UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

ANÁLISE IN SILICO DE ELEMENTOS TRANSPONÍVEIS DE DROSOPHILA WILLISTONI

PEDRO HENRIQUE MAPELLI

Orientadora: Maríndia Deprá

Trabalho de Conclusão De Curso a ser apresentado ao Instituto de Biociências – UFRGS, como requisito parcial para obtenção do título de bacharel do curso de Ciências Biológicas 2017/2

Porto Alegre, Janeiro de 2018

Artigo em preparação a ser enviado para o periódico *Genome Biology and Evolution*

In Silico Analysis of Transposable Elements in *Drosophila willistoni*

Pedro H Mapelli[1,*], Carolina Garcia[1], Vera LS Valente[1,2,3], Maríndia Deprá[1,2,3]

[1]Departamento de Genética, Universidade do Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul, Brazil

[2]Programa de Pós-Graduação em Biologia Animal, Universidade do Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul, Brazil

[3]Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade do Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul, Brazil

*Corresponding author: E-mail pedromapelli@gmail.com

Abstract

Transposable elements (TEs) are DNA sequences capable of moving within a genome. TEs, through their mobility, are a important aspect of organisms' genome arrangement, so they can be used as important tools to studying genome evolution and gene function. *Drosophila willistoni* is an model organism for TEs researches, as some of those elements existence and properties have been discovered in the *Drosophila* genus. *D. willistoni* had its genome entirely sequenced, being an interesting target for *in silico* studies. As such, this project had the objective of doing an *in silico* scan of four TE copies (*412, hobo, BuT2, Mar*). Homologue sequences of those four elements were detected. Copies were submitted to phylogenetic trees models to evaluate their proximity. The obtained results bring valuable insights of those TEs. Nevertheless, new approaches, such as *in situ* studies of the elements, are necessary to better understand the evolution of *D. willistoni* and its TEs.

Keywords: transposable elements, *in silico*, *Drosophila willistoni*.

Introduction

Transposable elements (TEs) are DNA sequences with the capability of movement within a genome, being present in virtually all organisms (Huang et al., 2012). Barbara McClintock first identified TEs in the 1950s in her studies with maize. Nowadays it is known that TEs are a considerable portion of organisms' genome: 12% in *Drosophila*, 45% in humans, 50% in maize and reaching up to 90% in some plants (Guerreiro, 2012).

TEs are separated by their mobility mechanism: class I encompasses those using RNA as an intermediate for transposition (retrotransposons) and class II those using DNA as an intermediate (Finnegan, 1990). The transposition of class I elements occurs via messenger RNA (mRNA) synthesis by host cell transcriptional machinery. Once the mRNA originated by the TE reaches the cell cytoplasm, transposition-related enzymes are synthesized. One of those enzymes, Reverse Transcriptase (RT), is responsible for the synthesis of a new DNA copy of the retroelement using the mRNA as a template, allowing the new DNA copy to be integrated in the host genome. This mechanism is called copy-and-paste, because the original TE is not detached from the host DNA, resulting in an increasing number of copies for each transposition event (Wicker et al., 2007).

On the other hand, transposition of class II elements normally occurs through a non-replicative mechanism. Their structure consists of a transposase gene capable of recognizing specific sequences in the element´s extremities. This allows the excision of the TE and insertion in a new location in the genome, thus being called cut-and-paste mechanism (Wicker et al. 2007).

Based on TEs observed distribution in different groups of organisms, two mechanisms of genetic transfer were proposed (Montchamp-Moreau et al., 1993): (1) vertical transfer, which occurs from one ancestral host to its offspring, separating TEs through speciation and (2) horizontal transfer, between two reproductively isolated species, resulting in the sometimes-observed

phylogenetic incongruence between host organisms and TEs. At first, TEs were considered junk DNA for their non-coding properties. Nevertheless, recent studies show that TEs, through their mobility, are a important mechanism for evolution in organisms and may be used as important tools to studying genome evolution and gene function (Muñoz-López & García-Pérez, 2010). In humans, TEs transposition has been linked to various diseases, from forms of cancer to sclerosis, and contributing both to neurologic development as well as to neurologic disorders (Ayarpadikannan & Kin, 2014). Even though studying TEs role in our species may be appealing, most of those elements are inactive in the human genome.

To fulfill this role of studying TEs, the genus *Drosophila* presents a series of characteristics that make it an interesting model organism. In *Drosophila melanogaster*, around 15% of its genome is constituted by TEs, 30% being active (Kaminker et al., 2002). Evidence suggests that some elements may have invaded this species as late as the twentieth century, in the first decades of the 1900s (Anxolabéhère et al., 1988). As so, *D. melanogaster* became one of the most studied organisms regarding TEs. Still, to better understand the evolution of those elements, an interesting approach would be to analyze their distribution in phylogenetically close species. As such, some researches in other *Drosophila* have been done, mostly about those who had their genome fully sequenced in the *Drosophila 12 Genomes Consortium* (2007).

Drosophila willistoni is the only Neotropical species included in this consortium. Its geographical range stretches from Mexico to northern Argentina, occurring in very different biomes and environments across the continent (Regner et al., 1996). The basic karyotype of *D. willistoni* consists of two pairs of metacentric chromosomes and an acrocentric pair (Dobzhansky, 1950), differing from other close species (fig 1). From a chromosomal perspective, is one of the most polymorphic species of the genus (Rohde & Valente, 2012), making it an interesting target for evolutionary genetics studies. Some researches were carried about TEs in *D.*

willistoni, such as *P, gypsy* elements (Sassi et al., 2005); and the TE *Galileo* (Gonçalves et al., 2014). Nevertheless, this genome is still underexplored, and the search and characterization of other TEs will allow a better understanding of evolution of both TEs and *Drosophila*.

The present study has the objective of doing an *in silico* scan of four TE copies (namely, *412, hobo, BuT2* and *Mar*) in the available *D. willistoni* genome, with the intuit of establishing the evolution relationships of said TEs copies. This will allow a better understanding about this genome composition and will be a basis for future analysis of *D. willistoni* chromossomic and gene evolution.

