Analysis of Novel Unknown D. melanogaster *Mutation* trex

by Sallianne Roher

In this study, a mutation of unknown identity with suspected association with the *D. melanogaster* gene *vestigial* was examined in order to determine its characteristics, mode of inheritance, and molecular nature and function. Flies of this mutation, appropriately named *trex*, display wings smaller in size with a crumpled appearance. Wild type mutant crosses (WTM1 and WTM2) were performed first to determine if the *trex* phenotype is dominant or recessive by crossing wild-type females with *trex* males, and then to determine whether it is inherited in an autosomal or sex-linked manner. Mapping crosses were performed using three marker genes (*purple*, *black*, and *brown*) in order to estimate *trex's* location within the genome from the resulting recombination frequencies. The association of *trex* with the given marker genes by genetic linkage was analyzed using chi-square analysis and a p-value test of significance. After extracting DNA from wild type, *trex*, and reciprocal cross progeny flies, polymerase chain reactions and gel electrophoresis were performed in order to deduce characteristics of the mutation of *trex*. Further, using BLAST bioinformatic analyses, wild-type and *trex* gene sequences and protein products were compared to analyze the molecular nature of *trex*. *trex* was found to be inherited in an autosomal recessive pattern as evidenced by the results of WTM1 and WTM2, both of which produced an F1 generation that displayed only the wild-type presentation of wing morphology. From the mapping crosses' recombination frequencies and statistical analysis, it was estimated that *trex* is located at 2:68 within the genome. Analysis by PCR and gel electrophoresis revealed that the *trex* mutation is apparent as the addition of genomic material in comparison to the wild-type molecular presentation. These findings were further supported by BLAST analysis, which revealed that the *trex* mutation was due to the insertion of retrotransposon 412 within the coding sequence. BLAST also revealed *trex* to be allelic to *vestigial*, which interacts with a protein produced by *scalloped* (*sd*), completing a transcription factor complex regulating wing development. When mutated, as in *trex* flies, *vestigial* is unable to perform its normal function and we see phenotypically abnormal wing formation.

Introduction:

D. melanogaster has long been utilized as a model organism in genetic research for its low cost and quick life and reproductive cycles. The genome of *Drosophila* has been studied and mapped extensively, allowing scientists to further understand mechanisms of inheritance. By understanding the function of *vestigial* and other genes in the *D. melanogaster* genome, we may have further understanding of their human orthologs like the vestigial-like gene family (VGLL) and the medically relevant effects of their mutations. vestigial aids in the specification of wing cells and the proliferation of such cells, leading to the fully formed wing and halteres observed in wild-type *D. melanogaster* (Simmonds et al. 1997). *vestigial* associates using a TDU motif to a transcription factor produced by the gene *scalloped* (*sd*), which has a TEA/ATSS-DNA-binding domain (TEAD) (Yamaguchi 2020). As a transcription factor complex, the *vg-sd* complex is then able to regulate the binding of RNA polymerase and thus the subsequent translation of genes pertaining to the specification of wing cells and their proliferation (Yamaguchi 2020). For example, the expression of *D. melanogaster* genes *cut* and *Serum Response Factor* (*SRF*) are activated by the *vg-sd* complex and then play integral roles in the development of normal wing morphology (Figure 1A) (Halter et al.1998). A mutation in the *vestigial* gene leads to abnormal wing/haltere development, as the DNA required for proper wing development is not expressed (Williams et al. 1990). As such, these flies have phenotypically smaller wings that have a crumpled appearance and lie perpendicular to the anteroposterior axis of the fly rather than lying flat.

In this study, two strains of *Drosophila melanogaster* were studied, one wild-type strain and a strain of an unidentified mutation with similar morphology to a *vestigial* mutation, in order to understand and explore the nature and characteristics of the given mutation. The mutant strain was given the name "*trex*" on account of the small and protruding shape of the wings, similar to the short, protruding arms of a T. rex (Figure 2B). Further, it was hypothesized that the *trex* mutation was associated with the *vestigial* gene because of the *vestigial*-like presentation of *trex* flies' phenotype. Each strain of flies was cultured, and their phenotypic differences were observed and compared. Reciprocal crosses were then performed between the strains to determine the mode of inheritance of the *trex* mutation. Mapping crosses were also performed to determine the chromosomal location of the *trex* gene. The molecular nature of the *trex* mutation was analyzed after extracting DNA from *trex*, wild-type, and reciprocal cross progeny and performing polymerase chain reaction on the samples to isolate the DNA fragments of interest. Samples were then subject to gel electrophoresis and comparison via BLAST analysis in order to determine the nature of the *trex* mutation on a molecular level and how it might lead to the mutant phenotype and its connection with possible human orthologs.

Materials and Methods:

a. Overview

True breeding strains of *trex* and wild-type (WT) *Drosophila* flies were raised and manipulated in controlled conditions, and then used in various crosses. At all stages of development, the flies were observed in order to compare the phenotypic differences between wild-type and *trex* flies. Crosses between the two strains were performed in order to determine the mode of inheritance of *trex* and the location of the gene within the *D. melanogaster* genome. The offspring of such crosses were analyzed using phenotype scoring and statistical analysis in order to reach statistical and qualitative conclusions about *trex*. DNA was extracted from WT flies, *trex* flies, and reciprocal cross progeny and then utilized in polymerase chain reactions and subsequent gel electrophoresis. *trex* DNA was then sequenced and subjected to bioinformatic analysis using nucleotide and protein BLAST comparison to the wild-type genome. All fly stocks were maintained within vials containing a cornmeal-based food source with anti-microbial factors along with a small amount of yeast for its added nutritive properties. Vials were kept within a 25-26°C incubator for the duration of the study except when stocks were being manipulated. Any manipulation of fly stocks occurred under sterile conditions, utilizing a CO_2 anesthetizing system and a stereomicroscope with an illuminator.

b. Reciprocal crosses performed to determine the mode of inheritance of trex

First, a wild-type marker cross (WTM1) was performed between wild-type females and *trex* males to determine the mode of inheritance of *trex* as it relates to its dominance or lack thereof. Five females and three males were subcultured together into a vial with cornmeal food and incubated at 25°C. Upon pupa formation, the parental flies were brooded into new vials. Hatched progeny were phenotypically scored every 8-10 hours and then removed from the sample in order to prevent the formation of an F2 generation. A virtual version of this cross between wild-type females and *trex* males was also performed on FlyLab JS with three batches of parental flies. Progeny of the virtual WTM1 crosses were also phenotypically scored.

