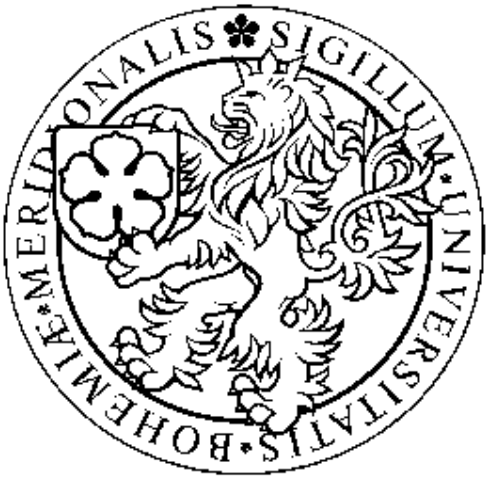
University of South Bohemia in České Budějovice

Faculty of Science

Department of Molecular Biology and Biochemistry



BACHELOR THESIS

Promoter analysis of *lin-3::gfp* transgene in *Caenorhabditis elegans*

Marion Sieber

Supervisor: Masako Asahina-Jindrová, PhD.

Institute of Parasitology, Academy of Sciences of the Czech Republic

Laboratory of Molecular Genetics of Nematodes

České Budějovice, 20.5.2011

Sieber, M., 2011:
Promoter analysis of *lin-3::gfp* transgene in *Caenorhabditis elegans*.
BSc. Thesis, in English – 20 pages, Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

ANNOTATION

In this work the promoter region of *lin-3::gfp* gene in a transgenic *C. elegans* strain was analyzed using PCR and sequencing.

I hereby declare under oath that I did the research work for this thesis by myself. I also declare that this thesis was written solely by me. Any additional sources and help from others are cited correctly.

I hereby declare that, in accordance with Article 47b of Act No. 111/1998 in the valid wording, I agree with the publication of my bachelor thesis, in full form to be kept in the Faculty of Science archive, in electronic form in publicly accessible part of the STAG database operated by the University of South Bohemia in České Budějovice accessible through its web pages.

Further, I agree to the electronic publication of the comments of my supervisor and thesis opponents and the record of the proceedings and results of the thesis defense in accordance with aforementioned Act No. 111/1998. I also agree to the comparison of the text of my thesis with the Theses.cz thesis database operated by the National Registry of University Theses and a plagiarism detection system.

České Budějovice, 20.5.2011

_____
Marion Sieber

This work is a part of a project which is financially supported by the grant GA ČR number 204/07/0948.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Masako Asahina-Jindrová, PhD., for giving me the opportunity to work in her laboratory and her kind support. I appreciate her confidence in me as well as her patience. Many thanks also to Nagagireesh Bojanala for his willingness to help me whenever I needed it. I am also obligated to Doc. RNDr. Marek Jindra, CSc. and his team for many helpful pieces of advice as well as for their warm welcome. It was a pleasure for me to work among all of them.

Furthermore I would like to use the opportunity to thank all my teachers, especially Prof. RNDr. Libor Grubhoffer, CSc. and Univ.-Prof. Mag. Dr. Norbert Müller, for making this unique study program possible.

Finally, my thanks go to my friends in Austria as well as in Czech Republic and to my family for their support.

AIM OF THE PROJECT

We have obtained a *C. elegans* strain carrying *lin-3::gfp* transgene from *Caenorhabditis* Genetic Center for *in vivo* expression study. However the information about the construct was insufficient and obscured, and the mapping of the construct was essential for further work. Therefore the goal of this project was to identify and map the exact promoter region of *lin-3* gene used to generate the transgenic worm.

CONTENTS

ABSTRACT

In *C. elegans, lin-3* activates the EGF signaling pathway responsible for proper development of the hermaphrodite vulva. As we are interested in the regulation of this process, a *C. elegans* strain carrying a *lin-3::gfp* transgene has been used for *in vivo* expression study. However the published and available information about the strain obtained from *Caenorhabditis* Genetic Center was insufficient and confusing. In this study, the promoter region of the transgene construct was amplified by PCR directly from the transgenic worm and sequenced. We were able to identify the region of the *lin-3* promoter used for the strain as well as the junction points between GFP vector and the promoter.

INTRODUCTION

C. elegans as a model organism

Caenorhabditis elegans (worm) is a non-parasitic soil nematode widely used as a model organism in developmental biology and genetics. As an adult it has an approximately one millimeter long, unsegmented and transparent body. The basic anatomy of *C. elegans* includes a mouth, pharynx, intestine, gonad, and collagenous cuticle. It is one of the simplest organisms possessing a nervous system.

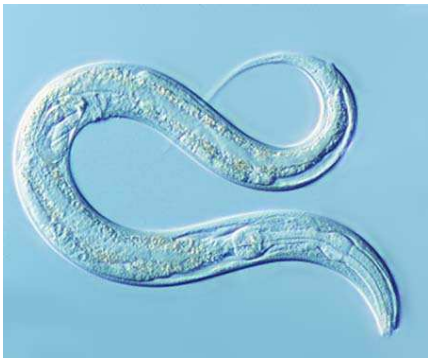


Figure 1: *C. elegans*

The body of an adult worm consists of a fixed number of somatic cells, 959 in hermaphrodites and 1031 in males. Cell linage has been thoroughly studied and has shown to be almost unchanging among individuals (Sulston and Horwitz, 1977). The worm's genome consists of five pairs of autosomes and one pair of sex chromosomes. In nature only 0.05% of the population are males (X0-sex chromosomes), while the vast majority are hermaphrodites (XX-sex chromosomes).

