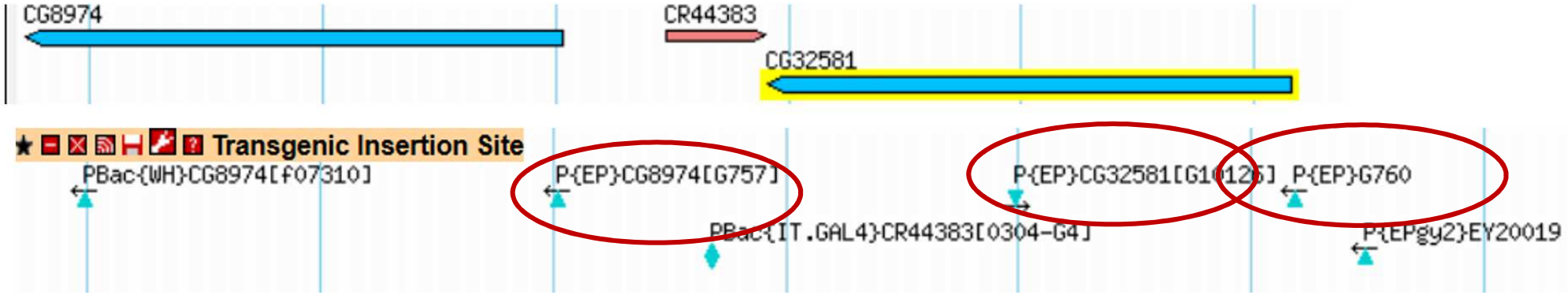
04/13/16

Olga Olejniczak

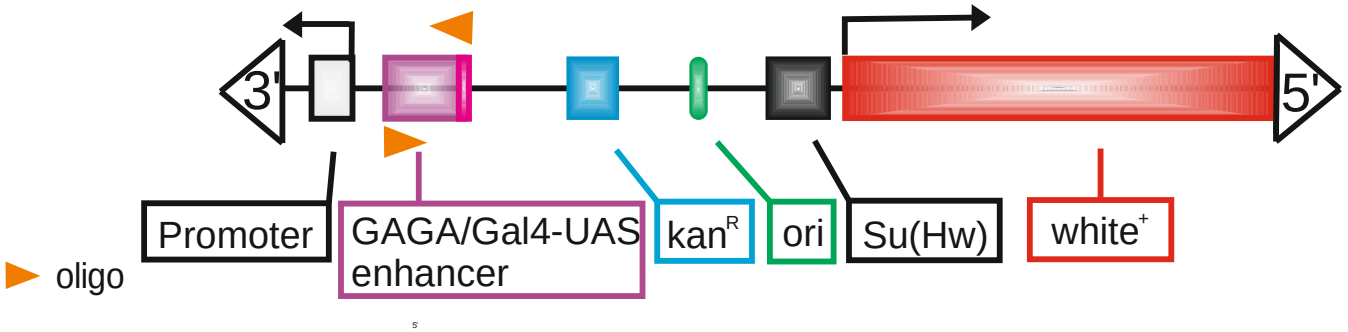
CG8974/CG32581 as an example of gene duplication



The purpose of this study is to map several P elements in the CG32581/8974 region. The reason for concern is that these genes are >95% identical, the result of a recent duplication event. Some of our functional data made us suspicious that one or more of these P's was mis-mapped.

Fly stocks with p-element P{EP}G760 upstream CG32581 (BL 26598), with p-element P{EP}CG8974^{G757} upstream CG8974 (BL 33461) and with p-element P{EP}CG32581^{G10126} for disruption (BL 32629) were ordered from Bloomington and DNA was isolated.