A second wild-type marker cross was performed (WTM2) using the same method as WTM1, except with *trex* females and wild-type males in order to determine the mode of inheritance as it relates to whether *trex* is inherited in an autosomal or sex-linked manner. Five females and three males were similarly subcultured into a vial with a cornmeal food source, incubated at 25°C, and were brooded into a new vial upon the formation of pupa within the original vial. The phenotypes of the progeny of this cross were scored every 8-10 hours and also removed from the sample in order to prevent the formation of an F2 generation. WTM2 was also reproduced virtually on FlyLab JS, similarly crossing three samples of *trex* females and wild-type males and phenotypically scoring progeny.

c. Mapping crosses performed to determine chromosomal location of trex

Three mapping crosses were performed between the *trex* gene and a marker gene (*black* (*bl*), *brown* (*bw*), and *purple* (*pr*)) in order to determine the

location of *trex* within the *D. melanogaster* genome. Female flies heterozygous for both the *trex* gene and the given marker gene were crossed with male flies homozygous for both *trex* and the marker gene. Each mapping cross was repeated three times with three separate cultures. Progeny of this cross were scored according to what phenotypic class they belonged to in order to determine the relative amounts of progeny with parental and recombinant phenotypes. The recombination frequency between *trex* and each marker gene was found using the equation $RF = \frac{recombinant\ progeny}{total\ progeny}$.

d. Chi-square statistical analysis of significance of genetic linkage of trex *and marker genes*

Statistical analysis of these progeny ratios was performed utilizing R statistical software to determine if the marker gene and *trex* were genetically linked. More specifically, a Chi-square test was performed utilizing the equation $X^2 = \Sigma(\frac{(Observed-Expected)^2}{Expected})$.

The value of this analysis was used for a p-value test of significance to evaluate the null hypothesis that the given marker gene and *trex* are not genetically linked and thus progeny display a 1:1:1:1 ratio of the phenotypic classes. Three degrees of freedom were utilized in this analysis, giving a corresponding critical value of 7.815 when p=0.05. Any Chi-square calculations giving a value larger than 7.815 were considered statistically significant and the null hypothesis was rejected with 95% confidence. In other words, for Chi-square evaluations giving a value greater than 7.815, the null hypothesis that the marker gene and *trex* are not genetically linked was rejected with 95% confidence, and genetic linkage between the genes was reasonably assumed. Based on genetic distances from statistically analyzed recombination frequencies, a prospective gene map was constructed to predict the location of *trex*.

e. DNA extraction

To begin the analysis of the molecular nature of the *trex* mutation, DNA was first extracted from wild-type flies, *trex* flies, and progeny from the reciprocal crosses (F1 generation flies from WTM1/2) so that it could be subjected to subsequent PCR and gel electrophoresis. Flies were macerated in a solution of 50mM Tris-HCl and 10mM ethylenediamine tetra acetic acid (EDTA) in order to disrupt cell membranes physically and chemically. Maceration physically allowed for the cells to be broken open, and EDTA served as a detergent to dissolve membrane lipids while Tris-HCl acted as a buffer. 20mM NaOH, which denatures DNA, and 1% SDS, which dissolves lipids and denatures proteins, were then added to the sample. This step was vital to disrupting the histones and other molecules which are bound to and surrounding DNA. 3M potassium acetate was added to precipitate excess material like lipids and proteins, which were then removed via centrifugation and extraction. The isolated solution of DNA was then treated with cold 100% isopropanol followed by another round of centrifugation to precipitate the DNA into a pellet form.

f. Polymerase chain reaction

The polymerase chain reaction was used to amplify the subsection of DNA where the *trex* gene is located in order to then analyze it via gel electrophoresis. A PCR master mix containing *Taq* polymerase, dNTPs, MgCl2 and buffer (containing Tris-HCl and potassium chloride) was used. *Taq* polymerase was utilized to construct complementary strands of DNA from the template strands, chosen especially for its ability to withstand the high temperatures of PCR thermocycling, as it comes from the thermostable archaebacterium *Thermus aquaticus*. dNTPs were necessary material for *Taq* polymerase to synthesize the new strands of DNA, while MgCl2 and buffer were used to stabilize the reaction and provide optimum conditions for the functionality of *Taq* polymerase. Forward and reverse primers were utilized to bind to the DNA segment of interest, providing a locus at which *Taq* polymerase could bind. These primer sequences were gactgcttggcagcaatgt and tccttggtttttgcagttcc, respectively. GAPDH primers, expressed constitutively in most cells, were also utilized to synthesize positive control PCR strands. Each sample for PCR contained equal volumes of solution composed of PCR master mix, a variable volume of DNA according to the concentration of the DNA sample, and a variable volume of sterile water to equalize PCR solution volume. For each DNA type (WT, *trex*, and WTM F1), a first sample was made containing forward and reverse mutant primers, a second sample was made containing GAPDH forward and reverse primers (positive control), and a third sample was made containing no

primers (negative control). From DNA extraction, WT DNA was isolated in a concentration of 0.1190 µg/µL, *trex* DNA was isolated in a concentration of 0.0565 µg/µL, and WTM F1 DNA was isolated in a concentration of 0.2193 µg/µL. As such, 2 µL of WT DNA, 3.5 µL of *trex* DNA, and 1 µL of WTM F1 were added to their respective samples so that there were nearly equal amounts of DNA in each PCR solution. Once the 9 respective samples of PCR solutions were prepared, they were run through PCR thermocycling of denaturation at 96°C, annealing at 57°C, and elongation at 72°C. This cycle was repeated about 30 times to exponentially amplify our target sequence.

g. Gel electrophoresis

The PCR amplified target sequences of DNA were visualized using agarose gel electrophoresis in order to make qualitative observations about the nature of the *trex* mutation and confirm the genetic identity of WTM F1. The electric field anode draws the negatively charged DNA through the gel, with smaller fragments traveling faster through the gel. A Tris-Acetate-EDTA (TAE) running buffer was used in the preparation of two 2% agarose gels in order to allow the flow of charge through the gel. Further, three dyes (xylene cyanole, bromophenol blue, and orange G) were used as indicators. A Ficoll loading buffer was utilized to improve the sedimentation of DNA samples within the gel. A control ladder and the 9 samples made by PCR were loaded into the 10 wells of the first agarose gel. TAs prepared additional *trex* and wild-type PCR products which were run through a second gel for the purpose of a standardized comparison. This second gel also contained a ladder and positive and negative controls. The prepared gels were subjected to 112-125 volts for 45 minutes.

h. Basic local alignment search tool (BLAST) analysis

The DNA fragments isolated and amplified by PCR were sequenced using a dideoxy chain termination method in which extracted DNA fragments were treated with dideoxynucleotides (ddNTPs) tagged with fluorescent labels of differing wavelength depending on the nitrogenous base identity of the ddNTP. ddNTPs do not contain a hydroxyl group and thus terminate the sequence at different lengths; then, the fluorescence of the different length chains is emitted and visualized using a program like FinchTV. Using FinchTV, the WT *D. melanogaster* genome from Flybase was screened for the *trex* primer sequences used in PCR in order to find the amplicon segment corresponding to where the *trex* mutation is housed. The WT sequence and *trex* sequence taken from FinchTV were then aligned and compared using a BLAST nucleotide analysis in order to determine the nature of the *trex* mutation on the genome level and to deduce what gene *trex* is allelic to within the *D. melanogaster* genome. Further, the *trex* mutation was investigated using a protein BLAST on Flybase. The resulting proteins of both the wild type and mutant were then compared by uploading both the WT and mutant coding sequences into Expasy, which configures a peptide sequence from a nucleotide sequence. A 3D model of the wild-type protein and its human ortholog was constructed and observed using Uniprot software.

