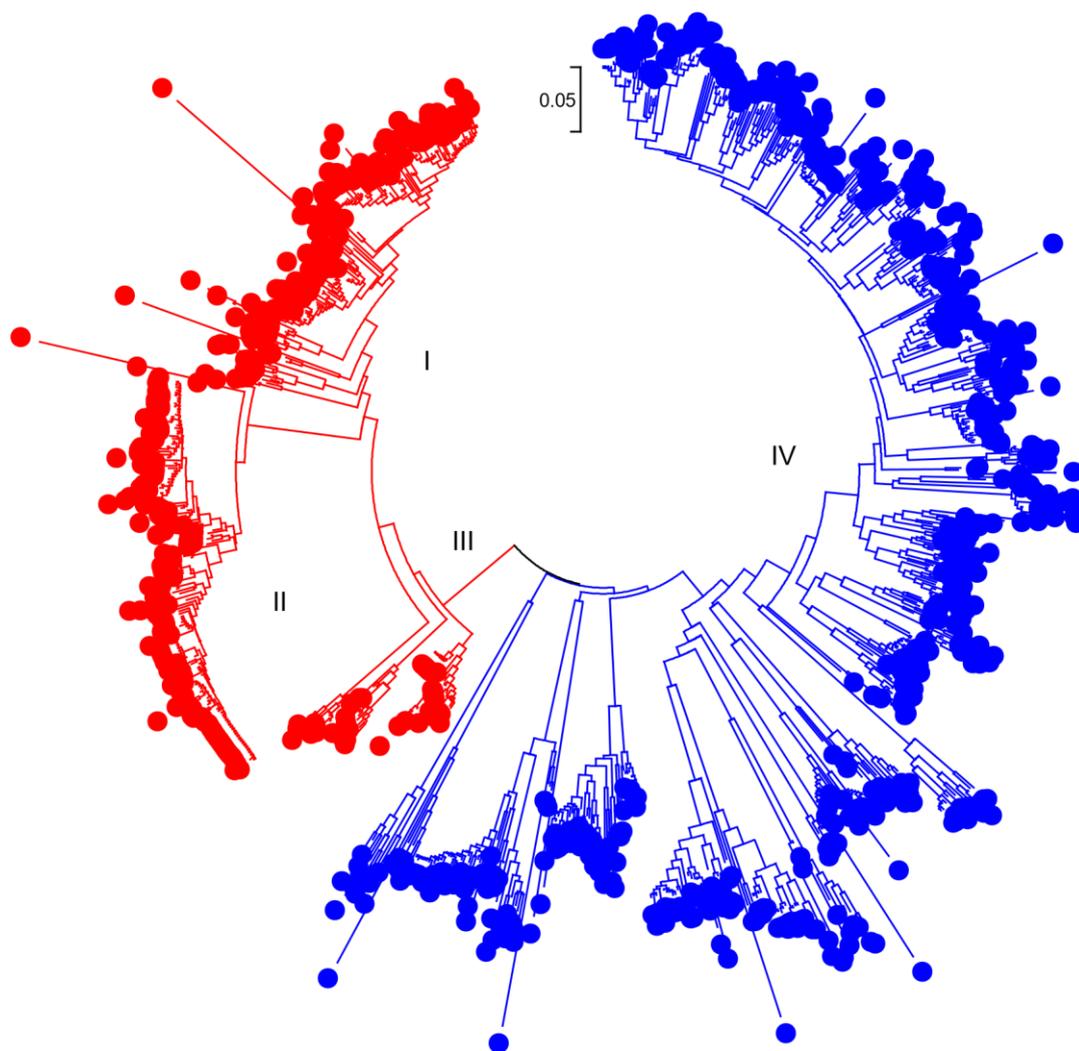
com a sequência 1.688 consenso de *D. oreana*, ambas descritas por Strachan e cols (1985). No total, foram mineradas 1.064 cópias do DNA satélite 1.688 em *D. erecta* e selecionadas para análise 949 cópias.

Com o BLAST utilizando a sequência 1.688 consenso de *D. yakuba*, foram obtidas 408 cópias 1.688, distribuídas em 73 *contigs* e dois *scaffolds* sem localização genômica definida. Com o BLAST utilizando a sequência 1.688 consenso de *D. oreana*, foram obtidas 541 cópias 1.688 em 85 *contigs* e quatro *scaffolds* sem localização genômica definida.

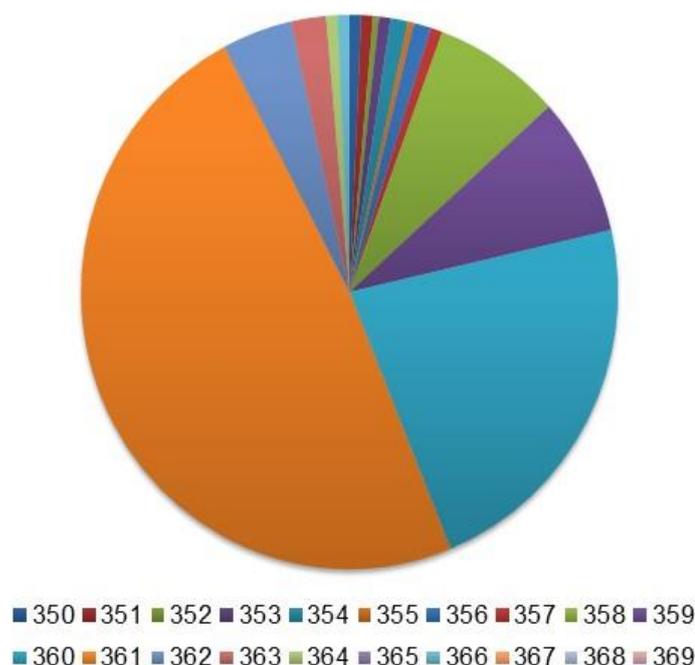
As 949 cópias do DNA satélite 1.688 obtidas de *D. erecta* foram adicionadas às 323 analisadas anteriormente por Kuhn e cols. (2012) e uma árvore NJ foi construída (Figura 17). De acordo com o padrão de ramificação da árvore NJ resultante, foi possível identificar quatro grupos. Os grupo I, II e III contêm 359 cópias heterocromáticas e o grupo II contém 590 cópias eucromáticas. Entre as 49 cópias de 199 pb, 18 agruparam no grupo III, enquanto que 31 agruparam junto às cópias do grupo IV.

A maioria das 359 cópias heterocromáticas do DNA satélite 1.688 em *D. erecta* apresentou um tamanho entre 360-369 pb e 350-359 pb com 224 cópias (62,39%) e 61 (16,99%) cópias para cada intervalo, respectivamente (Figura 18). A variante mais frequente presente no genoma de *D. erecta* é composta por 361 pb (138 cópias encontradas). Também foram encontradas 18 cópias da variante de 199 pb.

Foram identificadas sequências do DNA satélite 1.688 em 45 regiões cromossômicas eucromáticas de *D. erecta*, sendo 23 localizadas em regiões próximas a genes ortólogos de *D. melanogaster* (Anexo 4).



**Figura 17.** Árvore NJ construída com distâncias genéticas calculadas pelo método K2P e contendo todas as cópias estudadas do DNA satélite 1.688 de *D. erecta*. Ramos e círculos vermelhos representam cópias heterocromáticas. Ramos e círculos azuis representam cópias eucromáticas.



**Figura 18.** Distribuição da abundância de cópias heterocromáticas do DNA satélite 1.688 de *D. erecta* por tamanho das unidades de repetição (de 350 a 369 pb).

#### 4.6. DNA satélite 1.688 em outras espécies de *Drosophila*

Buscas por cópias 1.688 no banco genômico *Flybase* de *Drosophila* através de BLAST não revelaram cópias homólogas no genoma sequenciado de *D. ananassae* (grupo *melanogaster*) ou nos genomas sequenciados de *D. obscura* e *D. pseudoobscura*, que, juntamente com espécies do grupo *melanogaster*, também fazem parte do subgênero *Sophophora* (Figura 3).

No entanto, buscas utilizando a sequência consenso de *D. melanogaster* no banco de dados *Flybase.org* revelaram a presença de cinco cópias do DNA satélite 1.688 no genoma montado de *D. willistonii*, uma espécie distante do subgrupo *melanogaster* (Figura 3). Devido à importância desta descoberta, este resultado será discutido em detalhes no tópico 4.8.

#### 4.7. Análises interespecíficas do DNA satélite 1.688 (cópias heterocromáticas)

A Tabela 2 mostra o conteúdo dos nucleotídeos A, G,C e T das sequências do DNA satélite 1.688 entre as cinco espécies de *Drosophila* do subgrupo *melanogaster* e *D. willistoni*. É possível verificar que, embora exista variação, as cópias 1.688 são ricas em conteúdo A+T (aproximadamente 70% da unidade de repetição).

**Tabela 2.** Porcentagem de A,T, G e C nas sequências do DNA satélite 1.688 encontradas nas espécies de *Drosophila*.

Conteúdo A,G,C,T	A (%)	T (%)	G (%)	C (%)
<i>D. melanogaster</i>	35,3	34,6	13	17,1
<i>D. simulans</i>	33,7	37,2	15,3	13,8
<i>D. sechellia</i>	33,9	35,5	16,5	14,1
<i>D. yakuba</i>	32,1	38,8	14,3	14,8
<i>D. erecta</i>	33,8	36	15	15,2
<i>D. willistoni</i>	34,4	35	14	16,6

Foram selecionadas para estudo 4.775 cópias do DNA satélite 1.688 isoladas dos genomas sequenciados de *D. melanogaster*, *D.similans*, *D. sechellia*, *D. yakuba*, *D. erecta* e *D. willistoni*. Através de uma análise filogenética combinada com a análise *in silico* da localização das cadeias de cópias 1.688 a partir de sequências flangeadoras, foi possível classificar estas cópias em heterocromáticas ou eucromáticas (Tabela3). Devido ao grande número de sequências obtido, optamos por estudar somente as cópias 1.688 inferidas como pertencentes a heterocromatina e também incluímos quatro cópias 1.688 de *D. willstoni*.

A Tabela 4 mostra os valores de distância genética (método K2P) médios entre cópias 1.688 calculados para todas comparações intra- e interespecíficas. Os valores intraespecíficos variaram de 0,061 em *D. sechellia* a 0,191 em *D. erecta*. Os valores interespecíficos variaram de 0,110 entre cópias de *D. willistoni* e *D. melanogaster* até 0,292 entre as cópias de *D. sechellia* e *D. erecta*.

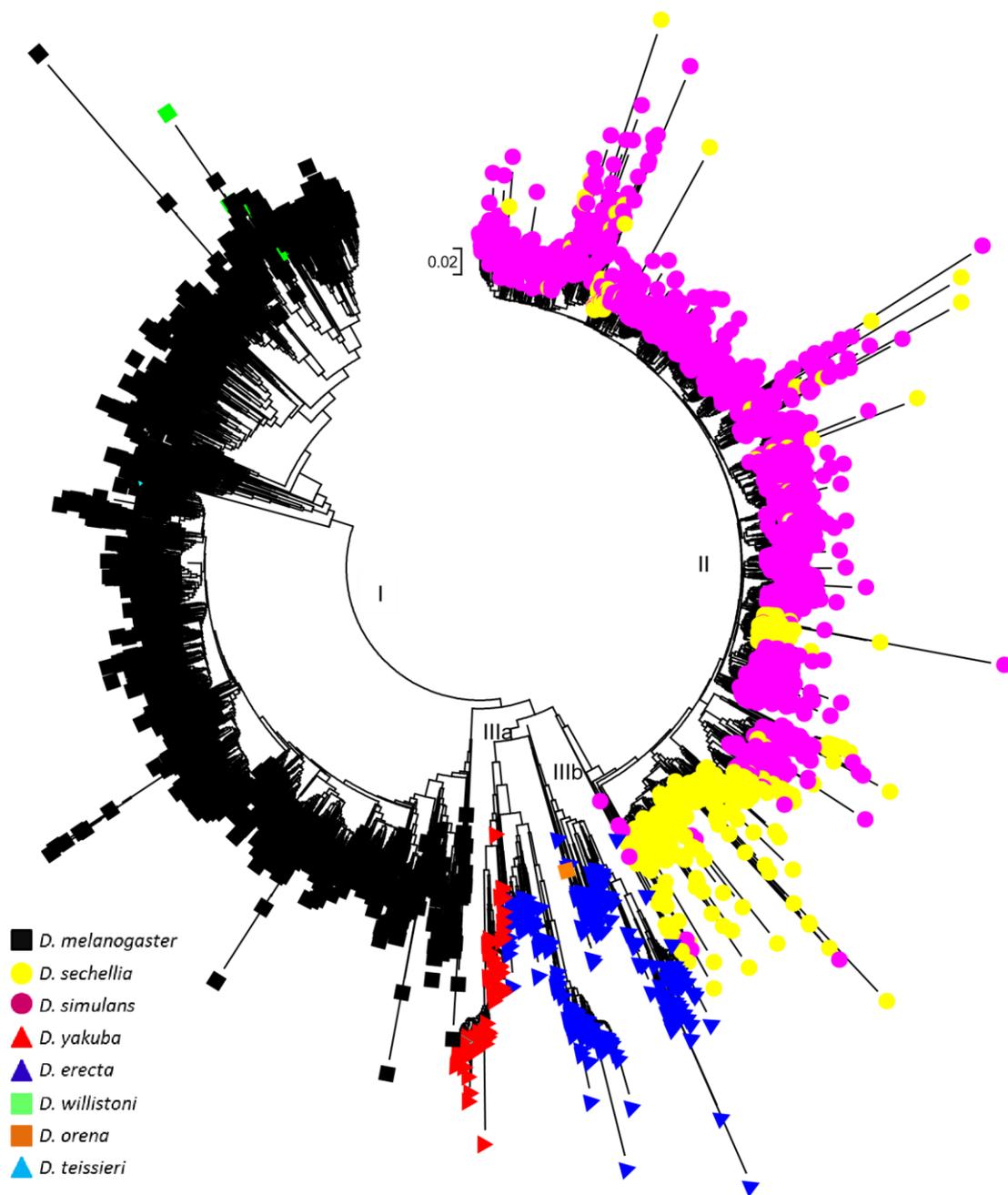
A Figura 19 mostra uma árvore NJ contendo todas as cópias 1.688

heterocromáticas obtidas, construída a partir de distâncias genéticas calculadas pelo método K2P. Com a topologia obtida, foi possível identificar três grupos principais. O grupo I contém todas as 1.795 cópias obtidas de *D. melanogaster* e as quatro cópias de *D. willistoni*. No entanto, estas cópias de *D. willistoni* não foram agrupadas em um ramo espécie-específico. O grupo II contém as 1.528 cópias de *D. simulans* e as 883 cópias de *D. sechellia*. Cópias do grupo II podem ser divididas em dois subgrupos. No subgrupo IIa foi alocada a maioria das cópias de *D. simulans*, juntamente com algumas cópias de *D. sechellia* que, em alguns casos, mostraram um agrupamento espécie-específico. No subgrupo IIb foi alocada a grande maioria das cópias de *D. sechellia*, juntamente com algumas poucas cópias de *D. simulans*. O grupo III também pode ser dividido em dois subgrupos. No subgrupo IIIa foram alocadas todas as cópias de *D. yakuba* juntamente com aproximadamente metade das cópias de *D. erecta*. Cópias destas duas espécies foram agrupadas em dois ramos espécie-específicos no subgrupo IIIa. No subgrupo IIIb, foram alocadas todas as demais cópias de *D. erecta*. Nesta árvore, também incluímos uma sequência 1.688 consenso de *D. teissieri* (Strachan e cols. 1985). Esta espécie não teve o genoma sequenciado, mas também faz parte do subgrupo *melanogaster*. Estas cópias foram agrupadas juntamente com cópias de *D. erecta* do subgrupo IIIb.