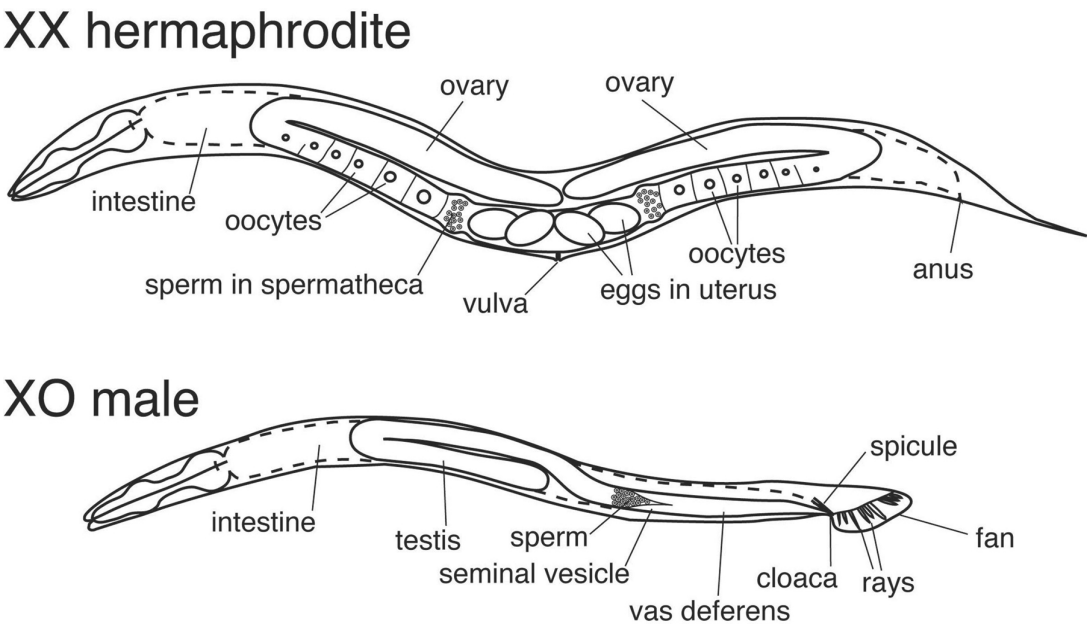


Figure 2: Body plan of both sexes

Hermaphrodites lay approximately 300 eggs. In case of fertilization by a male, the number of eggs can exceed 1000. After hatching from the egg, the animal passes through four larval stages before it molts one last time to become an adult. Although under optimal conditions the generation time is only 3 days, a worm can live 2-3 weeks as an adult. In case of low food supply or crowding the worm can choose an alternative developmental pathway leading to a larval stage called Dauer. These larvae are stress resistant and can survive without food for several months. As soon as they find nutrition they molt directly to the L4 larval stage.

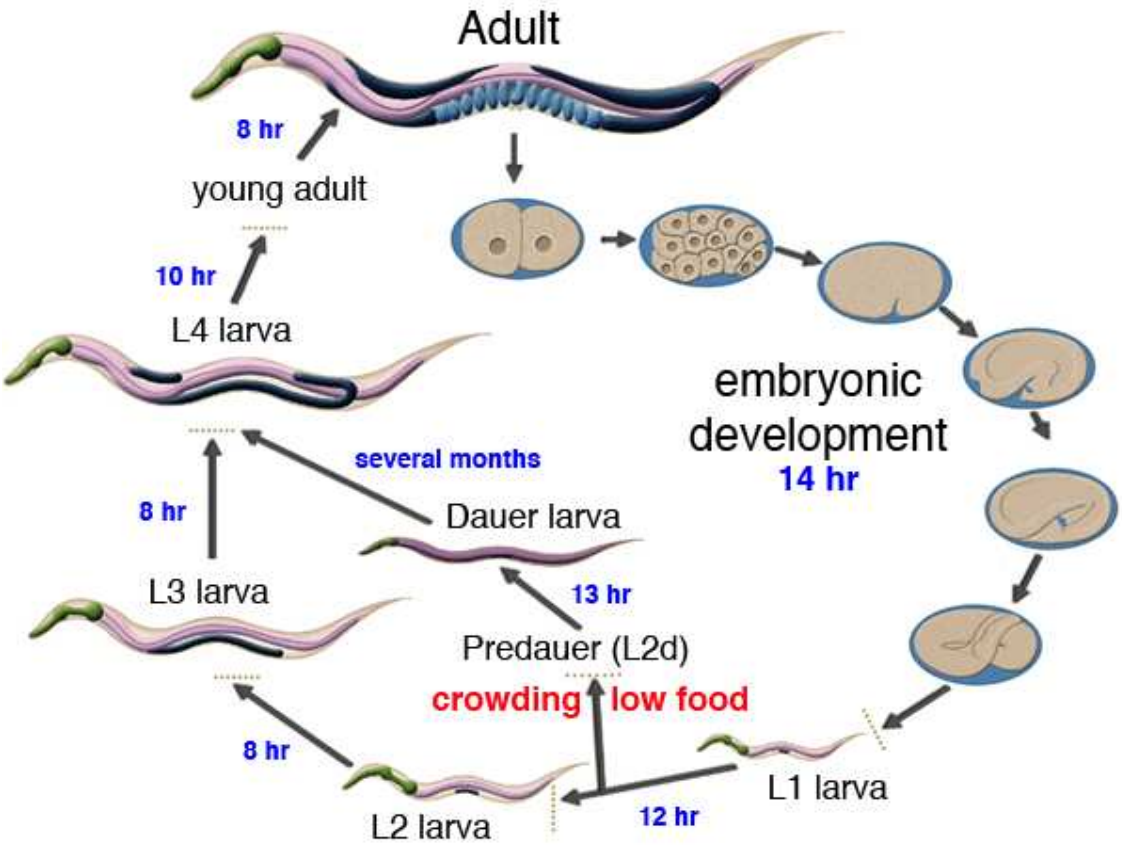


Figure 3: *C. elegans* life cycle showing both possible developmental pathways

Utilizing *C. elegans* as a model organism has several advantages. Firstly, it can be safely kept in laboratory culture (Brenner, S., 1974) and, as free living worms feed on bacteria, they can be easily nourished. They also have a short generation time which can be influenced by adapting incubation temperatures. Worms are very resistant and can even survive being frozen at -80 ℃ for a long time, providing a cost and time saving method to keep transgenic strains for further research. Additionally, their

transparent body allows the study of the living organism under the microscope in single cell resolution.

Green fluorescent protein (GFP)

GFP is a protein naturally occurring in jellyfish which exhibits a bright green fluorescence when exposed to blue light. As it does not harm the model organism, it is widely used as reporter of gene expression (Chalfie et. al., 1994). Vectors carrying the DNA-sequence encoding GFP has been intensively developed and used for promoter analyses and functional analyses of proteins. When a vector driven by a worm promoter is then injected into *C. elegans* gonad, the worm expresses GFP whenever and wherever the promoter is active. The choice of different promoters allows the scientist to monitor the worm's transcriptional activity.

PS4308*lin-3::gfp* transgenic strain

We obtained a C. elegans strain PS4308 carrying *lin-3::gfp* transgene from *Caenorhabditis* Genetic Center for *in vivo* expression study. The strain was generated by Hwang and Sternberg and in their paper is noted: "Two of the constructs contain either 10 kb (4) or 3 kb (3) of 5' upstream sequences from the first *lin-3* exon (E1). The other two contain either 4 kb (2) or 0.2 kb (1) of 5' upstream sequences from the putative second promoter (P2) in the fourth intron of *lin-3*." (Hwang and Sternberg, 2004) The article also indicates the type of vector used (*Δpes-10::gfp* enhancer assay vector) but unfortunately, the information regarding which of the four constructs mentioned above was used to generate PS4308 was lacking.