The most common approach for analysis of TEs in the genome is based on detecting homology to known or putative TEs sequences (Bergman and Quesneville, 2007). Four TEs were selected for this project:

Class I Transposons

412

412 is a LTR retrotransposon, which contains Long Terminal Repeats (LTR) flanking its internal coding region. This type of retrotransposon is very similar to retroviruses, although they do not encode viral capsid proteins and, therefore, are incapable of forming infectious particles (Bushman, 2002). Brookman et al. (1992) described a complex expression pattern during embryogenesis of this element in *D. melanogaster* male testes, being a useful marker for gonadal mesoderm. A *412 D. melanogaster* fragment of 885bp (GenBank access code X04132) was used as query (Blauth et al., 2011).

Class II Transposons

The *hobo*, *BuT2* and *Mar* elements are class II transposons, belonging to the *hAT* superfamily. The characteristics of this superfamily include the presence of short terminal inverted repeats (TIRs) and, because of the transposition process, generation of target site duplications (TSDs) (Feschotte & Pritham, 2007).

hobo

Three forms of *hobo* are found in *Drosophila* (Ortiz & Loreto, 2008). The first is the complete/canonical element, which contains a gene with the potential to encode a transposase enzyme. This form is still active in *D. melanogaster* and is related to the hybrid dysgenesis syndrome (Blackman et al., 1989). The second form is represented by those elements similar with the canonical element, but with deletions of variable lengths in the internal portion. The third form is known as *relic-hobo*, which possess 80% similarity with the canonical, with multiple rearrangements and inability to code a functional transposase. A 665bp fragment was used as query. This fragment was obtained by means of amplification from *D. willistoni hobo* transposon, employing *hobo* specific primers described in Deprá et al. (2009). After PCR, amplicons were cloned in pGem T Easy vector (Promega) and sequenced.

BuT2

The *BuT2* element is a 2775-bp long element that was found originally in *D. buzzatii* (Cáceres et al., 2001). It is implied that BuT2 is an active TE that is involved in multiple events of horizontal transfer in the *Drosophila* group (Rossatto et al., 2014). *D. buzzatii* canonical BuT2 nucleotide sequence fragment of 766bp (GenBank access code AF368884) was used as query (Rossatto et al, 2014).

Mar

Mar belongs in a groups of non-autonomous sequences (has to be mobilized by another TE enzymes) denominated miniature inverted-repeat transposable elements (*MITEs*). Like other members of *hAT* superfamily, they contain conserved TIRs and are flanked by TSDs, but differ by having short sequences with no coding capacity and AT-rich sequences in their inner region (Kuang et al., 2009). A *D. tropicalis* clone, labeled as a putative full-length Mar, consisting of 2487-bp obtained from Deprá et al. (2012) was used as query.

Methodology

In silico search

BLASTn was performed on Flybase´s *Drosophila willistoni* genome (http://flybase.org/blast/), searching for homologous sequences of the four queries cited above. The parameter set up to achieve the top list of significant TE copies was considered as: (1) Expectancy value <1E-9; (2) maximum of 20 hits for each scaffold for the phylogenetic tree analysis. All copies obtained were evaluated as present/absent in euchromatin scaffolds of *D. willistoni* according to Garcia et al. (2015). Another BLASTn was performed in GenBank (https://blast.ncbi.nlm.nih.gov/Blast.cgi) looking for potential horizontal transfer cases of those TEs from non-*Drosophilidae* organisms.

Sequence analysis

Nucleotide sequences obtained were aligned in MEGA7 by muscle tool (*hobo*, *BuT2*) or manually (*412*, *Mar*). Once aligned, the corresponding data of each TE were used to construct the phylogenetic tree of each TE, using the model indicated by the program. They were:

412- Maximum likelihood method using Generalized Time-Reversible model with Gamma parameter of 2.0. Since the number of fragments of 412 was too high, a new tree was constructed, using fragments with expectancy value of 1E-100 or lower, to refine the results. Forthe new tree, neighbor-joining method using Tamura three-parameter model with Gamma parameter of 1.0 was used.

hobo- Maximum likelihood method using Tamura-Nei method.

BuT2- Neighbor-joining method using Tamura three-parameter model with Gamma parameter of 2.4.

Mar- Maximum likelihood method using Generalized Time-Reversible model with Gamma parameter of 2.0.

Results

The four TEs studied were present in *D. willistoni* genome, each showing their particularities. A brief summary of the copies retrieved from the genome database is described in table 1.

412

The *412* query recovered 223 hits on Flybase. The biggest fragment has 789bp (80.7% similarity) and the smallest has 57bp (93% similarity) (Fig. 1). Two copies are present in euchromatin scaffolds. The generated phylogenetic tree (Fig. 2) shows the distinction between two groups of 412 TEs, and sub sequential divisions of those groups in smaller ones with high confidence. The query, originally from *D. melanogaster*, is isolated from both groups.

hobo

The *hobo* query recovered five hits on Flybase and one hit in GenBank coming from Medfly (*Ceratitis capitata*) genome. The biggest fragment has 434bp (94.7% similarity) and the smallest 63 (96.8% similarity) (Fig. 3). None of the copies was present in euchromatin. The phylogenetic tree (Fig.4) shows that the *Ceratitis capitata hobo* sequence is separated from the other *hobo* sequences of *D. willistoni*, which are similar to each other.

BuT2

The *BuT2* query recovered 51 hits on Flybase. The biggest fragment has 756bp (98.9% similarity) and the smallest has 34bp (100% similarity) (Fig. 5). Thirty-three copies are present in euchromatin scaffolds. The generated phylogenetic tree (Fig.6) shows clustering of few copies. However, most nucleotide sequences cannot be distinguished from each other.

Mar

The *Mar* query recovered 175 hits on Flybase. The biggest fragment has 1339bp (88.2% similarity) and the smallest has 36bp (100% similarity) (Fig. 7). Thirty-one copies are present in euchromatin scaffolds. The generated phylogenetic tree (Fig. 8) shows that most of the nucleotide

sequences cannot be separated with confidence, meaning a high resemblance between the copies and non-compatibility of the copies.

Discussion

This work presents findings about four TEs in *Drosophila willistoni* genome. Retrotransposons, in comparison to class II TEs, tend to conserve full-length copies and, through their copy-and-paste mechanism, have a greater number of copies within genomes (Lerat et al., 2003). Not surprisingly, we found that *412* element had the highest number of hits in BLASTn searches in comparison to *hobo*, *BuT2* and *Mar*. When analyzing the copies with expectancy value lower than 1E-100, only two euchromatin copies were found in *412*, same value found in *Drosophila sechellia* by Cizeron et al. (1998), which also found a greater number of complete copies of *412* in heterochromatin across all Drosophilidae. Cizeron et al. (1998) study also suggests that *412* was present in ancestors of *Drosophila* species, but there has been divergence of this element along the genomes of *Drosophilidae*. This explains the most complete copy found in this research having approximately 80% of similarity with the canonical *D. melanogaster* sequence.