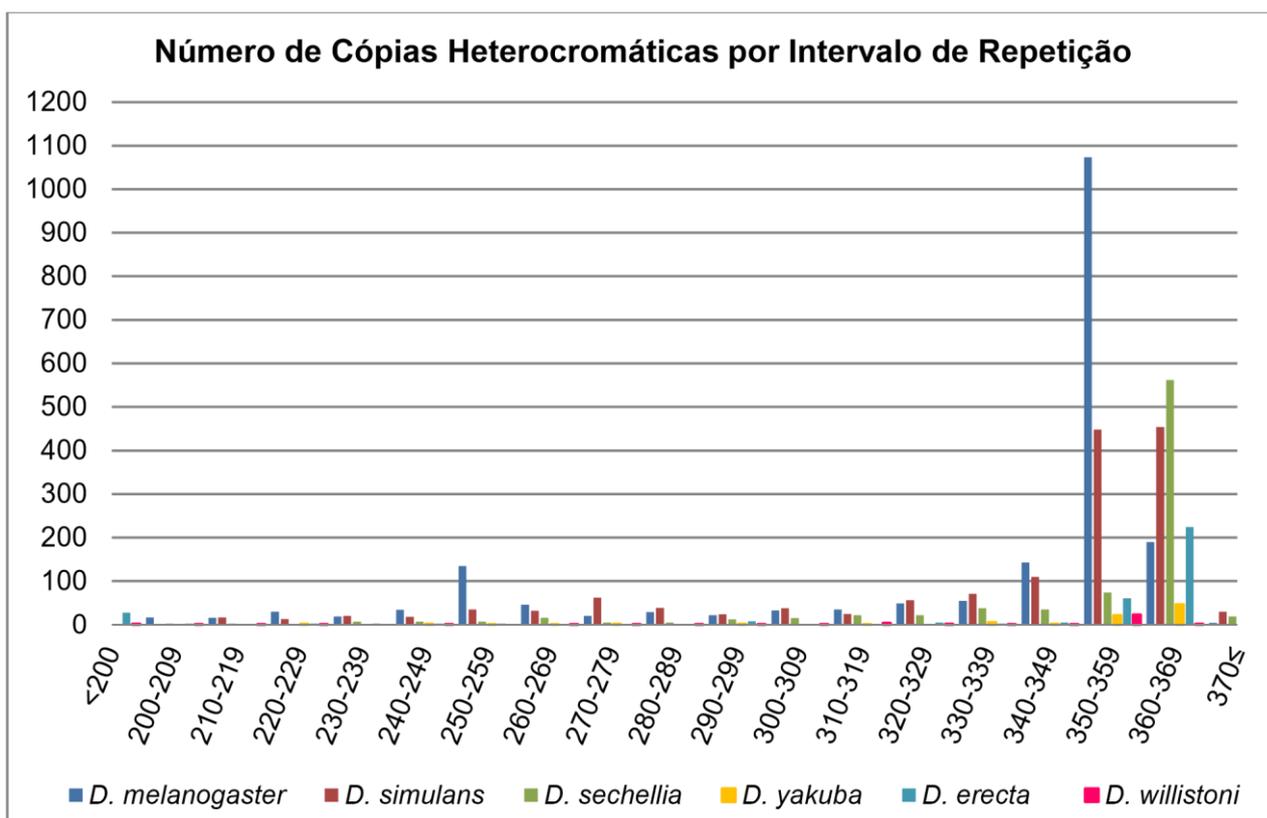
O tamanho de cópias heterocromáticas variou bastante entre as espécies analisadas, no entanto, a maior parte delas apresentou tamanho próximo ao característico da família 1.688, ou seja, de 359 a 361 pb. (Figura 20).

**Tabela 3.** Número de cópias de DNA satélite 1.688 nas espécies do gênero *Drosophila* analisadas de acordo com o domínio de cromatina.

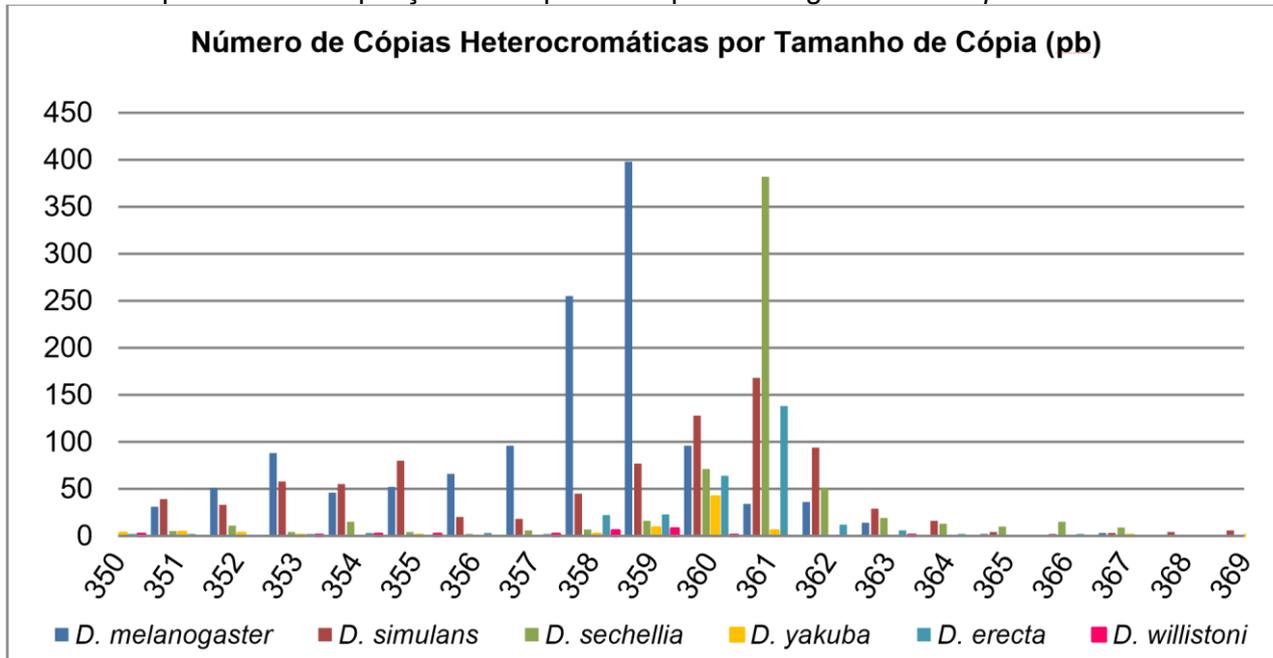
	<b>Cópias Heterocromáticas</b>	<b>Cópias Eucromáticas</b>	<b>Total</b>
<b><i>D. melanogaster</i></b>	1.816	345	2.161
<b><i>D. simulans</i></b>	1.528	127	1.655
<b><i>D. sechellia</i></b>	883	145	1.028
<b><i>D. yakuba</i></b>	89	922	1.011
<b><i>D. erecta</i></b>	359	590	949



**Figura 19.** Árvore NJ construída com distâncias genéticas calculadas pelo método K2P e contendo todas as cópias heterocromáticas do DNA satélite 1.688 estudadas no presente trabalho. Triângulo azul claro: sequência 1.688 Consenso *D. teissieri* (Strachan e cols. 1985); Quadrado laranja: sequência 1.688 Consenso *D. orena* (Strachan e cols. 1985).



**Figura 20.** Distribuição das cópias do DNA satélite 1.688 de acordo com o tamanho (em pb) de unidade por intervalo repetição de 10 pb nas espécies do gênero *Drosophila*.



**Figura 21.** Distribuição das cópias do DNA satélite 1.688 em pb de acordo com o tamanho de unidade de repetição nas espécies do gênero *Drosophila*.

**Tabela 4.** Divergência nucleotídica intra- e interespecífica das seis espécies analisadas utilizando o método Kimura-2-Parâmetros com  $\sigma=1,4\%$ .

	<i>D. melanogaster</i>	<i>D. sechellia</i>	<i>D. simulans</i>	<i>D. yakuba</i>	<i>D. erecta</i>	<i>D. willistoni</i>
<i>D. melanogaster</i>	0,119	-	-	-	-	-
<i>D. sechellia</i>	0,271	0,061	-	-	-	-
<i>D. simulans</i>	0,256	0,090	0,099	-	-	-
<i>D. yakuba</i>	0,289	0,282	0,260	0,127	-	-
<i>D. erecta</i>	0,280	0,292	0,278	0,234	0,191	-
<i>D. willistoni</i>	0,110	0,273	0,260	0,288	0,278	0,095

#### 4.8. O DNA satélite 1.688 de *D. willistoni*

Buscas por sequências 1.688 nos genomas de espécies não pertencentes ao subgrupo *melanogaster* através de BLAST com a sequência consenso de *D. melanogaster* encontraram um *hit* significativo e com baixo *e-value* ( $1.07728^{-40}$ ) no *scaffold* *scf2\_1100000006624* do genoma completamente sequenciado de *D. willistoni*. Este *scaffold* possui 1.053 pb e uma análise de sua sequência de nucleotídeos revelou que ele é composto, do início ao fim, por cinco cópias em tandem do DNA satélite 1.688, com tamanhos de 186 pb, 296 pb, 187pb, 355 pb e 29 pb.

Uma árvore NJ contendo todas as cópias 1.688 heterocromáticas obtidas neste trabalho mostrou que as quatro cópias do DNA satélite 1.688 de *D. willistoni* com tamanho maior que 180 pb se agruparam juntamente com cópias 1.688 de *D. melanogaster* (Figura 19).

Com o objetivo de visualizar melhor as relações filogenéticas entre as cópias 1.688 de *D. willistoni* e de *D. melanogaster*, uma nova árvore NJ foi construída contendo apenas cópias 1.688 destas duas espécies (Figura 22). Nesta árvore, foi possível observar que as cópias 1.688 de *D. willistoni* se agruparam com cópias 1.688 características do cromossomo 3 de *D. melanogaster*.

Como não existem sequências flanqueadoras da cadeia de cópias 1.688 no genoma montado de *D. willistoni*, fizemos uma busca pelas mesmas através de

BLAST nos *reads* originais de *D. willistoni* do banco de dados “Trace Archive File” ([www.ncbi.nlm.nih.gov/Trace](http://www.ncbi.nlm.nih.gov/Trace)). Para este BLAST, utilizamos uma das cópias 1.688 encontradas em *D. willistoni*. Como resultado, encontramos um *read* de 1.030 pb (gnl|ti|835059756) composto por uma sequência de 340 pb não homóloga ao DNA satélite 1.688 em sua extremidade 5’, seguida de uma cópia 1.688 de 162 pb. Este arquivo apresentou baixa qualidade de sequenciamento e não foi utilizada nas análises filogenéticas. Uma busca no GeneBank através de BLAST com esta sequência de 340 pb revelou que a mesma consiste de uma cópia parcial de um retrotransposon não-LTR chamado *Doc*. Não encontramos outras cópias do *Doc* no genoma montado de *D. willistoni* ou nas *reads* originais do sequenciamento desta espécie.

Os resultados do BLAST no GenBank com a cópia *Doc* de *D. willistoni* revelaram alta similaridade com cópias do *Doc* presentes em *D. melanogaster* (Anexo 5). De fato, os *hits* de menor *e-value* representam cópias do *Doc* presentes em *D. melanogaster*. Em especial, o terceiro *hit* consiste de um cosmídeo (GB: S50810.1) previamente sequenciado de *D. melanogaster* descrito por Slobodkin & Alatorsev (1992). Este cosmídeo tem 1.085 pb e contém um fragmento de 1.688 de 380 pb flanqueado por duas cópias do elemento transponível *Doc* em orientações opostas (Figura 23B).

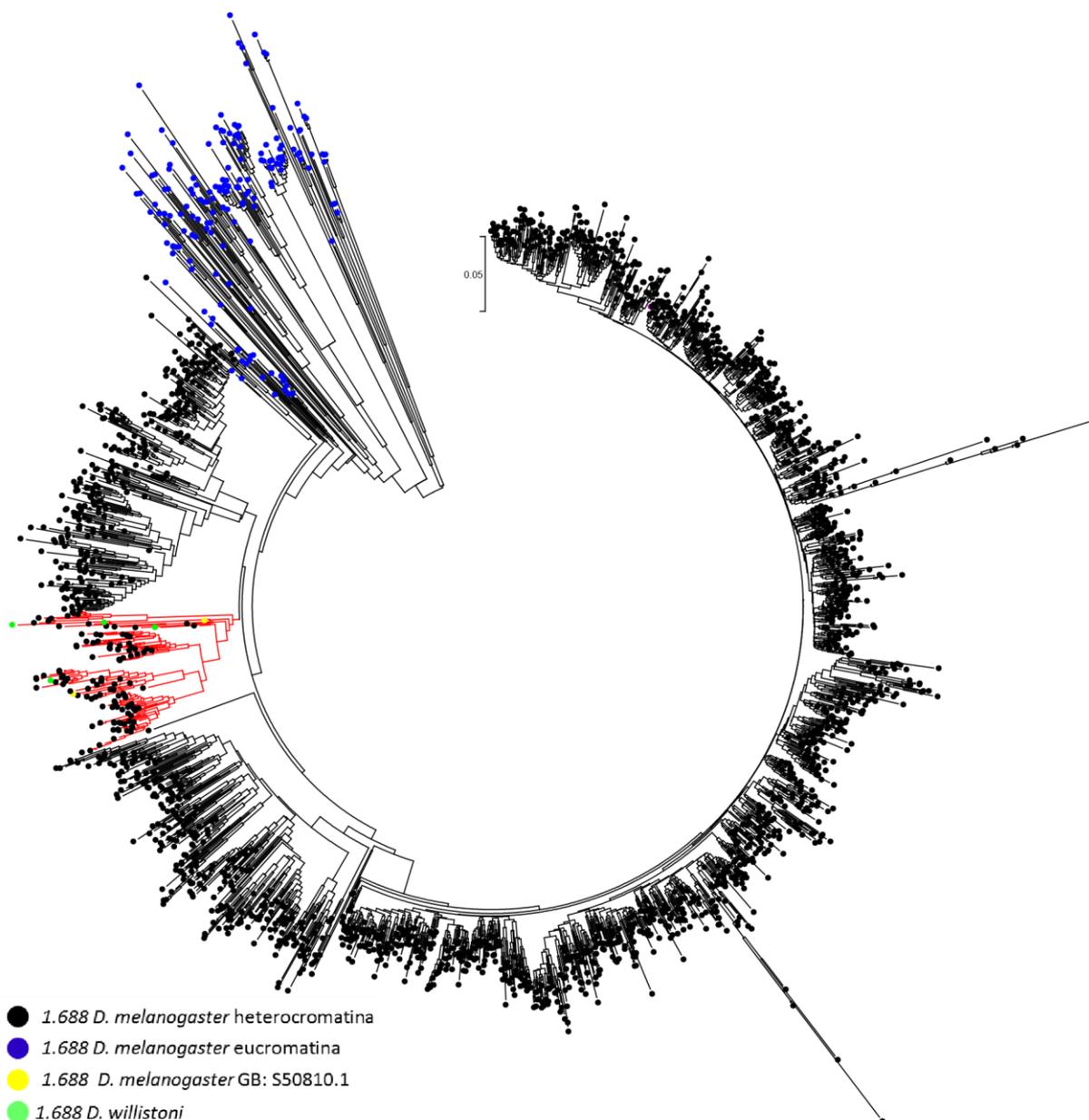
Para confirmar experimentalmente a presença do DNA satélite 1.688 em *D. willistoni*, um par de oligos Dmel359Fwd e Dmel359Rev foi construído a partir das sequências heterocromáticas 1.688 características de *D. melanogaster*. A posição destes oligos pode ser vista na Figura 23B. Devido à localização destes oligos na cópia 1.688, produtos de PCR só podem ser obtidos se as cópias 1.688 estão organizadas em tandem. Estes oligos foram utilizados em uma reação de PCR contendo o DNA total de *D. willistoni* e o DNA de *D. melanogaster* foi usado como controle. Como esperado, amplicons de aproximadamente 360 pb e 720 pb, correspondentes a monômeros e dímeros do DNAsat 1.688, foram observados na PCR de *D. melanogaster*. (Figura 24). Em *D. willistoni*, foi observado um único amplicon de 360 pb, confirmando a presença no genoma desta espécie de cópias

1.688, organizadas em tandem e muito similares às encontradas no genoma de *D. melanogaster*.