Results:

a. Phenotypic presentation of WT vs. trex *flies*

The phenotypic differences between WT and *trex* flies were observed and recorded in order to characterize the nature of the mutation and how it affects the phenotype as compared to WT flies. WT flies' wings run parallel to the anteroposterior axis and lay flat along the dorsal side of the abdomen of the fly (Figure 1). In contrast, *trex* wings protrude perpendicularly to the anteroposterior axis and have a crumpled appearance as opposed to lying flat (Figure 1). Such differences in wing anatomy are visibly present from the emergence of newly-eclosed flies from the pupa and remain the same through the adulthood of the flies. There are no visible differences between WT and *trex* first, second, or third instar larva nor the pupa they form. Further, there is no notable difference in the life cycle rate of development between WT and *trex* flies. The visible phenotypic difference in wing appearance appears only upon eclosion from the pupa. Upon comparison of males and females, no difference was observed between the manifestation of the wing mutation between males and females of the WT and *trex* strains of flies. The appearance of the wing mutation of the *trex* flies was visibly identical between males and females.

b. Reciprocal crosses yielded progeny of all wild-type phenotype

The WTM1 cross was performed in order to determine whether the *trex* mutation is inherited

in a dominant or recessive inheritance pattern as it involved crossed true-breeding WT females and *trex* males. The chromosome map of the WTM1 cross can be seen in Figure 2. A total of 307 progeny were scored for the WTM1 cross. It was observed that all progeny of the WTM1 cross displayed a wild-type phenotype (Table 1). Among the 307 progeny that were scored, 163 females and 144 males in total were scored, all of which displayed wild-type wing morphology (Table 1). The virtual WTM1 cross also yielded progeny all displaying a wild-type phenotype. There was a total of 1,506 female and 1,514 male progenies, for a total of 3,020 progeny, all of which were wild type (Table 2).

The WTM2 cross was performed to determine the inheritance pattern of *trex* as it relates to whether it is inherited via an autosomal or sex chromosome. The chromosome map of this cross can be seen in Figure 3. This was made possible by the fact that true breeding female *trex* flies were crossed with true breeding wild-type males. By analyzing the results of this cross, more specifically as it relates to the sex of the progeny, the autosomal versus sex-linked nature of this mutation was made apparent. A total of 350 progeny, 183 of which were female and 167 of which were male, were scored for the WTM2 (Table 3). All progeny of the WTM2 cross displayed a wild-type phenotype (Table 3). The virtual WTM2 cross produced a total of 1,484 female and 1,489 male progeny, all of which (2,973 total) displayed the wild-type phenotype as well (Table 4).

c. Mapping crosses

The mapping crosses were performed to deduce the approximate location of the *trex* gene. The frequency of recombination between *trex* and the given marker gene was calculated in order to estimate the genetic distance between *trex* and the given marker gene, and the progeny were statistically analyzed using a Chi-square test and p-value test of significance. The null hypothesis of each cross was that the given marker gene and *trex* were not genetically linked and therefore assort independently. This hypothesis predicts a 1:1:1:1 ratio of each of 4 phenotypic classes, two parental phenotypic classes and two recombinant phenotypic classes.

In Mapping Cross A, females heterozygous for both *brown* (*bw*) and *trex* were crossed with males homozygous for both *bw* and *trex*. 2,902 progeny were scored, of which 1,489 were female and 1,413 were male. Of the progeny scored, 2,144 displayed a parental phenotype, while 758 displayed a recombinant phenotype (Table 5). The recombination frequency between *bw* and *trex* was calculated to be 26.1%. The calculated Chi-square value for this cross was found to be 663.6, which corresponds to a p-value of 2.2e-16 (Table 6).

In Mapping Cross B, female flies heterozygous for both *black* (*bl*) and *trex* were crossed with male flies homozygous for *bl* and *trex*. 1,519 female and 1,501 male flies were scored for this cross for a total of 3,020 scored flies (Table 7). 2,590 flies of the total displayed a parental phenotype, while only 430 displayed a recombinant phenotype (Table 7). The recombination frequency was calculated to be 14.2%, and the Chi-square value of this cross was 1,545.3, which gives a p-value of 2.2e-16 (Table 8).

Lastly, Mapping Cross C involved crossing female flies heterozygous for *purple* (*pr*) and *trex* with male flies homozygous *pr* and *trex*. A total of 3,024 progeny were collected and scored based on the phenotypic presentation, including 1,058 females and 1,516 males (Table 9). Of the total progeny, 2,694 displayed a phenotype of the parents and 330 displayed a recombinant phenotype, giving a recombination frequency of 10.9% (Table 9). The Chi-square value calculated for this cross was 1,850.4, leading to a p-value of 2.2e-16 (Table 10).

d. PCR and gel electrophoresis

DNA segments amplified using PCR thermocycling were viewed after being subjected to gel electrophoresis size-based separation of DNA. Gel electrophoresis allows for the separation of DNA fragments based on size, as the positively charged anode pulls the negatively charged DNA fragments toward it through the gel. Smaller fragments travel faster through the agarose gel and are visualized as a band that has traveled closer to the anode in the given amount of time. Larger fragments move slower through the gel and are visible as bands further from the anode. Under a UV viewing light, 3 of the 10 wells used for the first agarose gel in gel electrophoresis produced traveling bands through the agarose gel that were visible. The first well containing the ladder solution produced the appropriate ladder as expected. The third well containing a positive PCR control with GAPDH primers produced a band that

traveled to a location corresponding to about 100 base pairs (Figure 5). The fifth well containing our experimental sample of trex traveled to a location corresponding to about 1,000 base pairs (Figure 5). No other wells produced visible bands in the gel. The second gel prepared using TA samples did not produce any results. The negative control lanes in wells 8, 9, and 10 rightfully displayed no bands, as the samples contained no primers and thus the DNA contained was not amplified by PCR. However, it was given by TAs that wells 2, 3, and 4 with the positive control sample should have *all* produced bands of about 100 bp. Further, wells 6 and 7 containing wild-type and WTM F1 DNA, respectively, should have produced bands traveling significantly farther than the band produced from well 6 containing *trex* DNA.

e. BLAST analysis

The bioinformatic BLAST analysis of the *trex* mutation allowed for the understanding of *trex* on a molecular level by revealing the sequence and structure-level mutation which forms the basis of the *trex* phenotype. Upon screening the wild-type genome for the forward and reverse primer sequences in order to locate the wild-type amplicon, the reverse primer was not found within the wild-type genome. The nucleotide BLAST comparison of the *trex* sequence produced by dideoxy chain termination and the wild-type sequence starting with the reverse primer gave a 94% agreement in identity of 425 base pairs, as shown in Figure 6A. The comparison of *trex* and wild-type DNA sequences show an insertion of genetic material after the 425 base pair alignment (Figure 6C). The inserted DNA in the *trex* sequence was analyzed using a nucleotide BLAST on Flybase, aligning with over 25 *D. melanogaster* retrotransposons with an e-value of 0 (Figure 9A). The wild-type amplicon segment corresponding to *trex* (Figure 10) was analyzed using a protein BLAST and was found to be allelic to the *D. melanogaster* gene *vestigial* (*vg*) as hypothesized. Further, when the wild-type amino acid sequence retrieved from Expasy (Figure 10) was subjected to a protein BLAST specifically searching for human orthologs, it was found to have a high correspondence with the human gene *vestigial-like family member 2* (*VGLL-2*) (Figure 11).