On the other hand, the *hobo* element had the lowest number of copies with only five, one being from another organism, namely *Ceratitis capitata*. Torti et al. (2005) named the *Ceratitis capitata hobo*-like sequence as *cchobo*. In the same work, it was shown that this sequence has 99.7% similarity with the *D. melanogaster* canonical *hobo* sequence, suggesting a horizontal transfer between those two species. The *hobo* phylogenetic tree obtained in the present project reinforces this theory, since the two *Drosophila* have more divergent *hobo* sequences.

The *BuT2* fragment search revealed one full-length copy of the *D. buzzatii* canonical element. All the other copies were restricted to the 1-250bp sequence of the query, implying a high conservation of the region. Rossatto et al. (2014) suggests that *BuT2* has an active transposase in *D. buzzatii*, but the element in *D. willistoni* contains a nonsense mutation resulting

in a stop codon in the species. Nevertheless, over 60% of the copies obtained were present in euchromatin, implying a potential activity of those conserved elements.

A putative full-length *Mar* element from *D. tropicalis* was used as query in the BLASTn searches. Even though most results were fragments of approximately 100bp in the extremities, four copies with more than 1000bp were found. Thirty-one euchromatin copies were detected, all of them being fragments of conserved *TIRs* and *TSDs*. Deprá et al. (2012) suggests that most, if not all, of those elements are inactive, but their association with host genes imply a role of *Mar MITEs* in gene regulation and genome organization.

Those findings, although preliminary, bring valuable insights of what to expect and how to carry out further studies to better develop the understanding of TEs. *In situ* hybridization on polytene chromosomes of the *D. willistoni* is expected to confirm the results obtained in this analysis.

References:

Anxolabéhère D., Kidwell M. G., Periquet G. 1988. Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of *Drosophila melanogaster* by mobile *P* elements. Molecular Biology Evolution. 5: 252–269.

Ayarpadikannan S, Kim H-S. 2014. The Impact of Transposable Elements in Genome Evolution and Genetic Instability and Their Implications in Various Diseases. Genomics & Informatics. 12(3):98-104.

Bergman CM, Quesneville H. 2007. Discovering and detecting transposable elements in genome sequences, Briefings in Bioinformatics. 8(6):382–392

Blackman RK, Koehler MM, Grimaila R, Gelbart WM. 1989. Identification of a fully functional *hobo* transposable element and its use for germ-line transformation of *Drosophila*. EMBO Journal. 8:211–217

Blauth ML, Bruno RV, Abdelhay E, Valente VLS. 2011. Spatiotemporal transcription of the *P* element and the *412* retrotransposon during embryogenesis of *Drosophila melanogaster* and *D. willistoni*. Genetics and Molecular Biology. 34(4):707-710.

Brookman JJ, Toosy AT, Shashidhara LS, White RA. 1992. The *412* retrotransposon and the development of gonadal mesoderm in *Drosophila*. Development. 116:1185-1192

Bushman F. 2002. Lateral DNA transfer: mechanisms and consequences. Cold Spring Harbor Laboratory Press, New York, 448 pp.

Cáceres M, Puig M, Ruiz A. 2001. Molecular Characterization of Two Natural Hotspots in the *Drosophila buzzatii* Genome Induced by Transposon Insertions. Genome Research. 11(8):1353-1364.

Cizeron G, Lemeunier F, Loevenbruck C, Brehm A, Biémont C. 1998. The retrotransposable element *412* in *Drosophila* species. Molecular Biology Evolution. 15:1589-1599.

Deprá M, Valente VL, Margis R, Loreto EL. 2009. The *hobo* transposon and *hobo*-related elements are expressed as developmental genes in *Drosophila*. Gene. 448(1):57-63.

Deprá M, Ludwig A, Valente VL, Loreto EL. 2012. *Mar*, a MITE family of *hAT* transposons in *Drosophila*. Mobile DNA. 3:13.

Dobzhansky T. 1950. Genetics of Natural Populations. XIX. Origin of Heterosis through Natural Selection in Populations of *Drosophila pseudoobscura*. Genetics. 35(3):288-302.

Feschotte C, Pritham EJ. 2007. DNA transposons and the evolution of eukaryotic genomes. Annual Review of Genetics. 41:331–368.

Finnegan DJ. 1990. Transposable elements and DNA transposition in eukaryotes. Current Opinion in Cell Biology. 2(3):471-477.

Garcia C, Delprat A, Ruiz A, Valente VLS. 2015. Reassignment of *Drosophila willistoni* Genome Scaffolds to Chromossome II Arms. G3: Genes|Genomes|Genetics. 5(12):2559-2566.

Gonçalves JW, Valiati VH, Delprat A, Valente VLS, Ruiz A. 2014. Structural and sequence diversity of the transposon *Galileo* in the *Drosophila willistoni* genome. BMC Genomics. 15(1):792.

Guerreiro MPG. 2012. What makes transposable elements move in the *Drosophila* genome? Heredity. 108(5):461-468.

Huang CRL, Burns KH, Boeke JD. 2012. Active Transposition in Genomes. Annual review of genetics. 46:651-675.

Kaminker et al. 2002. The transposable elements of the *Drosophila melanogaster* euchromatin: A genomics perspective. Genome Biology. 3:1-20.

Kuang et al. 2009. Identification of miniature inverted-repeat transposable elements (*MITEs*) and biogenesis of their siRNAs in the Solanaceae: new functional implications for *MITEs*. Genome Research. 19:42-56.

Lerat E, Rizzon C, Biemont C. 2003. Sequence divergence within transposable element families in the *Drosophila melanogaster* genome. Genome Research. 13(8):1889-1896.

Montchamp-Moreau, C., S. Ronsseray, M. Jacques, M. Lehmann, and D. Anxolabéhère. 1993. Distribution and conservation of sequences homologous to the *1731* retrotransposon in *Drosophila*. Molecular Biology Evolution.10:791–803.

