

Experimental validation of transposable element insertions using the Polymerase Chain Reaction (PCR)

Miriam Merenciano¹, Marta Coronado-Zamora¹, and Josefa González¹

¹Institute of Evolutionary Biology, CSIC, UPF. Barcelona 08003, Spain

Correspondence: josefa.gonzalez@csic.es

Abstract

Transposable elements (TEs), also known as transposons, are repetitive DNA sequences, present in virtually all organisms, that can move from one genomic position to another. TEs can be a source of mutations with important consequences for organisms. Despite their interest, its repetitive nature has made their study very challenging. However, the emergence of new sequencing technologies that allow obtaining long-read sequences, has improved the *in silico de novo* detection and annotation of TEs. The *de novo* annotation of TEs has already been performed in several organisms including the fruit fly *Drosophila melanogaster*. Yet, experimental validation can be used to confirm the presence of TEs in specific *D. melanogaster* natural populations. Here, we present a step-by-step protocol to experimentally validate by polymerase chain reaction (PCR) the presence and/or absence of TEs in natural populations of *D. melanogaster*. This detailed protocol has been implemented in the participant high schools of the *Citizen Fly Lab* activity that is part of the international citizen science project *Melanogaster: Catch the Fly!* (<https://melanogaster.eu>). Specifically, the students collaborate with the scientists of the European *Drosophila* Population Genomics Consortium (DrosEU) in the experimental validation of new genetic variants, previously identified using bioinformatic techniques.

Keywords: citizen science, *Drosophila melanogaster*, primer design, DNA extraction, DNA amplification, electrophoresis.

Introduction

Transposable elements (TEs) are DNA repetitive fragments with the ability to move along the genome [1, 2]. TEs are virtually present in all eukaryotic and in almost all prokaryotic genomes, and they usually represent a considerable fraction of the genome, although their abundance is highly variable from one species to another: from less than 1% in some yeasts to more than 85% in some plants [3-5]. Because of their mobile and repetitive nature, TEs constitute a relevant source of mutations that have been suggested to play an important role in gene regulation and genome evolution [6]. Mutations caused by TEs can have no negative consequences on the organism –for example, when they are inserted in an intergenic region without disrupting any genic or regulatory sequence–, but others can have important consequences, even causing diseases such as hemophilia A in humans [7]. The deficiency of the clotting *Factor VIII* (F8) that characterizes hemophilia A, can

be caused by the insertion of a TE in the gene producing the factor F8, disrupting its function. However, TE insertions can also be relevant for adaptation, and in fact, there is growing evidence demonstrating how TEs can be involved in the capacity of organisms to cope with environmental stress [8-13]. The textbook example of evolution by natural selection of the industrial melanism in the peppered moth was discovered to be, indeed, caused by a TE [14]. The insertion of a TE within the *cortex* gene leads to its overexpression, causing the overproduction of melanin and conferring the moths the dark coloration that allowed them to camouflage in soot-darkened trees during the Industrial Revolution. Another example of an adaptive TE insertion can be found in the fruit fly *Drosophila melanogaster*, where a TE insertion, named *FBti0019386*, near an immune-related gene confers protection against the bacterial pathogen *Pseudomonas entomophila* [15]. This TE contains a regulatory region that is activated during bacterial infection and induces the expression of its nearby immune-response gene. Therefore, TEs provide, through different mechanisms, the raw material on which natural selection can act. However, despite their relevance, they have been systematically understudied due to methodological limitations regarding their detection and annotation [16]. The availability of whole-genome sequences and recent advances in sequencing technologies (mainly long-read sequencing techniques) allows a better annotation of TE insertions [16]. In *D. melanogaster*, TEs have been annotated *in silico* in different natural populations [16, 17]. Experimental validation helps confirm whether a TE is present, absent or polymorphic in a specific population before further analyzing that particular TE insertion.

To experimentally validate *in silico* annotated TE insertions, we launched the citizen science activity *Citizen Fly Lab*, which is part of the international citizen science project *Melanogaster: Catch the Fly!* The *Citizen Fly Lab* activity involves high-school students and teachers in the validation of TEs using the polymerase chain reaction (PCR). This technique allows making numerous copies of a specific region of the genome *in vitro*. Even though it was described in the 1980s [18], and it is universally used as the reference method for DNA amplification [19], it has recently become more relevant for citizens because it is one of the diagnostic techniques for COVID-19 [20]. Basically, in the PCR method, a pair of short sequences of DNA –called primers– are designed to hybridize in a specific region of the DNA that will be amplified for its posterior evaluation.

Here, we describe the specific steps for TE experimental validation by PCR in natural populations of *D. melanogaster*. For that, we will use the *FBti0019386* TE insertion to exemplify the whole process. This TE was detected and annotated in the *D. melanogaster* reference genome, called ISO-1, and it is available at the FlyBase database [21]. Until recently, the ISO-1 was the only reference genome

available for *D. melanogaster*. However, a new genome browser containing additional 51 high-quality reference genomes together with high-quality *de novo* annotation of TE insertions is now available: *DrosOmics* (<http://gonzalezlab.eu/drosomics>) [16, 22-24]. Briefly, the procedures detailed in this protocol are: primer design, genomic DNA extraction, PCR reaction, electrophoresis, and interpretation of the results, including a troubleshooting section to help solve the most common problems when carrying out a PCR.

The *Citizen Fly Lab* activity, and this protocol, represent a valuable opportunity for high school students and teachers to experience a practical approach to the research process of an ongoing cutting-edge scientific project, from a very dynamic and participative perspective. This particular implication in which participants are directly involved in different stages of the scientific process, allows young people to internalize especially complex concepts of the life sciences, relevant to their daily lives, at the same time that they contribute to expedite the scientific process, and to advance science. The *Citizen Fly Lab* activity could be a powerful resource to engage society and especially young people in collaborative and international scientific efforts, and to empower school teachers to promote STEAM careers with especial interest in girls through formal education and citizen science. If your school wants to join the *Citizen Fly Lab* activity, please contact the corresponding author.

Materials

Reagents

- Primers (Custom DNA Oligos)
- Agarose
- SYBR Safe DNA Gel Stain
- DNA Ladder (1 kb)
- DNase/RNase-Free Distilled Water
- Proteinase K
- DreamTaq Green PCR Master Mix (2X)

Buffers

- **Squashing buffer:** 10 mM Tris-HCl (pH 8.2), 1 mM EDTA, 25 mM NaCl. Weight 15.75 mg Tris-HCl, 3.72 mg EDTA, and 14.61 mg NaCl; and dissolve it in 9 mL distilled water. Adjust volume to 10 mL. Store at room temperature.

