Domain Engineering Sciences
Rte du Rawyl 47
CH-1950 Sion 2
Phone+41 27 606 85 11
Fax +41 27 606 85 75
info@hevs.ch

www.hevs.ch

Università degli Studi di Palermo
Facoltà di Scienze MM. FF. NN.
Dipartimento di Biologia Cellulare e dello Sviluppo
"A. Monroy"

Degree Course Life Technologies

Option Biotechnology

Diploma 2008

Alexandre Super

cis-regulatory sequences driving the expression of Ectodermin in *Paracentrotus lividus* sea urchin embryo

Professor **Prof. Dr. Sergio Schmid**

Expert **Prof. Dr. Giovanni Spinelli**

.....

Lieu, date de la remise du rapport

HES-SO Valais

SI	TV	EE	IG	EST
X	X			

Données du travail de diplôme
Daten der Diplomarbeit

FO 1.2.02.07.DB
mam/30/06/2008

Confidentiel / *Vertraulich* ☐ oui / ja ☐ non / nein

☐ FSI ☒ FTV	Année académique / *Studienjahr* 2007/2008	No PS / *Nr. PS* bt/2008/66
Mandant / *Auftraggeber* ☐ HES—SO Valais ☐ Industrie ☒ Ecole hôte	Mandataire / *Beauftragter* Alexandre Super	Lieu d'exécution / *Ausführungsort* ☐ HES—SO Valais ☐ Industrie ☒ Ecole hôte
Professeur / *Dozent* Prof. Dr. Sergio Schmid	Expert / *Experte* *(données complètes)* Prof. Dr. Giovanni Spinelli, Dipartimento di Biologia Cellulare e dello Sviluppo "Alberto Monroy", Università degli Studi di Palermo	

Titre / *Titel*
Cis-regulatory sequences driving the expression of Ectodermin in Paracentrotus lividus sea urchin embryo

Description et Objectifs / *Beschreibung und Ziele*

Ectodermin is Smad4 Ubiquitin Ligase essential for the specification of ectoderm germ layer during the early embryonic stages and for cell proliferation.

It is an inhibitor of TGF-β signaling during development, homeostasis and cancer.

A better understanding of its regulation could lead to novel therapeutic approaches against cancer involving these signaling pathways.

- Identify Ectodermin gene sequence in the sea urchin genome
- Identify the temporal and spatial expression patterns during embryogenesis
- Identify putative cis-regulatory elements by in-silico analysis
- Design promoter GFP-fusion constructs
- Perform microinjection of those contructs on sea urchin embryos and quantify expression of the transgene in space and time
- Perform Q-PCR to assay expression patterns of Ectodermin

Signature ou visa / *Unterschrift oder Visum*	Délais / *Termine*
Resp. de l'unité *Leiter der Einheit:*	Attribution du thème / *Ausgabe des Auftrags:*
Professeur/Dozent:	Remise du rapport / *Abgabe des Schlussberichts:*
Etudiant/Student:	Exposition publique / *Ausstellung Diplomarbeiten:*
	Défense orale / *Mündliche Verfechtung:*

Rapport reçu le / *Schlussbericht erhalten am* ……..……. Visa du secrétariat / *Visum des Sekretariats* ……..

Super, Alexandre **2008**

cis-regulatory sequences driving the expression of Ectodermin in *Paracentrotus lividus* sea urchin embryo

Résumé

Objectifs

L'Ectodermin est une ubiquitin ligase qui favorise la dégradation du facteur de transcription Smad4 (Dupont et al., 2005), régulant les processus de différentiation et de prolifération cellulaires. Un gène codant pour cette enzyme a été identifié dans l'espèce d'oursin Paracentrotus lividus. Des expériences préliminaires (RT-PCR et ISH) ont révélés qu'il était probablement impliqué dans la spécification de l'ectoderme oral dans l'embryon.
L'objectif de ce projet est de révéler la function et les caractéristiques d'expression de cette enzyme pendant l'embryogenèse. Pour cela, il est tout d'abord nécessaire d'analyser in-silico la séquence d'ADN pour trouver les sites de cis-régulation putatifs. Des transgènes promoteur-GFP doivent également être réalisés et injectés afin d'observer les motifs spatiaux et temporels d'expression. L'expression ectopique de l'Ectodermin, engendrée par l'injection d'ARNm synthétique, devrait permettre de formuler des hypothèses préliminaires quant à la fonction de cette enzyme dans les processus de l'embryogenèse.