Muñoz-López M, García-Pérez JL. 2010. DNA Transposons: Nature and Applications in Genomics. Current Genomics. 11(2):115-128.

Ortiz MF, Loreto EL. 2008 Characterization of new *hAT* transposable elements in twelve *Drosophila* genomes. Genetica. 135:67-75.

Regner LP, Pereira MS, Alonso CE, Abdelhay E, Valente VL. 1996. Genomic distribution of *P* elements in *Drosophila willistoni* and a search for their relationship with chromosomal inversions. The Journal of heredity. 87(3):191-198.

Rohde C, Valente VLS. 2012. Three decades of studies on chromosomal polymorphism of *Drosophila willistoni* and description of fifty different rearrangements. Genetics and Molecular Biology. 35(4):966-979.

Rossatto DO, Ludwig A, Deprá M, Loreto ELS, Ruiz A, Valente VLS. 2014. *BuT2* is a Member of the Third Major Group of *hAT* Transposons and is Involved in Horizontal Transfer Events in the genus *Drosophila*. Genome Biology and Evolution. 6(2):352-365

Sassi AK, Herédia F, Loreto ELS, Valente VLS, Rohde C. 2005. Transposable elements *P* and *gypsy* in natural populations of *Drosophila willistoni*. Genetics and Molecular Biology. 28:734-739.

Wicker et al. 2007. A unified classification system for eukaryotic transposable elements. Nature Reviews. Genetics. 8(12):973-982.

Table 1. Summary of the results.

Figure 1. *412* TE query, compared to longest and shortest copies found in Flybase.

Figure 2. Evolutionary relationships of taxa for *412* TE.

The evolutionary history was inferred using the Neighbor-Joining method [Saitou & Nei, 1987]. The optimal tree with the sum of branch length = 1.20688542 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (10000 replicates) are shown next to the branches (Felsenstein, 1985). (next to the branches). The evolutionary distances were computed using the Tamura 3-parameter method (Tamura, 1992) and are in the units of the number of base substitutions per site. The rate variation among sites was modeled with a gamma distribution (shape parameter = 1). The analysis involved 58 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 914 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar, et al., 2015).

Figure 3. *hobo* TE query, compared to longest and shortest copies found in Flybase

Figure 4. Molecular Phylogenetic analysis by Maximum Likelihood method for *hobo* TE.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model (1993). The tree with the highest log likelihood (-1803.5176) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 7 nucleotide sequences. Codon positions included

were 1st+2nd+3rd+Noncoding. There were a total of 697 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2015).

Figure 5. *BuT2* TE query, compared to longest and shortest copies found in Flybase

Figure 6. Evolutionary relationships of taxa for *BuT2* TE.

The evolutionary history was inferred using the Neighbor-Joining method [Saitou & Nei, 1987]. The optimal tree with the sum of branch length = 2.20634342 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (10000 replicates) are shown next to the branches (Felsenstein, 1985). (next to the branches). The evolutionary distances were computed using the Tamura 3-parameter method (Tamura, 1992) and are in the units of the number of base substitutions per site. The rate variation among sites was modeled with a gamma distribution (shape parameter = 1). The analysis involved 49 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 854 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar, et al., 2015).

Figure 7. *Mar* TE query, compared to longest and shortest copies found in Flybase

Figure 8. Molecular Phylogenetic analysis by Maximum Likelihood method for *Mar* TE.

The evolutionary history was inferred by using the Maximum Likelihood method based on the General Time Reversible model (Nei & Kumar, 2000). The tree with the highest log likelihood (-6899.5292) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the

Maximum Composite Likelihood (MCL) approach. A discrete Gamma distribution was used to model evolutionary rate differences among sites (2 categories (+*G*, parameter = 2)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site (next to the branches). The analysis involved 95 nucleotide sequences. There were a total of 2652 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2015).

References:

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution.4:406-425.

Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution.39:783-791.

Nei M, Kumar S. 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York.

Tamura K. 1992. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. Molecular Biology and Evolution. 9:678-687.

Tamura K, Nei M. 1993 Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution. 10:512-526

Kumar S, Stecher G, Tamura K. 2015. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular Biology and Evolution.

Table 1

Transposons	query base pairs	number of copies found Flybase + ncbi	biggest fragment (positives/total)	smallest fragment (positive/total)
412	885	223	637/789	53/57
hobo	655	4+1	411/434	61/63
BuT2	766	51	748/756	34/34
Mar	2487	175	1182 / 1339	36/36

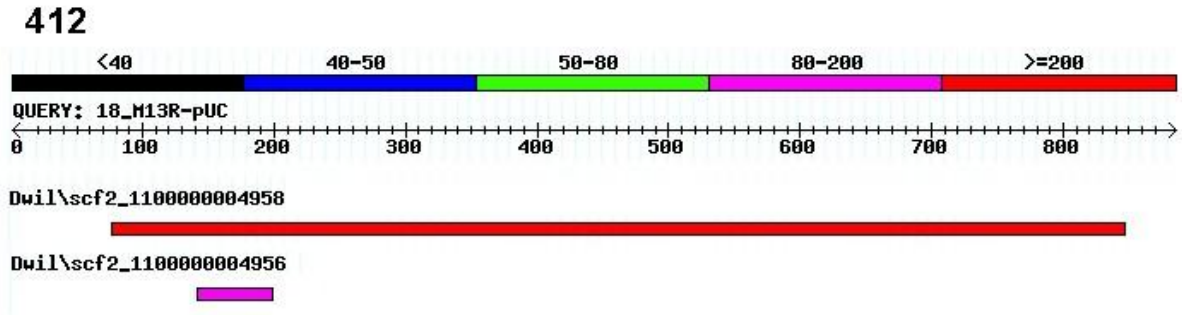
412
<40
40-50
50-80
80-200
>=200
QUERY: 18_M13R-pUC
0
100
200
300
400
500
600
700
800
Dwil\scf2_1100000004958
Dwil\scf2_1100000004956

Figure 1.