Discussion:

a. trex *is inherited in an autosomal recessive pattern*

WTM1 and Virtual WTM1 produced only phenotypically wild-type progeny, which points toward a recessive pattern of inheritance of *trex*. Because the strains of flies that were crossed were true breeding, the results of this cross reveal which phenotype exhibits itself as dominant over the other. As none of the progeny of this cross displayed a *trex* phenotype, it was concluded that *trex* is recessive to a WT phenotype. On the other hand, the WTM2 cross and Virtual WTM2 crosses were done to determine whether *trex* is an autosomal or sex-linked mutation. Similar to WTM1, all progeny for both WTM2 and Virtual WTM2 displayed a wild-type phenotype. If *trex* was inherited in a sex-linked pattern, all males of the F1 generation would display a *trex* phenotype, as the mutation would be carried on the X chromosome passed to them by the female *trex* flies of the parental generation. However, no such *trex* male flies appeared in the F1 generation. As such, it can be concluded that *trex* is inherited via an autosomal chromosome within the *D. melanogaster* genome.

b. trex *is predicted to be located at 2:68 within the* Drosophila *genome*

Statistical analysis of all three mapping crosses (A, B and C) gave Chi-square values much larger than the critical value of 7.815. By these values the null hypothesis that *trex* and each marker gene are not linked genetically is rejected, as the results show a significant deviance from a 1:1:1:1 phenotypic ratio. Based on this rejection of the null hypothesis, it is feasible that *trex* and the given marker genes are, in fact, genetically linked and can be reasonably compared in order to predict a location of *trex* within the *Drosophila* genome. The recombination frequencies may be arranged to predict the location of *trex* within chromosome 2 of the *D. melanogaster* genome. Based on the information that black is 14 centiMorgans from *trex*, purple is 11 centiMorgans from *trex*, and brown was found to be approximately 26 centiMorgans from *trex*, it was predicted that *trex* is approximately located at 2:68 (Figure 2).

c. trex *contains an insertion of genetic material*

The gel electrophoresis results of a 1000 bp band of *trex* DNA as compared to a positive control band of 100 base pairs do not lead to conclusive results about

the nature of the *trex* mutation. However, as given by TA information, the wild-type and reciprocal cross progeny DNA samples should have traveled much farther than the *trex* band of DNA. Because it traveled a shorter distance, it can be concluded that *trex* is much larger than the WT sequence because it was pulled slower through the agarose gel as opposed to the WT band. As a larger DNA fragment, *trex* must then contain extra inserted DNA as compared to the WT sequence.

d. trex *contains retrotransposon 412 and is allelic to* vestigial

When screening the *D. melanogaster* genome primer sequences in order to determine where the wild-type amplicon was located, the reverse primer was not found, as indicated in the results. With the results of the gel electrophoresis in consideration, it was hypothesized that the reverse primer was located within an inserted sequence as part of the *trex* mutation. The hypothesis of an insertion mutation as the molecular basis of *trex* was upheld by the nucleotide BLAST results, which showed an insertion in the *trex* sequence after the 425 bp alignment of the sequences (Figure 6C). The results of the nucleotide BLAST on Flybase gave an alignment of the insertion sequence with retrotransposon 412. As such, it can be concluded that the *trex* mutation arises from a retrotransposon 412 insertion into the wild-type sequence. Further, the wild-type nucleotide sequence corresponding to the unknown mutant gene *trex* was translated into the amino acid sequence via Expasy (Figure 10) and was subjected to a protein BLAST analysis and was found to correspond to the *D. melanogaster* gene *vestigial*. It can be concluded then that *trex* is allelic to *vestigial*. Because of this transposon, which is partially located within the coding portion of the wild-type sequence, the normal structure of the protein produced by *vestigial* is disrupted. As such, it cannot perform its normal function as a co-transcription factor acting with *scalloped*, as described in the introduction. This malfunction is visibly apparent in the phenotypic presentation of *trex* flies. Retrotransposon 412 causes a malfunction of the vg-sd complex, such that it is unable to properly regulate wing/haltere genes, leading to malformed wing morphology.

f. trex *has a human ortholog vestigial-like protein 2*

As indicated in the results, as an allele of the *D. Melanogaster* gene *vestigial, trex* displays a high level of concordance with the human gene *VGLL-2*. *VGLL-2* can be concluded to be a human ortholog of *trex*. This gene is a member of the vestigial-like family of proteins discussed in the introduction, and as such, acts as a co-transcription factor to transcription factors with a TEAD-binding domain (Yamaguchi 2020). *VGLL-2* specifically may be important in regulating the distribution of skeletal muscle fibers in humans by regulating myocyte transcription factors with TEAD-binding domains (Honda et al. 2017). This ortholog allows us to understand how vestigial also functions regulating gene expression through its interactions with TEAD-binding domain-containing transcription factors through a TDU motif. The mutation of *VGLL-2* is associated with various disorders of the skeletal muscle and associated tissues, such as rhabdomyosarcomas (e.g., pleotropic rhabdomyosarcoma and spindle cell rhabdomyosarcoma) (Furlong et al. 2001). The lack of genetic regulation by *VGLL-2* leads to a dysregulation in gene expression in skeletal muscle development and subsequent malformations and irregular development of such tissues (Furlong et al. 2001).

g. Conclusions

Overall, it was found that *trex* is an allele of the well-researched *D. melanogaster* gene *vestigial*. *trex* is inherited in an autosomal recessive fashion and is located at 2:68 within the *D. melanogaster* genome. The *trex* phenotype, visible as small, crumpled wings that lie perpendicular to the anteroposterior axis of the fly, arises as a result of the insertion of retrotransposon 412. Otherwise, *trex* has a correspondence of about 425 base pairs with 94% agreement with the wild-type *D. melanogaster* nucleotide sequence. Further, as an allele of *vestigial*, *trex* has the human ortholog gene *VGLL-2*.

References

Baena-Lopez LA, Garcia-Bellido A. 2006. Control of growth and positional information by the graded vestigial expression pattern in the wing of Drosophila melanogaster. Proc. Natl. Acad. Sci. U.S.A. 103(37):13734-13739.

Furlong MA, Mentzel T, & Fanburg-Smith JC. 2001. Pleomorphic rhabdomyosarcoma in adults: a clinicopathologic study of 38 cases with emphasis on morphologic variants and recent skeletal muscle-specific markers. Modern Pathology: An Official Journal of the United States and Canadian Academy of Pathology, Inc. 14(6):595–603. https://doi.org/10.1038/modpathol.3880357

Halder G, Polaczyk P, Kraus ME, Hudson A, Kim J, Laughon A, Carroll S. 1998. The vestigial and scalloped proteins act together to directly regulate wing-specific gene expression in Drosophila. Genes Dev. 12(24):3900-3909.