In order to detect **p-elements, primers for the UAS enhancer** sequence were designed (the sequence of enhancer was taken from flybase: http://flybase.org/reports/FBrf0100830.html)



Primer F CCCCCTCGAATGTTCTCTCT
Primer R AGGCCTAAGCTTGATGACCTC
PRODUCT SIZE: 100

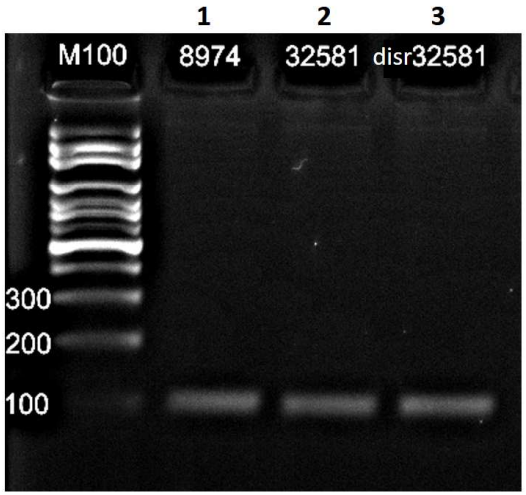
Products from the amplification performer with using primers for detection of the enhancer sequence within p-elements

3 distinct fly stocks:

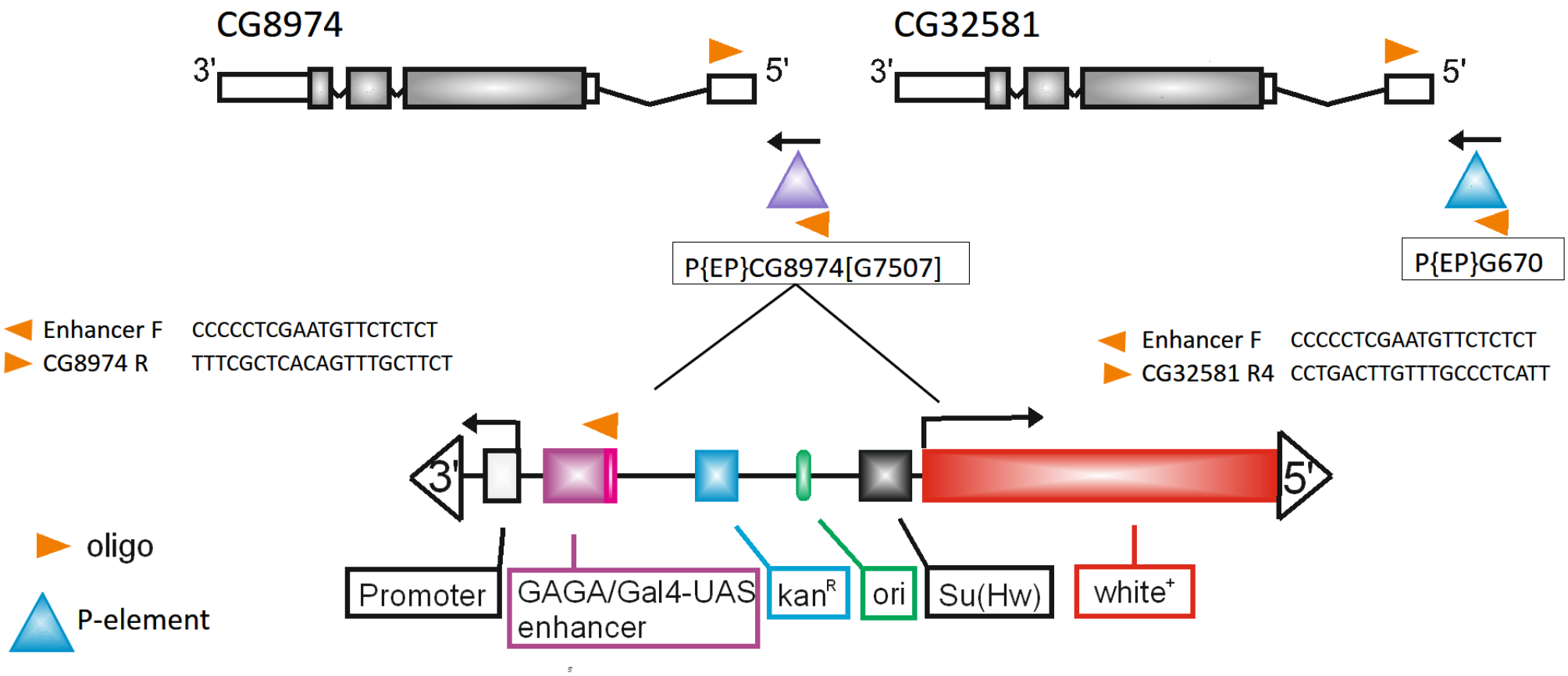
1 → 8974 – with p-element inserted upstream CG8974