Résultats

L'analyse in-silico a permis de révéler plusieurs facteurs de régulation potentiels tels que Twist, TCF, GATA, Cdx, Nkx ou encore HNF. L'implication ou non de ces sites devra être déterminée par des expériences complémentaires.
Trois transgènes (contenant des fragments de promoteur de 1.1, 1.45 et 2.5kb) ont été réalisés et injectés. Jusqu'à présent aucune expression de GFP n'a pu être mesurée dans les embryons injectés. Des investigations in-silico complémentaire ont révélés qu'une séquence de 2.5kb adjacente au terminateur du gène pourrait être impliquée dans la régulation de l'expression génétique.
L'expression ectopique de l'Ectodermin a induit des altérations remarquables au niveau des structures du squelette de la larve pluteus.

Mots-clés

Ectodermin, oursin, Paracentrotus lividus, embryogenèse, skeletogenèse, ectoderme oral, microinjection

Abstract

Objectives

Ectodermin is a Smad4 ubiquitin ligase that is essential for ectoderm differentiation in embryos and control of cell proliferation (Dupont et al., 2005). Ectodermin-like gene has been identified in Paracentrotus lividus sea urchin. Preliminary RT-PCR and ISH give hints that Ectodermin is probably involved in the oral ectoderm specification in P. lividus.

The goal of this project is to explore the function and the expression characteristics of this enzyme during the sea urchin embryogenesis. For this purpose, in-silico analysis of the putative cis-regulatory elements has to be performed. Promoter-GFP transgenes have to be designed and injected in sea urchin zygotes in order to assess spatial and temporal expression patterns. Ectodermin mRNA has to be synthesized to induce ectopic expression of the enzyme in the embryo, providing insights of its function.

Results

In-silico analysis of the promoter region has revealed putative sites for factors such as Twist, TCF, GATA, HNF, Cdx and Nkx that could be linked to cis-regulatory mechanisms after further experiments.

Three promoter-GFP transgenes (with promoter fragments of 1.1, 1.45 and 2.5kb) have been designed, constructed and injected. So far no GFP expression could have measured in injected embryos. After additional in-silico analysis of the genome, it has appeared that a 2.5kb non-coding region following the terminator of the gene could be as well involved in the cis-regulation.

Injection of functional Ectodermin mRNA has induced noticeable alterations in the skeleton structures at the pluteus larva stage.

Keywords

Ectodermin, sea urchin, Paracentrotus lividus, embryogenesis, skeletogenesis, oral ectoderm, microinjection

Summary

Nomenclature

AP Alkaline Phosphatase

BCIP 5-Bromo-4-Chloro-Indolyl-Phosphate

BMP Bone Morphogenic Protein

Cdx Caudal homeobox

GATA GATA-binding transcription factor

GFP Green Fluorescent Protein

GRN Gene Regulatory Network

HNF Hepatocyte Nuclear Factor

Hox Homeobox

IPG Injection Pressure Generator

ISH In-Situ Hybridization

MFSW Millipore-Filtered Sea Water

MT Myc Tag

NBT Nitro Blue Tetrazolium

Nkx Nk homeobox

PMC Primary Mesenchyme Cell

RING Really Interesting New Gene (type of zinc finger domain)

RT-PCR Reverse Transcriptase Polymerase Chain Reaction

Smad Contraction of SMA (*C. elegans*) and MAD (*D. melanogaster*)

SMC Secondary Mesenchyme Cell

TCF Transcription Factor

TGF-β Transforming Growth Factor Beta

1. Introduction

Embryonic development is a phenomenon strictly regulated by signaling proteins organized in pathway and networks [1]. These signals act as temporal and spatial switches on the expression of the genes involved in the differentiation and growth of the embryonic tissues. Regulatory events during the embryogenesis are not linear sequences of events driving activation or repression of the transcription. The regulatory genes have many *cis*-regulatory elements modulating their expression in function of their tissue localization and the stage of the development. This property of the gene regulatory networks (GRNs) allows them to trigger a huge amount of genetic events across lifetime with a limited set of transcription factors. The versatility of the regulation mechanisms make them highly conserved through evolution. For example homeodomains of the fruit fly and the chicken share such high similarity that the replacement of the fly gene by the chicken homolog ensures complete functionality.