- **TBE buffer 1X:** 0.13 M Tris base, 45mM Boric acid, 2.5 mM EDTA (pH 8.0). Weight 10.8 g Tris base and 5.5 g Boric acid and dissolve it in 900 mL distilled water. Add 4 mL of 0.5M EDTA (pH 8.0). Adjust volume to 1L. Store at room temperature.

Equipment

- 10 *Drosophila melanogaster* flies
- 1.5 mL microcentrifuge tubes
- 0.2 mL microcentrifuge tubes
- Pipette tips
- Erlenmeyer flask
- Beaker
- 100 mL test tube
- Gloves
- Ice bucket
- Thermocycler
- Electrophoresis chamber
- Electrophoresis power supply
- Microcentrifuge, fixed speed
- Transilluminator
- Set of three adjustable-volume micropipettes: 100-1000 μ l, 20-200 μ l, and 2-20 μ l
- Microwave
- Fridge / freezer
- Thermoblock (optional)

Methods

1. Primer design

Timing: 30 minutes

Design two primer pairs to experimentally validate by PCR the presence and/or absence of the selected TE insertion (**Figure 1**). We will exemplify the process using the *FBti0019386* TE insertion, a TE that is 346 bp long. The first pair of primers will assay the presence of the TE insertion and consists of a "Left" (L) primer, which binds the TE sequence, and a "Right" (R) primer that binds the right flanking region of the insertion. With this primer pair, we expect a band only when the TE is present (**Figure 1A**). On the other hand, the other primer pair will assay for the absence of the TE

insertion and consists of a “Flanking” (FL) primer which binds the left flanking region of the TE and the R primer mentioned above. In this case, the absence of a TE insertion should give a band. If the TE insertion we are trying to validate is small (<800 bp approximately), the FL and R primers could give a longer band indicating the presence of the insertion (**Figure 1B**). However, we assume that this longer band is unlikely to be amplified if the TE sequence is >800 bp.

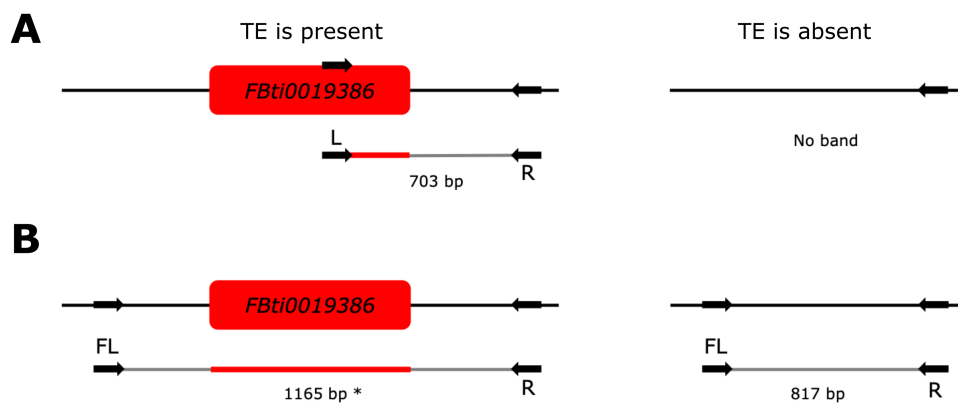


Figure 1 - Schematic representation of the *FBti0019386* TE insertion and the primers to experimentally validate its presence and/or absence. **A) Region spanned by the “Left” (L) and “Right” (R) primer pair in flies with the *FBti0019386* insertion (left panel). No band is expected in flies without the insertion (right panel). **B**) Region spanned by the “Flanking” (FL) and “Right” (R) primer pair in flies with the *FBti0019386* insertion (left panel) and without the insertion (right panel). *This band is unlikely to be amplified for TEs whose length is >800 bp.**

1.1. Get the nucleotide sequence of the region where the TE is annotated

We are going to use the *DrosOmics* genome browser (<http://gonzalezlab.eu/drosomics>) [24] to get the nucleotide sequence of the region where the TE under study is inserted in the reference *D. melanogaster* genome. In the *DrosOmics* browser, the identification name of each TE can be encoded in three ways: (1) “FBti” (from **FlyBase transposable insertion**) followed by a seven-digit number (e.g.: *FBti0019386*), which identifies TEs annotated in the *D. melanogaster* reference genome available also at FlyBase [21]; (2) “ISO-1” followed by a digit and the TE family (e.g.: *ISO1_36_INE-1*) identifies TEs that have been discovered *de novo* in the reference genome (called ISO-1); and (3), a coordinate, indicating the region where the TE was discovered, and the TE family name (e.g.: *2R_14873922_14873929_roo*), which identifies TE insertions discovered *de novo* in non-reference genomes. The following step-by-step guide shows how to use the browser *DrosOmics* to obtain the nucleotide sequence of the region where the *FBti0019386* TE is inserted from the *D. melanogaster* reference genome (ISO-1; [24]).

1. Access the *DrosOmics* browser (<http://gonzalezlab.eu/drosomics>) and close the views of the two *D. melanogaster* genomes that are displayed by default (**Figure 2A**, the two vertical red arrows point the "x" to close them).
2. Open the genome where the TE is predicted to be inserted by going to the main menu and clicking in "Add > Linear genome view" (**Figure 2B**, step 1). A new window will appear with a drop-down menu displaying all available genomes in the browser. Following the example, we open the *D. melanogaster* reference genome by selecting "*D. melanogaster* (ISO-1)" under "Assembly" and clicking the "open" button (**Figure 2B**, steps 2 and 3).
3. Activate the TE track. Click the "Open track selector" button and a track selector panel will be displayed on the right, showing all the available tracks that can be activated (**Figure 2C**, step 1). Activate the "TE annotations" track under the "Genome annotations" category (**Figure 2C**, step 2). Following the example, we click on "TE annotations (ISO-1)".
4. Find the TE in the genome. Introduce the name of the TE under study in the search box (**Figure 2D**). The genomic coordinates of the insertion will appear on the search box, as well as in the right panel when clicking on the TE insertion (**Figure 2E**). In the example, for the *FBti0019386* TE, its sequence comprises the nucleotides 16,189,464-16,189,810 of chromosome arm 3R of the reference genome (**Figure 2E**).
5. Download the DNA sequence of the region where the TE is annotated in FASTA format adding 500 bp to each flanking region. Introduce the new coordinates in the search box (**Figure 3A**). Following this example, the new coordinates are 3R:16,188,964-16,190,310. To obtain the sequence, you have to click-and-drag the overview scale bar of the whole display (**Figure 3B**, from position 1 to 2), and a new menu will appear that will allow you to get the selected sequence in FASTA format ("Get sequence"). Check that the starting and ending coordinates correspond to the region that you want to download. Finally, copy the nucleotide sequence obtained (**Figure 3C**).