Honda, M, Hidaka K, Fukada S, et al. Vestigial-like 2 contributes to normal muscle fiber type distribution in mice. 2017. Sci Rep 7:7168. https://doi.org/10.1038/s41598-017-07149-0

Simmonds A, Hughes S, Tse J, Cocquyt S, Bell, J. 1997. The effect of dominant vestigial alleles upon vestigial-mediated wing patterning during development of Drosophila melanogaster. Mechanisms of Development. 67(1):17–33.

Tolwinski NS. 2017. Introduction: Drosophila-A Model System for Developmental Biology. Journal of Developmental Biology, 5(3):9. https://doi.org/10.3390/jdb5030009

Williams JA, Atkin AL, Bell JB. 1990. The functional organization of the vestigial locus in Drosophila melanogaster. Mol. Gen. Genet. 221(1):8-16.

Williams JA, Bell JB, Carroll SB. 1991. Control of Drosophila wing and haltere development by the nuclear vestigial gene product. Genes Dev. 5(12B):2481-2495.

Yamaguchi N. 2020. Multiple roles of vestigial-Like family members in tumor development. Frontiers in Oncology. 10:1266.

Figures and Tables

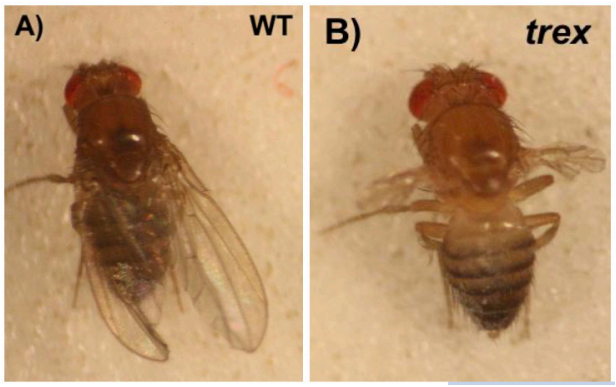

Figure 1. Dorsal comparison of female Drosophila melanogaster wing morphology in wild-type and trex mutant. [A] Depicts a female fly with a wild-type phenotype of wings of normal size and shape, flat in appearance, allowing view of vein arrangement and running parallel to the anteroposterior axis of the fly. Wings appear to lie mostly flat along the dorsal side of the abdomen of the fly. [B] Depicts a *trex* mutant female fly displaying wings that protrude perpendicular to the anteroposterior axis of the fly. Wings are shorter and crumpled in appearance, so that wing veins are not easily discernible.

Wild Type Marker Cross 1 (WTM1)				
Female parental genotype: +/+ Male parental genotype: *trex/trex*				
Class	Phenotype	Female	Male	Total
1	+/+	163	144	307
2	*trex/trex*	0	0	0
	Total	163	144	307

Table 1. WTM1 cross phenotype scoring. A total of 307 progeny were scored, including 163 females and 144 males. All progeny displayed a wild-type phenotype, indicating a recessive inheritance pattern of the *trex* mutation.

Virtual Wild Type Marker Cross 1 (WTM1)				
Female parental genotype: +/+ Male parental genotype: *trex/trex*				
Class	Phenotype	Female	Male	Total
1	+/+	511+498+497	504+493+517	3,020
2	*trex/trex*	0	0	0
	Total	1,506	1,514	3,020

Table 2. Virtual WTM1 cross phenotype scoring. A total of 3,020 progeny were scored, including 1,506 female and 1,514 male progenies. All progeny displayed a wild-type phenotypic presentation, indicating a recessive pattern of inheritance of *trex* in agreement with the actual WTM1 cross displayed in Table 1.

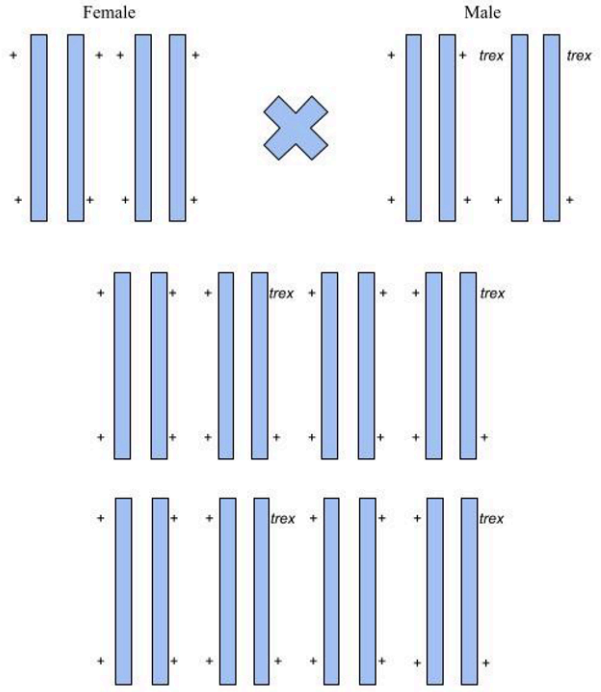

Figure 2. Chromosome map of Wild Type Marker Cross 1 (WTM1). The female parental fly displays a homozygous genotype for the wild-type genotype, while the male parental fly shows a homozygous genotype for the *trex* genotype. When crossed, all progenies display a wild-type phenotype, though progeny have inherited one *trex* allele from the male parental genome. As such, it can be concluded that *trex* is recessive to the wild-type phenotype.

Wild Type Marker Cross 2 (WTM2)				
Female parental genotype: *trex/trex* Male parental genotype: +/+				
Class	Phenotype	Female	Male	Total
1	+/+	183	167	350
2	*trex/trex*	0	0	0
	Total	183	167	350

Table 3. WTM2 cross phenotype scoring. A total of 350 progeny were scored, of which 183 were female and 167 were male. All progenies were of the wild-type class, including all male offspring, indicated an autosomal inheritance pattern of the *trex* mutation.

Virtual Wild Type Marker Cross 2 (WTM2)				
Female parental genotype: *trex/trex*				
Male parental genotype: +/+				
Class	Phenotype	Female	Male	Total
1	+/+	494+488+502	473+513+503	2,973
2	*trex/trex*	0	0	0
	Total	1,484	1,489	2,973

Table 4. Virtual WTM2 cross phenotype scoring. A total of 2,973 progeny were scored, including 1,484 female and 1,489 male progenies. All progeny displayed a wild-type phenotypic presentation, indicating a recessive pattern of inheritance of *trex* in agreement with the actual WTM2 cross displayed in Table 3.