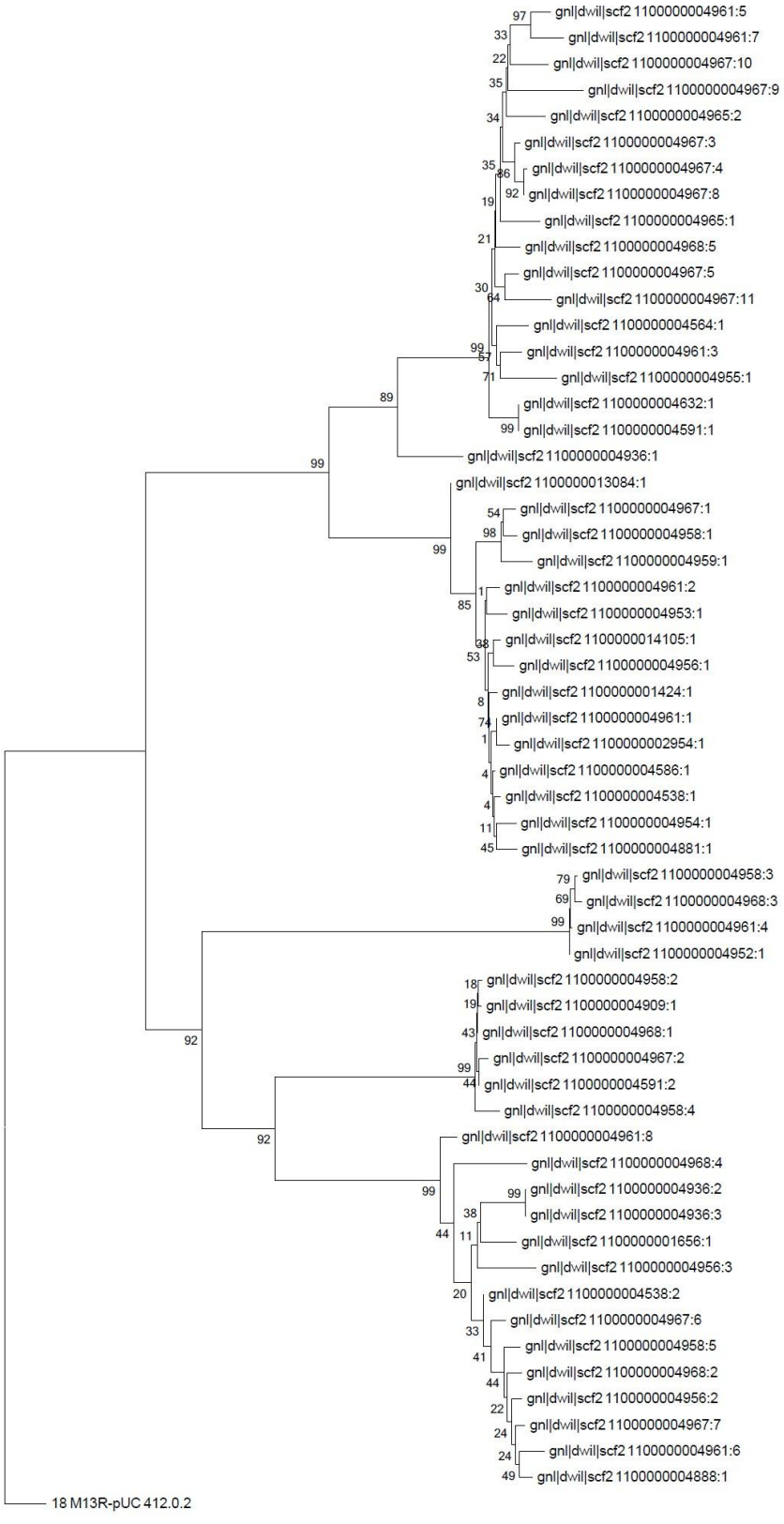
gnl|dwil|scf2 1100000004961:5
gnl|dwil|scf2 1100000004961:7
gnl|dwil|scf2 1100000004967:10
gnl|dwil|scf2 1100000004967:9
gnl|dwil|scf2 1100000004965:2
gnl|dwil|scf2 1100000004967:3
gnl|dwil|scf2 1100000004967:4
gnl|dwil|scf2 1100000004967:8
gnl|dwil|scf2 1100000004965:1
gnl|dwil|scf2 1100000004968:5
gnl|dwil|scf2 1100000004967:5
gnl|dwil|scf2 1100000004967:11
gnl|dwil|scf2 1100000004564:1
gnl|dwil|scf2 1100000004961:3
gnl|dwil|scf2 1100000004955:1
gnl|dwil|scf2 1100000004632:1
gnl|dwil|scf2 1100000004591:1
gnl|dwil|scf2 1100000004936:1
gnl|dwil|scf2 1100000013084:1
gnl|dwil|scf2 1100000004967:1
gnl|dwil|scf2 1100000004958:1
gnl|dwil|scf2 1100000004959:1
gnl|dwil|scf2 1100000004961:2
gnl|dwil|scf2 1100000004953:1
gnl|dwil|scf2 1100000014105:1
gnl|dwil|scf2 1100000004956:1
gnl|dwil|scf2 1100000001424:1
gnl|dwil|scf2 1100000004961:1
gnl|dwil|scf2 1100000002954:1
gnl|dwil|scf2 1100000004586:1
gnl|dwil|scf2 1100000004538:1
gnl|dwil|scf2 1100000004954:1
gnl|dwil|scf2 1100000004881:1
gnl|dwil|scf2 1100000004958:3
gnl|dwil|scf2 1100000004968:3
gnl|dwil|scf2 1100000004961:4
gnl|dwil|scf2 1100000004952:1
gnl|dwil|scf2 1100000004958:2
gnl|dwil|scf2 1100000004909:1
gnl|dwil|scf2 1100000004968:1
gnl|dwil|scf2 1100000004967:2
gnl|dwil|scf2 1100000004591:2
gnl|dwil|scf2 1100000004958:4
gnl|dwil|scf2 1100000004961:8
gnl|dwil|scf2 1100000004968:4
gnl|dwil|scf2 1100000004936:2
gnl|dwil|scf2 1100000004936:3
gnl|dwil|scf2 1100000001656:1
gnl|dwil|scf2 1100000004956:3
gnl|dwil|scf2 1100000004538:2
gnl|dwil|scf2 1100000004967:6
gnl|dwil|scf2 1100000004958:5
gnl|dwil|scf2 1100000004968:2
gnl|dwil|scf2 1100000004956:2
gnl|dwil|scf2 1100000004967:7
gnl|dwil|scf2 1100000004961:6
gnl|dwil|scf2 1100000004888:1
18 M13R-pUC 412.0.2
0.05