Indeed the action of these pathways is not limited to the early stages of the organism development; studies reveal their implication in tissue degenerative diseases or abnormal cell proliferation pathologies such as cancer. Considering this, a better understanding of the gene regulation mechanisms during the embryogenesis could lead to significant progresses in designing novel therapeutic strategies against such pathologies.

Late studies on *Xenopus laevis* (African clawed frog) have revealed a new player involved in the ectodermal fate of the embryonic cells, a protein so called Ectodermin [2]. Ectodermin is a RING-type ubiquitin ligase interacting with the Smad4 receptor in cell nuclei. Smad4 belongs to a family of transcription factors regulated by TGF-β proteins. By promoting degradation of Smad4, Ectodermin attenuates the cell response to the mesoderm-inducing signals of TGF-β proteins Nodal and BMP4 (figure 1).

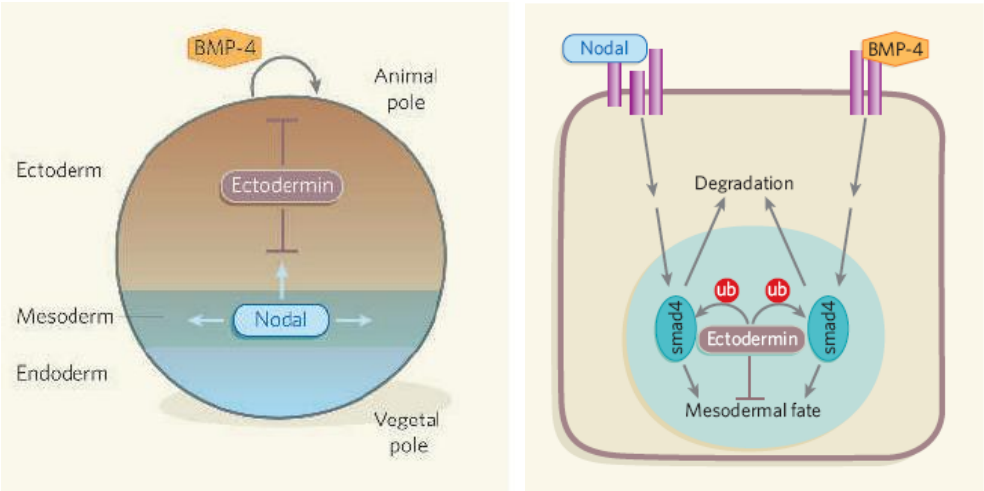


Figure 1 : Ectodermin function in *X. laevis* embryos [3]

The TGF-β signaling pathway is well-known in cancer biology for its anti-proliferative action. TGF-β/Smad4 signals prevent the cells to escape cytostasis after the final differentiation. Ectodermin could

be a key switch in malignant cell growth as many types of tumors conserve a wild-type Smad4. The experiments of Dupont [2] on colorectal cancer cells C32 expressing high level of Ectodermin and wild type Smad4 showed that depletion of Ectodermin leads to significant recovery of the TGF-β signaling functions.

Nodal factors have been implicated in several developmental processes in the chordates, such as mesoderm and endoderm formation or patterning of the embryo along anterior-posterior or left-right axis [4]. More recently, presence of *nodal*-related genes has been reported in nonchordates organisms like sea urchin [5]. In sea urchin embryo, Nodal signal leads to the differentiation of the oral ectoderm during the early stages (60- to 128-cells) of the embryogenesis (figure 2). Experiments showed that overexpression or inihibition of Nodal leads to severe disturbance in the oral-aboral polarity. Embryos overexpressing Nodal present radialized phenotype (similar to embryos treated with nickel) with multiple disorganized spicules growing along the vegetal-animal axis. To the opposite, inhibition of Nodal expression leads to the total absence of oral ectoderm and no spicules formation. So Nodal is one of the leading elements in the apparition of oral ectoderm in sea urchin, even if most of the process is still unknown.