2 → 32581 – line with p-element upstream CG2581

3 → disr32581 – line with p-el inserted within the exon 2

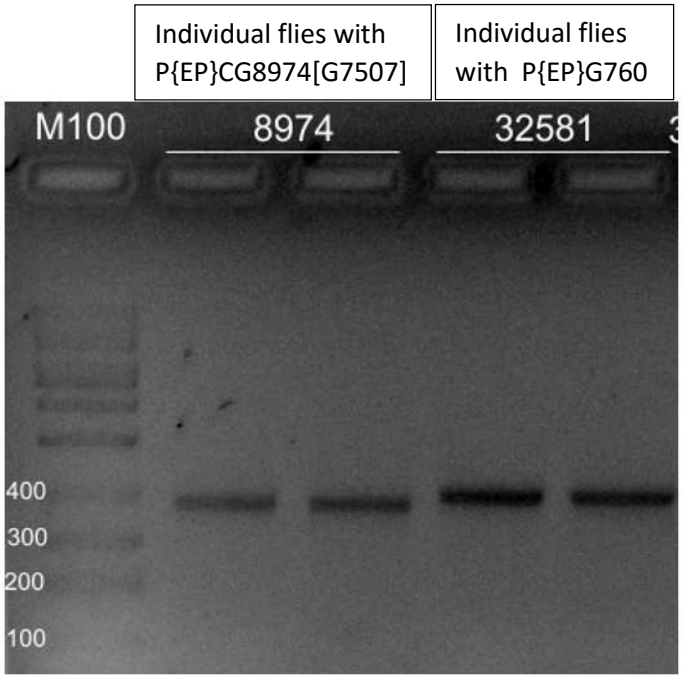


Detection of p-elements for overexpression of CG32581 and CG8974



DNA was isolated from individual flies with overexpressed CG8974 gene, with p-element P{EP}CG8974[G7507] upstream gene or with CG32581 gene overexpressed, with p-element P{EP}G760 inserted upstream gene sequence.

One primer for the enhacer within p-element and one specific primer for either CG8974 or CG32581 where used to find out if p-elements are mapped properly:



- Both p-elements for overexpression were detected

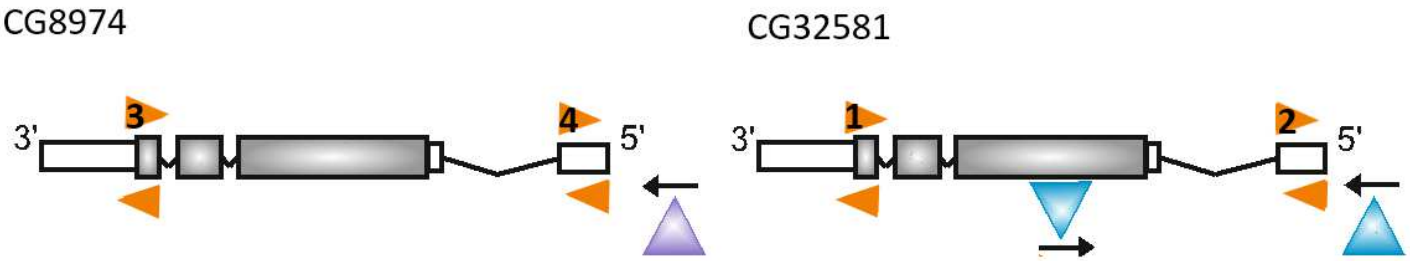
The next amplification was performed on DNA isolated from the same flies with potential overexpression of either CG8974 or CG32581 using gene-specific primers, in order to investigate if:

- flies with CG8974 overexpressed (p-element inserted upstream gene) have CG32581 expressed too:
  lane 1 - primers for the 3' end of CG32581