Fig.2

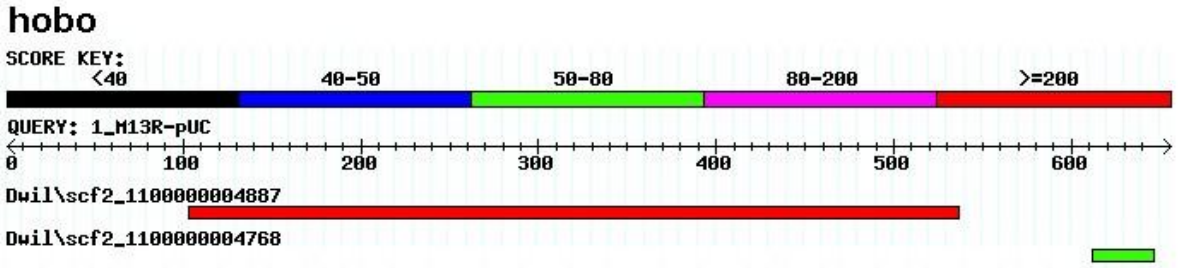
hobo
SCORE KEY:
<40
40-50
50-80
80-200
>=200
QUERY: 1_M13R-pUC
0
100
200
300
400
500
600
Dwil\scf2_1100000004887
Dwil\scf2_1100000004768

Figure 3.

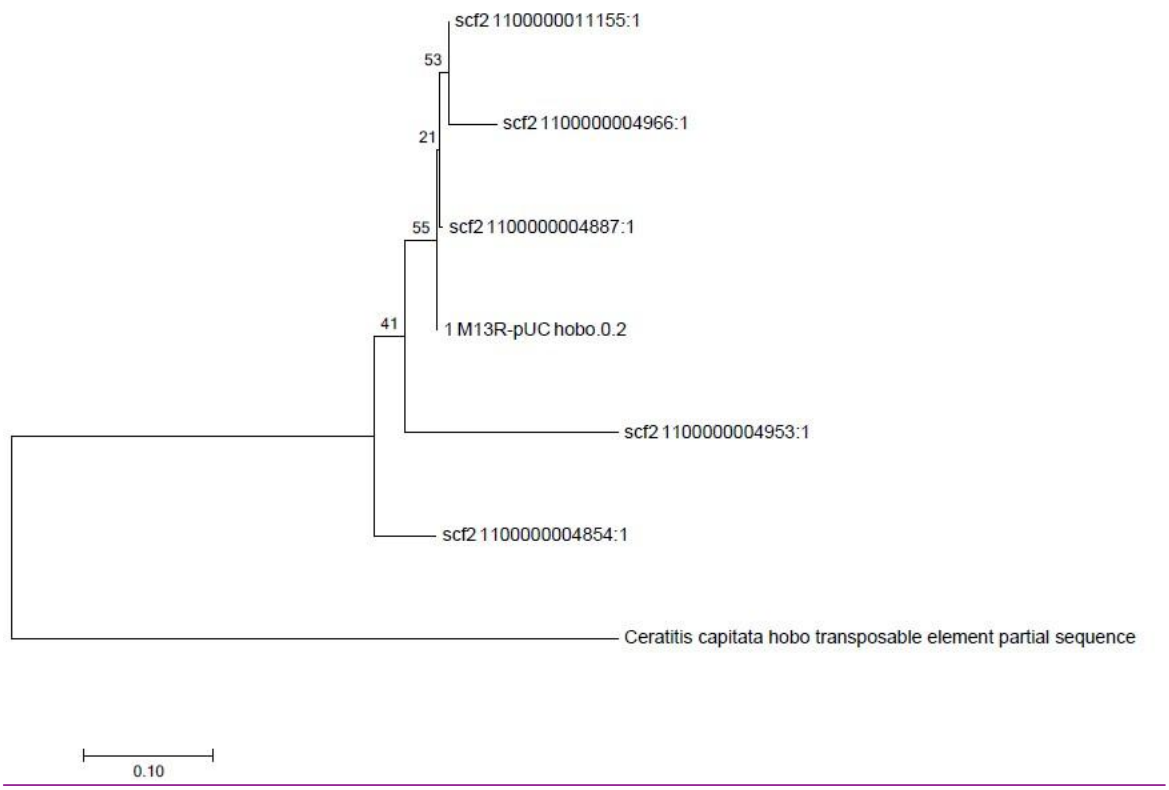
scf2 1100000011155:1
53
scf2 1100000004966:1
21
55
scf2 1100000004887:1
41
1 M13R-pUC hobo.0.2
scf2 1100000004953:1
scf2 1100000004854:1
Ceratitis capitata hobo transposable element partial sequence
0.10

Figure 4

Figure 5.