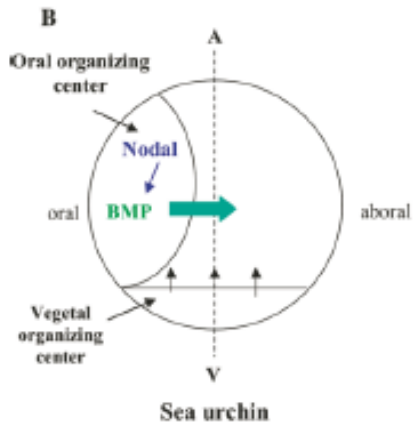


Figure 2 : Nodal pathway in sea urchin embryo [5]

Two sequences for an Ectodermin-like enzyme have been identified in the *Paracentrotus lividus* genome (data not published), but the function, spatial and temporal occurrence of this protein are still unknown in sea urchins. Given the state of knowledge on *X. laevis* Ectodermin and the Nodal pathway in sea urchin embryos, it has been proposed that Ectodermin could act as a repressor on Nodal signal and block the differentiation of the oral region in the sea urchin embryo.

Sea urchin as a model organism of developmental biology

Developmental biology relies on the use of several model species to study and explain the mechanisms leading to the formation of a complete and functional organism. Evolutionary conserved mechanisms among species and genres allow the construction of developmental models by cross-examination of the results. The model organisms are selected for the simplicity of their genomes, their short development

time, and their easy maintenance and reproduction. Historically, the first organism used and probably the most famous is the fruit fly (*Drosophila melanogaster*). Today a large variety of organisms are used as model such as the chicken (*Gallus gallus*), the african clawed frog (*Xenopus laevis*), the zebrafish (*Danio rerio*), the roundworm (*Caenorhabditis elegans*) and various species of sea urchins.

Sea urchins present many advantages to be used as model organism for experiments. They are commonly found in all the seas of the world, so research centers close enough to the sea don't need any breeding facilities for maintaining a sufficient amount of specimens for experiments. Males and females eject sperm and eggs in their environment for reproduction and the embryos grow in the plain water, absorbing the nutriments they found there. In consequence, reproducing these conditions in-vitro requires only fresh sea water and a temperature-controlled room, neither expensive feeding medium, nor special attachment-matrix is required for accomplish the embryogenesis. Development is quite fast, approximately 45 hours from the zygote to the pluteus larva stage in *Paracentrotus lividus*. Genetic modifications can be easily introduced by microinjection, as the eggs are only protected by a thin jelly-coat removable by chemical treatment, and observed as the embryos remain transparent until the very late stages of the development when the calcification of the shell occurs. Development is highly synchronous, that means that all the embryos of one fertilized batch will develop with the same time course. In addition, sea urchin embryos are very robust, so sterility is not a major issue during the experiments.

Paracentrotus lividus (purple sea urchin) is endemic specie of the Mediterranean Sea and the eastern Atlantic Ocean. It has an average diameter of 7 centimeters and harbors various dark color phenotypes of brown, purple and green. Males and females are hardly distinguishable as size and color cannot be used for discrimination. Optimal temperature for development is around 18°C. Collect of fertile urchins is a seasonal activity in Sicily, going from December to May, assuming that the weather is not too hot, or that the winter storms didn't washed away all the gametes stocked by the specimens. *P. lividus* feeds mainly on algae and little invertebrates but can be easily maintained in tank of sea water with a filtration system during six months to one year without external feeding.

Development course of an embryo [6]

In-vitro, fertilization is a very fast event that lasts no more than one or two minutes. After fertilization, the cortical reaction occurring to prevent polyspermia causes the expansion of the vitelline layer around the zygote, providing a visual indicator of a successful fertilization.

The cleavage of the zygote starts after two hours with the first cell-division. Next divisions will follow every twenty minutes. Cleavage in sea urchin is reductive, that means that the number of cells increases after each division without affecting the size of the embryo. The cleavages start to be unequal after the fourth, producing four micromeres surrounded by four macromeres at the vegetal pole (figure 3).

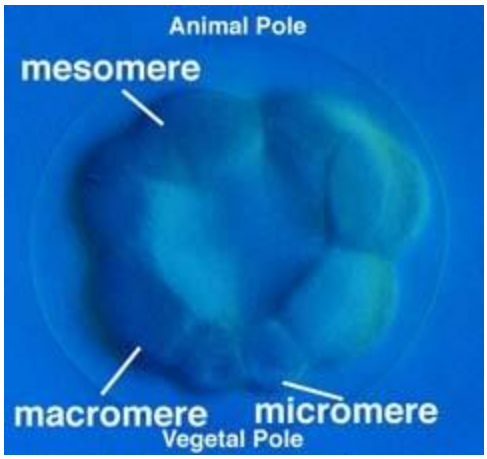


Figure 3 : 16-cells stage sea urchin embryo.