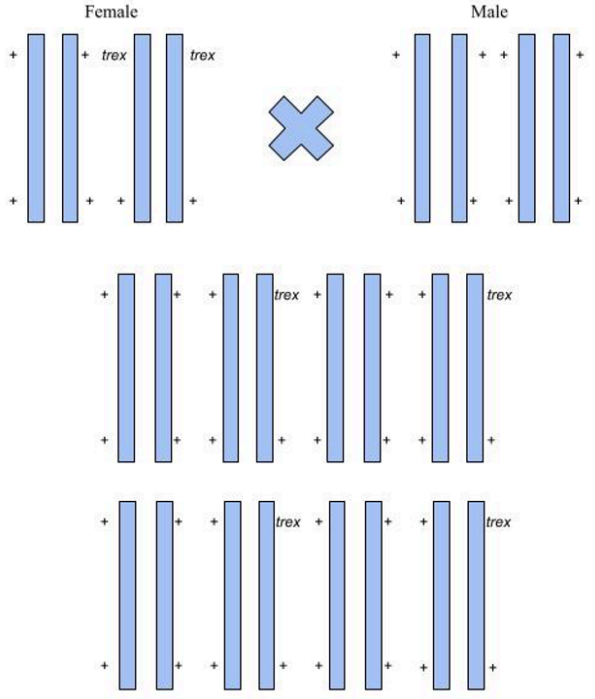

Figure 3. Chromosome map of Wild Type Marker Cross 2 (WTM2). The female parental fly displays a homozygous genotype for the *trex* genotype while the male parental fly shows a homozygous genotype for the wild-type genotype. When crossed, all progenies display a wild-type phenotype, though progeny have inherited one *trex* allele from the female parental genome. As such, it can be concluded that *trex* is inherited in autosomal fashion. If *trex* was a sex-linked mutation, all male progeny would display a *trex* phenotype, as they would all carry the *trex* gene on their only X chromosome inherited from the female parental fly.

Mapping Cross A			
Female parental genotypes: +/*bw*, *trex*/+			
Male parental genotype: *bw/bw*, *trex/trex*			
Class	Phenotype	Female	Male
1	+/+	185+175+204	159+162+167
2	*bw*/+	62+80+63	69+69+51
3	+/*trex*	62+76+46	58+63+59
4	*bw/trex*	175+176+185	176+187+193
	Total:	1,489	1,413

Table 5. Mapping cross A phenotypic classes. A total of 2,902 progeny were scored, 2,144 of which displayed a parental phenotype (+/+ or *bw/trex*) and 758 of which displayed a recombinant phenotype (*bw*/+ or +/*trex*), giving a recombination frequency of 26.1%.

Mapping Cross A Chi-Square Analysis						
Class	Phenotype	Observed	Expected	O - E	$(O - E)^2$	$(O - E)^2/E$
1	+/+	1,052	725.5	326.5	106,602.30	146.9
2	*bw*/+	394	725.5	-331.5	109,892.30	151.5
3	+/*trex*	364	725.5	-361.5	130,682.30	180.1
4	*bw/trex*	1,092	725.5	366.5	134,322.30	185.1
	Total	2,902	2,902			$X^2 = 663.6$

Table 6. Mapping cross A Chi-square analysis. Due to the null hypothesis that *trex* and *bw* are not genetically linked, a 1:1:1:1 ratio was utilized as the expected value of progeny. As such, a Chi-square value of 663.6 was calculated with the formula $X^2 = \sum(\frac{(Observed-Expected)^2}{Expected})$.

With 3 degrees of freedom, this Chi-square value coincides with a p-value of less than 2.2e-16 and thus the null hypothesis is rejected.

Mapping Cross B			
Female parental genotypes: +/*bl* , *trex* /+			
Male parental genotype: *bl/bl, trex/trex*			
Class	Phenotype	Female	Male
1	+/+	219+199+213	231+237+208
2	*trex*/+	45+38+41	31+31+32
3	+/*bl*	30+38+30	32+45+37
4	*trex/bl*	224+226+216	202+207+208
	Total:	1,519	1,501

Table 7. Mapping Cross B phenotypic classes. A total of 3,020 progeny were scored, 2,590 of which displayed a parental phenotype (+/+ or *trex/bl*) and 430 of which displayed a recombinant phenotype (*trex*/+ or +/*bl*), giving a recombination frequency of 14.2%.

Mapping Cross B Chi-Square Analysis						
Class	Phenotype	Observed	Expected	O - E	$(O - E)^2$	$(O - E)^2/E$
1	+/+	1,307	755	552	304,704	403.6
2	*trex*/+	218	755	-537	288,369	381.9
3	+/*bl*	212	755	-543	294,849	390.5
4	*trex*/*bl*	1,283	755	528	278,784	369.3
	Total	3,020	3,020			X^2 =1,545.3

Table 8. Mapping cross B Chi-square analysis. Due to the null hypothesis that *trex* and *bl* are not genetically linked, a 1:1:1:1 ratio was utilized as the expected value of progeny. As such, a Chi-square value of 1,545.3 was calculated with the formula $X^2 = \sum(\frac{(Observed-Expected)^2}{Expected})$.

With 3 degrees of freedom, this Chi-square value coincides with a p-value of less than 2.2e-16 and thus the null hypothesis is rejected.

Mapping Cross C			
Female parental genotypes: *pr*/+, +/*trex*			
Male parental genotype: *pr*/*pr*, *trex*/*trex*			
Class	Phenotype	Female	Male
1	+/+	238+239+236	218+216+227
2	*pr*/+	33+28+22	32+34+28
3	+/*trex*	29+29+13	21+27+34
4	*pr*/*trex*	239+193+209	227+220+232
	Total:	1,508	1,516

Table 9. Mapping Cross C phenotypic classes. A total of 3,024 progeny were scored, 2,694 of which displayed a parental phenotype (+/+ or *pr*/*trex*) and 330 of which displayed a recombinant phenotype (*pr*/+ or +/*trex*), giving a recombination frequency of 10.9%.

Mapping Cross C Chi-Square Analysis						
Class	Phenotype	Observed	Expected	O - E	$(O - E)^2$	$(O - E)^2/E$
1	+/+	1,374	756	618	381,924	505.2
2	*pr*/+	177	756	-579	335,241	443.4
3	+/*trex*	153	756	-603	363,609	481
4	*pr*/*trex*	1,320	756	564	318,096	420.8
	Total	3,024	3,024			X^2 =1,850.4

Table 10. Mapping cross C Chi-square analysis. Due to the null hypothesis that *trex* and *pr* are not genetically linked, a 1:1:1:1 ratio was utilized as the expected value of progeny. As such, a Chi-square value of 1,850.4 was calculated with the formula $X^2 = \sum(\frac{(Observed-Expected)^2}{Expected})$.

With 3 degrees of freedom, this Chi-square value coincides with a p-value of less than 2.2e-16 and thus the null hypothesis is rejected.

Gene	Location of gene	Predicted location of Unknown
bw	2-103	103-26 = 77
bl	2-49	49+14 = 63
pr	2-54	54+11 = 65

Average Location = (77+63+65)/3 = 68.3 so the predicted location is 2:68

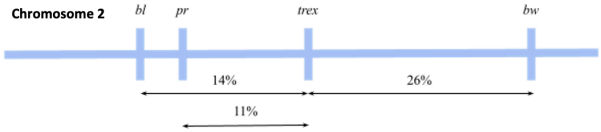

Figure 4. Chromosome map. Chromosome 2 of *D. melanogaster* is depicted with bl being located at 2-49 with an estimated genetic distance of 14 cM from *trex*, *pr* being located at 2-54 with an estimated genetic distance of 11 cM from *trex*, and bw being located at 2-103 with an estimated genetic distance of 26 cM from *trex*. Based on these locations and their estimated distances from *trex*, *trex* is predicted to be located at 2-68.