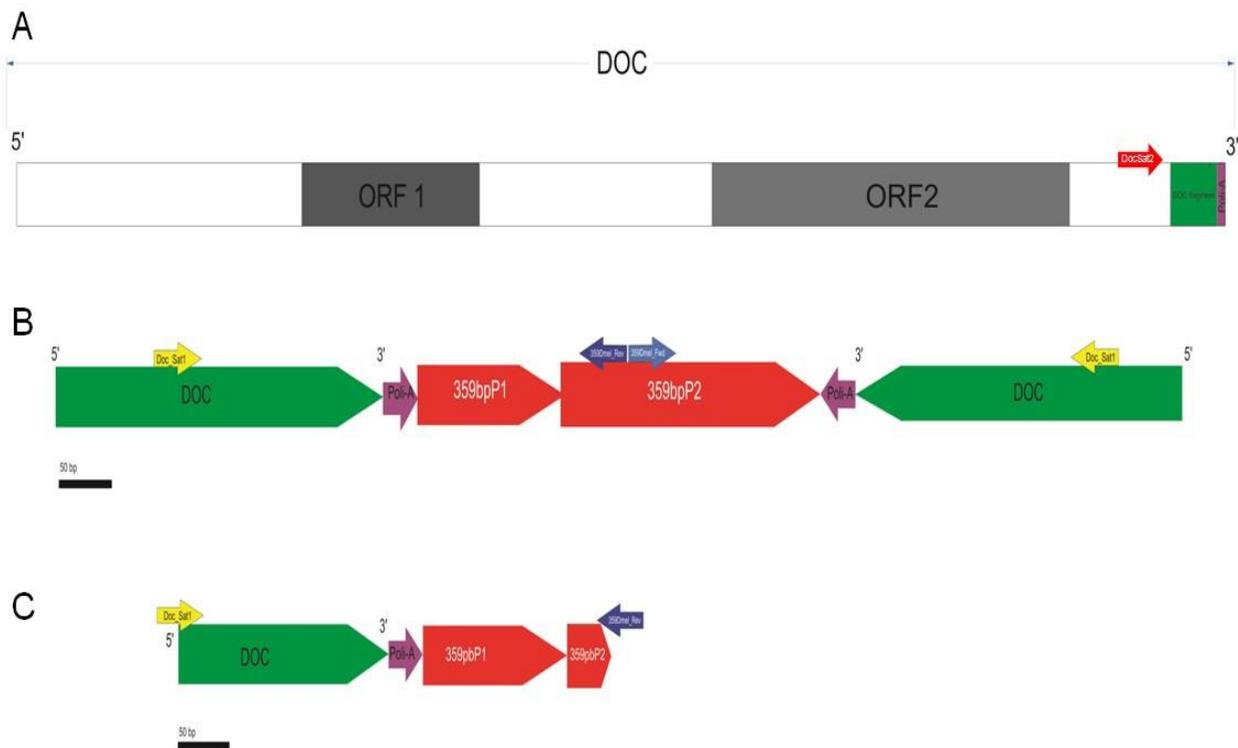
Para confirmar experimentalmente a associação do DNA satélite 1.688 com o *Doc* em *D. willistoni*, realizamos uma PCR no DNA total desta espécie contendo um oligo (Dmel359Rev) construído a partir de sequências do DNAsat 1.688 de *D. melanogaster* e um oligo (*DocSat1*) construído a partir da extremidade 3' de um elemento *Doc* canônico presente no banco de dados *repbase*.

(<http://www.girinst.org/repbase/>). A posição destes oligos em um arranjo *Doc*+DNAsat 1.688 pode ser vista na Figura 23B. Como controle, utilizamos o DNA de *D. melanogaster*. Em *D. melanogaster*, a PCR resultou em três amplicons com tamanhos aproximados de 400, 700 e 1.000pb (Figura 25). Em *D. willistoni*, a mesma PCR resultou em dois amplicons com tamanhos aproximados de 360, e 400 pb (Figura 25). Portanto, estes resultados confirmaram a associação do DNA satélite 1.688 com o retroelemento *Doc* de *D. willistoni*.

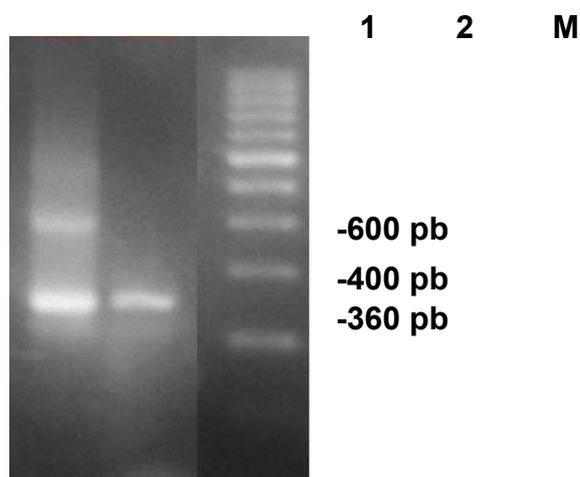
A obtenção de um amplicon de tamanho similar (~400 pb) em *D. melanogaster* e *D. willistoni* levantou a hipótese de que o mesmo poderia consistir de um mesmo arranjo *Doc*+1.688 presente no genomas das duas espécies. Para testar esta hipótese, o DNA da banda de 400 pb obtido com a PCR de *D. melanogaster* e *D. willistoni* foi eluído do gel e clonado. Foram selecionados para o sequenciamento dois clones de *D. melanogaster* (Dmel*Doc*1.688/1-2) e um clone de *D. williston* (Dwill*Doc*1.688/1). A análise das sequências confirmou que todos os clones consistem de arranjos *Doc*+1.688. No entanto, o alinhamento de sequências apresentado na Figura 26 mostra que o arranjo *Doc*+1.688 presente no clone Dwill*Doc*1.688/1 de *D. willistoni* não é idêntico ao arranjo *Doc*+1.688 dos clones Dmel*Doc*1.688/1-2 e do arranjo *Doc*+1.688 presente no GenBank (GB: S50810.1).



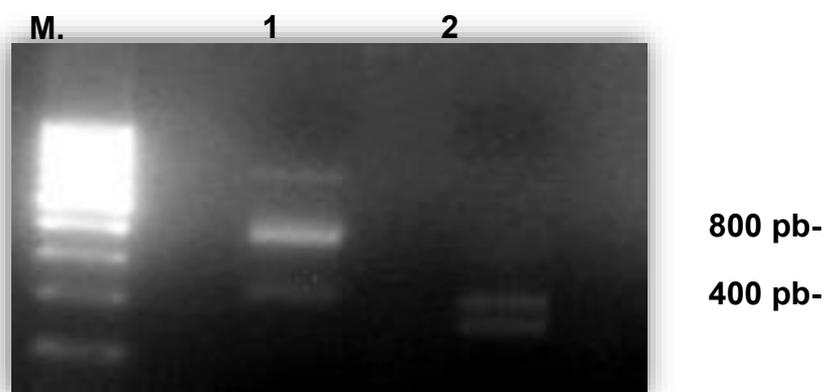
**Figura 22.** Árvore filogenética de NJ utilizando o método K2P contendo as cópias do DNA satélite 1.688 presentes no genoma montado de *D. melanogaster* e *D. willistoni*. Ramos em vermelho identificam um agrupamento de cópias do cromossomo 3 de *D. melanogaster*.



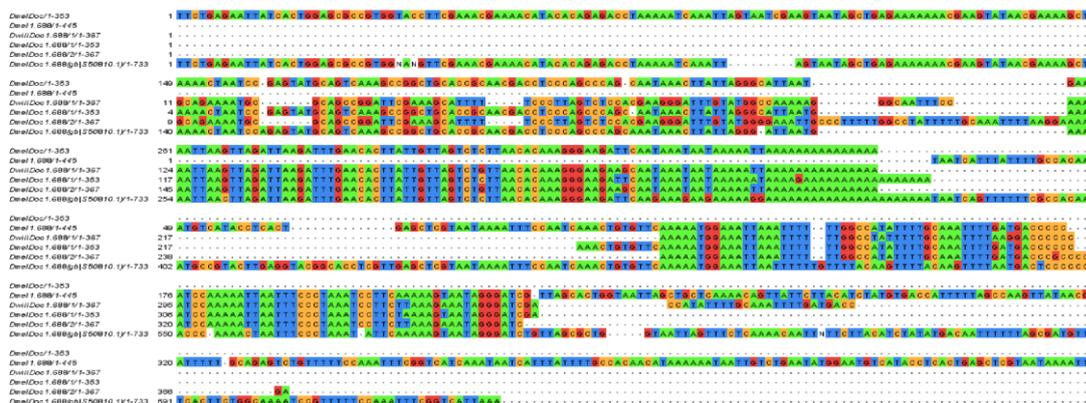
**Figura 23. A.** Representação esquemática do retrotransposon não-LTR *Doc* de acordo com banco de dados Repbase **B.** Representação esquemática do cosmídeo contendo sequências do elemento *Doc* associadas a sequências do DNA satélite 1.688 (GenBank: S50810.1). **C.** Representação esquemática de fragmentos de DNA amplificados por PCR utilizando os oligos *Docsat1* e *Dmel360pbRev* em *D. melanogaster* e *D. willistoni*. As setas amarelas representam o oligonucleotídeo *DocSat1* e as setas Azuis representam o conjunto de o lignonucleotídeos *Dmel359pb*.



**Figura 24.** Gel de agarose mostrando o padrão de bandas amplificadas do DNA satélite 1.688 utilizando o conjunto de oligos Dmel359Fwd/Rev nas espécies *D. melanogaster* (1) e *D. willistoni* (2). **M** = Marcador de peso molecular 200 pb (Promega).



**Figura 25.** Gel de agarose mostrando o padrão de bandas amplificadas utilizando o conjunto de oligos Docsat1/Dmel359R nas espécies *D. melanogaster* (1) e *D. willistoni* (2). **M.** Marcador de peso molecular 200 pb (Promega).



**Figura 26.** Alinhamento nucleotídico contendo em ordem: fragmento da sequência consenso do elemento transponível *Doc* (Repbase); sequência consenso do DNA satélite 1.688 de *D. melanogaster*; Arranjos *Doc-1* amplificados por PCR em *D. willistoni* (DwilDoc1.688-1); em *D. melanogaster*; (DmelDoc1.688-1/2). sequência cosmídica contendo arranjo *Doc-1.688* presente no cosmídeo descrito por Slobodkin & Alator (GenBank: S50810.1).

## 5. DISCUSSÃO

Trabalhos anteriores relataram a presença do DNA satélite 1.688 em seis espécies do subgrupo *melanogaster*: *D. melanogaster*, *D. simulans*, *D. mauritiana*, *D. orena*, *D. yakuba* e *D. teissieri* (Carlson & Brutlag 1977; Strachan e cols. 1985; Lohe & Brutlag 1986). No presente trabalho, relatamos pela primeira vez cópias do DNA satélite 1.688 em outras duas espécies do subgrupo *melanogaster*: *D. sechellia* e *D. erecta*. Não encontramos cópias 1.688 no genoma de *D. ananassae*, espécie que faz parte do grupo *melanogaster*. Portanto, os dados indicam que o DNA satélite 1.688 surgiu na espécie ancestral do subgrupo *melanogaster*.

Strachan e col. (1985) realizaram o primeiro estudo evolutivo sobre o DNA satélite 1.688 no subgrupo *melanogaster*. As espécies estudadas foram *D. melanogaster*, *D. simulans*, *D. mauritiana*, *D. orena*, *D. yakubae* e *D. teissieri*. (Figura 2). O número de cópias analisadas por espécie variou de quatro a nove, todas elas obtidas através de digestão do DNA genômico com as enzimas de restrição *EcoRI* ou *HaeIII*. Este trabalho resultou em vários dados importantes sobre a evolução de DNAs satélites como: (i) a constatação de que cópias 1.688 evoluíram por evolução combinada, ou seja, a variação intraespecífica entre cópias 1.688 é menor que a variação interespecífica; (ii) a verificação de que a homogeneização de mutações espécie-específicas ocorre através de um processo gradual e (iii) a constatação de que a filogenia das cópias 1.688 reflete a filogenia das espécies, sugerindo que a homogeneização das cópias 1.688 ocorreu de maneira tempo-dependente. No entanto, considerando a abundância deste DNA satélite nos genomas destas espécies, apenas poucas cópias foram analisadas.

O presente trabalho teve como objetivo estudar novamente o DNA satélite 1.688 nas espécies de *Drosophila* do subgrupo *melanogaster*, mas através da análise de genomas sequenciados de *Drosophila*.

Com a mineração de cópias do DNA satélite 1.688 nos bancos de dados genômicos de *Drosophila*, estudamos um total de 6.849 cópias do DNA satélite 1.688 de *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakubae* e *D. erecta*. Dentre estas espécies, o DNA satélite 1.688 de *D. sechellia* e *D. erecta* ainda não havia sido estudado.

As análises filogenéticas das sequências das cópias 1.688 em todas as espécies estudadas do subgrupo *melanogaster* mostraram árvores contendo ramos separados para cópias eucromáticas e heterocromáticas (Figuras 6, 11, 13, 15 e 17). A localização eucromática das cópias 1.688 também foi posteriormente confirmada através da análise de sequências flanqueadoras destas cópias, que foram comparadas com o genoma montado de *D. melanogaster* como referência.