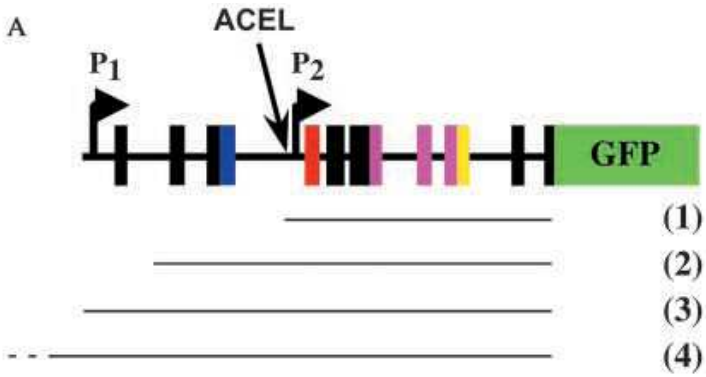


Figure 4: *lin-3::gfp* transgene constructs described in the paper of Hwang and Sternberg. The arrows (P_1 and P_2) indicate the two possible starting points for *lin-3* transcription. GFP stands for the GFP-containing vector and ACEL represents an anchor cell specific enhancer of *lin-3*.

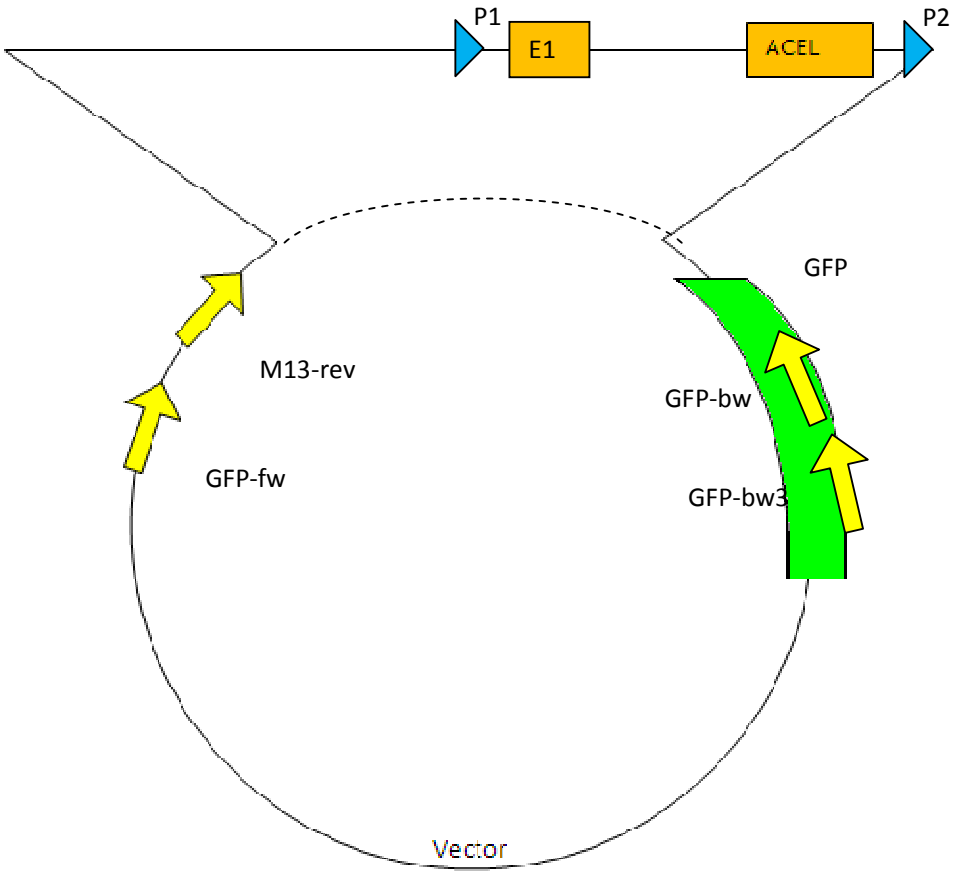


Figure 5: Graph of our starting knowledge about the transgene drawn in vector shape. Blue arrows (P1 and P2) indicate potential transcription start sites in the promoter. Yellow arrows indicate primers. E1: the first exon, ACEL: anchor cell specific enhancer of lin-3

LIN-3 as a ligand of the EGF-signaling pathway

The epidermal growth factor (EGF) family is involved in cell proliferation, differentiation and survival, and many other crucial processes. As its signaling pathway is highly conserved among species, knowledge gained by its study in relatively simple model organisms is in many cases also valid for the human body. In *C. elegans lin-3*, a member of this family which activates the EGF-signaling pathway in vulval precursor cells (VPCs) is responsible for the induction of vulva formation.

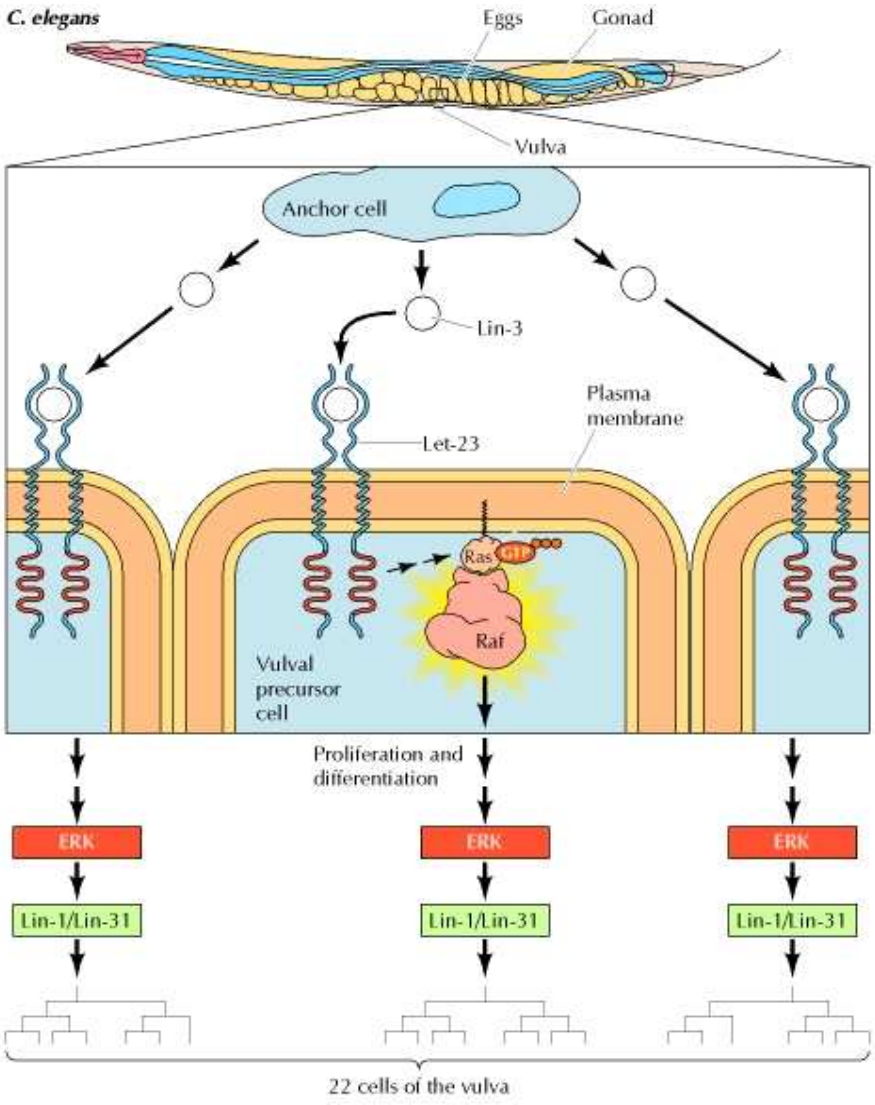


Figure 6: LIN-3, secreted by the gonadal anchor cell (AC), binds to the LET-23 (EGF) receptor expressed in vulval precursor cells (Aroian et. al., 1990). This binding leads to a cascade of reactions inducing three VPCs to proliferate and differentiate into the 22 cells of the vulva. (adapted from: The Cell, 2nd Edition, 2000)