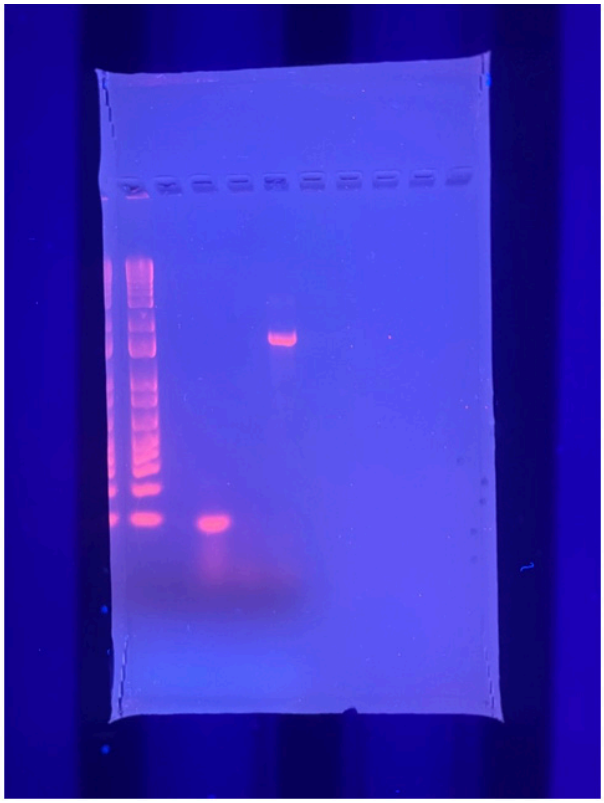

Figure 5. Gel electrophoresis under ultraviolet viewing light. The lane farthest left displays the control ladder allowing the standardization of bp size of genes corresponding to the distance that bands travel. The second band shown is associated with a positive control sample and has traveled a distance associated with a band size of 100 bp. The third band, farthest right, is associated with the sample with trex and has traveled a distance corresponding to 100 base pairs.

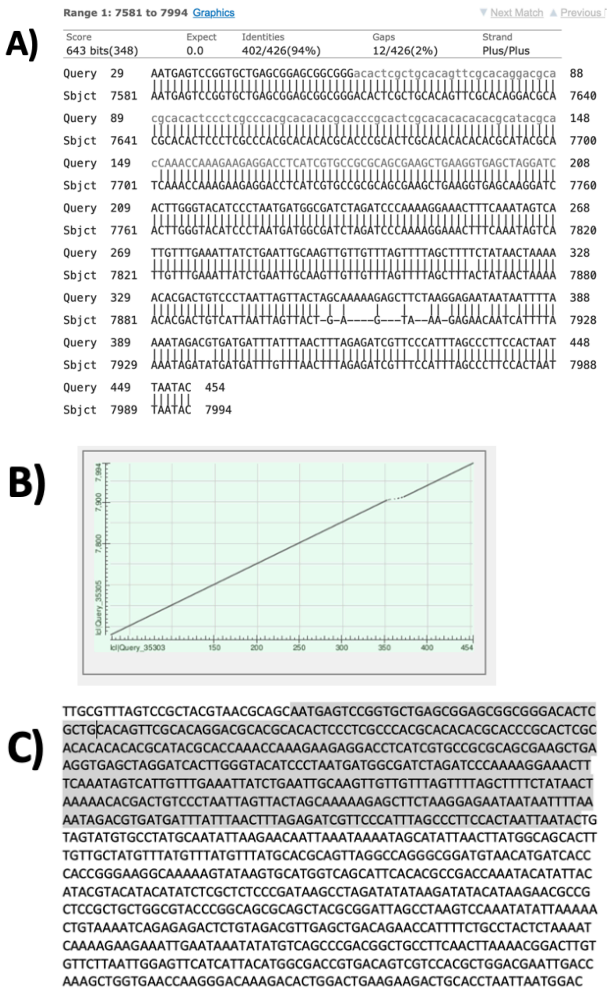

Figure 6. Comparison of trex and wild-type nucleotide sequences. [A] Displays the nucleotide BLAST alignment of the wild-type amplicon and trex sequences. The alignment includes about 425 base pairs and is in 94% agreement. This correspondence is illustrated in [B], which displays a dot plot graph of the alignment between the respective sequences. The sequence in [C] is the wild-type sequence retrieved from FinchTV after dideoxy chain-termination sequencing. The highlighted region shows the 425 bp sequence in agreement with the wild-type genome. The sequence that follows and is unhighlighted is an inserted sequence not found in the wild-type genome.

ATGGCAGTGTCCTGCCCCGAAGTTATGTACGGTGCCTACTATCCATATCTGTACGGGCGCGCTGGAACAAGTCGATCGTTCTACCA
GTATGAGAGGTTCAACCAGGATTTGTACTCCTCCTCGGGCGTCAACTTGGCCGCCTCATCGAGTGCCTCGGGCAGCTCCCCACTCGC
CCTGCAGCCCCATCCTGCCGCCATCGGTGAGCGCAAATGCCGCTGCGGCAGTGGCAGCGGCGGCCCACAATAGTGCAGCGGCGGCA
GTGGCGGTCGCAGCGAATCAGGCCAGCTCCTCCGGCGGAATCGGAGGCGGTGGACTCGGCGGCCTGGGAGGACTGGGCGGCGGACC
GGCGAGCGGACTGCTTGGCAGCAATGTGGTGCCCGGGAGCAGTAGTGTCGGGAGCGTGGGGCTGGGAATGAGTCCGGTGCTGAGCG
GAGCGGCGGGACACTCGCTGCACAGTTCGCACAGGACGCACGCACACTCCCTCGCCCACGCACACACGCACCCGCACTCGCACACA
CACACGCATACGCATCAAACCAAAGAAGAGGACCTCATCGTGCCGCGCAGCGAAGCTGAAGCACGCCTGGTGGGCTCCCAACAACA
TCAGCATCATAATGAATCATCCTGCTCCTCCGGTCCGGACTCCCCGCGCCACGCCCACTCGCACAGCCATCCGCTGCACGGTGGCG
GCGGAGCGACGGGCGGTCCATCTTCCGCGGGTGGTACGGGCTCGGGGGGCGGCGACGGAGGCGGCACGGGGGCAATACCCAAGAAT
CTGCCAGCACTGGAGACGCCGATGGGCAGCGGAGGCGGCGGATTGGCCGGCAGTGGCCAGGGACAGGCTCAATATCTATCGGCCTC
CTGCGTTGTTCACCAACTACTCGGGCGACACGGCCAGCCAGGTGGACGAGCACTTTTCCCGCGCCCTCAACTACAACAACAAGG
ATTCTAAAGAGAGCAGCAGCCCGATGTCGAATCGAAATTTCCCACCGTCGTTCTGGAACAGCAATTACGTGGCCCATACCCGCG
CCCACACCACCAGGTTAGCGACTTGTATGGCACCGGCGACGGACACCGGCTACGCCACCGATCCGTGGGTGCCGCATGCGGCCCA
CTATGGTTCCTATGCCCACGCAGCGCACGCCCATGCCGGCCCACGCCCACGCCTACCACCACAACATGGCCCAGTACGGCAGTCTCC
TCAGGCTGCCCCAGCAGTACGCCAGCCACGGTTCCAGGCTGCACCACGACCAGCAGACGGCCCACGCCCTGGAGTACTCCAGCTAT
CCCACAATGGCAGGCCTGGAAGCGCAGGTGGCGCAAGTACAGGAATCGTCGAAGGATCTTTACTGGTTCTAA

Figure 7. Coding sequence of WT. The coding sequence of the wild-type genome is included above. This sequence contains the mutant forward primer but does not contain the reverse primer. As the reverse primer was not found, it was hypothesized that the primer may be included in the *trex* mutation as an insertion into the wild-type genome.