CG32581 For5	TGATATTAGAAGGCTACACGAAGA
CG32581 Rev5	TGCACTATTGGAACGAGCAG

lane 2 – primers for the 5' end of CG32581

CG32581_For4	TCAATTGATTGACTTGTCCACACA
CG32581_Rev4	CCTGACTTGTTTGCCCTCATT



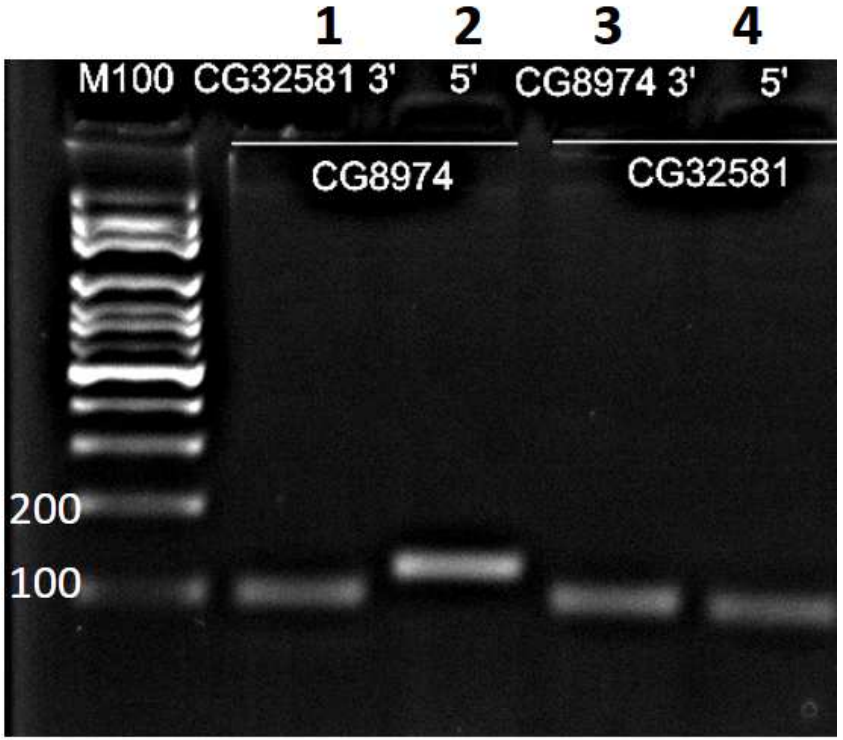
- flies with CG32581 gene overexpressed contain CG8974 gene

lane 3 – primers for the 3' end of CG8974

CG8974 For2	CTGCTAATCGGGGAACAAGG
CG8974 Rev2	TATGCATAGAACAGCCACAATATC

lane 4 – primers for the 5' end of CG8974

CG8974_F	ACATAACACACTTGTCCACACA
CG8974_R	TTTCGCTCACAGTTTGCTTCT



P-element P{EP}CG8974^{G757} (BL 33461) **for CG8974** overexpression is insterted properly, within the very start of the CG8974 transcript:

```
Sequence recovered from both 5' and 3' ends of
P element.
The P element insertion position is 75
In the 209 bases.
This insertion position refers to the first
base of the 8 base target recognition sequence
```

Source: Flybase

Recovered sequence from inverse PCR (Flybase)