MATERIALS AND METHODS

Growing worms and synchronization

E. coli bacteria (OP50) were grown on agarose gel plates overnight at 20 ℃. Healthy adult worms were transferred one by one to fresh plates with a worm picker. After approximately 3 days when a culture with many eggs was reached, worms and eggs were bleached with a bleaching solution (1 ml = 285 µl SAVO + 180 µl 4M NaOH + 535 µl H_2O) to synchronize the age of the worms as only the eggs survive the procedure. The eggs were put on fresh plates and after 3 days a culture of healthy adults was obtained.

Extraction of DNA from worms

The worms were transferred to 1.5 ml tubes using 1 ml S-basal gelatin worm buffer (0.59 m% NaCl, 5 V% 1M K_3PO_4, 0.1 V% Cholesterol 5 µg/ml in propanoic acid, 0.5 g/l gelatine). After spinning and removing the supernatant, 100 µl lysis buffer (10 mM Tris-HCl, pH 8.2, 50 mM KCl, 2.5 mM $MgCl_2$, 0.45% Tween20, 0.05 % gelatin, 0.2 mg/ml Proteinase K) were added to break the cuticle. The mixture was frozen at -80 ℃. Before use, the lysate was thawed, vortexed and frozen again for 3 times. Then it was warmed to 65 ℃ for 3 h and heated to 9 5 ℃ for 10 minutes.

Primer design for PCR and sequencing

Based on our knowledge about the vector and the inserted gene, the following primers were designed to subdivide the promoter and gain information about the length of the insertion.

Table1: List of primers.*lin-3* primers bind to the promoter region, GFP and M13 primers bind to the vector

Name	Sequence	Length
lin3-fw1	TGTGACCCTGAAAACTGT	18
lin3-fw2	GCATCCTTCTACTCTTTATGCC	22
lin3-fw3	CCCTTCGTGGTTTCGTCAA	19
lin3-fw4	CACTACGGTGACGGTTCA	18
lin3-bw1	GGCATAAAGAGTAGAAGGATGC	22
lin3-bw2	AGGAGGAAATGGTAGCGGG	19
lin3-bw3	TACAGTTGAGTTTCTGGG	18
lin3-bw4	TTGGAGTGAATCCAGAGG	18
lin3-bw5	GAAATGCGAACGCAGGAATAG	21
M13-rev	GGAAACAGCTATGACCATG	19
GFP-fw	GTGGAATTGTGAGCGGATAACA	22
GFP-bw	GACAACTCCAGTGAAAAG	18
GFP-bw3	CTTGTAGTTCCCGTCATC	18

Polymerase Chain Reaction (PCR)

PCR conditions were optimized to amplify DNA fragments of different regions of the transgene with many different sets of primers. A 50 µl mixture containing 5 µl buffer (10x), 4 µl dNTPs (stock of 2.5 mM of each nucleotide), 2 µl of each primer (10 µM stock), 2 µl DNA (worm lysate), 0.4 µl ExTaq polymerase (Takara) and 34 µl Milli-Q water was prepared. After the PCR reaction, PCR-products were analyzed by gel electrophoresis on 0.8% agarose gel (SERVA) in TAE buffer. 10 µl of each PCR sample were mixed with 2 µl loading dye and the mixture was placed in the wells of the gel. The 1 Kb Plus DNA Ladder (Invitrogen) was used to identify the length of the fragments. After the electrophoresis the gel was stained with ethidium bromide and a picture was taken (see Results and Discussion section). From the sizes of the bands, the length of the insertion was estimated and the most suitable primer-sets were chosen for the sequencing.

Then, two primer combinations (G: GFP-fw and lin3-bw1; H: lin3-fw2 and GFP-bw3) were used to amplify the desired regions. For each set we prepared 2 mixtures of 50 µl, containing 5 µl buffer (10x), 1.5 µl dNTPs, 1.5 µl of each primer, 0.4 µl Platinum Pfx DNA proofreading polymerase (Invitrogen), 1 µl DNA (worm lysate), 1 µl

$MgSO_4$ (5mM) and 38 µl Milli-Q water. A third set of primers fw1 and bw1 from *nhr-25* gene (5600 bp product – genomic) was used as a PCR control.

The optimal annealing temperature was determined by gradient PCR (Mastercycler gradient, Eppendorf) and conventional PCR was carried out with the following program (Biometra, TPersonal). 5 µl of each sample were loaded together with 1 µl loading dye on agarose gel.

Initial denaturation	94 ℃	5 min.	
Denaturation	94 ℃	30 sec.	Cycling 35 x
Annealing	57.2 ℃	30 sec.	
Extension	68 ℃	3 min.	
Final extension	68 ℃	5 min.	
Cooling	4 ℃	until stop	

Cleaning PCR-product

The PCR-product was cleaned with QIAquick PCR purification kit (QIAGEN). Because high concentration of the sample was needed, only 30 µl elution buffer were used. To estimate the concentration of DNA after the purification, the sample was run on a gel again.

Sequencing reaction

For sequencing reactions the primers below were used. The column called “Template” refers to the PCR product used, which is further specified (see Results and Discussion section).

Table 2: Primers and templates used for sequencing

Template	Primer	Template	Primer
G	lin3-fw2	H	M13-rev
G	lin3-fw1	H	lin3-bw3
G	GFP-bw	H	lin3-bw1

For each primer a mixture of 5 µl PCR-product (approx. 1 µg DNA), 4 µl sequencing reaction buffer (2.5x), 3.2 µl primer (1 µM stock), 4 µl Big-Dye Terminator kit (Applied Biosystems) and 2.8 µl Milli-Q water was prepared. The reaction was carried out in the Biometra TPersonal using the following temperature program:

Initial Denaturation	96 ℃	1 min.	
Denaturation	96 ℃	10 sec.	} Cycling 26 x
Elongation	60 ℃	4 min.	

Purification

For purification the product of the sequencing reaction was mixed with 80 µl isopropanol (80%) and the mixture was left in the dark for 15 minutes. After spinning at 13 000 g, 4 ℃ for 30 minutes the pellet was washed with 200 µl isopropanol (80%) and dried.

Sequencing

Sequencing was performed and analyzed using ABI PRISM 310 Genetic Analyzer.