1.2. Obtain the primer sequences from the sequence region of the TE

There are many available online tools that allow you to easily design primer pairs to amplify a specific DNA region. In this protocol, we use the primer-BLAST online tool (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>; [25]).

1. In the primer-BLAST webpage, paste the nucleotide sequence obtained in the previous step inside the "PCR Template" box (**Figure 4A**). Delimit the region where the primers should be designed in the "Range" box (**Figure 4B**). We will start creating the FL-R primer pair that assays for the absence of the TE insertion. For the FL-R pair, primers should be designed in

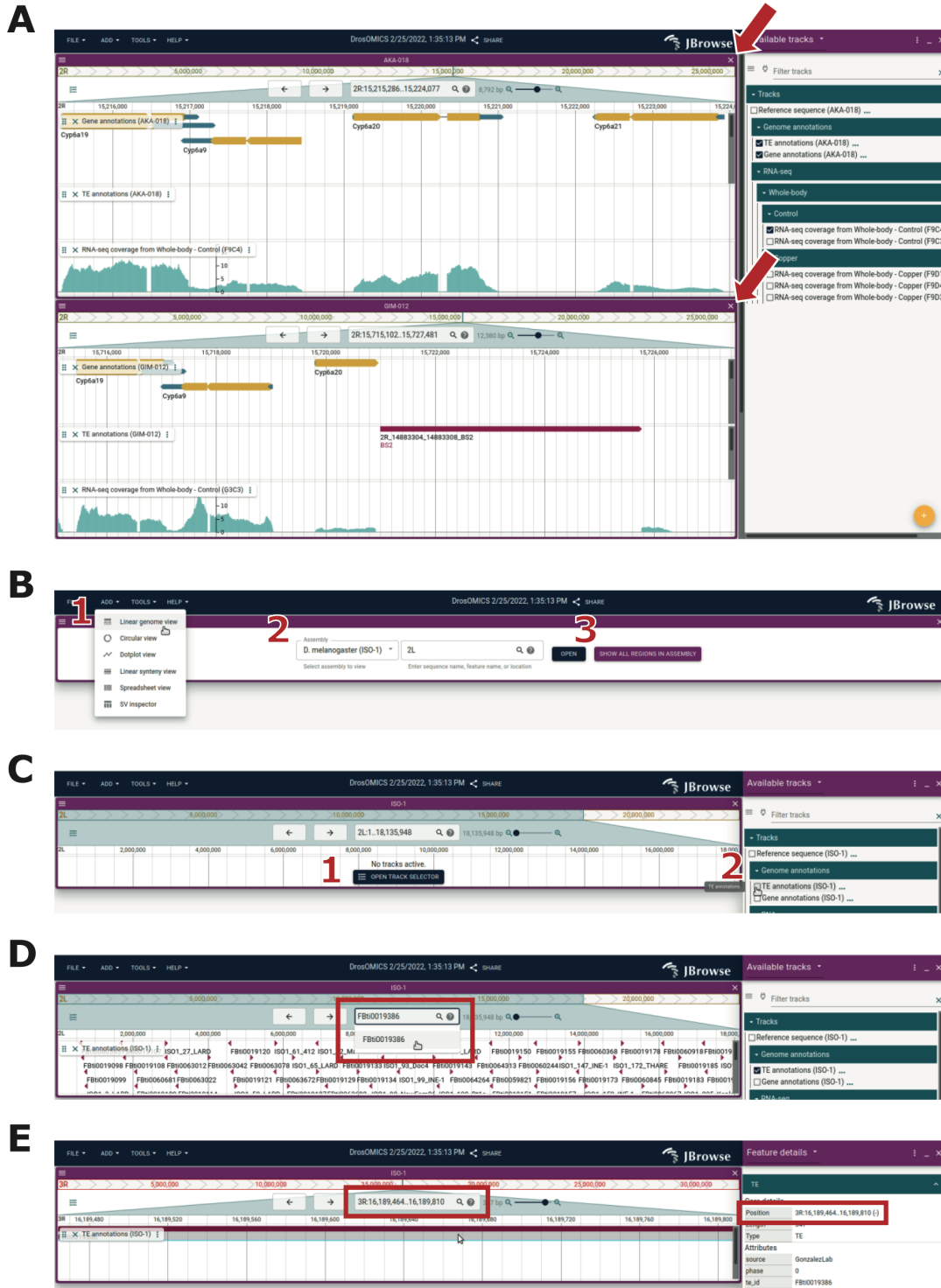


Figure 2 - Genomic coordinates of the region where the TE is annotated in the *DrosOmics* browser. Screenshot of the *DrosOmics* browser (<http://gonzalezlab.eu/drosomics>) [24]. **A)** Default view of the *DrosOmics* browser displaying two genomes. The two vertical red arrows point to the “x” to close them. **B)** The three steps to open a new genome. **C)** The two steps to open a track from the track selector. **D)** Using the search box to look for a TE. **E)** View of the selected TE, highlighting its coordinates.