O presente trabalho relatou pela primeira vez a presença de cópias 1.688 eucromáticas nos genomas de *D. simulans*, *D. sechellia*, *D. yakuba* e *D. erecta*. Este resultado indica que cópias eucromáticas do DNA satélite 1.688 estavam presentes no ancestral das espécies do subgrupo *melanogaster*. Foram encontradas muito mais cópias heterocromáticas do que eucromáticas nos genomas de *D. melanogaster*, *D. simulans* e *D. sechellia* (Tabela 3). Este padrão está de acordo com modelos que preveem um maior acúmulo de sequências repetitivas em regiões cromossômicas com baixa recombinação, como a heterocromatina (Charlesworth e cols. 1994). De fato, há vários exemplos que corroboram estes modelos. Por exemplo, estudos realizados em *Drosophila* mostraram que o número de inserções de elementos transponíveis em regiões heterocromáticas é muito superior ao número de inserções em regiões eucromáticas (Fiston-Lavier e cols. 2001; Rizzon e cols. 2002). Segundo estes trabalhos, a baixa densidade de cópias presentes nas regiões eucromáticas estaria relacionada a uma seleção contra os efeitos mutagênicos dos TEs em uma região com alta densidade de genes.

Por outro lado, as espécies *D. yakuba* e *D. erecta* apresentaram uma quantidade de cópias heterocromáticas inferior à de cópias eucromáticas (Tabela 3). No momento, ainda não é possível afirmar se este resultado é real ou se reflete lacunas de montagem dos genomas sequenciados destas duas espécies. No entanto, é interessante mencionar que os padrões de abundância de cópias 1.688 eucromáticas e heterocromáticas no genoma destas cinco espécies estão de acordo com a filogenia das espécies.

### 5.1. O DNA satélite 1.688 de *D. melanogaster*

Recentemente, Kuhn e cols. (2012) reportaram a análise de 323 cópias do DNA satélite 1.688 distribuídas nas regiões eucromáticas e heterocromáticas dos cromossomos X, 2 e 3 do genoma de *D. melanogaster*. No presente trabalho, ampliamos consideravelmente o número de cópias analisadas para 2.161, correspondendo a 14% das 15.000 cópias estimadas para este espécie.

Todas as novas sequências 1.688 de *D. melanogaster* aqui estudadas foram mineradas do Cromossomo U, um arquivo aleatoriamente concatenado de sequências não montadas de *D. melanogaster*. A maior parte das sequências que compõem o Cromossomo U é oriunda de arranjos heterocromáticos que não puderam ser organizados em nenhum outro cromossomo (He e cols. 2012). De fato, inferimos que 1.793 cópias 1.688 devem ser heterocromáticas, enquanto apenas 43 cópias são eucromáticas.

De acordo com os dados obtidos, pudemos inferir que aproximadamente 72% das cópias mineradas do Cromossomo U localizam-se no cromossomo X (Figura 9). As cópias do cromossomo 3 representaram 25% e as cópias do cromossomo, 2, 3%. Estes dados estão de acordo com os previamente descritos na literatura, que mostraram através de hibridação *in situ* em cromossomos, que cópias do DNA satélite 1.688 localizam-se principalmente na heterocromatina do cromossomo X e, em menor abundância, em regiões heterocromáticas dos cromossomos 3 e 2 (Losada & Villasante 1996; Abad e cols. 2000), sendo os blocos com cópias 1.688 do cromossomo 3 mais extensos que os do cromossomo 2.

Nossos resultados indicaram a existência de variabilidade com relação ao tamanho das cópias heterocromáticas do DNA satélite 1.688 em *D. melanogaster*. De acordo com a distribuição das cópias em relação ao tamanho da repetição e cromossomo de origem, foi possível observar que as cópias mais frequentes dos cromossomos X, 3 e 2 são compostas por 359, 353 e 257 pb, respectivamente. Estes dados corroboram estudos anteriores que sugeriram que a família 1.688 poderia ser subdividida em três subfamílias cromossomo-específicas. Cópias de 359 pb seriam características do cromossomo X, cópias de 353-356 pb seriam características do

cromossomo 3 e cópias de 254 pb seriam típicas do cromossomo 2 (Hsieh & Brutlag 1979; Carlson & Brutlag 1979; Abad e cols. 2000; Kuhn e cols. 2012). Portanto, nossos dados confirmam a existência destas subfamílias em *D. melanogaster* e alteram o tamanho principal das cópias do cromossomo 2, de 254 para 257 pb.

As 1.315 cópias heterocromáticas do cromossomo X se agruparam em um único grande ramo na árvore filogenética da Figura 8. Esta baixa variabilidade pode ser resultado da organização destas cópias em um grande bloco localizado na região de heterocromatina pericentromérica do braço esquerdo do cromossomo X. Assume-se frequentemente que mecanismos de homogeneização atuam de maneira mais eficiente entre cópias com alta similaridade e presentes em uma mesma cadeia. Em outras palavras, a presença de um grande número de cópias 1.688 em um único e extenso arranjo deve ter favorecido a ação de eventos de recombinação desigual, que conseqüentemente, resultaram no processo de homogeneização.

As cópias do cromossomo 3 se agruparam em três diferentes ramos na árvore filogenética da Figura 8. Este padrão pode ser explicado pela organização destas cópias em três diferentes arranjos da heterocromatina do cromossomo 3, situação que pode ter levado as cadeias à homogeneização local. As cópias de *D. melanogaster* identificadas na heterocromatina do braço direito do cromossomo 2 agruparam com as cópias heterocromáticas do cromossomo 3. Este resultado sugere que estas duas únicas cópias no braço direito do cromossomo 2 podem ter-se originado num evento recente de dispersão de cópias do cromossomo 3 para o cromossomo 2 (Figura 8).

Entre as cópias 1.688 mineradas do Cromossomo U, identificamos 43 cópias eucromáticas. Este resultado pode estar relacionado à possível grande extensão destas cadeias, o que pode ter dificultado o processo de sua montagem no genoma. A ausência de regiões flangeadoras nestas cadeias 1.688 dá suporte a esta hipótese. Destas 43 cópias, apenas quatro foram identificadas no cromossomo X, enquanto que 39 foram localizadas no cromossomo 2. Experimentos de hibridação *in situ* fluorescente (FISH) em cromossomos metafásicos serão realizados para estabelecer a exata localização citológica destas cadeias.

## 5.2. O DNA satélite 1.688 de *D. simulans*

A presença do DNA satélite 1.688 em *D. simulans* já tinha sido relatada por Strachan e cols. (1985) e Lohe & Brutlag (1987). No entanto, apenas nove cópias heterocromáticas haviam sido previamente caracterizadas (Strachan e cols. 1985).

Foi estimado que 17% do genoma de *D. simulans* é composto por DNAs satélites (Bosco e cols. 2007). Trabalhos anteriores sugeriram que o DNA satélite 1.688 nesta espécie representa 0,11% do genoma (Lohe & Brutlag 1987). No presente trabalho foram caracterizadas 1.655 cópias do DNA satélite 1.688 de *D. simulans*, o que significa que a contribuição do DNA satélite 1.688 para o tamanho do genoma desta espécie é de no mínimo 0,33%.

Das 1.655 cópias 1.688 caracterizadas, 1.528 foram identificadas como heterocromáticas e 127 como eucromáticas. O número de cópias heterocromáticas mineradas em *D. simulans* foi similar ao encontrado em *D. melanogaster* (1.859 cópias).

Ferree & Barbash (2009) demonstraram por FISH em cromossomos mitóticos que o DNA satélite 1.688 de *D. simulans* está concentrado em um único um bloco na heterocromatina pericentromérica do cromossomo X. No presente trabalho, encontramos cópias 1.688 montadas em arquivos referentes aos cromossomos X e 2. No entanto, a árvore filogenética da Figura 11 mostra que cópias 1.688 destes cromossomos não se agruparam em ramos distintos, sugerindo ausência de organização cromossomo-específica.

As cópias heterocromáticas do 1.688 mineradas de *D. simulans* não demonstraram um tamanho mais evidente. Embora a variante de 361 pb seja a mais frequente, variantes de 360 e 359 pb também são abundantes. Este padrão pode estar relacionado com as características específicas do sequenciamento do genoma desta espécie, que incluiu sete populações diferentes. Neste caso, é possível que exista uma diferença interpopulacional na abundância das variantes principais com relação ao tamanho.

Encontramos pela primeira vez cópias 1.688 eucromáticas de *D. simulans*, várias das quais fazem parte de cadeias eucromáticas já descritas em *D.*

*melanogaster*, indicando que estas cadeias se localizam no mesmo loco cromossômico.

### **5.3. O DNA satélite 1.688 de *D. sechellia***

Fazemos aqui o primeiro relato e caracterização do DNA satélite 1.688 em *D. sechellia*, cujo genoma deve conter 24% de DNAs satélites (Bosco e cols. 2007). De acordo com nosso trabalho, a contribuição do DNA satélite 1.688 para o tamanho do genoma desta espécie é de no mínimo 0.2%.

Foram caracterizadas 1.028 cópias do DNA satélite 1.688 em *D. sechellia*, sendo que 883 foram identificadas como heterocromáticas e 145 como eucromáticas. As cópias heterocromáticas de *D. sechellia* possuem um tamanho mais frequente de 361 pb.

As análises filogenéticas demonstraram que cópias 1.688 heterocromáticas de *D. sechellia* formaram um grupo separado das cópias eucromáticas (Figura 13). No entanto, a ausência de dados sobre a localização cromossômica das cópias heterocromáticas impossibilita qualquer inferência sobre a presença ou não de subfamílias cromossomo-específicas.

No presente trabalho, encontramos pela primeira vez cópias 1.688 eucromáticas em *D. sechellia*. Várias destas cópias fazem parte de cadeias eucromáticas já descritas em *D. melanogaster*, indicando que estas cadeias se localizam no mesmo loco cromossômico.

### **5.4. O DNA satélite 1.688 de *D. yakuba***

A presença do DNA satélite 1.688 em *D. yakuba* já tinha sido relatada por Strachan e cols. (1985). No entanto, apenas seis cópias heterocromáticas foram previamente caracterizadas (Strachan e cols. 1985).

Foi estimado que 23% do genoma de *D. yakuba* é composto por DNAs satélites (Bosco e cols. 2007). No presente trabalho foram caracterizadas 1.011 cópias do DNA satélite 1.688 de *D. yakuba*. Isto significa que a contribuição do DNA satélite 1.688 para o tamanho do genoma desta espécie é de no mínimo 0.13%. No entanto, apenas 89 cópias parecem ser heterocromáticas.

Das 89 cópias heterocromáticas, 42 possuem tamanho de 360 pb, que consiste no tamanho das sequências 1.688 de *D. yakuba* analisadas por Strachan e cols (1985). Na árvore filogenética da figura 15, estas 89 cópias se agruparam juntamente com a sequência 1.688 consenso descrita por Strachan e cols. (1985) para esta espécie. Portanto, as cópias 1.688 heterocromáticas demonstraram um padrão de composição nucleotídica e tamanho similar ao previamente descrito para esta espécie.

No presente trabalho, encontramos pela primeira vez cópias 1.688 eucromáticas de *D. yakuba*. Foram encontradas 922 cópias eucromáticas organizadas em diversas cadeias localizadas em diferentes regiões genômicas.

O estudo das cópias eucromáticas do 1.688 de *D. yakuba* revelou duas principais subfamílias, uma contendo cópias de 190 pb e outra contendo cópias de 360 pb. As cópias eucromáticas de 190 pb provavelmente se formaram através de eventos de recombinação desigual, que resultaram na deleção de um fragmento de 170 pb. Trabalhos realizados anteriormente não haviam relatado nenhuma cópia 1.688 de 190 pb em *D. yakuba* ou qualquer outra espécie do subgrupo *melanogaster*, com exceção de *D. teissieri*, cuja cópia heterocromática consenso descrita por Strachan e cols. (1985) possui tamanho de 196 pb. De fato, *D. teissieri* e *D. yakuba* são espécies próximas filogeneticamente (Figura 2).

As cópias do DNA satélite 1.688 no subgrupo *melanogaster* foram previamente descritas a partir de fragmentos de aproximadamente 360 pb gerados após a digestão do DNA genômico com as enzimas de restrição *EcoRI* e *HaeIII* (Strachan e cols. 1985). No entanto, as cópias de 190 pb de *D. yakuba* encontradas no presente trabalho não possuem sítios de clivagem para estas enzimas. Este fato provavelmente impossibilitou sua detecção em trabalhos anteriores. Nossos dados sugerem que houve uma amplificação de variantes de 190 pb na eucromatina de *D. yakuba*.

Notavelmente, foram descritas 28 inversões cromossômicas em *D. yakuba* (Ranz e cols. 2007), o maior número de inversões já descrito no gênero *Drosophila*. Sabe-se que a recombinação ectópica é um mecanismo importante para a geração de inversões cromossômicas. Portanto, é possível que o grande número de inversões cromossômicas presentes em *D. yakuba* esteja relacionado com eventos de recombinação entre cópias 1.688 eucromáticas. De fato, Ranz e cols.(2007)

mostraram que 15 inversões cromossômicas de *D. yakuba* possuem regiões de quebra flanqueadas por sequências repetitivas. No entanto, estas sequências não foram caracterizadas.

### **5.5. O DNA satélite 1.688 de *D. erecta***

Neste trabalho apresentamos a primeira caracterização do DNA satélite 1.688 em *D. erecta*, cujo genoma possui 9% de DNAs satélites (Bosco e cols. 2007). Estudos anteriores estimaram a contribuição do DNA satélite 1.688 0,24% do tamanho do genoma (Lohe & Brutlag 1987). Nossos resultados demonstram que o percentual de contribuição mínima do DNA satélite 1.688 para o genoma de *D. erecta* é de exatamente 0,24%.

Foram caracterizadas 949 cópias do DNA satélite 1.688 de *D. erecta*, das quais 359 cópias são heterocromáticas e 590 cópias são eucromáticas. A variante mais representativa deste DNA satélite em *D. erecta* possui 361 pb. No entanto, também identificamos uma variante de 199 pb. Portanto, devem existir pelo menos duas subfamílias do DNA satélite 1.688 no genoma desta espécie. Estudos anteriores descreveram somente a presença de variantes de 360 pb do 1.688 nesta espécie.

Na árvore filogenética das cópias 1.688 de *D. erecta* (Figura 17) foram observados três ramos heterocromáticos principais. Dois destes ramos são compostos por sequências de 360 pb, enquanto que um terceiro é composto por cópias de 199 pb. Este resultado sugere que as cópias do 1.688 de *D. erecta* devem estar localizadas em três arranjos heterocromáticos distintos, possivelmente localizados em diferentes cromossomos. A divisão das cópias 1.688 heterocromáticas de *D. erecta* em três agrupamentos filogenéticos explica a grande variabilidade intraespecífica calculada para estas cópias (19% em média). Experimentos de FISH em cromossomos metafásicos serão realizados para testar esta hipótese.

No presente trabalho, encontramos pela primeira vez cópias 1.688 eucromáticas de *D. erecta*. Várias delas fazem parte de cadeias eucromáticas já descritas em *D. melanogaster*, indicando que estas cadeias se localizam no mesmo loco cromossômico.