Data processing

Using BioEdit (Hall, 1999), the sequencing results were aligned with the corresponding genomic sequence from the database and the sequence of the vector. This way we were able to determine the junction between vector and insertion as well as the enzymes used for cloning.

RESULTS AND DISCUSSION

Amplification of *lin-3* promoter in the transgene

Using vector specific primers (M13-rev and GFP-bw), the whole insertion was attempted to amplify from the worm lysate. However it did not give any PCR product though these primer set was working on other pure plasmid DNA. Therefore smaller fragments were amplified using different combinations of primers (Table 1, Table 3). The figure below shows the position of the *lin-3* primers in all four possible constructs described by Hwang and Sternberg. Construct (1) and (2) carry 0.2 kb and 4 kb of *lin-3* gene from a position between ACEL and the putative promoter 2 (P2) respectively. Construct (3) and (4) carry 3 kb and 10 kb *lin-3* gene upstream of the putative promoter 1 (P1) respectively.

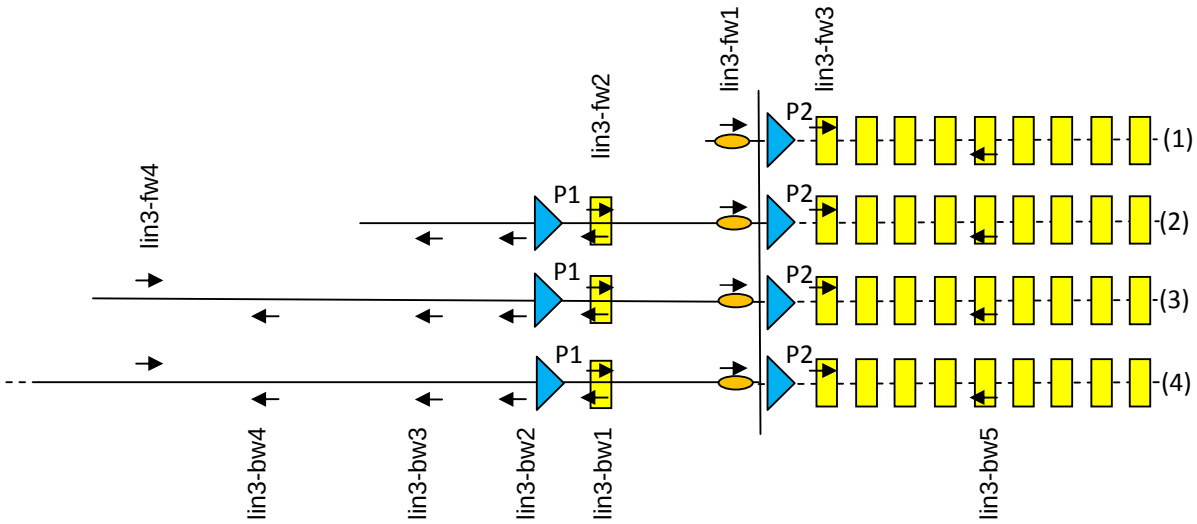


Figure 7: Primer position in all four constructs. The blue arrows (P1 and P2) indicate potential transcription start sites, yellow boxes are exons of *lin-3* and orange ovals represent the ACEL. The vertical line shows the presumable junction point to the GFP vector.

As preliminary experiments showed that the junction at the 3' end has to be relatively close after the ACEL, the last bp of the ACEL was chosen as starting point for the following calculations. The expected fragment length for the listed primer sets (ID A-J in Table 3), was calculated for each of the four possible constructs. The very right column in Table 3 shows the PCR products confirmed on the agarose gel (Figure 8).

Table 3: Primer combinations and their expected sizes of the products

ID	Primer set		Expected fragment length				Amplified length
	fw	bw	(4)	(3)	(2)	(1)	
A*	lin3-fw4	lin3-bw4	1626	1626	1626	1626	1600
B	M13-rev	lin3-bw4	10200	3200	-	-	-
C*	lin3-fw1	lin3-bw5	1800	1800	1800	1800	1800
D	M13-rev	lin3-bw5	-	-	-	-	-
E*	lin3-fw3	lin3-bw5	1671	1671	1671	1671	1650
F	lin3-fw3	GFP-bw	-	-	-	-	-
G	lin3-fw2	GFP-bw3	2430	2430	2430	-	2400
H	GFP-fw	lin3-bw1	13190	9190	2630	-	2600
I	GFP-fw	GFP-bw3	15660	8660	5100	1300	5000
J	M13-rev	GFP-bw3	15630	8630	5070	1270	5000

* These primer sets amplify endogenous *lin-3* gene in the worm genome. They were used to ensure that those primers were working properly.

PCR products amplified with primer-sets I and J showed the match with the expected sizes of the construct (2) and differed from other constructs. This result indicates that the 4 kb fragment was most likely the one driving the transgene in the worm PS4308 strain. The primer-sets G and H gave robust and more specific bands compared to I and J that should have the full insert. Therefore primer-sets G and H were used to amplify the template to further sequence the region. Figure 9 depicts primers in the vector and *lin-3* gene.

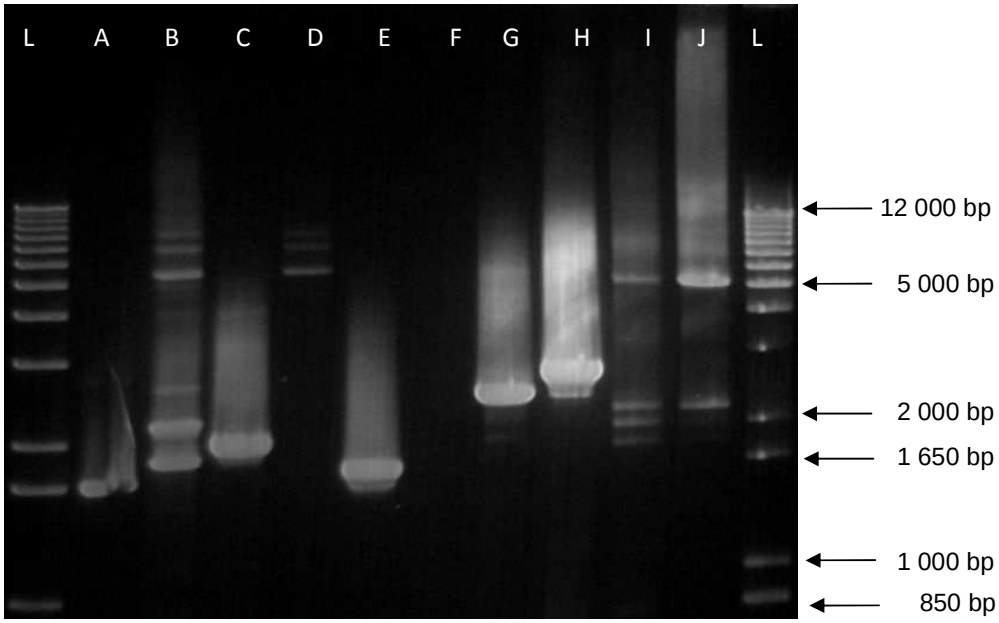


Figure8: PCR products with various sets of primers. Sizes are in agreement with the Construct (2). Primer-sets G and H gave robust and much more specific bands than I or J. L: 1 kb Plus ladder.