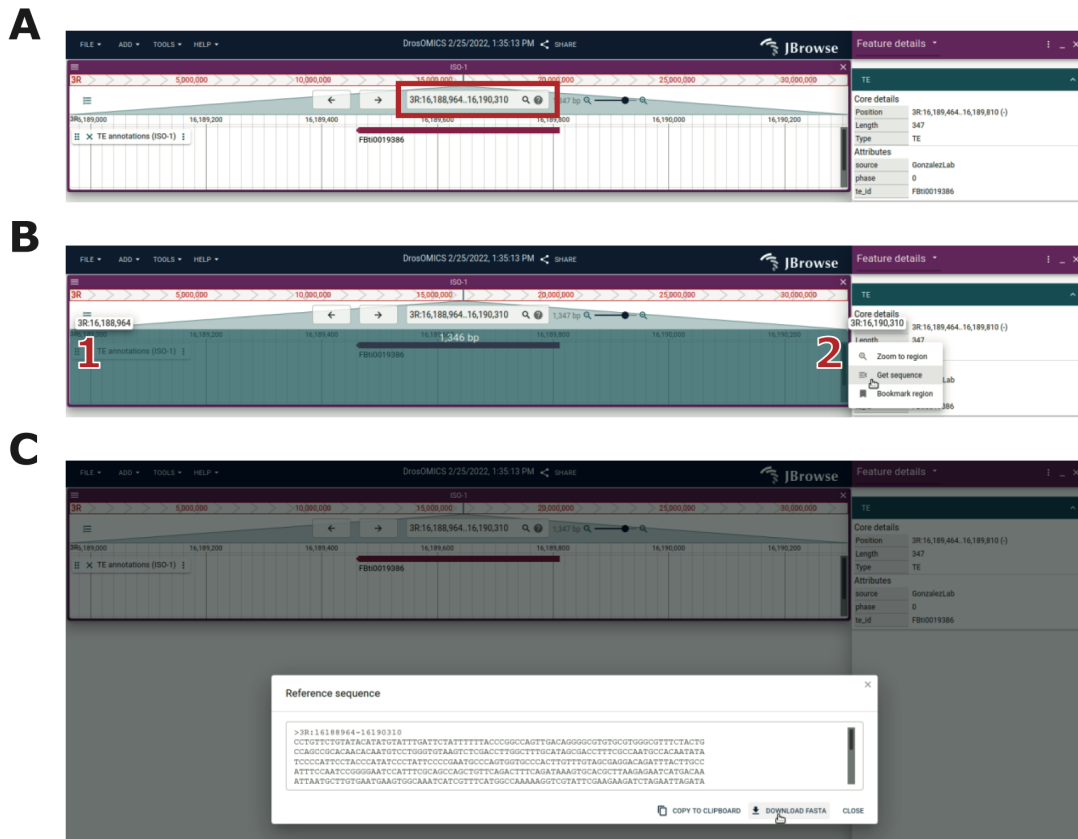


Figure 3 – Nucleotide sequence of the region where the TE is annotated in the *DrosOmics* browser. Screenshot of the *DrosOmics* browser (<http://gonzalezlab.eu/drosomics>) [24]. **A)** View of the TE region after updating the coordinates to include ± 500 bp to the TE flanking regions. **B)** Selecting a genome region: click-and-drag the scale bar to select a region, from position 1 to 2. Check that the starting and ending coordinates correspond to the region that you want to obtain. **C)** Obtaining the DNA sequence in FASTA format.

the first and last 250 bp of the TE flanking regions, respectively, to get a visible band if the TE is not present (**Figure 1B**). Hence, the nucleotide range for the “Forward” (FL) primer should be 1-250, while the range for the “Reverse” (R) primer should be the last 250 nucleotides of the pasted sequence.

Note: increase the value of the “Max PCR product size” if the sequence pasted exceeds 1000 bp (**Figure 4C**).

2. Click the “Get primers” button at the end of the webpage. Choose any of the primer pairs obtained with a product length >500 bp. In our example, FL primer will be 5'-CGCACAAACAATGTCCTGG-3' and R primer will be 5'-GGAATCAATCACATCAACCCGT-3' spanning 1165 bp if the TE is present. Since the length of the TE is 346 bp, FL-R primer will be spanning 817 bp if the TE is absent (**Figures 1B and 5**).

NIH U.S. National Library of Medicine National Center for Biotechnology Information Log in

Primer-BLAST A tool for finding specific primers

Finding primers specific to your PCR template (using Primer3 and BLAST).

Primers for target on one template Primers common for a group of sequences

Retrieve recent results Publication Tips for finding specific primers Save search parameters Reset page

PCR Template

Enter accession, gi, or FASTA sequence (A refseq record is preferred) Clear

A `CGATAAATTAAGTTGTAACAACTACTTTCGAAAGCAATAAAAAATTATTACAGTA
TTTTCACGATAAGGCGAATAACAAACCATTAACTTGTATGTGCGGGTACAAACTATT
CCTCAAAATGTTATCAGTTATCTGATTACGGGTTGATGTTGATTCACAAATAAAGT
TGTAAAGATTGAGTTATTTAAGATAACCGCGTGAACATTTCTTTTGTAAATTTA
TTTACGGTCGTTCAATAAAGATTGC`

B Range Clear

From To

Forward primer 1 250

Reverse primer 1119 1369

Or, upload FASTA file Choose file No file chosen

Primer Parameters

Use my own forward primer (5'->3' on plus strand) Clear

Use my own reverse primer (5'->3' on minus strand) Clear

C Max

PCR product size Min 70 Max 2000

of primers to return 10

Primer melting temperatures (T_m) Min 57.0 Opt 60.0 Max 63.0 Max T_m difference 3

Exon/intron selection

A refseq mRNA sequence as PCR template input is required for options in the section

Exon junction span No preference

Figure 4 - Introduction of the nucleotide sequence in the primer design online tool to design the FL-R primer pair. Screenshot of the primer-BLAST webpage (<https://www.ncbi.nlm.nih.gov/tools/primer-blast>). **A)** "PCR template" box where the nucleotide sequence of the TE and 500 bp of each flanking region must be introduced. **B)** "Range" box to delimit the region where the primers should be designed. **C)** "Maximum PCR product size" box. If the sequence pasted exceeds 1000 (the default value), increase its value as in the figure.

NIH U.S. National Library of Medicine National Center for Biotechnology Information Log in

Primer-BLAST JOB ID: Bw3YH55K1EMbzFpAoVWEYRBGprAH93ag

Primer-BLAST Results

Input PCR template: lcl|Query_1

Range: 1 - 1347

Specificity of primers: Primer pairs are specific to input template as no other targets were found in selected database: Refseq mRNA (Organism limited to Homo sapiens)

Other reports: Search Summary

Graphical view of primer pairs

Query: l:1..1347 (1,347 nt)

Detailed primer reports

Primer pair	Sequence (5'→3')	Template strand	Length	Start	Stop	T _m	GC%	Self complementarity	Self 3' complementarity
Primer pair 1									
Forward primer	GGCGAACAACATGCTCTCTGG	Plus	22	1250	1229	60.04	55.00	5.00	3.00
Reverse primer	GGAATCAATCAATCAACCCCT	Minus	22	1250	1229	58.98	45.45	3.00	2.00
Product length	1165								
Primer pair 2									
Forward primer	TATTCCCGAATGCCAGTG	Plus	20	183	202	59.82	55.00	5.00	3.00
Reverse primer	TTCAGCCGACATTAGAAAGT	Minus	21	1192	1172	57.87	42.86	8.00	2.00
Product length	1010								
Primer pair 3									

Figure 5 - FL-R primers. Screenshot of the primer-BLAST webpage (<https://www.ncbi.nlm.nih.gov/tools/primer-blast>). List of primer pairs obtained, highlighting the ones that will be further used.