SEQUENCE

CCGGTGGAGACGGGCTGTTATACAGGCAGCTATTTACCGATAGGGTAAAGGTCTATCGGC

▼

CTCGGTCAGGTATCGTCAGTGCAGTTGTTTTGAGGTGCTAAGACGTCAATTGATAAATTC

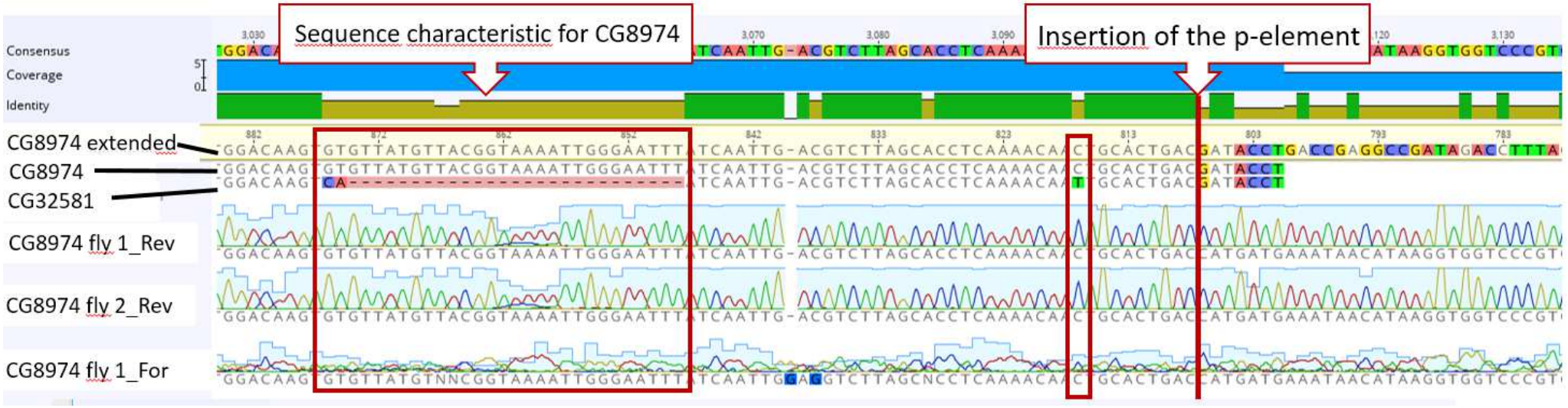
CCAATTTTACCGTAGCATAACACACTTGTCCACACAATCAATTTAGCAACATTCTCTGAT

TCAGTTTAACAATACTTAGCACGGCAAGT

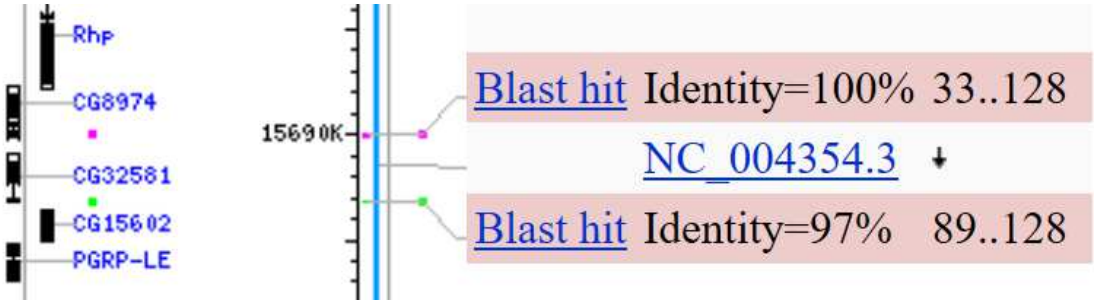
▼ P-element

Sequencing of the product of amplification performed with using 1 primer for enhancer within p-element and 1 primer for 5'end of either CG8974 gene revealed that p-element P{EP}CG8974^{G757} (BL 33461) is properly inserted:

Reversed sequences



BLAST result:



P-element P{EP}G760 (BL 26598) previously mapped upstream CG32581 is really upstream of CG8974:

```
The P element insertion position is 1 in the 626 bases.
This insertion position refers to the first base of
the 8 base target recognition sequence.
```

Source: Flybase

Sequence characteristic for CG8974 not CG32581:

Forward: **AAATTCCCAATTTTACCGTAACATAACAC**

Reversed: **GTGTTATGTTACGGTAAAATTGGGAATTT**

Recovered sequence from inverse PCR (Flybase)