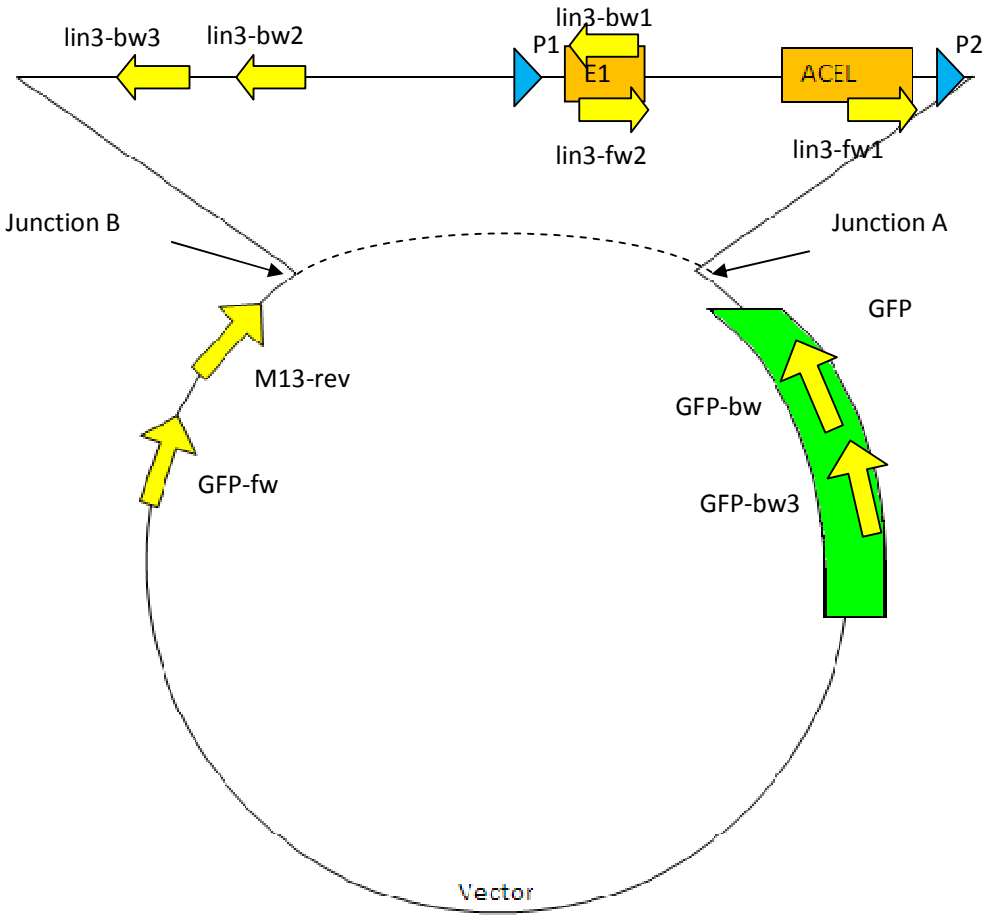


Figure 9: Primer positions in the construct.

Verification of the construct

To verify the construct used to generate PS4308 strain, PCR products were further analyzed by sequencing. As the PCR fragments shown above were not clean enough for sequencing, the annealing temperature was optimized using gradient PCR for the primer sets G and H. Primer set G amplifies the 3' end of the insert containing the Junction A and the primer set H amplifies the 5' end of the insert containing the Junction B (Figure 9).

Optimization of annealing temperature

To determine the optimal annealing temperature, gradient PCR was performed for the primer combinations G and H in the temperature range between 57 ℃ and 65 ℃. The optimal band for both primer-sets was achieved at 57.2 ℃ (Figure 10).

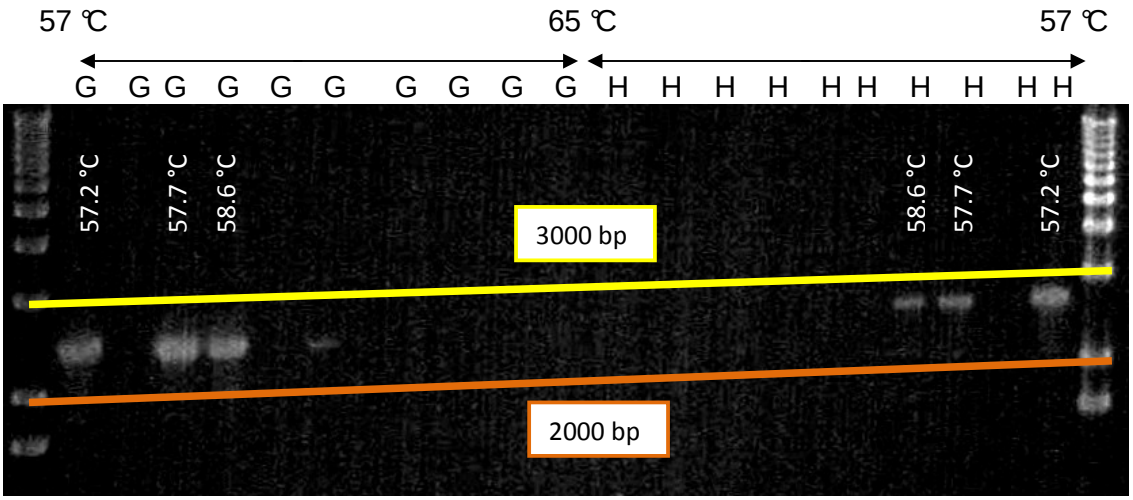


Figure 10: Optimization of annealing temperature by gradient PCR. PCR products of the primer set G on the left and H on the right. 57.2 ℃ gives nice bands for both primer sets.

Amplification of templates for sequencing

Conventional PCR was performed with the condition optimized by the gradient PCR. The primer-sets G and H amplify products with correct sizes, 2400 and 2600 bp respectively (Figure 11).

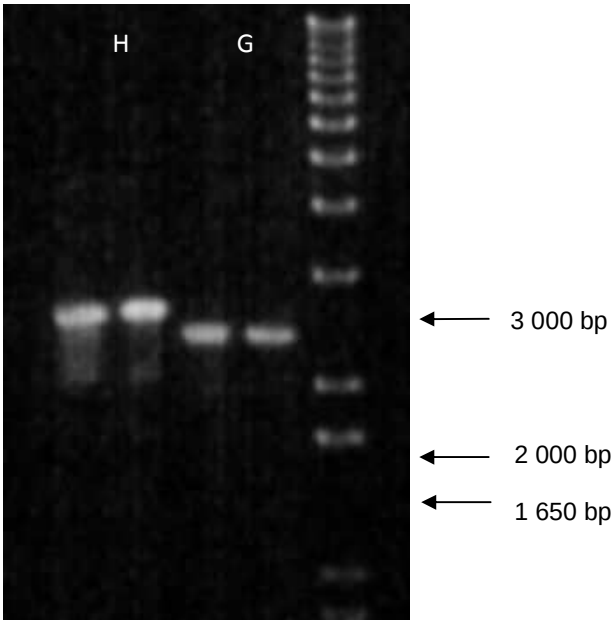


Figure 11: Optimized PCR products for sequencing. Robust and clean PCR products using primer-sets G and H were obtained by conventional PCR.