- Next, we will design the pair of primers that assay the presence of the TE: the L-R primer pair. The L primer should be designed within the last 300 bp of the TE sequence. Hence, keeping the same nucleotide sequence pasted in the "PCR Template" box, the nucleotide range for

the L primer should be considered depending on the TE size. In our example, the TE length is 346 bp, so the L primer range is 546-846. Since the R primer is already designed, paste its sequence in the "Use my own reverse primer (5'->3' on minus strand)" (Figure 6).

Figure 6 - Introduction of the nucleotide sequence in the primer design online tool to design the L-R primer pair. Screenshot of the primer-BLAST webpage (<https://www.ncbi.nlm.nih.gov/tools/primer-blast>). **A)** Range box to delimit the region where the L primers should be designed. **B)** "Use my own reverse primer" box where the sequence of the R primer must be introduced.

4. Click the "Get primers" button at the end of the webpage. Choose any of the primer pairs obtained. In our example, the L primer will be 5'- ATTTTGTACGTCAGCGCAGC-3', and the L-R primer pair will span 703 bp (Figure 1).

1.3. Positive control

Positive controls are used to technically validate an experiment, *i.e.*, to confirm that all the technical steps of the experiment were correctly performed. Our PCR positive control will amplify a region of the genome that is always present in all fly populations (for instance, a housekeeping gene). In our case, the positive control primers amplify the *Actin5C* gene (*Act5C*), a gene involved in various types of cell motility and ubiquitously expressed in all eukaryotic cells. To amplify this gene, the following primer pair is needed: the forward primer 5'-GCGCGTTACTCTTTCACCA-3' and the reverse primer 5'-ATTCCCAAGAACGAGGGCTG-3'.

The expected band size obtained after the PCR reaction is 222 bp.

1.4. Buy primers

Primers can be obtained from different companies (Merck, IGT, ...). To buy primers from Merck, paste the sequence of your primers, chose a name for each one, and specify the synthesis scale to 0.025 μ mol in the following webpage:

<https://www.sigmaaldrich.com/configurators/tube?product=standard>

You can add as many primers as you need to your shopping cart.

Note: your organization does not need to have an account to order, but you need to be registered to place an order through the Merck website.

Go to the "Login/Sign Up" tab on the site header and create your profile. It will give you the option to affiliate your profile with your organization's account. If your organization does not have an existing account, you will need to select "Continue Registration" and complete that process. After which, you can submit an order request.

2. Genomic DNA extraction

Timing: 1 hour

In this step, you will extract genomic DNA from a pool of 10 *D. melanogaster* flies from a single strain. Work with gloves.

1. Anesthetize 10 flies using a CO₂ pad or ice and put them in an empty labeled 1.5 mL microcentrifuge tube. Flies can be anesthetized covering completely the tube containing them in ice until they do not move.
2. Add 10 μ l of squashing mix (1% proteinase K in squashing buffer) in the 1.5 mL microcentrifuge tube containing the 10 individuals of *D. melanogaster*.

Note: keep proteinase K on ice while it is out of the freezer. Vortex the proteinase K before adding it to the squashing mix. Keep the squashing mix at room temperature during the procedure and vortex it before use.

3. Homogenize the samples with a homogenizer or a pipette tip at room temperature until no fly body parts are visible.

Note: This step is crucial for the success of the DNA extraction. Make sure no fly body parts are visible.

4. Add an additional 90 μ l of squashing buffer and mix gently. Make sure the tube is properly closed.
5. Incubate the tube at 37°C for 30 minutes in a thermoblock or in a thermocycler to lyse the cells and digest the proteins attached to the DNA.

6. Incubate the tube at 96°C for 2 minutes to denaturalize the proteinase K and thus stop the digestion reaction.
7. Spin the tubes in a centrifuge.
8. The genomic DNA is ready. Use immediately in a PCR reaction or store it at -20°C until further use.

3. PCR reaction

(Timing: 2.5 hours approximately - depending on the PCR program)

In this step, you will prepare the PCR reactions. You will have to thaw all the reagents needed to prepare the PCR reactions and the genomic DNA prepared in the previous step. Work on ice and use gloves.

1. Prepare three reaction mixes using the reagent volumes described in the tables (**Tables 1-3**). In our example, they will be named as “*FBti0019386* FL-R”, “*FBti0019386* L-R”, and “MIX C+”. Prepare enough mix for all the DNA samples you are analyzing adding an extra volume (usually for one extra reaction), so that all the tubes can be filled properly. In this way, (slight) pipetting inaccuracies can be compensated.

Note: “DreamTaq Green PCR Master Mix (2X)” is a ready-to-use solution containing the DNA polymerase, an optimized buffer with MgCl₂, deoxynucleotides (dNTP), and loading dye for the electrophoresis step. This pre-mixed formulation saves time and reduces contamination due to a reduced number of pipetting steps required for PCR set up. It is capable of robust amplification up to 6 kb from genomic DNA.

Note: remember that FL, L, and R primers are specific for each TE insertion. In case you are validating more than one TE insertion, prepare as many different FL-R and L-R mixes as TEs you want to validate labeling them properly so as not to confuse them.