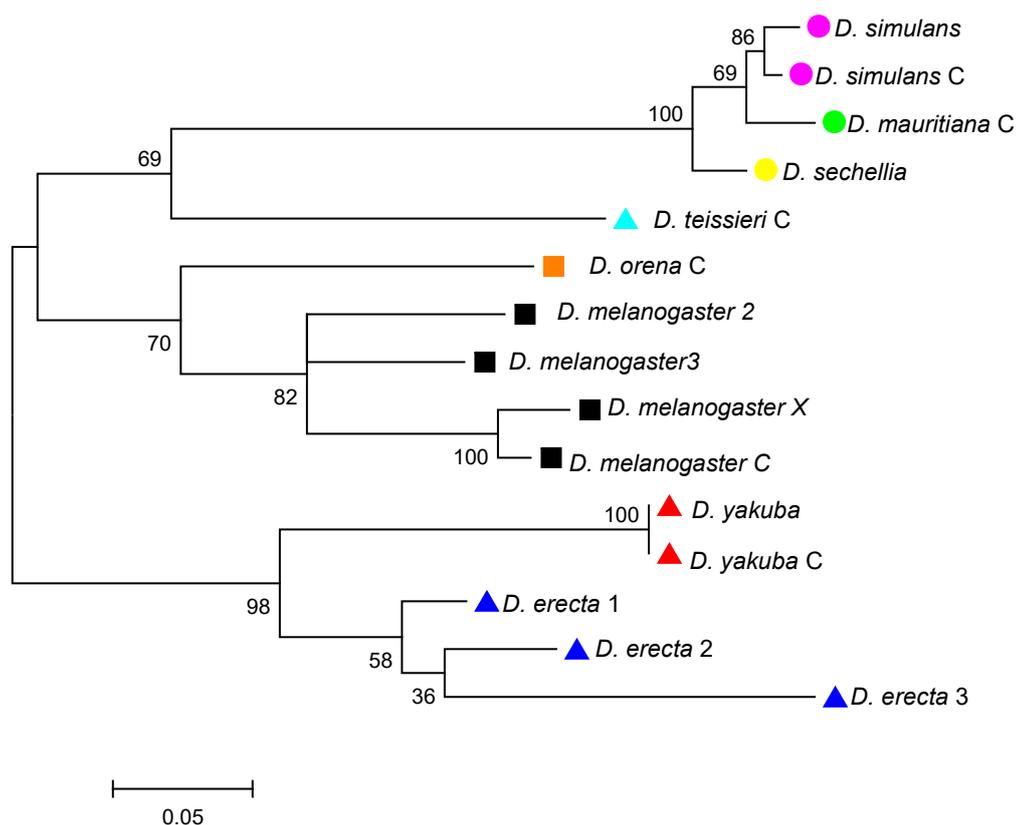
### **5.6. Análises interespecíficas das Cópias 1.688 heterocromáticas**

Estudamos 4.779 cópias 1.688 heterocromáticas extraídas dos genomas sequenciados de cinco espécies de *Drosophila* do subgrupo *melanogaster*. Nossos resultados mostraram que a variabilidade nucleotídica interespecífica entre cópias 1.688 heterocromáticas é maior que a variabilidade intraespecífica para todas as espécies analisadas deste subgrupo (Tabela 4). Conseqüentemente, a grande maioria das cópias de cada espécie foi agrupada em ramos espécie-específicos na árvore filogenética da Figura 19. Este padrão é frequentemente observado em famílias de DNAs repetidos em tandem e é conhecido como evolução combinada (Dover 1982; Charlesworth e cols. 1994; Plohl 2012).

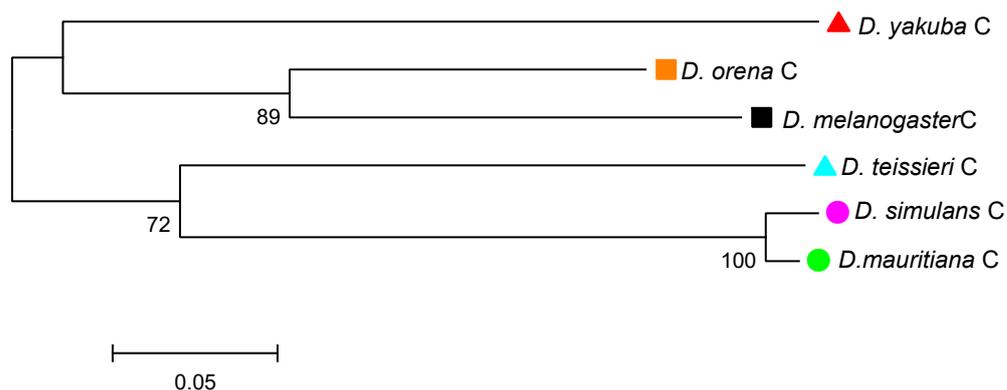
Nossos dados também demonstraram que as sequências heterocromáticas do DNA satélite 1.688 evoluíram em um padrão de evolução combinada. No entanto, o padrão de agrupamento das cópias do 1.688 está apenas parcialmente de acordo com as relações filogenéticas inferidas para as espécies do subgrupo *melanogaster* (Figuras 19 e 27). Esta incongruência entre a filogenia das cópias e a filogenia das espécies indica, principalmente em relação a *D. teissieri* e *D. orena*, diferentes taxas de evolução das cópias 1.688 nestas espécies.

Adicionalmente, também construímos uma árvore filogenética contendo as sequências consenso 1.688 descritas por Strachan e cols. (1985). Similarmente, a topologia da árvore gerada mostrou-se apenas parcialmente de acordo com a filogenia das espécies (Figuras 19 e 28). Estes dados mostram que a hipótese de que o DNA satélite 1.688 evolui de maneira gradual e tempo-dependente não é válida para todas as espécies.

As cópias 1.688 de *D. simulans* e *D. sechellia* foram agrupadas em um mesmo ramo principal, indicando uma grande similaridade entre as cópias destas espécies. De fato, as cópias apresentaram um valor de distância genética interespecífico muito inferior aos valores interespecíficos encontrados nas comparações com as demais espécies (Tabela 4). Embora seja observada uma tendência de cópias de *D. simulans* e *D. sechellia* se agruparem em ramos espécie-específicos, a separação das cópias em apenas dois grupos não é evidente. (Figura 19).



**Figura 27.** Árvore de Máxima Verossimilhança utilizando o modelo de substituição T92 com as seqüências do DNA satélite 1.688 heterocromáticas representativas de cada espécie juntamente com seqüências consenso do 1.688 (indicadas com C) descritas por Strachan e cols. (1985).



**Figura 28.** Árvore de Máxima Verossimilhança utilizando o modelo de substituição T92 com as seqüências consenso do DNA satélite 1.688 descritas por Strachan e cols. (1985).

*D. simulans* e *D. sechellia* são espécies muito próximas filogeneticamente, sendo *D. sechellia* originada de um evento de efeito fundador envolvendo uma linhagem ancestral de *D. simulans* no arquipélago de Seychelles (Tsacas e cols.

1981). O padrão de agrupamento filogenético apresentado pelas cópias heterocromáticas do DNA satélite 1.688 em *D. simulans* e *D. sechellia* (Figura 19) pode estar relacionado apenas com a recente separação entre as duas espécies e o tempo de divergência (~1 milhão de anos) entre elas pode não ter sido suficiente para a ocorrência de uma homogeneização completa (fixação) de variantes diferentes do DNA satélite 1.688 em seus genomas.

Adicionalmente, Garrigan e cols. (2012) inferiram a ocorrência de processos de introgressão envolvendo *D. simulans* e *D. sechellia* entre 4-5 mil anos. A homogeneização de variantes específicas depende do isolamento reprodutivo das espécies (Dover 1982). Logo, processos recentes de introgressão também podem ter contribuído para a ausência de homogeneização completa para variantes espécies-específicas nos genomas destas duas espécies.

*Drosophila mauritiana* é a espécie filogeneticamente mais próxima de *D. simulans* e *D. sechellia*. Foi proposto que *D. mauritiana* também se originou a partir de uma linhagem ancestral de *D. simulans* através de efeito fundador nas ilhas Maurício (Tsacas e cols. 1974). *D. mauritiana* não teve seu genoma sequenciado até o presente. Entretanto, cópias 1.688 de *D. mauritiana* foram estudadas por Strachan e cols. (1985) e também se mostraram muito semelhantes em relação às cópias de *D. simulans*. Novamente, um tempo de divergência recente (~0,9 milhões de anos) de separação entre as espécies pode explicar esta similaridade. Na árvore filogenética apresentada na Figura 27, a cópia 1.688 consenso de *D. mauritiana* foi agrupada, como esperado, juntamente com o ramo de cópias de *D. simulans* e *D. sechellia*.

Em *Drosophila erecta*, cópias heterocromáticas foram agrupadas em dois grupos espécie-específicos, conforme árvore filogenética apresentada na Figura 19. Um destes grupos se situa em um ramo maior que também contém cópias de *D. yakuba*, agrupadas de maneira espécie-específica. Este resultado demonstra uma similaridade entre as sequências de *D. erecta* e *D. yakuba*. Como já foi dito, a distribuição das cópias em *D. erecta* em dois ramos indica que provavelmente estas cópias estão organizadas em dois arranjos diferentes, no mesmo cromossomo ou em cromossomos diferentes. A similaridade entre cópias destas duas espécies indica que possivelmente as cópias do DNA satélite 1.688 estavam presentes em um único

arranjo no ancestral comum das espécies *D. erecta* e *D. yakuba*. Posteriormente, cópias do 1.688 de *D. erecta* devem ter colonizado outras duas regiões genômicas e passado por um processo de homogeneização envolvendo diferentes variantes.

### **5.8. Evidência de Transferência Horizontal do DNA satélite 1.688 de *D. melanogaster* para *D. willistoni***

Com os dados obtidos no presente trabalho, mostramos que o DNA satélite 1.688 se originou na espécie ancestral do subgrupo *melanogaster* (subgênero *Sophophora*). Apesar disso, inesperadamente encontramos cinco cópias em tandem (entre parciais e completas) do DNA satélite 1.688 no genoma de *Drosophila willistoni*, uma espécie também pertencente ao subgênero *Sophophora*, mas filogeneticamente distante do subgrupo *melanogaster* (Figura 3). De fato, *D. willistoni* e espécies do subgrupo *melanogaster* compartilharam um ancestral comum há mais de 35 milhões de anos (Figura 3). Não há trabalhos na literatura relatando a presença de cópias 1.688 no genoma de *D. willistoni*.

Análises filogenéticas mostraram que as cópias do DNA satélite 1.688 de *D. willistoni* se agruparam juntamente com as cópias do cromossomo 3 de *D. melanogaster*, indicando uma alta similaridade entre estas sequências provenientes de espécies filogeneticamente distantes (Figuras 19 e 22). A presença de cópias 1.688 características de *D. melanogaster* no genoma de *D. willistoni* foi confirmada experimentalmente através de PCR (Figura 24).

Transferência horizontal (HT) é a troca de material genético entre espécies isoladas reprodutivamente (Gilbert e cols. 2010). De acordo com a literatura, eventos de transferência horizontal podem ser inferidos de acordo com três tipos de evidências: (i) alta similaridade nucleotídica, (ii) filogenia incongruente e (iii) distribuição descontínua do elemento na filogenia do grupo (Loreto e cols. 2008). Nós encontramos estes três tipos de evidência em nossos dados. Portanto, sugerimos que um evento de HT foi responsável pela transferência de cópias 1.688 de *D. melanogaster* para o genoma de *D. willistoni*.

Através de uma análise dos *reads* originais do sequenciamento de *D. willistoni*, conseguimos identificar uma sequência de 340 pb flanqueando apenas uma

extremidade das cópias 1.688 de *D. willistoni*. Uma busca por sequências similares no GenBank através do BLAST revelou que esta sequência corresponde a parte de um elemento transponível chamado de *Doc*.

*Doc* foi descrito como um típico retrotransposon não-LTR com estrutura similar à encontrada em outros elementos não-LTR presentes no genoma de *D. melanogaster*, como os retrotransposons *F*, *G*, *jockey* *oI* factor (Driver 1989). No entanto, o *Doc* possui distribuição restrita a apenas duas espécies do gênero, *D. melanogaster* e *D. simulans* (Biémont & Cizeron 1999). Portanto, a presença do *Doc* em *D. willistoni* é aqui relatada pela primeira vez. Não encontramos outras cópias de *Doc* no genoma de *D. willistoni* além do fragmento encontrado associado às cópias 1.688.

A sequência do *Doc* de *D. willistoni* foi utilizada para busca de sequências similares no Genbank através de BLAST. Os *hits* mais significativos mostraram cópias do *Doc* de *D. melanogaster*. Em especial, o terceiro *hit* mais significativo (GenBank: S50810.1) mostrou justamente uma cópia do *Doc* de *D. melanogaster* terminando abruptamente com cópias 1.688 características de *D. melanogaster*. Este arranjo entre o *Doc* e cópias 1.688 foi descrito por Slobodkin & Alatortsev (1992), juntamente com o relato de outros três diferentes arranjos envolvendo as mesmas sequências. Portanto, estes resultados sugerem que o evento de transferência horizontal também envolveu uma cópia do *Doc* de *D. melanogaster* adjacente a uma cadeia de cópias 1.688.

Um estudo anterior realizado por Pasyukova & Nuzhdin (1993) mostrou que o elemento *Doc* está presente em todos os cromossomos de *D. melanogaster* e que mais da metade das novas inserções deste elemento tiveram como alvo o braço esquerdo do cromossomo 3. Mais recentemente, Kaminker e cols. (2002) descreveram a presença de 55 cópias do elemento *Doc* no genoma de *D. melanogaster* e mostraram que 19 delas estão localizadas no braço esquerdo do cromossomo 3. Portanto, a transferência de cópias 1.688 do cromossomo 3 de *D. melanogaster* para *D. willistoni* foi provavelmente mediada por um elemento *Doc* presente no mesmo cromossomo. Isto explica o agrupamento das cópias 1.688 de *D. willistoni* com as cópias do cromossomo 3 de *D. melanogaster* na árvore filogenética da Figura 22.

Através de experimentos de PCR, conseguimos amplificar o arranjo *Doc* + cópias 1.688 presente no genoma de *D. willistoni*. Um clone correspondente a este arranjo foi sequenciado (*DwillDoc1.688-1*). No alinhamento nucleotídico da Figura 26, é possível observar que o arranjo *Doc+1.688* encontrado em *D. willistoni* é muito similar aos arranjos *Doc+1.688* encontrados em *D. melanogaster*.

Um estudo anterior mostrou que inserções do *Doc* em *D. melanogaster* geram duplicações de sítio alvo (TSDs) com tamanho variando de 0-14 pb (Driver 1989). Esta variação de tamanho indica que o *Doc* se utiliza de estratégias de inserção que não envolvem proteínas codificadas por ele mesmo (Driver 1989). No presente estudo, não encontramos as sequências flanqueadoras do arranjo *Doc+1.688* em *D. willistoni*, fato que impossibilitou a verificação da presença de TSDs.

Apesar dos retrotransposons não-LTR serem bastante abundantes no genoma de eucariotos, há poucos relatos de eventos de transferência horizontal envolvendo estes elementos. Até o presente, foram relatados apenas 14 eventos de HT envolvendo retrotransposons não-LTR em vertebrados e invertebrados. Por outro lado, foram relatados mais de 200 casos de HT envolvendo transposons de DNA e retroelementos com LTR em eucariotos (Schaake cols. 2010).

A razão para a baixa quantidade de eventos de HT envolvendo elementos nãoLTRs provavelmente está relacionada à ausência de intermediários de dupla fita DNA durante o processo de transposição. Retrotransposons não-LTR utilizam como mecanismo de transposição a conversão de transcritos de RNA em DNA complementar e são inseridos novamente no genoma em um sítio genômico de forma replicativa (Han 2010). Os intermediários de fita dupla presentes em transposons de DNA e retrotransposons LTR são moléculas mais estáveis que os transcritos de elementos não-LTR, sendo assim mais suscetíveis a eventos de HT (Loreto e cols. 2008; Schaak e cols. 2010).