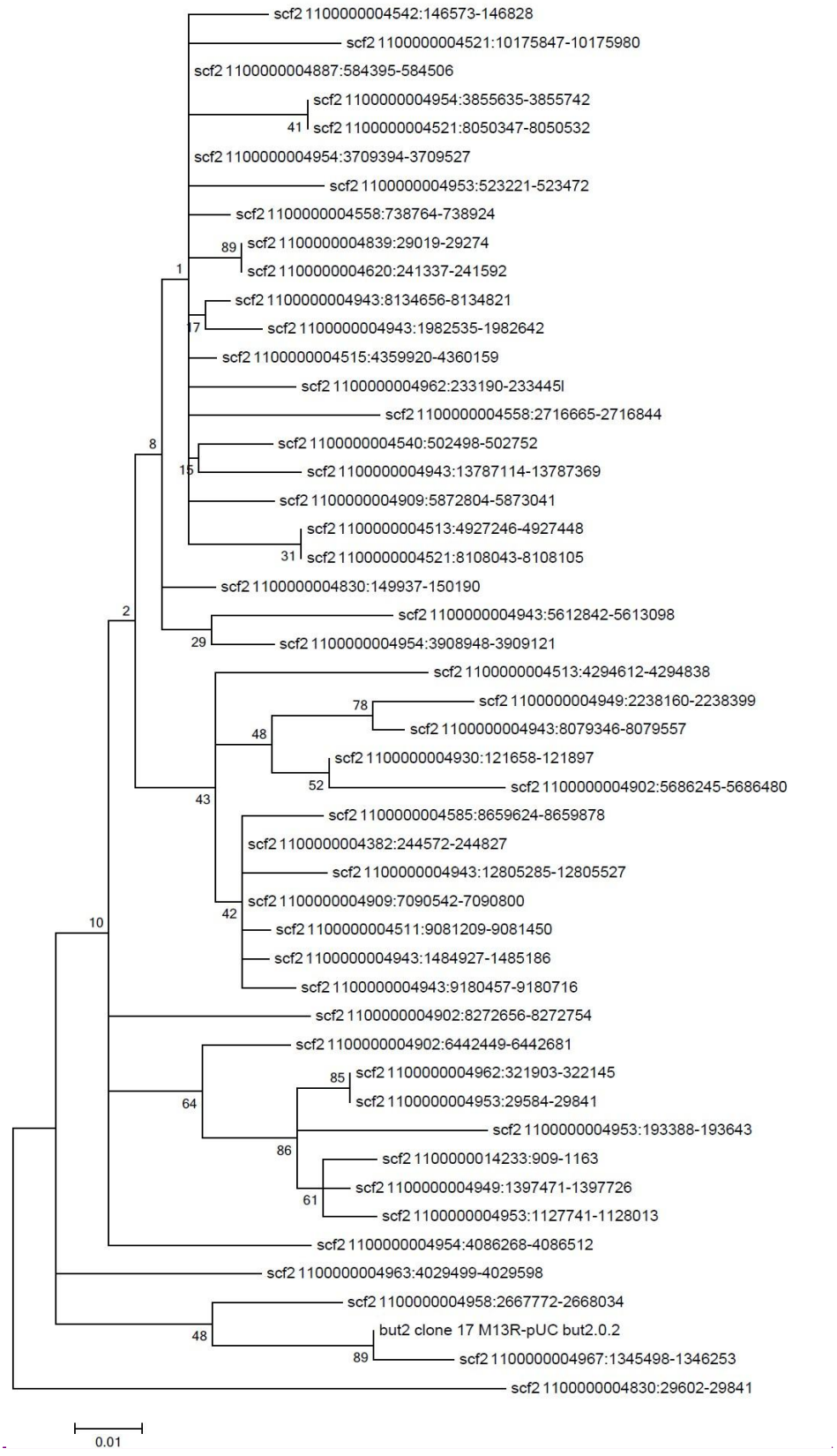
scf2 1100000004542:146573-146828
scf2 1100000004521:10175847-10175980
scf2 1100000004887:584395-584506
scf2 1100000004954:3855635-3855742
41
scf2 1100000004521:8050347-8050532
scf2 1100000004954:3709394-3709527
scf2 1100000004953:523221-523472
scf2 1100000004558:738764-738924
89
scf2 1100000004839:29019-29274
1
scf2 1100000004620:241337-241592
scf2 1100000004943:8134656-8134821
17
scf2 1100000004943:1982535-1982642
scf2 1100000004515:4359920-4360159
scf2 1100000004962:233190-233445l
scf2 1100000004558:2716665-2716844
8
scf2 1100000004540:502498-502752
15
scf2 1100000004943:13787114-13787369
scf2 1100000004909:5872804-5873041
scf2 1100000004513:4927246-4927448
31
scf2 1100000004521:8108043-8108105
scf2 1100000004830:149937-150190
2
scf2 1100000004943:5612842-5613098
29
scf2 1100000004954:3908948-3909121
scf2 1100000004513:4294612-4294838
78
scf2 1100000004949:2238160-2238399
48
scf2 1100000004943:8079346-8079557
scf2 1100000004930:121658-121897
52
scf2 1100000004902:5686245-5686480
43
scf2 1100000004585:8659624-8659878
scf2 1100000004382:244572-244827
scf2 1100000004943:12805285-12805527
scf2 1100000004909:7090542-7090800
42
10
scf2 1100000004511:9081209-9081450
scf2 1100000004943:1484927-1485186
scf2 1100000004943:9180457-9180716
scf2 1100000004902:8272656-8272754
scf2 1100000004902:6442449-6442681
85
scf2 1100000004962:321903-322145
64
scf2 1100000004953:29584-29841
scf2 1100000004953:193388-193643
86
scf2 1100000014233:909-1163
scf2 1100000004949:1397471-1397726
61
scf2 1100000004953:1127741-1128013
scf2 1100000004954:4086268-4086512
scf2 1100000004963:4029499-4029598
scf2 1100000004958:2667772-2668034
48
but2 clone 17 M13R-pUC but2.0.2
89
scf2 1100000004967:1345498-1346253
scf2 1100000004830:29602-29841
0.01

Figure 6

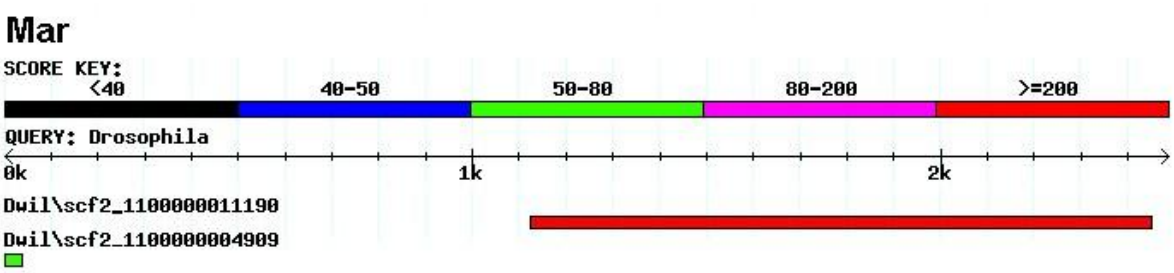
Mar
SCORE KEY:
<40
40-50
50-80
80-200
>=200
QUERY: Drosophila
0k
1k
2k
Dwil\scf2_1100000011190
Dwil\scf2_1100000004909

Figure 7.