Table 1 - PCR reaction mix for the amplification of *FBti0019386* FL-R

<i>FBti0019386</i> FL-R MIX	
Reagent	Volume for 1 reaction
DreamTaq PCR master mix (2X)	4 µl
FL primer (10 pmol/µl)	0.5 µl
R primer (10 pmol/µl)	0.5 µl
H ₂ O	13 µl

Table 2 - PCR reaction mix for the amplification of *FBti0019386* L-R

<i>FBti0019386</i> L-R MIX Reagent	Volume for 1 reaction
DreamTaq PCR master mix (2X)	4 µl
L primer (10 pmol/µl)	0.5 µl
R primer (10 pmol/µl)	0.5 µl
H ₂ O	13 µl

Table 3 - PCR reaction mix for the amplification of the positive control (C+)

MIX C+ Reagent	Volume for 1 reaction
DreamTaq PCR master mix (2X)	4 µl
F primer <i>Act5C</i> (10 pmol/µl)	0.5 µl
R primer <i>Act5C</i> (10 pmol/µl)	0.5 µl
H ₂ O	13 µl

2. Prepare a 1/10 dilution of the genomic DNA extracted in the previous step.
Note: the 1/10 dilution can be prepared by adding 2 µl of DNA to 18 µl distilled water in a 1.5 mL microcentrifuge tube.
3. For every TE and strain, label five 0.2 mL microcentrifuge tubes to prepare the following PCR reaction samples:
4. Sample FL-R: PCR reaction containing the previously extracted DNA and FL-R mix.
5. Sample FL-R C-: PCR reaction without DNA as a negative control and FL-R mix. No PCR bands are expected with this sample.
6. Sample L-R: PCR reaction containing the previously extracted DNA and L-R mix.
7. Sample L-R C-: PCR reaction without DNA and L-R mix. This is the PCR negative control: no PCR bands are expected with this sample.
8. Sample C+: PCR reaction containing the previously extracted DNA and the mix C+ containing the *Actin5C* primers. This is the PCR positive control: a 222 bp band is expected.
9. Add 18 µl of each mix into the corresponding 0.2 mL microcentrifuge tubes.
10. Add 2 µl of the diluted DNA into the 0.2 mL microcentrifuge tubes previously filled with mix.
Note: remember not to put DNA in the negative control (C-) samples.
11. Spin the samples in the centrifuge.
12. Put the samples in the thermocycler ensuring that all the tubes are properly closed. PCR cycling conditions can be found in **Table 4**.

Table 4 - PCR cycling conditions

Steps	Temperature	Time	Cycles
Initial denaturation	94°C	120 sec	1
Denaturation	94°C	30 sec	30 cycles
Annealing	57°C*	30 sec	
Extension	72°C	90 sec**	
Final extension	72°C	1 min	1
Hold	4°C-15°C	forever	

*The annealing temperature is the temperature at which the primers bind to the complementary template region. We recommend using an annealing temperature of 57°C. Increasing the annealing temperature may help get rid of unspecific band amplifications (see **Table 5** for *Troubleshooting*).

**Set the extension time depending on the maximum amplification size expected and considering that usually PCR polymerases expand 1 kb in 1 minute. So, taking the validation of *FBti0019386* as an example, the biggest band expected would be 1165 bp. Thus, the extension time is 1.5 minutes.

13. Run the PCR program.

14. When the PCR program finishes, proceed with the electrophoresis step or store the samples at -20°C until further use.

4. Electrophoresis

(Timing: 1 hour)

In this step, you will cast a 1% agarose gel, load the gel with DNA samples and ladder, and separate them using gel electrophoresis. You will need to have your PCR samples prepared and ready to load into the gel.

1. Prepare the gel tray for casting. Place a comb in the gel tray.

Note: consider the number of wells that you need to fill when installing the comb.

2. Prepare enough volume of a 1% agarose gel to fill your tray in an Erlenmeyer flask that can hold 2-4 times the volume to be prepared.

Note: 1% refers to the percentage of agarose in the volume of buffer. The gel percentage is calculated as (grams of agarose / milliliters of buffer) × 100%. For instance, for a total volume of 50 mL, add 5 mg of agarose to 50 mL 1X TBE buffer. The standard percentage of agarose for a gel is usually about 1-2%. The percentage of agarose that is appropriate depends on

the size of DNA fragments you are expecting to separate. The percentage of agarose determines how well the DNA separates, and the resolution of the final gel.

3. Heat the flask in the microwave for 60 seconds. If the agarose is not completely dissolved, return the flask to the microwave and continue heating in 30-second intervals until the agarose is in solution.
4. Remove the flask from the microwave with a heat-resistant glove and swirl it very gently.
Note: The microwaved solution can become superheated and foam over quickly when agitated. Wear appropriate protection and use caution.
5. Wait 2-5 minutes until the solution starts to cool down.
6. Add SYBR Safe DNA Gel Stain to the solution and mix by swirling the flask. SYBR Safe DNA Gel Stain provides highly sensitive staining for the visualization of DNA on agarose gels. It has been specifically formulated as a less hazardous alternative to ethidium bromide that can use UV or blue light excitation. For a total volume of 50 mL agarose solution, add 1 μ L SYBR Safe DNA Gel Stain to the solution.
7. Pour the solution with the gel stain slowly into the gel tray and check for leakages.
8. Wait about 20-30 minutes for the gel to solidify at room temperature. Once the gel has solidified, remove the comb carefully. Make sure not to pierce the gel.
9. Pour TBE 1X over the gel until the gel is fully covered. The buffer should reach about 2-3 mm above the gel.
10. Load 10 μ L of your sample into the gel. Repeat this for each sample, taking care to use a new pipette tip each time.
Note: For each of your samples, you should have a note recording which well it has been loaded into.
11. Load 5 μ L of the DNA ladder into the gel.
12. Connect the gel tray to the power and run it for 30-40 minutes at a maximum voltage of 120V to separate the DNA fragments.
13. Once your run is complete, disconnect the gel tray. Open the lid and discard the buffer. You can dispose of it in the sink. Your gel is now ready for visualization via transillumination.
14. Visualize the gel in a transilluminator.
15. Take a picture of the results.
16. Discard the agarose gel in a bin for agarose residues and clean the electrophoresis tray and the comb with distilled water.

Results interpretation and expected outcomes

In this section, we mention the steps for a correct interpretation of the results and we briefly describe the different results that can be obtained.