No gênero *Drosophila*, foram relatados quatro eventos de HT envolvendo quatro retrotransposons não-LTR, sendo eles o *I* factor, *F*, *jockey* e *Doc* (revisado por Loreto e cols. 2008). Em especial, Sánchez-Garcia e cols.(2005) descreveram um evento de HT envolvendo o elemento *Doc* entre *D. melanogaster* e *D. yakuba*. Foi detectada a presença de somente uma cópia do *Doc* de *D. melanogaster* no genoma de *D. yakuba*.

Similarmente, encontramos apenas uma cópia do *Doc* de *D. melanogaster* no genoma de *D. willistoni*.

Estudos anteriores relataram eventos de HT entre *D. melanogaster* e *D. willistoni* envolvendo diferentes TEs, como os elementos P (Daniels e cols. 1990) e o retrotransposon LTR *copia* (Jordan & McDonald 1999; Rubin e cols. 2011). Portanto, a existência de outros casos de HT relatados para estas espécies reforça nossa hipótese de que um evento de HT foi responsável pela presença do arranjo *Doc+1.688* de *D. melanogaster* no genoma de *D. willistoni*.

Os mecanismos através dos quais as sequências de DNA são transferidas horizontalmente de uma espécie para outra ainda são pouco conhecidos. Alguns mecanismos foram propostos após a descrição de eventos de HT de elementos transponíveis e genes associados a vírus, bactérias ou moléculas de DNA circular extracromossômico (revisado Keeling & Palmer 2008). O fenômeno de transdução é um processo de transposição bem descrito para elementos L1 em mamíferos, em especial humanos, em que um único segmento de DNA à jusante do L1 é mobilizado como parte de eventos errôneos de retrotransposição (Pickeral e cols. 2000). Neste tipo de evento, frequentemente há uma perda de parte da região 5' do TE, gerando cópias incompletas. Por exemplo, diversos elementos L1 presentes no genoma de humanos possuem cópias com diferentes rearranjos nas regiões 5', incluindo fragmentos invertidos ou contendo grande deleções (Hutchison e cols. 1989). Por outro lado, as regiões 3' não apresentam danos na sequência. Não há relatos de transdução em outras espécies além de humanos e camundongos.

Essa perda da porção 5' do retrotransposon *Doc*, juntamente com os resultados obtidos no presente trabalho, sugerem que, através do mecanismo de transdução, uma cópia do *Doc* foi transcrita juntamente com cópias *1.688* adjacentes e este transcrito foi posteriormente transferido horizontalmente para o genoma de *D. willistoni*. Este seria, portanto, o primeiro caso de transdução relatado em não-mamíferos e em *Drosophila*.

Há vários casos descritos na literatura que mostram que, após uma HT envolvendo um TE, este pode amplificar-se no genoma do novo hospedeiro (Kidwell & Lisch 2000). Encontramos apenas uma cópia do arranjo *Doc+1.688* de *D.*

*melanogaster* em *D. willistoni*. Portanto, aparentemente não ocorreu amplificação do *Doc* no genoma de *D. willistoni* após a HT.

Yang & Nuzhdin (2003) relataram que cópias do *Doc* de *D. melanogaster* contendo sequências 5' truncadas não possuem mais capacidade de transposição devido à perda da região promotora da RNA polimerase. Isto poderia explicar porque o elemento *Doc* não sofreu amplificação no genoma de *D. willistoni*. O mesmo argumento poderia explicar a presença de apenas uma cópia do *Doc* de *D. melanogaster* em *D. yakuba* após a HT (Sánchez-Garcia e cols.2005).

Em resumo, há vários relatos envolvendo HT de elementos transponíveis em *Drosophila* e em outros organismos. Neste trabalho, relatamos pela primeira vez um exemplo de transferência horizontal de DNA satélites em eucariotos. No entanto, nossos dados sugerem que este evento de HT foi associado a um elemento transponível através de um processo de transdução.

## 6. CONCLUSÕES

A família de DNA satélite 1.688 foi encontrada em todas as cinco espécies estudadas do subgrupo melanogaster, o que indica que a mesma estava presente na espécie ancestral do subgrupo melanogaster. Portanto, a família

1.688 possui pelo menos 10 milhões de anos.

Cópias 1.688 foram encontradas nas regiões heterocromáticas e eucromáticas dos cromossomos de todas as cinco espécies. Portanto, os eventos de dispersão de cópias heterocromáticas para eucromatina ou vice-versa ocorreram na espécie ancestral do subgrupo melanogaster.

As cópias heterocromáticas foram muito mais abundantes do que as cópias eucromáticas nos genomas de *D. melanogaster*, *D. simulans* e *D. sechellia*. De fato, existem vários exemplos que mostram que DNAs repetitivos tendem a se acumular nas regiões heterocromáticas dos cromossomos. Surpreendentemente, as espécies *D. yakuba* e *D. erecta*, que são filogeneticamente próximas, apresentaram uma quantidade de cópias heterocromáticas inferior à de cópias eucromáticas. Para *D. yakuba*, esta amplificação de cópias 1.688 pode estar de alguma forma relacionada ao grande número de rearranjos cromossômicos relatados para esta espécie.

O tamanho das cópias 1.688 se mostrou variável tanto intra como interespecificamente. A variante mais abundante consiste de repetições com tamanho aproximado de 360 pb. A variante menos abundante consiste de repetições com tamanho aproximado de 190 a 199 pb, dependendo da espécie analisada.

Nas espécies do subgrupo melanogaster, as cópias heterocromáticas 1.688 apresentaram uma variação intra-específica menor que a variação interespecífica, evidenciando um padrão de evolução combinada. No entanto, as filogenias obtidas com as cópias 1.688 não foram totalmente congruentes com as relações filogenéticas das espécies, indicando, em alguns casos, diferentes taxas de evolução.

Sugerimos que as poucas cópias 1.688 encontradas no genoma de *D. willistoni* foram resultado de um evento de transferência horizontal envolvendo cópias

1.688 de *D. melanogaster*. Este é o primeiro relato de HT envolvendo DNAs satélites em eucariotos. No entanto, propomos que este evento de HT foi mediado pelo retrotransposon Doc.

## 7. PERSPECTIVAS

No presente trabalho relatamos o primeiro caso de transferência horizontal envolvendo sequências de DNA satélite em eucariotos. Futuros estudos deverão ser realizados para se entender a abrangência deste fenômeno e de seu impacto na evolução dos DNAs satélites e dos genomas.

Futuras análises sobre a localização cromossômica do DNA satélite *1.688* em todas as espécies do subgrupo *melanogaster* precisam ser realizadas para que sejam confirmadas as inferências feitas no presente trabalho e para que sejam acrescentados mais elementos para a discussão da evolução molecular do DNA satélite *1.688*.

Em relação às cópias *1.688* eucromáticas, futuros estudos de análises das sequências flanqueadoras servirão para identificar exatamente quais cópias presentes nos genomas das espécies estudadas podem estar ligadas a processos de regulação gênica, principalmente quando estas estiverem posicionadas em *cis*.

Os resultados obtidos no presente trabalho também levantaram a hipótese de que a expansão no número de cópias *1.688* eucromáticas no genoma de *D. yakuba* possa estar entre os fatores responsáveis pelo alto número de inversões cromossômicas relatadas para esta espécie. Futuros estudos das regiões dos pontos de quebra destas inversões poderão confirmar (ou não) a associação das inversões com as cópias *1.688*.

## 8. REFERÊNCIAS BIBLIOGRÁFICAS

ABAD JP, AGUDO M, MOLINA I, LOSADAA, RIPOLL P, VILLASANTE A.

**Pericentromeric regions containing 1.688 satellite DNA sequences show antikinetochore antibody staining in prometaphase chromosomes of *Drosophila melanogaster*.** Mol. Gen. Genet. 264(4):371-377.2000.

ADAMS, M. D.; CELNIKER, S. E.; HOLT, R. A.; EVANS, C. A.; GOCAYNE, J. D. e cols.  
**The genome sequence of *Drosophila melanogaster*.** Science 287: 2185-2195.  
(2000).

ALTSCHUL SF, GISH W, MILLER W, MYERS EW, LIPMAN DJ. **Basic local alignment search tool.** J. Mol. Biol. 215: 403–410. (1990).

ARAVIN, A. A., LAGOS-QUINTANA, M., YALCIN, A., ZAVOLAN, M., MARKS, D., SNYDER, B., TUSCHL, T. **The Small RNA Profile during *Drosophila melanogaster* Development.** *Developmental cell*, 5(2), 337-350. (2003).

BACHMANN, L.; SPERLICH, D.; **Gradual evolution of a specific satellite DNA family in *Drosophila ambigua*, *D. tristis*, and *D. obscura*.** Mol. Biol. Evol. 10, 647–659. (1993).

BACHMANN L, SCHIBEL JM, RAAB M, SPERLICH D. **Satellite DNA as a taxonomic marker.** Biochemical Systematics and Ecology 21: 3–11. (1993).

BARTOLOMÉ, C.; BELLO, X.; MASIDE, X. **Widespread evidence for horizontal transfer of transposable elements across *Drosophila* genomes.** Genon. Biol, v. 10, p. R22. (2009).

BENSON G. **Tandem repeats finder: a program to analyze DNA sequences.** Nucleic Acids Res. 27(2): 573-580. (1999).

BIÉMONT, C.; VIEIRA, C. **Junk DNA as an evolutionary force.** Nature. v. 443, p. 521-524. (2006).

BOSCO, G; CAMPBELL, P; LEIVA-NETO, JT; MARKOW, TA. **Analysis of *Drosophila* species genome size and satellite DNA content reveals significant differences among strains as well as between species.** Genetics 177(3):1277-1290. (2007).

BROWN, DD; WENSINK, PC; JORDAN, E. **A comparison of the ribosomal DNAs of *Xenopus laevis* and *Xenopus mulleri*: evolution of tandem genes.** J. Mol. Biol. 63:57–63. (1972).

CESARI, M.; LUCHETTI, A.; PASSAMONTI, M.; SCALI, V.; MANTOVANI, B. **PCR amplification of the Bag320 satellite family reveals the ancestral library and past gene conversion events in *Bacillus rossius* (Insecta Phasmatodea).** Gene 312, 289–295. (2003).

CHARLESWORTH, B.; SNIEGOWSKI, P.E.; STEPHAN, W. **The evolutionary dynamics of repetitive DNA in eukaryotes.** Nature 37, 215–220. (1994).

CHENG, Z., et al. **Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon.** Plant Cell 14, 1691–1704. (2002).

COHEN, S; SEGAL, D. **Extrachromosomal circular DNA in eukaryotes: possible involvement in the plasticity of tandem repeats.** Cytogenet. Genome Res. 124:327-338. (2009).

DANIELS, S. B., PETERSON K.R., STRAUSBAUGH L. D., KIDWELL M. G., CHOVNICK A., **Evidence for horizontal transmission of the P transposable element between *Drosophila* species.** Genetic 124 339-355. (1990).

DE LA HERRÁN, R; FONTANA, F; LANFREDI, M; CONGIU, L; LEIS, M, et al: **Slow rates of evolution and sequence homogenization in an ancient satellite DNA family of sturgeons.** Mol Biol Evol;18:432– 436.(2001).

DIBARTOLOMEIS, S. M; TARTOF, K. D; JACKSON; F. R. **A superfamily of *Drosophila* satellite related (SR) DNA repeats restricted to the X chromosome euchromatin.** *Nucleic Acids Res.* 20(5): 1113-1116. (1992).

DOVER, G. A. **Molecular drive: a cohesive mode of species evolution.** *Nature* 299, 111–117. (1982).

DOVER, G. A. **Molecular drive in multigene families: how biological novelties arise, spread and are assimilated.** *Trends Genet.* 2, 159–165. (1986).

DRIVER A., LACEY S.F., CULLINGFORD T.E., MITCHELSON A., O'HARE K. **Structural analysis of Doc transposable elements associated with mutations at the white and suppressor of forked loci of *Drosophila melanogaster*.** *Mol Gen Genet.* (1989).

DROSOPHILA 12 GENOMES CONSORTIUM. **Evolution of genes and genomes on the *Drosophila* phylogeny.** *Nature* 450(8): 203-218. (2007).

EDGAR RC; **MUSCLE: multiple sequence alignment with high accuracy and high throughput,** *Nucleic Acids Res.* 32(5), 1792-97. (2004).

FERREE, P.M; BARBASH, D.A. **Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in *Drosophila*.** *PLoS Biol.* 7(10):e1000234. (2009).

FINNEGAN D.J.: **Transposable elements: how non-LTR retrotransposons do it.** *Curr Biol*, 7:R245-R248 (1997).

FISTON-LAVIER A.S., SINGH N.D., LIPATOV M, PETROV D.A. ***Drosophila melanogaster* recombination rate calculator.** *Gene* 463 (2010).

GILBERT C, CORDAUX R. **Horizontal transfer and evolution of prokaryote transposable elements in eukaryotes.** *Genome Biol Evol.* (2013).

GILBERT, C., SCHAACK, S., PACE II, J. K., BRINDLEY, P. J., FESCHOTTE, C. **A role for host-parasite interactions in the horizontal transfer of transposons across phyla.** *Nature*, 464(7293), 1347-1350. (2010).

GRAY, D. M., and D. M. SKINNER. **A circular dichroism study of the primary structures of three crab satellite DNAs rich in AT pairs.** *Biopolymers* 13:843-852. 1974.

GREGORY, TR; JOHNSTON, JS. **Genome size diversity in the family Drosophilidae.** *Heredity* 101(3):228-38. 2008.

GRENIER, E; CASTAGNONE-SERENO P; ABAD, P. **Satellite DNA Sequences as Taxonomic Markers in Nematodes of Agronomic Interest.** *Parasitology* vol. 13. (1997).

GUTKNECHT, J; SPERLICH, D; BACHMANN, L. **A species specific satellite DNA family of Drosophila subsilvestris appearing predominantly in B chromosomes.** *Chromosoma* 103:539-544. (1995)

HAN J.S. **Review Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions.** *Mobile DNA* 1:15 (2010).

HEIKKINEN, E.; LAUNONEN, V.; MULLER, E.; BACHMANN, L. **The pVb370 BamHI satellite DNA family of the Drosophila virilis group and its evolutionary relation to mobile dispersed genetic pDv elements.** *J. Mol. Evol.* 41, 604–614. (1995).

HENIKOFF, S.; AHMAD, K.; MALIK, H.S. **The centromere paradox: stable inheritance with rapidly evolving DNA.** *Science* 293, 1098–1102. (2001).

HESLOP-HARRISON, J.S.; MURATA, M.; OGURA, Y.; SCHWARZACHER, T.; MOTOYOSHI, F. **Polymorphisms and genomic organization of repetitive DNA from centromeric regions of Arabidopsis chromosomes.** *Plant Cell* 11, 31–42. (1999).

HESLOP-HARRISON, J.S.; BRANDES, A.; SCHWARZACHER, T. **Tandemly repeated DNA sequences and centromeric chromosomal regions of Arabidopsis species.** Chromosom. Res. 11, 241–253. (2003).

HOSKINS, R.A.; SMITH, C.D.; CARLSON, J.W. et al. **Heterochromatic sequences in a Drosophila whole-genome shotgun assembly.** Genome Biol. 3(12):1-16. (2002).