After eight hours, the morula transforms into blastula and swims freely in the water. The vegetal pole thickens to form vegetal plate. From this region, the primary mesenchyme cells (PMCs) start to ingress in the blastula cavity. The PMCs will give rise to the larva skeleton in later stages.

Following the ingression of PMCs, the vegetal plate invaginates to form the archenteron, during the early gastrula stage (around 24 hours after fertilization). The ingression of the secondary mesenchyme cells (SMCs) from the tip of the forming archenteron leads to its extension toward the oral ectoderm, near the apical plate.

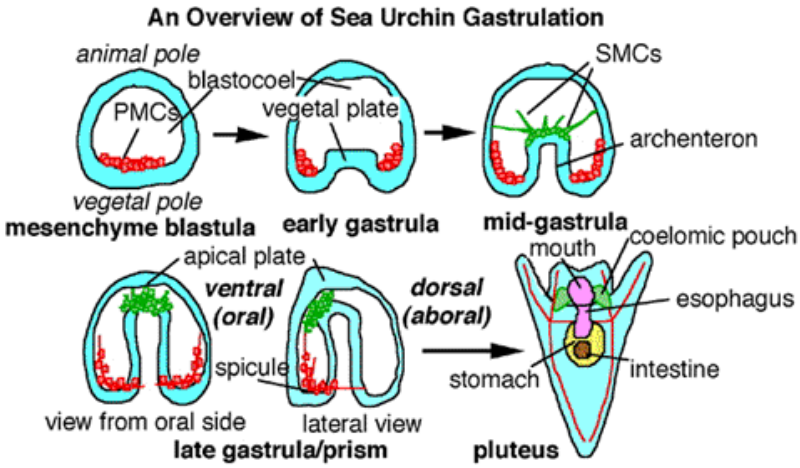


Figure 4 : Overview of the sea urchin development from blastula to pluteus stage.

When the mouth is formed by the fusion of the archenteron and the oral ectoderm, the embryo reaches the prism-stage and can start to feed. During this stage, the spicules, generated by the mineralization of the PMCs, start to radiate from the oral side to form the primitive skeleton of the larva.

Embryos achieved their transformation into pluteus larvae two days after the fertilization, by deploying four mineralized arms around the oral region. They will stay at this stage, swimming freely and feeding, several weeks before their final metamorphosis into sea urchins.