1. Check the DNA ladder. All the DNA bands from the ladder must be correctly separated to ensure that there were no problems during the electrophoresis step.
2. Check the control samples (C- and C+). Negative control samples are not expected to give any band, while positive control samples are expected to give a single band of 222 bp. No further conclusions can be obtained if control samples do not give the expected results (see **Table 5** for *Troubleshooting*).
3. Check the other samples (FL-R and L-R). Different results can be obtained:
 - a) FL-R band and L-R band of the size expected if the TE insertion is present. In that case, the presence of the TE insertion has been validated in the tested strain (**Figure 7A**).
 - b) No FL-R band and L-R band of the size expected if the TE insertion is present. In that case, the presence of the TE insertion has been validated in the tested strain (**Figure 7B**).
 - c) FL-R band of the size expected if the TE insertion is absent and no L-R band. In that case, the TE insertion is not present in the tested strain (**Figure 7C**).
 - d) Two FL-R bands, one of the size expected if the TE insertion is present and another one of the size expected if the TE insertion is absent, and L-R band of the size expected if the TE insertion is present. In that case, the presence of the TE insertion has been validated in the tested strain, although it is polymorphic (some alleles have the insertion and others do not) (**Figure 7D**).
 - e) FL-R band of the size expected if the TE insertion is absent, and L-R band of the size expected if the TE insertion is present. In that case, the presence of the TE insertion has been validated in the tested strain, although it is polymorphic (**Figure 7E**).
 - f) No FL-R or L-R band. No results have been obtained (see **Table 5** for *Troubleshooting*) (**Figure 7F**).
 - g) Presence of non-expected bands (see **Table 5** for *Troubleshooting*).

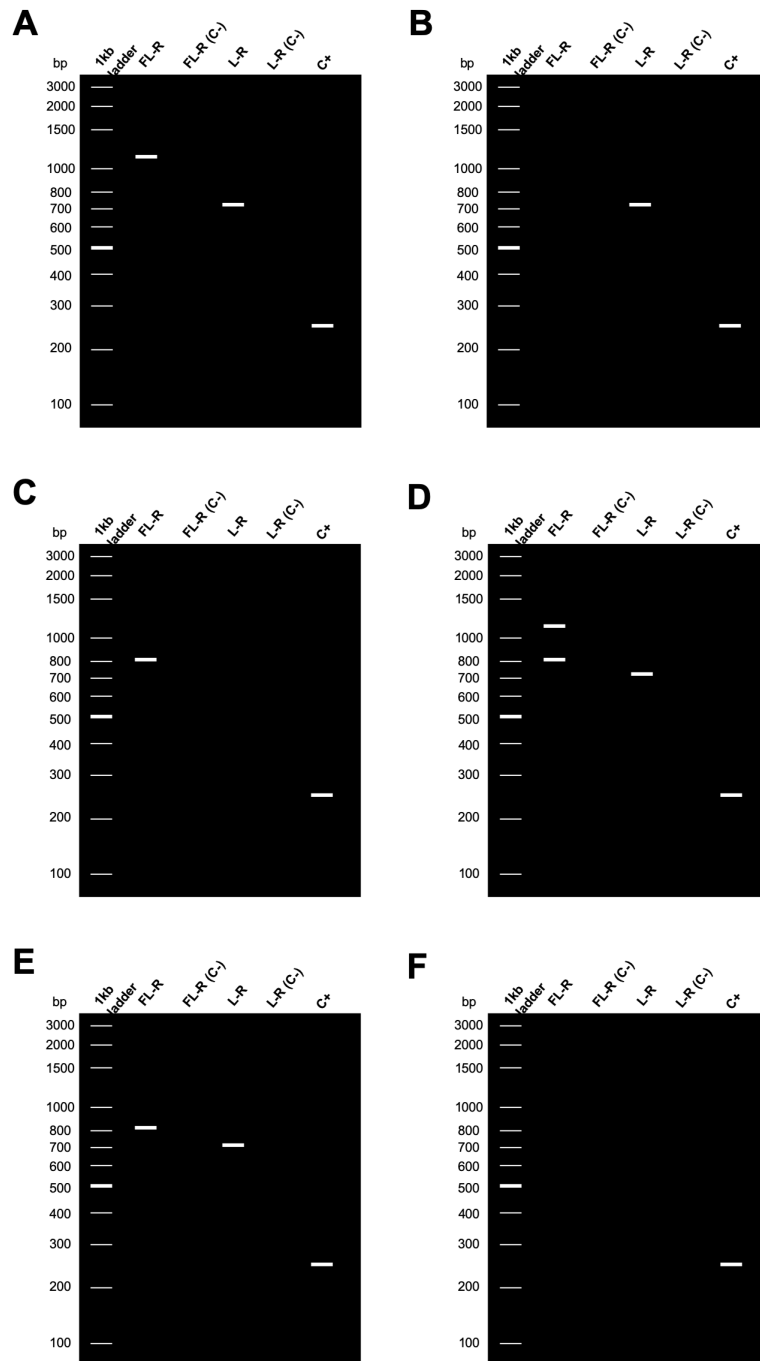


Figure 7 - Possible results that can be obtained from the PCR. Note that the band sizes resulting from FL-R and L-R pairs correspond to the *FBti0019386* example. **A)** The TE insertion is present in the tested strain: FL-R band and L-R band of the expected sizes (1165 and 703 bp, respectively, **Figure 1**). **B)** The TE insertion is present in the tested strain: L-R band of the expected size (703 bp, **Figure 1**). **C)** The TE insertion is absent in the tested strain: FL-R band of the expected size if the TE insertion is absent (817 bp), and no L-R band (**Figure 1**). **D)** The TE insertion is present and it is polymorphic in the tested strain: two FL-R bands, one of the size expected if the TE insertion is present (1165 bp) and another one of the size expected if the TE insertion is absent (817 bp), and L-R band of the size expected if the TE insertion is present (703 bp) (**Figure 1**). **E)** The TE insertion is present and it is polymorphic in the tested strain: FL-R band of the expected size if the TE insertion is absent (817 bp), and L-R band of the size expected if the TE insertion is present (703 bp). **F)** Inconclusive PCR, control samples show the expected results but no FL-R nor L-R bands (see **Table 5** for *Troubleshooting*).

Troubleshooting

Table 5 - Troubleshooting

Problem	Possible cause	Solution
No bands on gel	Insufficient template or too much template was added	Titrate the amount of template required for your system.
	Primer concentration was too low or was unbalanced	Make sure primer concentration is within the desirable range and that concentration of both PCR primers is the same.
	Contamination with nucleases	Repeat the DNA extraction. Be sure to follow precautions against contamination.
	Reagents were not thawed properly before use	Thaw and thoroughly mix all reagents before setting up reactions.
	Primer annealing temperature was too high	Decrease annealing temperature in 2 °C increments.
	Extension time was too short	Increase extension times. This can be especially important for long PCR required for the validation of long TE insertions.
	Thermal cycler was not at correct temperature	Repair/calibrate the thermal cycler or use a different machine.