HOSKINS, R.A., et al. **Sequence finishing and mapping of Drosophila melanogaster heterochromatin.** Science 316, 1625–1628. (2007).

HSIEH, T; BRUTLAG, D. **Sequence and sequence variation within the 1.688 g/cm<sup>3</sup> satellite DNA of *Drosophila melanogaster*.** J. Mol. Biol. 135(2):465-81. (1979).

HUTCHISON, C. A., S. C. HARDIES, D. D. LOEB, W. R. SHEHEE, M. H. EDGELL,. **LINEs and related retroposons: Long interspersed repeated sequences in the eukaryotic genome.** Mobile DNA pp. 593–617 (1989).

JORDAN, I. KING, MATYUNINA L. V., JOHN F. MCDONALD J. F. **Evidence for the recent horizontal transfer of long terminal repeat retrotransposon.** Proceedings of the National Academy of Sciences 96.22 (1999).

KAMINKER J.S., BERGMAN C. M., BRENT KRONMILLER B, JOSEPH, CARLSON J., SVIRSKAS R, PATEL S., FRISE E, WHEELER D.A., LEWIS S.E., RUBIN G. M., ASHBURNER M., CELNIKER S. E. **The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective.** Genome Biology, 3 (12) . (2002).

KEELING P. J., PALMER J. D. **Horizontal gene transfer in eukaryotic evolution.** Nature Reviews Genetics 9, 605–618 (2008).

KELLER, A. ***Drosophila melanogaster* 's history as a human commensal.** *Current Biology*, 17(3), R77-R81. (2007).

KIDWELL M. G. **Evolution of hybrid dysgenesis determinants in *Drosophila melanogaster***. *Proceedings of the National Academy of Sciences*. Vol. 80 no 616551665. (1983).

KIDWELL M. G., LISCH D. R. **Transposable elements and host genome evolution**. *Trends Ecol Evol*. Mar;15(3):95-99.(2000).

KING, K.; JOBST,J; HEMLEBEN,V. **Differential homogenisation and amplification of two satellite DNAs in the genus *Cucurbita* (Cucurbitaceae)**. *J. Mol. Evol.*, 41, 996-1005. 1995.

KIMURA M. **A Simple Method for Estimating Evolutionary Rates of Base Substitutions Through Comparative Studies of Nucleotide Sequences**. *J. Mol. Evol.* 16, 111--120 (1980).

KOUKALOVA, B; MORAES, AP; RENNY- BYFIELD, S; MATYASEK, R; LEITCH, AR; KOVARIK A. **Fall and rise of satellite repeats in allopolyploids of *Nicotiana* over c. 5 million years**. *New Phytol*;186:148– 160. 2010.

KUHN, GCS; SENE, FM; MOREIRA-FILHO, O; SCHWARZACHER, T; HESLOPHARRISON, JS. **Sequence analysis, chromosomal distribution and long-range organization show that rapid turnover of new and old pBuM satellite DNA repeats leads to different patterns of variation in seven species of the *Drosophila buzzatii* cluster**. *Chromosome Res.* 16: 307-324. 2008.

KUHN, G.CS.; TEO, C. H; SCHWARZACHER, T; HESLOP- HARRISON, JS: **Evolutionary dynamics and sites of illegitimate recombination revealed in the interspersion and sequence junctions of two nonhomologous satellite DNAs in cactophilic *Drosophila* species**. *Heredity*;102:453– 464. 2009

KUHN, GCS; KÜTTLER, H; MOREIRA- FILHO, O; HESLOP- HARRISON, JS: **The 1.688 repetitive DNA of *Drosophila*: concerted evolution at different genomic scales and association with genes**. *Mol Biol Evol*;29:7– 11. 2012.

KORYAKOV, DE; ALEKSEYENKO, AA; ZHIMULEV, IF. **Dynamic organization of the betaheterochromatin in the *Drosophila melanogaster* polytene X chromosome.**

Mol. Gen. Genet. 260(6):503-9. 1999.

LACHAISE D, CARIOU ML, DAVID JR, LEMEUNIER F, TSACAS L, ASHBURNER M. **Historical Biogeography of the *Drosophila-Melanogaster* Species Subgroup.**

Evolutionary Biology (1988).

LACHAISE, D., HARRY, M., SOLIGNAC, M., LEMEUNIER, F., BENASSI, V., AND CARIOU, M.L. **Evolutionary novelties in islands: *Drosophila santomea*, a new melanogaster sister species from Sao Tome.** Proc. Roy. Soc. B. 267, 1487–1495.

(2000).

LI, W. **Molecular Evolution**, Sinauer, Sunderland, MA. (1997).

LIEBHABER S.A, KAN Y.W. **Differentiation of the mRNA transcripts originating from the alpha 1- and alpha 2-globin loci in normals and alpha-thalassemics.** *J Clin Invest.*68:439–446. (1981).

LOSADA, A.; VILLASANTE, A. **Autosomal location of a new subtype of 1.688 satellite DNA of *Drosophila melanogaster*.** Chromosome Res. 4(5):372-83. 1996.

LOHE, A. R.; BRUTLAG, D. L. **Multiplicity of satellite DNA sequences in *Drosophila melanogaster*.** Proc. Natl. Acad. Sci. USA 83: 696-700.1993

LOHE, A. R., ROBERTS, P.A. **Evolution of satellite DNA sequences in *Drosophila*.** In: Verma R. S, editor. **Heterochromatin: molecular and structural aspects.**

Cambridge: Cambridge University Press. 1988.

LOHE, A. R., ROBERTS, P.A. **Evolution of DNA in heterochromatin: the *Drosophila melanogaster* sibling species subgroup as a resource.** *Genetica* 109:

125–130, 2000.

LOHE, AR; HILLIKER, AJ; ROBERTS, PA. **Mapping simple repeated DNA sequences in heterochromatin of *Drosophila melanogaster*.** Genetics

134(4):1149-74. 1993.

LORETO, E.L.S.; CARARETO, C.M.A.; CAPY, P. **Revisiting horizontal transfer of transposable elements in *Drosophila***. *Heredity*. v. 100, p. 545-554. 2008

LUCHETTI A, CESARI M, CARRARA G, CAVICCHI S, PASSAMONTI M, SCALI M, MANTOVANI B. **Unisexuality and Molecular Drive: Bag320 Sequence Diversity in *Bacillus* Taxa (Insecta Phasmatodea)**. *J Mol Evol*. 2003 May;56(5):587-96

MARSHALL, L.G., CLARKE, L., **A novel cis-acting centromeric DNA element affects *S. pombe* centromeric chromatin structure at a distance**. *J. Cell. Biol.* 128, 445–454. (1995).

MASUMOTO, H; MASUKATA, H; MURO, Y; NOZAKI, N; OKAZAKI, T: **A human centromere antigen (CENP- B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite**. *J Cell Biol*;109:1963– 1973. 1989.

MEŠTROVIĆ, N.; PLOHL, M.; MRAVINAC, B.; UGARKOVIĆ, Đ. **Evolution of satellite DNAs from the genus *Palorus*—experimental evidence for the library hypothesis**. *Mol. Biol. Evol.* 15, 1062–1068. 1998.

MILLER, W.J.; NAGEL, A.; BACHMANN, J.; BACHMANN, L. **Evolutionary dynamics of the SGM transposon family in the *Drosophila obscura* species group**. *Mol. Biol. Evol.* 17, 1597–1609. 2000.

MODI, W.S.; IVANOV, S.; GALLAGHER, D.S. **Concerted evolution and higherorder repeat structure of the 1.709 (satellite IV) family in bovids**. *J.Mol. Evol.* 58, 460–465. 2004.

MORAN J.V., MORRISH T.A. **Chromosomes: Noncoding DNA (Including Satellite DNA)** eLS. 2003.

MRAVINAC, B.; PLOHL, M.; UGARKOVIĆ, Đ. **Conserved patterns in the evolution of *Tribolium* satellite DNAs**. *Gene* 332, 169–177. 2004.

MRAVINAC, B.; UGARKOVIĆ, Đ.; FRANJEVIĆ, D.; PLOHL, M. **Long inversely oriented subunits form a complex monomer of *Tribolium brevicornis* satellite DNA.** *J. Mol. Evol.* 60, 513–525. 2005.

NEI.M. **Molecular Evolutionary Genetics.** Columbia University Press, New York. (1987).

NIJMAN, I.J.; LENSTRA, J.A. **Mutation and recombination in cattle satellite DNA: a feedback model for the evolution of satellite DNA repeats.** *J. Mol. Evol.*, 52, 361-371. 2001.

O'HARE, K., ALLEY M. R. K, CULLINGFORD T. E., DRIVER A., SANDERSON M. J. **DNA sequence of the Doc retroposon in the white-one mutant of *Drosophila melanogaster* and of secondary insertions in the phenotypically altered derivatives white-honey and white-eosin.** *Mol. Gen. Genet.* 225: 17-24. (1991).

OHTA, T.; DOVER, G.A. **The cohesive population genetics of molecular drive.** *Genetics* 108, 501–521. 1984.

ORGEL, L.E.; CRICK, F.H. **Selfish DNA: the ultimate parasite.** *Nature* 284, 604–607. 1980.

PASYUKOVA, E G.; NUZHIDIN, S V. **Doc and copia instability in an isogenic *Drosophila melanogaster* stock.** *Molecular and General Genetics MGG*, v. 240, n. 2, p. 302-306, (1993).

PICARIELLO, O.; FELICIELLO I; BELLINELLO, R.; CHINALI G. **S1 satellite DNA as a taxonomic marker in brown frogs: molecular evidence that *Rana graeca graeca* and *Rana graeca italica* are different species** *Genome*.(2002).

PICKERAL O.K. MAKALOWSKI W. BOGUSKI M. S . BOEKE J.D. **Frequent human genomic DNA transduction driven by LINE-1 retrotransposition.** *Genome Res.* 10: 411-415 (2000).

PLOHL M, LUCHETTI A, MESTROVIĆ N, MANTOVANI B. **Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin.** *Gene* 409(1-2): 72-82. 2008.

POWEL JR **Progress and prospects in evolutionary biology: The *Drosophila* model.** Oxford University Press, NY. Shiels C, Coutelle C, Huxley C (1997). Contiguous arrays of satellites 1,3, and b form a 1.5-Mb domain on chromosome 22p. *Genomics* 44: 35-44. 1997.

RICHARD GF, KERREST A, DUJON B **Comparative genomics and molecular dynamics of DNA repeats in eukaryotes.** *Microbiol. Mol. Biol. Rev.* 72(4):686-727. 2008.

RIZZON C., MARAIS G., GOUY M., BIEMONT C. **Recombination Rate and the Distribution of Transposable Elements in the *Drosophila melanogaster* Genome** *Genome Res.* 2002 12: 400-407.

RUBIN, P. M.; LORETO, E. L. S.; CARARETO, C. M. A. ; VALENTE, V. L. S. **The copia retrotransposon and horizontal transfer in *Drosophila willistoni*.** *Genetical Research* , p. 1-6,( 2011).

SÁNCHEZ-GARCIA, A.; MASIDE, X.; CHARLESWORTH, B. **High rate of horizontal transfer of transposable elements in *Drosophila*.** *Trends Genet.*, v. 21, n. 4, p. 200-203. (2005).

SCHMIDT, T.; HESLOP-HARRISON, J.S. **Genomes, genes and junk: the large scale organization of plant chromosomes.** *Trends Plant Sci.* 3, 195–199. (1998).

SCHAACK S., GILBERT C., FESCHOTTE C. **Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution .***Trends in Ecology & Evolution* (2010).

SCOTT, A. F., HEATH, P., TRUSKO, S., BOYER, S. H., PRASS, W., GOODMAN, M., SLIGHTOM, J. L. **The sequence of the gorilla fetal globin genes: evidence for multiple gene conversions in human evolution.** *Molecular Biology and*

*Evolution*, 1(5), 371-389. (1984).

SHIELS, C.; COUTELLE, C.; HUXLEY, C. **Contiguous arrays of satellites 1, 3, and beta form a 1.5-Mb domain on chromosome 22p.** *Genomics* 44, 35–44. (1997).

SILVA, J.C.; LORETO, E.L.S.; CLARK, J.B. **Factors that affect the horizontal transfer of transposable elements.** *Curr. Issues Mol. Biol.*, v. 6, p. 57-72. (2004).

SLAMOVITS, C.H.; COOK, J.A.; LESSA, E.P.; ROSSI, M.S. **Recurrent amplifications and deletions of satellite DNA accompanied chromosomal diversification in South American Tuco-tucos (Genus *Ctenomys*, Rodentia: Octodontidae): a phylogenetic approach.** *Mol. Biol. Evol.* 18, 1708–1719. (2001).

SLOBODKIN, I. V.; ALATORTSEV, V. E. **Insertion of the Doc mobile element into *Drosophila* satellite DNA.** *Genetika*, v. 28, n. 10, p. 169 (1992).

SMITH, G.P. **Evolution of repeated DNA sequences by unequal crossover.** *Science* 191, 528–535. 1976.

STRACHAN, T; WEBB, D; DOVER, G. **Transition stages of molecular drive in multiplecopy DNA families in *Drosophila*.** *EMBO J.* 4:1701-1708. 1985

SUN, X; WAHLSTROM, J; KARPEN, G. **Molecular structure of a functional *Drosophila* centromere.** *Cell* 91(7):1007-19. 1997.

TAMURA K, PETERSON D, PETERSON N, STECHER G, NEI M, AND KUMAR S **MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods.** *Molecular Biology and Evolution* (2011).

THROCKMORTON, L. H. **The phylogeny, ecology, and geography of *Drosophila*.** Pp. 421–469 in R. C.King: *Invertebrates of Genetic Interest*. Plenum Press, New York. (1975).

TSACAS, L. **Drosophila teissieri, nouvelle espèce africaine du groupe melanogaster et note sur deux autres espèces nouvelles pour l'Afrique.** Bulletin de la Société Entomologique de France 76, 35–45. (1971).

TSACAS, L., AND DAVID, J. **Une septième espèce appartenant au sous-groupe Drosophila melanogaster Meigen: Drosophila orena spec. nov.** Cameroun (Diptera: Drosophilidae). Beiträge zur Entomologie 28, 179–182. (1978).

TSACAS, L., AND LACHAISE, D. **Quatre nouvelles espèces de la Côte d'Ivoire du genre Drosophila, groupe melanogaster, et discussion de l'origine du sous-groupe melanogaster** (Diptera : Drosophilidae). Annales de l'Université d'Abidjan E7, 193–211. (1974).

TSACAS, L., AND DAVID, J.R. **Drosophila mauritiana n.sp. du groupe melanogaster de l'Île Maurice.** Bulletin de la Société Entomologique de France 79, 42–46. (1974).

TSACAS, L., AND BAECHLI, G. **Drosophila sechellia, n.sp., huitième espèce du sous-groupe melanogaster des Îles Seychelles (Diptera, Drosophilidae).** Revue française d'entomologie 3, 146–150. (1981).

UGARKOVIĆ, Đ.; PLOHL, M. **Variation in satellite DNA profiles—causes and effects.** EMBO J. 21, 5955–5959. 2002.

UGARKOVIC D **Functional elements residing within satellite DNAs.** EMBO J. 6(11): 1035- 1039. 2005.