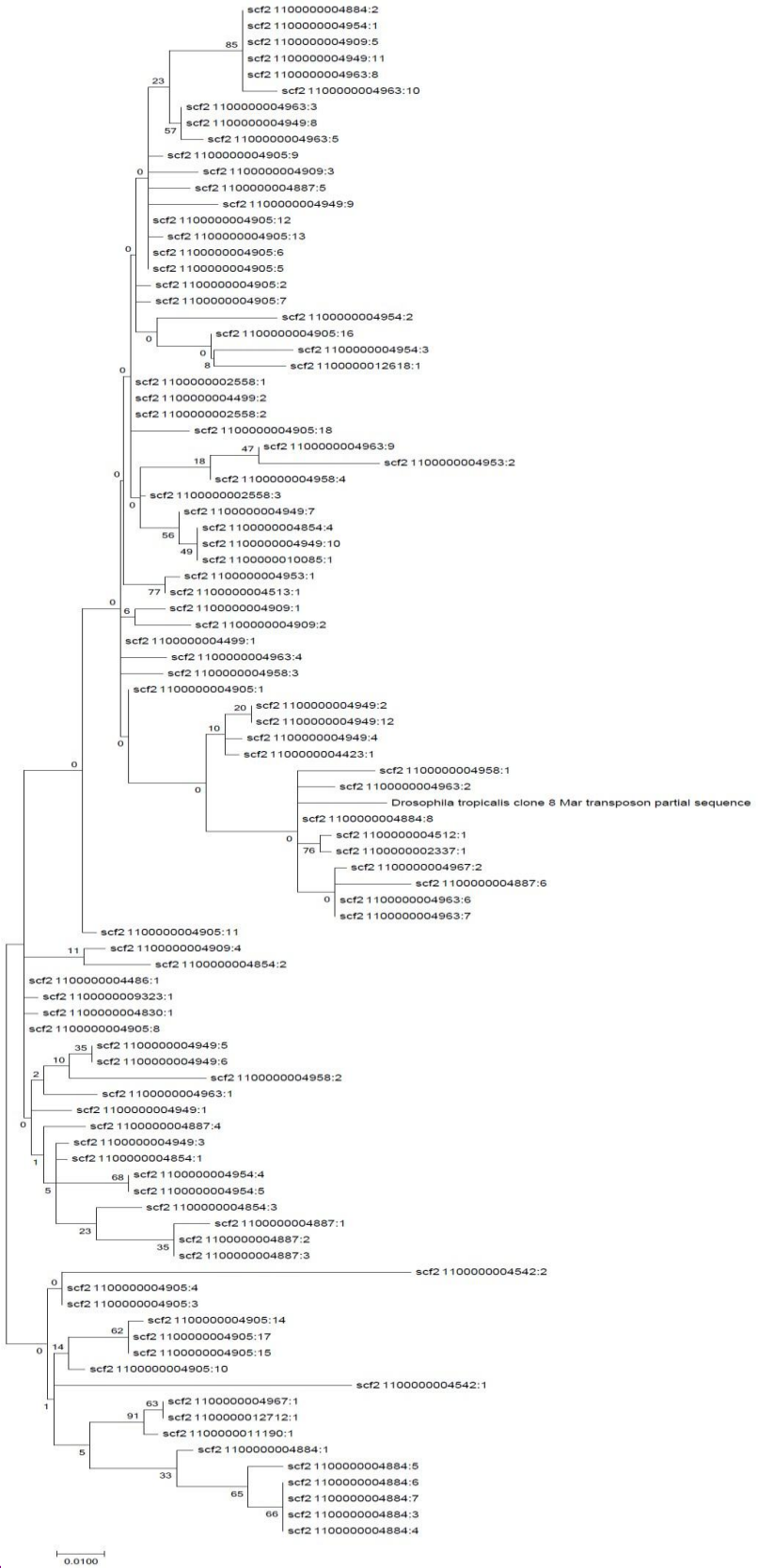
scf2 1100000004884:2
scf2 1100000004954:1
scf2 1100000004909:5
scf2 1100000004949:11
scf2 1100000004963:8
scf2 1100000004963:10
scf2 1100000004963:3
scf2 1100000004949:8
scf2 1100000004963:5
scf2 1100000004905:9
scf2 1100000004909:3
scf2 1100000004887:5
scf2 1100000004949:9
scf2 1100000004905:12
scf2 1100000004905:13
scf2 1100000004905:6
scf2 1100000004905:5
scf2 1100000004905:2
scf2 1100000004905:7
scf2 1100000004954:2
scf2 1100000004905:16
scf2 1100000004954:3
scf2 1100000012618:1
scf2 1100000002558:1
scf2 1100000004499:2
scf2 1100000002558:2
scf2 1100000004905:18
scf2 1100000004963:9
scf2 1100000004953:2
scf2 1100000004958:4
scf2 1100000002558:3
scf2 1100000004949:7
scf2 1100000004854:4
scf2 1100000004949:10
scf2 1100000010085:1
scf2 1100000004953:1
scf2 1100000004513:1
scf2 1100000004909:1
scf2 1100000004909:2
scf2 1100000004499:1
scf2 1100000004963:4
scf2 1100000004958:3
scf2 1100000004905:1
scf2 1100000004949:2
scf2 1100000004949:12
scf2 1100000004949:4
scf2 1100000004423:1
scf2 1100000004958:1
scf2 1100000004963:2
Drosophila tropicalis clone 8 Mar transposon partial sequence
scf2 1100000004884:8
scf2 1100000004512:1
scf2 1100000002337:1
scf2 1100000004967:2
scf2 1100000004887:6
scf2 1100000004963:6
scf2 1100000004963:7
scf2 1100000004905:11
scf2 1100000004909:4
scf2 1100000004854:2
scf2 1100000004486:1
scf2 1100000009323:1
scf2 1100000004830:1
scf2 1100000004905:8
scf2 1100000004949:5
scf2 1100000004949:6
scf2 1100000004958:2
scf2 1100000004963:1
scf2 1100000004949:1
scf2 1100000004887:4
scf2 1100000004949:3
scf2 1100000004854:1
scf2 1100000004954:4
scf2 1100000004954:5
scf2 1100000004854:3
scf2 1100000004887:1
scf2 1100000004887:2
scf2 1100000004887:3
scf2 1100000004542:2
scf2 1100000004905:4
scf2 1100000004905:3
scf2 1100000004905:14
scf2 1100000004905:17
scf2 1100000004905:15
scf2 1100000004905:10
scf2 1100000004542:1
scf2 1100000004967:1
scf2 1100000012712:1
scf2 1100000011190:1
scf2 1100000004884:1
scf2 1100000004884:5
scf2 1100000004884:6
scf2 1100000004884:7
scf2 1100000004884:3
scf2 1100000004884:4
0.0100

Figure 8