Problem	Possible cause	Solution
	Electrophoresis not working properly	Make sure that the agarose gel was correctly prepared and that the electrophoresis is working at the correct voltage.
Extra, nonspecific bands on gel	Primers hybridized to a secondary site on the DNA template	Increase the annealing temperature by increments of 2°C. Design new primers that are less specific for the secondary site.
	Too much template was added	Titrate the amount of template for your system.
	Primer concentration was too high	Titrate the amount of primers for your system.
	Thermal cycler was not at correct temperature	Repair/calibrate the thermal cycler or use a different machine.
Excessive smearing of amplified DNA	Too much template was added	Titrate the amount of template for your system.
	Too many cycles were performed	Reduce the number of cycles below 30.

Problem	Possible cause	Solution
	The annealing temperature was too low	Increase the annealing temperature by increments of 2°C.
	Primer concentration was too high	Titrate the amount of primers for your system.
	Extension time was too long	Decrease extension time in small increments.
	Template DNA quality was low	Impure DNA can fail to amplify properly. Use freshly prepared DNA.
	Too much DNA was loaded on gel	Re-run with less PCR sample amount.
Primer dimers visible	Too much primer was added/ primer concentration too high	Titrate the amount of primers for your system.
	Primers possess complementary overlapping sequence	Design primers to avoid self-complementary internal sequences.

Acknowledgments

We thank Roberto Torres for contributing to the manuscript. We thank the IES Milà i Fontanals for allowing us to carry out the pilot *Citizen Fly Lab* activity in their premises. We thank teachers of the citizen science project *Melanogaster: Catch The Fly!* (MCTF!) for implementing the protocol in their schools. The MCTF! citizen science project is co-directed by *La Ciència Al Teu Món* (<https://lacienciaalteumon.cat>) and the Evolutionary and Functional Genomics Lab (<https://gonzalezlab.eu>) in collaboration with DrosEU (<https://droseu.net>) and has received funding from the Fundación Española para la Ciencia y la Tecnología (FECYT)-Ministerio de Ciencia e Innovación (MICINN), the Fundación General CSIC (FGCSIC) and the European Research Council (ERC grant H2020-ERC-2014-CoG-647900). DrosEU is funded by an ESEB STN award.

References

1. Bourque G, Burns KH, Gehring M, et al (2018) Ten things you should know about transposable elements 06 Biological Sciences 0604 Genetics. *Genome Biol* 19:1-12
2. Wells JN, Feschotte C (2020) A Field Guide to Eukaryotic Transposable Elements. *Annu Rev Genet* 54:539-561
3. Debarry JD, Kissinger JC (2011) Jumbled Genomes: Missing Apicomplexan Synteny. *Mol Biol Evol* 28:2855-2871
4. Hua-Van A, Le Rouzic A, Boutin TS, et al (2011) The struggle for life of the genome's selfish architects. *Biol Direct* 6:19
5. Guio L, González J (2019) New Insights on the Evolution of Genome Content: Population Dynamics of Transposable Elements in Flies and Humans. *Methods Mol Biol* 1910:505-530
6. Branco MR, Chuong EB (2020) Crossroads between transposons and gene regulation. *Philos Trans R Soc B* 375:
7. Kazazian HH, Wong C, Youssoufian H, et al (1988) Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* 332:164-166
8. Casacuberta E, González J (2013) The impact of transposable elements in environmental adaptation. *Mol Ecol* 22:1503-1517
9. Dubin MJ, Mittelsten Scheid O, Becker C (2018) Transposons: a blessing curse. *Curr Opin Plant Biol* 42:23-29
10. Rech GE, Bogaerts-Márquez M, Barrón MG, et al (2019) Stress response, behavior, and development are shaped by transposable element-induced mutations in *Drosophila*. *PLOS*

- Genet 15:e1007900
11. Schrader L, Schmitz J (2019) The impact of transposable elements in adaptive evolution. *Mol Ecol* 28:1537–1549
 12. Hunter RG (2020) Stress, Adaptation, and the Deep Genome: Why Transposons Matter. *Integr Comp Biol* 60:1495–1505
 13. Baduel P, Quadrana L (2021) Jumpstarting evolution: How transposition can facilitate adaptation to rapid environmental changes. *Curr Opin Plant Biol* 61:102043
 14. Hof AEV t, Campagne P, Rigden DJ, et al (2016) The industrial melanism mutation in British peppered moths is a transposable element. *Nat* 2016 5347605 534:102–105
 15. Ullastres A, Merenciano M, González J (2021) Regulatory regions in natural transposable element insertions drive interindividual differences in response to immune challenges in *Drosophila*. *Genome Biol* 2021 221 22:1–30
 16. Rech GE, Radío S, Guirao-Rico S, et al (2021) Population-scale long-read sequencing uncovers transposable elements contributing to gene expression variation and associated with adaptive signatures in *Drosophila melanogaster*. *bioRxiv* 2021.10.08.463646
 17. Mohamed M, Dang NTM, Ogyama Y, et al (2020) A Transposon Story: From TE Content to TE Dynamic Invasion of *Drosophila* Genomes Using the Single-Molecule Sequencing Technology from Oxford Nanopore. *Cells* 2020, Vol 9, Page 1776 9:1776
 18. Mullis KB (1990) Target amplification for DNA analysis by the polymerase chain reaction. *Ann Biol Clin (Paris)* 48:579–582
 19. Garibyan L, Avashia N (2013) Polymerase chain reaction. *J Invest Dermatol* 133:1–4
 20. Carter LJ, Garner L V., Smoot JW, et al (2020) Assay Techniques and Test Development for COVID-19 Diagnosis. *ACS Cent Sci* 6:591–605
 21. Larkin A, Marygold SJ, Antonazzo G, et al (2021) FlyBase: updates to the *Drosophila melanogaster* knowledge base. *Nucleic Acids Res* 49:D899–D907
 22. Long E, Evans C, Chaston J, Udall JA (2018) Genomic Structural Variations Within Five Continental Populations of *Drosophila melanogaster*. *G3 Genes|Genomes|Genetics* 8:3247–3253
 23. Chakraborty M, Emerson JJ, Macdonald SJ, Long AD (2019) Structural variants exhibit widespread allelic heterogeneity and shape variation in complex traits. *Nat Commun* 2019 10:1–11
 24. Coronado-Zamora M, Salces-Ortiz J, Gonzalez J (2022) DrosOmics: the comparative genomics browser to explore omics data in natural strains of *D. melanogaster*. *bioRxiv* 2022.07.22.501088

25. Ye J, Coulouris G, Zaretskaya I, et al (2012) Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13:134