Sequencing PCR products

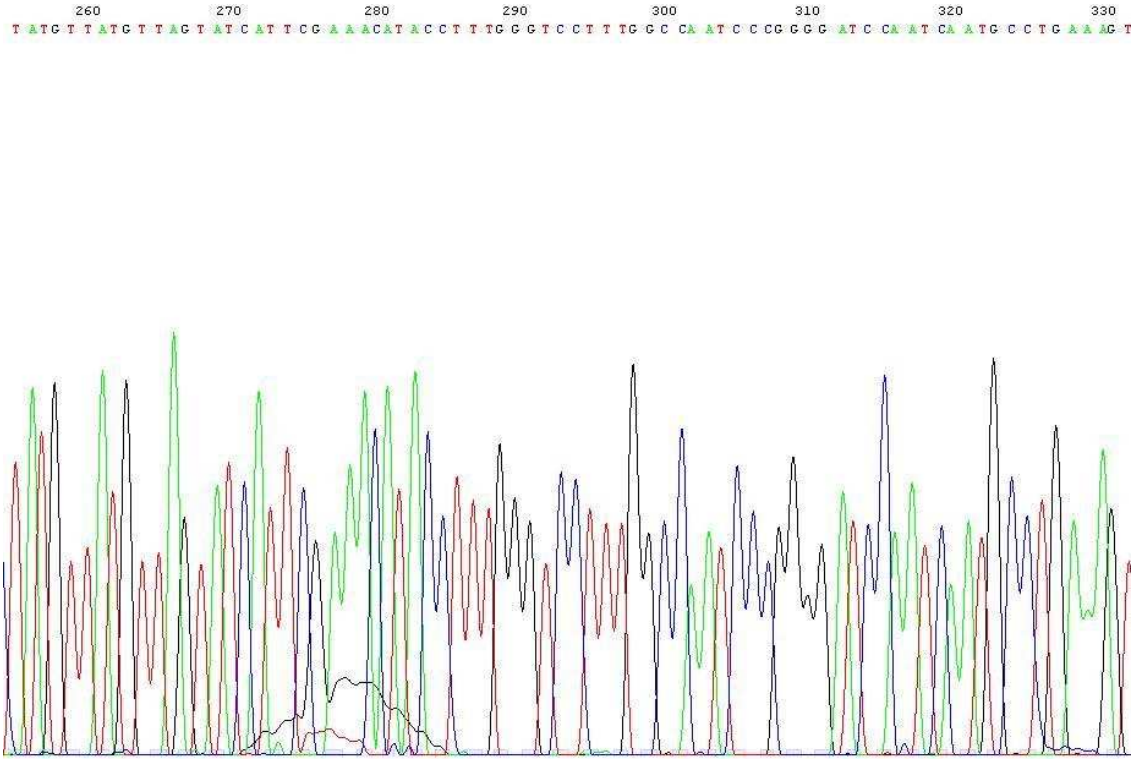


Figure 12: Example of sequencing results

Primers (M13-rev, lin3-bw3, lin3-bw1, lin3-fw2, lin3-fw1, GFP-bw) were used for sequencing. Approximately the first 500 to 800 bp were legible reliably. The unreliable sequences were cut and some of the unidentified nucleotides (N) were manually assigned by reading the raw sequencing data.

Using the ClustalW alignment algorithm (Larkin et. al., 2007) integrated in BioEdit we were able to identify the exact position of the junction between vector and the *lin-3* insertion. The distance between the two cloning sites in the *lin-3* gene corresponds to the length calculated from the PCR results shown before.

The alignments of the junctions were shown below (Figure 13). The mismatches around the junction (gaps) were seen and it might be due to the use of restriction enzymes for cutting and ligation of vector and *lin-3* gene. Nevertheless the almost perfect matching before and after the junction region allowed for a precise identification of the junction sites.

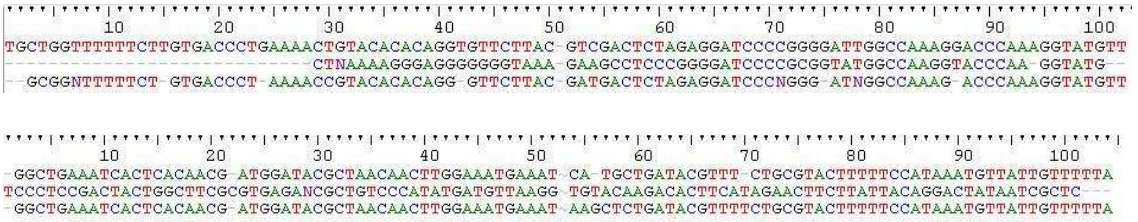


Figure 13: Alignment around the junction A (top) and B (bottom). The first line represents in both cases the chimera sequence obtained by artificially putting *lin-3* and GFP vector together. The other two lines show the sequencing results from the two different primers reaching the junction. For junction A these are lin3-fw1 and GFP-bw were used. For junction B M13-rev and lin3-bw3 were used. The quality of the match was very low for lin3-fw1, because the junction was too close to the primer. The result from the M13-rev primer also showed many mismatches. Therefore the sequences obtained from these two primers were excluded from further analysis.

Hwang and Sternberg (2004) mentioned that junction B lies approximately 4 kb upstream of the second putative promoter, which is shown to be right after the ACEL. Therefore our results for the position of junction B are in accordance with the literature. This was not the case for junction A. Figure 14 from the paper shows that the transgene contains several exons between the ACEL and the junction to the GFP vector.

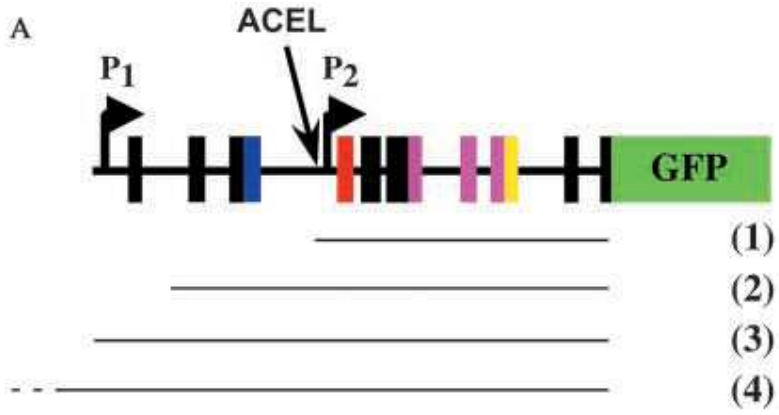


Figure 14: Constructs described in the paper of Hwang and Sternberg. Construct (2) seemed to be the one in the strain PS4308 which we obtained.