TGTAGTATGTGCCTATGCAATATTAAGAACAATTAAATAAAATAGCATATTAACTTATGGCAGCACTTTG
TTGCTATGTTTATGTTTATGTTTATGCACGCAGTTAGGCCAGGGCGGATGTAACATGATCACCCACCGGG
AAGGCAAAAAGTATAAGTGCATGGTCAGCATTCACACGCCGACCAAATACATATTACATACGTACATAC
ATATCTCGCTCTCCCGATAAGCCTAGATATATAAGATATACATAAGAACGCCGCTCCGCTGCTGGCGTAC
CCGGCAGCGCAGCTACGCGGATTAGCCTAAGTCCAAATATATTAAAAACTGTAAAATCAGAGAGACTCT
GTAGACGTTGAGCTGACAGAACCATTTTCTGCCTACTCTAAAATCAAAAGAAGAAATTGAATAAATATAT
GTCAGCCCGACGGCTGCCTTCAACTTAAAACGGACTTGTGTTCTTAATTGGAGTTCATCATTACATGGCG
ACCGTGACAGTCGTCCACGCTGGACGAATTGACCAAAGCTGGTGAACCAAGGGACAAAGACACTGGAC
TGAAGAAGACTGCACCTAATTAATGGAC

Figure 8. Transposon sequence. The sequence included above corresponds to the sequence inserted into the *trex* fly genome after the 425 base pair alignment with the wild-type sequence. This is the sequence that was subjected to a nucleotide BLAST within Flybase.

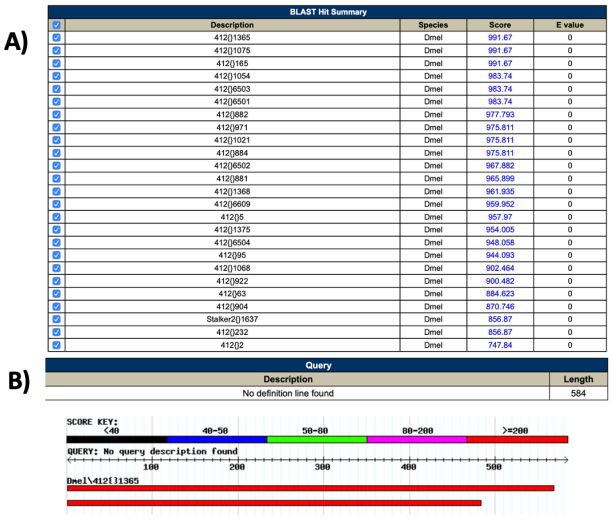

Figure 9. Nucleotide BLAST of transposon mutation sequence. [A] displays the results of the nucleotide BLAST results of the inserted sequence found in the sequenced *trex* DNA fragment. There are 25 transposon sequences that align with the query search with an e-value of 0. [B] displays the correspondence of the inserted sequence with retrotransposon 412 found across the *D. melanogaster* genome.

MAVSCPEVMYGAYYPYLYGRAGTSRSFYQYERFNQDLYSSSGVNLAASSSASGSSHSP
CSPILPPSVSANAAAAVAAAAHNSAAAAVAVAANQASSSGGIGGGGLGGLGGLGGGPA
SGLLGSNVVPGSSSVGSVGLGMSPVLSGAAGHSLHSSHRTHAHSLAHAHTHPHSHTHTH
THQTKEEDLIVPRSEAEARLVGSQQHQHHNESSCSSGPDSPRHAHSHSHPLHGGGGATG
GPSSAGGTGSGGGDGGGTGAIPKNLPALETPMGSGGGGLAGSGQGQAQYLSASCVVFT
NYSGDTASQVDEHFSRALNYNNKDSKESSSPMSNRNFPPSFWNSNYVHPIPAPTHHQVS
DLYGTATDTGYATDPWVPHAAHYGSYAHAAHAHAAHAHAYHHNMAQYGSLLRLPQ
YASHGSRLHHDQQTAHALEYSSYPTMAGLEAQVAQVQESSKDLYWF

Figure 10. Amino acid sequence of vestigial. The coding portion of the nucleotide sequence of the gene *vestigial,* which was found to be allelic to *trex,* was entered into Expasy in order to retrieve the amino acid sequence that corresponds.

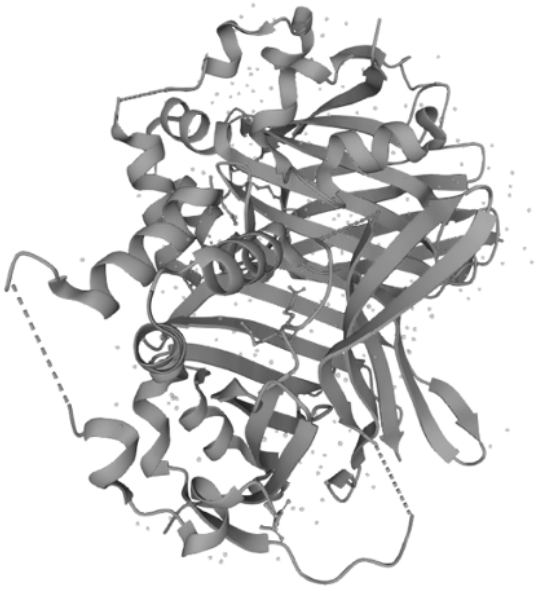

Figure 11. Protein structure of vestigial. The structure of *vestigial*, the gene allelic to *trex,* is included above from a Uniprot analysis of the amino acid sequence of the wild-type genome retrieved from Expasy.

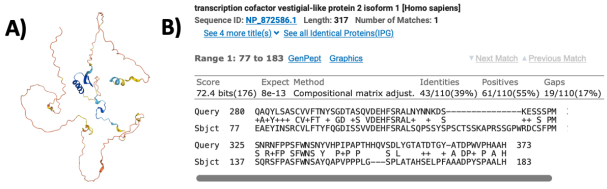

Figure 12. Human ortholog vestigial-like protein 2. [A] shows the protein structure of *VGLL-2*, the human ortholog to *vestigial* and also *trex*. The alignment of these genes as orthologs is shown in [B] as the protein BLAST alignment of *vestigial* and its orthologs within the human genome.