SEQUENCE

GGTAAAGGTCTATCGGCCTCGGTCAGGTATCGTCAGTGCAGTTGTTTTGAGGTGCTAAGA

CGTCAATTGATAAATTCCCAATTTTACCGTAACATAACACACTTGTCCACACAATCAATT

TAGCAACATTCTCTGATTCGGTTTAACAATACTTAGCACGGCAAGTAGAAGCAAACAGTG

AGCGAAAACAAACAATAAAATGAGGGCAAACAAGTCAGGTTTTGCAAACAATACACTTAC

AATGTAAGTAATTTTGAGAAGTTATACTATTTGCTGTGCCACAGGAAGTTATATATAGTT

TAATTAAGTTATAATGCGTTTAAACGCAATCCCTGGAACTCGATGTATATGAGCATATAA

TTCCCCTTGACTCCAATAGGAAGTGCTCCATGAAAACATTATGTGACATCTATACGTACT

GTATAAATAAGCATATAATCGGTTCCAATGTACATATGTAACAGGTAGATGGAAACCTTC

TAGCCCTATATACACGCGAGTGTTTGAGTGTGTAGATGTGTACATAAGTACGCACATAAG

TCCAGGTACATATGCAGGTATACCGTCTTGTTTGCCAACGCAATTCANACAAAATTTGAC

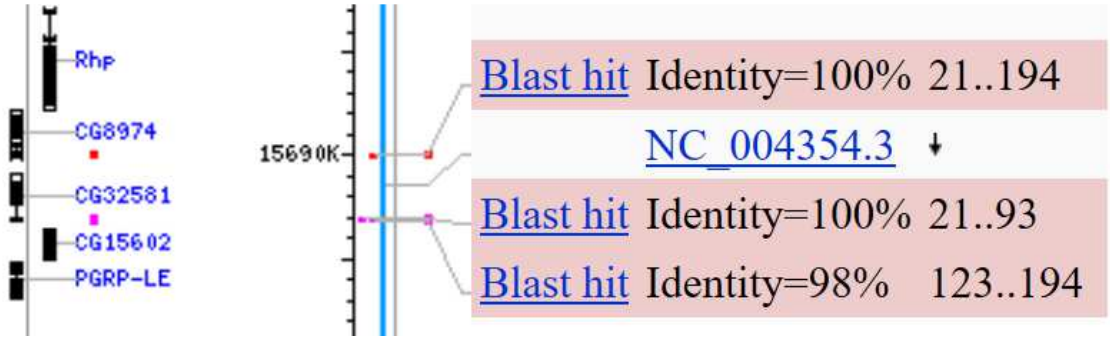
TTTTTTGTTTAATTTATGTGGGTGCA

P-element

Sequencing of the product of amplification performed with using 1 primer for enhancer within p-element and 1 primer for 5'end of CG32581 revealed that p-element upstream CG32581 (P{EP}G760) is mismapped and inserted upstream CG8974 instead:

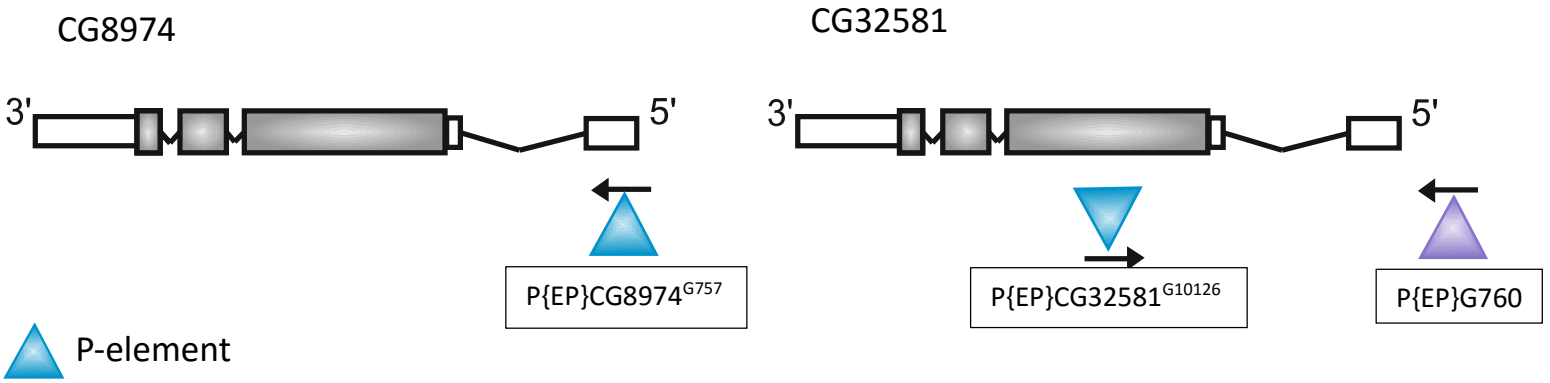
Reversed sequences

BLAST result:



Result: P-element P{EP}G760 (BL 26598) should be insterted upstream CG32581 but in fact is inserted upstream CG8974.

Location of p-elements taken from Flybase:



Where are p-elements inserted in practice?