Quite contrary to Figure 14, we found the junction A to be within the last nucleotides of the ACEL. Therefore also the putative second promoter (P2) mentioned before is most likely not present in the transgenic line we are working with.

By looking at the junctions, we were able to identify the restriction enzymes used for cutting the vector as well as the *lin-3* gene. As shown in Figure 15 we found that the vector was cut with SalI for junction A and HindIII for junction B. For the 5' junction, HindIII was used for cutting the GFP vector and SphI was used to cut the *lin-3* gene. The HindIII site from the vector was filled while the gene site was chewed. Then both blunt ends were ligated. For the 3' end of *lin-3*, a restriction site might have been introduced during the amplification of the gene by PCR and blunted. The vector was cut by SalI blunted and ligated.

Junction A

lin-3	**CACACAGGTGTTCTTAC**
Vector	**gtcGACTCTAGAGGATCCCCGGGGA**
Sequencing GFP-BW	**CACACAGG_GTTCTTACGATGACTCTAGAGGATCCCNGGG_A**

Junction B

Vector	**CAACTTGGAAATGAAATAAGCTt**
lin-3	**gcatgCTGATACGTTT_CTGCGTAC**
Sequencing lin3-bw3	**CAACTTGGAAATGAAATAAGCTCTGATACGTTTTCTGCGTAC**

Figure15: Scheme of interface between vector and the *lin-3* insert. Shaded background shows the restriction enzyme sites in the vector (SalI in the Junction A, HindII in the Junction B) and in *lin-3* (a natural SphI site in the *lin-3* promoter). The small letters indicate nucleotides which can no longer be seen in the sequencing results due to the cloning.

Figure 16: Sequences of restriction enzymes used for cloning. From left to right: SalI, HindIII, SphI

CONCLUSION

In this study, the transgene was amplified directly from the transgenic worm by PCR and sequenced. We provided the evidence that we received a transgenic strain having the 4 kb fragment of *lin-3* gene, one of the constructs mentioned in the original paper (Hwang and Sternberg, 2004). Moreover the complete sequence of the insert was identified by aligning the sequencing results with the database information. The positions of the junctions between the GFP vector and the insert were confirmed. By analyzing the sequencing results in the region of the junctions the restriction enzymes used for cutting and ligation could be identified. The cloning sites were HindIII / SphI at the 5' junction and SalI for the 3' junction. Surprisingly, although it was mentioned in the original paper that the insert contains full coding genomic region of *lin-3*, this region did not exist in the construct and it was cut right after the ACEL region. Revised construct information is shown in Figure 17.

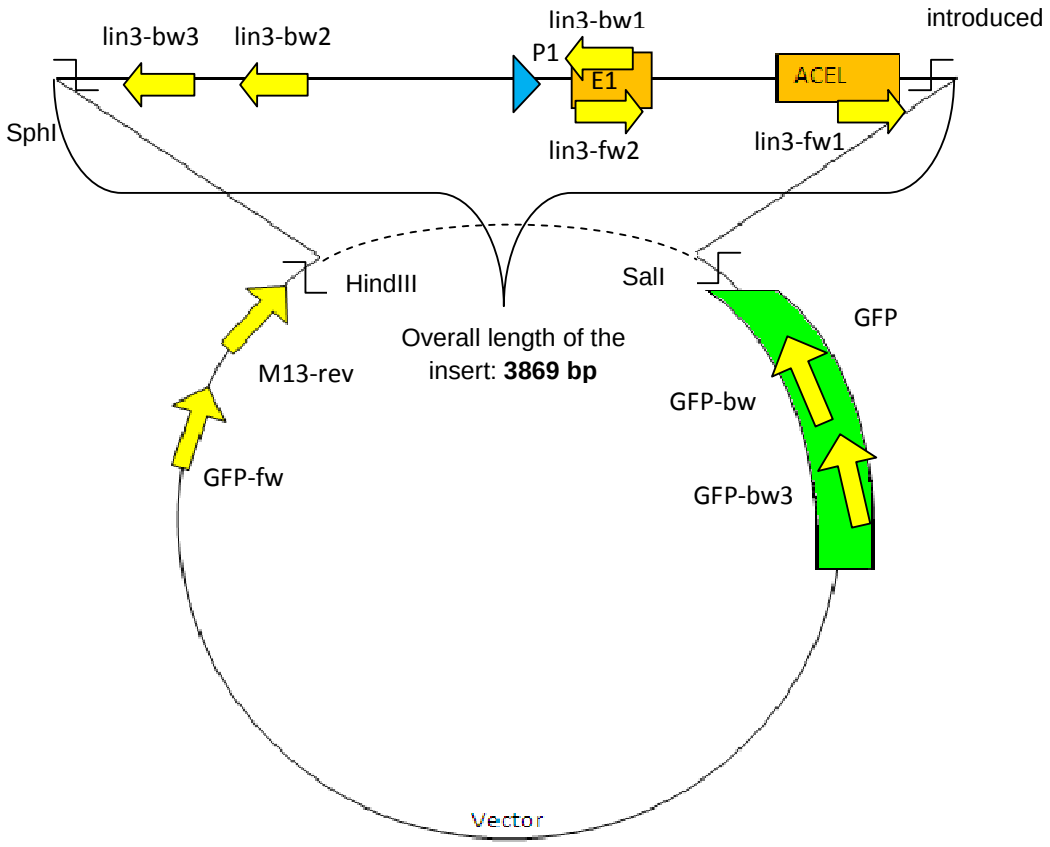


Figure 17: Revised transgene construct in the worm strain PS4308.

References

Aroian, R. V., Koga, M., Mendel, J. E., Ohshima Y. and Sternberg, P. W. (1990). The let-23 gene necessary for *Caenorhabditis elegans* vulval induction encodes a tyrosine kinase of the EGF receptor subfamily. Nature 348: 693–699.

Brenner, S. (1974). The genetics of *Caenorhabditits elegans*. Genetics 77, 71-94.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W. and Prasher, D.C. (1994). Green fluorescent protein as a marker for gene expression. Science 263, 802-805.

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98

Hwang, B. J. and Sternberg, P. W. (2004). A cell-specific enhancer that specifies *lin-3* expression in the *C. elegans* anchor cell for vulval development. Development 131, 143-151.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. and Higgins, D. G. (2007). ClustalW and ClustalX version 2. Bioinformatics 23(21): 2947-2948.

Sulston, J. E., Horvitz, H. R. (1977). Post-embryonic cell lineages of the nematode *Caenorhabditis elegans*. Dev. Biol. 56 (1): 110–56.

BIBLIOGRAPHY

Altun, Z. F. and Hall, D. H. (2005). Handbook of *C. elegans* Anatomy. In Worm Atlas. http://www.wormatlas.org/handbook/contents.htm

www.wormbook.org

www.wikipedia.org