USAKIN, L., ABAD, J., VAGIN, V.V., DE PABLOS, B., VILLASANTE, A., GVOZDEV, V.A. **Transcription of the 1.688 satellite DNA family is under the control of RNA interference machinery in Drosophila melanogaster ovaries.** Genetics 176(2): 1343--1349.(2007).

VOLPE TA, KIDNER C, HALL IM, TENG G, GREWAL SI, MARTIENSSEN RA). **Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi.** Science 13;297(5588): 1833-7. 2002.

WILLARD, H.F.; WAYE, J.S. **Chromosome-specific subsets of human alpha satellite DNA: analysis of sequence divergence within and between chromosomal subsets and evidence for an ancestral pentameric repeat.** J Mol. Evol. 25, 207–214. 1987.

WARING, GL; POLLACK, JC. **Cloning and characterization of a dispersed, multicopy, X chromosome sequence in *Drosophila melanogaster*.** Proc. Natl. Acad. Sci. U.S.A. 84(9): 2843-2847. 1987.

YANG, H.P., NUZHDIN, S.V. **Fitness costs of Doc expression are insufficient to stabilize its copy number in *Drosophila melanogaster*.** Mol. Biol. Evol. 20(5). (2003).

ZIMMER, EA; MARTIN, SL; BEVERLY, SM; KAN, YW; WILSON, AC. **Rapid duplication and loss of genes coding for the alpha chains of hemoglobin.** Proc. Natl. Acad. Sci. USA 77: 2158-2162. 1980.

## 9. ANEXOS

**Anexo 1.** Tabela contendo posição genômica de arranjos eucromáticos de cópias do DNA satélite 1.688 em *D. simulans* e localização genômica de arranjos ortólogos em *D. melanogaster*.

<b>Scaffold em <i>D. simulans</i></b>	<b>Cromossomo</b>	<b>Posição citológica em <i>D.simulans</i></b>	<b>Posição citológica em <i>D. melanogaster</i></b>	<b>Gene (próximos ou ortólogos)</b>
<b>Sim2L-B1</b>	2L	21	27C	<b>CG11289</b>
<b>Sim2L-B3</b>	2L	27	21E-F	<b>CG3544</b>
<b>Sim2L-B4</b>	2L	30	30 F	<b>CG42367</b>
<b>Sim2L-B5</b>	2L	30	30 F	<b>Cpr30F</b>
<b>Sim3R-B2</b>	3R	84-93		<b>CG33631-33630</b>
<b>Sim3R-B6</b>	3R	84-93		<b>CG7918</b>
<b>Sim3R-B7</b>	3R	84-93		<b>CG33631-33630</b>
<b>Sim-X-B2</b>	X	10B-11A		<b>CG42251</b>
<b>Sim-X-B3</b>	X	11B-14A		<b>CG15756</b>
<b>Sim-X-B4</b>	X	10B-11A		<b>Lsp1alpha</b>
<b>Sim-X-B5</b>	X	11A-11B		<b>CG15754</b>
<b>Sim-X-B6</b>	X	11A-11B		<b>Ade5</b>
<b>Sim-X-B7</b>	X	11A-11B		<b>CG42251</b>
<b>Sim-X-B9</b>	X	11B		<b>Ade5</b>
<b>Sim-X-B15</b>	X	11B-14A		<b>CG15754</b>
<b>Sim-X-B29</b>	X			<b>GD15682</b>

**Anexo 2.** Tabela de localização dos arranjos eucromáticos do DNA satélite 1.688 encontrados em *D. sechellia* em relação a genes ortólogos a *D. melanogaster*.

<b>Sequência de <i>D. sechellia</i></b>	<b>Cromossomo</b>	<b>Posição em <i>D. sechellia</i></b>	<b>Posição em <i>D. melanogaster</i></b>	<b>Gene próximo ou ortólogo</b>
<b>Contig3384</b>	X	8-9	13E	Desert
<b>Contig3374</b>	X	8-9	13E	Desert
<b>Scaffold-4</b>	X	3D3-6E2		CG16782/CG33691
<b>Scaffold-10</b>	X	1A1-3D3		CG3038/CG14269
<b>Scaffold-15</b>	X	8D9-10B6		BCL7like/Tim8
<b>Scaffold-20</b>	X	11F7-12F1,2		CG32635/NetA
<b>Scaffold-21</b>	X	11A9-11F4,7		CG2577/CG1716
<b>Scaffold-34</b>	X	8A5,B2-8D1		Lim1/Reg3
<b>Scaffold-35</b>	X	7E1,2-8A3		nAcRalpha-7E /CG1789
<b>Scaffold-36</b>	X	10B10-10F2		CG15196/CG10352

**Anexo 3.1.** Tabela de localização dos arranjos eucromáticos do DNA satélite 1.688 encontrados no cromossomo X de *D. yakuba* em relação a genes ortólogos a *D. melanogaster*. (H) = heterocromatina e (E) = eucromatina

<b>Bloco com sequências do DNA satélite 1.688 extraído do cromossomo X de <i>D. yakuba</i></b>	<b>Gene ortólogo em <i>D. melanogaster</i></b>	<b>Localização cromossômica em <i>D. melanogaster</i></b>
1	Sem Gene Próximo (SGP)	HC /(Chr U)
2	CG- TYN	E/ X
3	CG43906	E/ X
4	CG 32816	E/ X
5	CG 32816	E/ X
6	CG 17896	E/X
8	CG 13366	E/X
9	SGP	E/X
10	CG FZ3	E/X
12	CG 43867	E/X
13	SGP	E/X
17	SGP	E/X
19	SGP	E/X
20	GG DAAM	E/ X
21	CG-SSX	E/ X
22	SPG	E/ X
23	CG14777	E/ X
24	GENE - INC	E/X
25	GENE- Mur28	E/X
27	SGP	E/X
28	CG 42666	E/X
29	CG 42666	E/X
33	SGP	E/X
35	SGP	E/X
36	CG 15894	E/X
37	~CG33664	E/ X
38	SGP	E/ X
39	CG 16989	E/ X
40	CG 44329	E/ X
41	CG 42699	E/X
42	CG 33080	E/X
43	SGP	E/X
44	SGP	E/X
45	CG 32762	E/X

46	GENE SPR	E/X
48	GENE 32700	E/X
49	GENE RG	E/X
50	Ge 42594	E/ X

52	SGP ~CG 12682	E/ X
53	GENE Cbp80	E/ X
54	GENE 32773	E/ X
56	GENE Bi	E/X
57	GENE norpA	E/X
58	GENE Fas2	E/X
59	GENE Kirre	E/X
60	GENE Kirre	E/X
61	Não encontrado	E/X
62	GENE Kirre	E/X
63	"	E/X
64	"	E/X
65	GENE Kirre	E/X
66	"	E/X
67	GENE CSW	E/X
68	"	E/X
69	EIF2B-epsilon	E/X
70	CG 14817	E/X
71	"	E/X
72	"	E/X
73	CG 43759	E/X
74	CG 2556	E/X
75	CG 43921	E/X
76	"	E/X
77	Smr (Chromatin binding)	E/X
80	GENE Rad	E/X
83	G. Px. Rp515Aa	E/X
84	GENE SyT12	E/X
85	~GENE -SeT2	E/X
86	GENE - Neto	E/X
87	GENE inaD	E/X
88	GENE inaE	E/X
89	SGP	E/X
90	SGP	E/X
91	SGP~P.Bac-F07322	E/X
92	SGP	E/X
93	SGP	E/X
94	GENE NetA	E/X

95	SGP	E/X
97	Flo-2	E/X
98	GENE Eag	E/X
99	CG 43737	E/X
100	T.E. - INE-1	E/X
101	GENE RUT	E/X
103	~CG 9106	E/X
104	GENE Cng1	E/X
105	SGP	E/X
106	SGP	E/X
107	CG9170	E/X
108	SGP	E/X
109	CG34015	E/X
110	SGP	E/X
112	CG42339	E/X
113	CG 43901	E/X
114	CG 42582	E/X
115	CG 17003	E/X
116	~CG14220	E/X
118	SGP	E/X
119	SGP	E/X
120	SGP	E/X
122	CG1431	E/X
123	SGP ~INE-1(3544)	E/X
124	CG 15738	E/X
125	CG 13003	E/X
126	CG 5445	E/X
127	SGP	E/X
128	CG32547	E/X

Tabela de localização dos arranjos eucromáticos do DNA satélite 1.688

**Anexo 3.2.**

encontrados no cromossomo 2L de *D. yakuba* em relação a genes ortólogos a *D. melanogaster*.

<b>Blocos com sequências do 1.688 extraídos do cromossomo 2L de <i>D. yakuba</i></b>	<b>Gene ortólogo em <i>D. melanogaster</i></b>	<b>Ramo Agrupamento filogenético das sequências/ cromossomo em <i>D. melanogaster</i></b>
1	GENE USH	E/2L
2	CG2839	E/2L
3	SGP~ T.E. Roo	E/2L
4	SGP	E/2L
6	CG 9962	E/2L
7	SGP	E/2L
10	SGP	E/2L
11	SGP ~RpL40	E/2L
12	SGP~P.Bac {}03851	E/2L
13	SGP	E/2L
14	"	E/2L
15	GENE RapGAP1 ~TE Burdock	E/2L
16	CG 5171	E/2L
17	~P.Bacf 04167	E/2L
18	SGP ~ INE	E/2L
19	PvF2	E/2L
20		E/2L
21	Ca-beta	E/2L
22	SGP	E/2L
23	CG 6443	E/2L
24	CG31760	E/2L
25	Kek4	E/2L
26	Fu2	E/2L
27	SnRNA U2:34ABb	E/2L

**Anexo 3.3** encontrados no cromossomo 2R de *D. yakuba* em relação a genes ortólogos a *D. melanogaster*.

Tabela de localização dos arranjos eucromáticos do DNA satélite 1.688

<b>Blocos com sequências do DNA satélite 1.688 extraídos do cromossomo 2R de <i>D. yakuba</i></b>	<b>Gene ortólogo em <i>D. melanogaster</i></b>	<b>Ramo Agrupamento filogenético das sequências/ cromossomo em <i>D. melanogaster</i></b>
6	CG30148	E/2R
7	CG43277	E/2R
16	SGP	E/2R
17	P.Bac{RB}e 02213/ Mi{MIC}03827	E/2R
18	Liprin-gamma	E/2R
19	CG43107/ P.Bac{RB}e 00758	E/2R
20	CR44376	E/2R
21	SGP	E/2R
22	CG13579	E/2R

#### Anexo 3.4.

encontrados no cromossomo 3L de *D. yakuba* em relação a genes ortólogos a *D. melanogaster*.

<b>Bloco com sequências do DNA satélite 1.688 extraído do cromossomo 3L de <i>D. yakuba</i></b>	<b>Gene ortólogo em <i>D. melanogaster</i></b>	<b>Ramo Agrupamento filogenético das sequências/ cromossomo em <i>D. melanogaster</i></b>
1	SGP	E/3L
2	GENE p13OCAS	E/3L
3	GENE Glut1	E/3L
4	GENE Ru	E/3L
6	GENE ATPsyn-b	E/3L
7	CG4476	E/3L
8	CDNA BT00258	E/3L
9	SGP	E/3L
10	SGP	E/3L
11	CG33912/ ~INE-1	E/3L
13	CG11357	E/3L
14	CG32236	E/3L

Tabela de localização dos arranjos eucromáticos do DNA satélite 1.688

15	RNAseq- exonjunction	E/3L
16	GENE Con	E/3L
17	GENE Con	E/3L
18	SGP	E/3L
19	CG34991	E/3L
20	SGP/ ~P.Bac{WH} f 05239	E/3L

**Anexo 4.** encontrados nos arquivos sem localização cromossômica definida de *D. erecta* em relação a genes ortólogos a *D. melanogaster*.

<b>Sequência de <i>D. erecta</i></b>	<b>Gene próximo/ortólogo em <i>D. melanogaster</i></b>	<b>Ramo Agrupamento filogenético das sequências/ cromossomo em <i>D. melanogaster</i></b>
<b>Contig2869</b>	GG16346/-	E/X
<b>Contig 7464</b>	-/Gene Klrre	E/X
<b>Contig 780</b>	-/ Gene CNN	E/X
<b>Contig 826</b>	-/ Gene CNN	E/X
<b>Contig 1025</b>	-/ Gene CNN	E/X
<b>Contig 1272</b>	-/ Gene CNN	E/X
<b>Contig 503</b>	-/ Gene CNN	E/X
<b>Contig 2164</b>	-/TEJockey	E/X
<b>Scaffold 1442</b>	-/ Gene CNN	E/X
<b>Scaffold4847</b>	-/TE Jockey	E/X
<b>Scaffold 4644.3</b>	GG12794/ TE Appl	E/X
<b>Scaffold4644.4</b>	GG12802/CG13367	E/X
<b>Scaffold4644.7</b>	GG12806/ Gene sdk	E/X
<b>Scaffold4644.8</b>	GG12806l/Gene sdk	E/X
<b>Scaffold4644.9</b>	GG12809/Gene Fz3	E/X
<b>Scaffold4644.11</b>	GG12736 /CG3713	E/X
<b>Scaffold4644.12</b>	GG12727/CG3690	E/X
<b>Scaffold4644.14</b>	GG12724/ Gene TfIIAS-2	E/X
<b>Scaffold4644.15</b>	GG1720/ CG3655	E/X
<b>Scaffold4644.16</b>	GG12719/-	E/X
<b>Scaffold4644.17</b>	GG112719/-	E/X

Tabela de localização dos arranjos euromáticos do DNA satélite 1.688

<b>Scaffold4644.18</b>	GG12830 /CG11378	E/X
<b>Scaffold4644.19</b>	GG12711/CG11403	E/X
<b>Scaffold4644.20</b>	GG12710/Gene Atf3	E/X
<b>Scaffold4644.21</b>	GG12708/Gene Daam	E/X
<b>Scaffold4644.22</b>	GG12853	E/X
<b>Scaffold4644.23</b>	GG12853	E/X
<b>Scaffold4644.25</b>	GG12857 /CG14777	E/X
<b>Scaffold4644.27</b>	GG12877/ CG 42666	E/X
<b>Scaffold4644.28</b>	GG12877/CG42666	E/X
<b>Scaffold4644.29</b>	GG12908/ Gene csw	E/X
<b>Scaffold 4585.2</b>	snoRNA:GG25968/ Gene link	E/2R
<b>Scaffold4585.4</b>	GG22747/ CG 5888	E/2L
<b>Scaffold4585.6</b>	GG22838	E/2R
<b>Scaffold4585.7</b>	GG20003/Gene egl	E/2R
<b>Scaffold4585.8</b>	GG19944/CG30419	E/2R

<b>Scaffold4820.1</b>	GG11920/CG2267	E/3R
<b>Scaffold4820.2</b>	GG 11760	E/3R
<b>Scaffold4820.4</b>	GG11734/CG 15529	E/3R
<b>Scaffold4820.5</b>	GG11734/CG 15529	E/3R
<b>Scaffold 4820.7</b>	GG 12094	E/3R
<b>Scaffold 4820.8</b>	GG 16730	E/3R
<b>Scaffold 4820.9</b>	GG11368	E/3R
<b>Scaffold 4820.10</b>	GG 12529/ CG 5391	E/3R

**Anexo 5.** Resultado do BLAST no GenBank com a cópia *Doc* de *D. willistoni* revelaram alta similaridade com cópias do *Doc* presentes em *D. melanogaster*