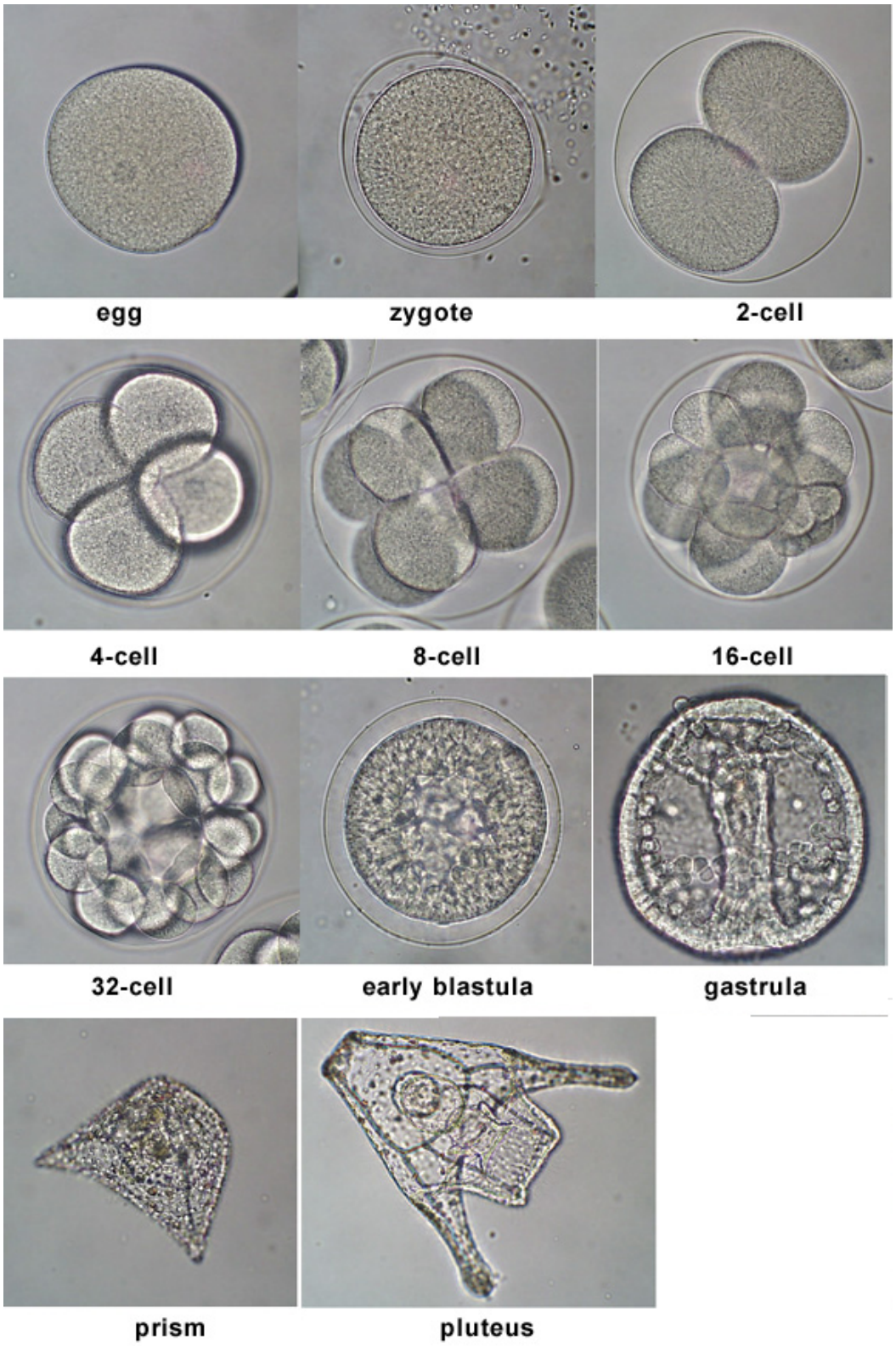


Figure 5 : Overview of the embryogenesis from the egg to the pluteus larva.

2. Objectives

The long-term goal of this project is to obtain a complete understanding of the function of the Ectodermin during the development of *P. lividus* embryos. To achieve this objective, it is necessary to collect data about the spatial and temporal expression of the enzyme in the embryo, and to quantify this expression. For this task, In Situ Hybridization (ISH) and RT-PCR have been already performed at the successive stages of the development.

As Ectodermin is supposed to be an enzyme promoting the degradation of the Smad4 transcription factor, the exploration of its connections with known GRNs in sea urchin embryos and the *cis*-regulatory elements driving its expression has to be performed. Putative transcription factors in-silico analysis will be carried on the non-coding regions flanking the two *ectodermin* copies in *P. lividus* genome sequence. This analysis will help to define what fragment of these regions could be determinant in driving Ectodermin expression.

In order to obtain insights of the non-coding region involved in the *cis*-regulation, fragments of the upstream region ("promoter" region) will be cloned in a pHE-GFP vector, fused in frame with a coding sequence for the Green Fluorescent Protein (GFP) marker. The promoter-GFP transgene will be injected in fertilized eggs, and observation of the development should reveal expression patterns dependent of the regulatory modules present on the fragment of promoter cloned [7].

Ectopic expression of Ectodermin will be induced by injecting synthetic mRNA of the gene in fertilized eggs. This should reveal insights of the function of this enzyme during the embryogenesis.

3. Materials & Methods

All the protocols written in this section are taken and adapted from *Molecular Cloning – A laboratory Manual* 2nd Edition [8].

3.1. *In-silico* analysis of *ectodermin*

Paracentrotus lividus genome sequence was already available. Full length genome has been analyzed trough GENSCAN online software (http://genes.mit.edu/GENSCAN.html) to detect putative sequences of Ectodermin.

Putative cis-regulatory elements of 5' and 3' flanking sequences of Ectodermin have been analyzed with TESS online software (http://www.cbil.upenn.edu/cgi-bin/tess/tess). Results have been compared to existing literature and TRANSFAC database (http://www.cbil.upenn.edu/cgi-bin/tess/tess?RQ=NBqt).

Alignments of sequences have been performed on the GPL software GENtle v1.9.4 of the University of Cologne (http://gentle.magnusmanske.de/) with the ClustalW algorithm included, and the ClustalW online software (http://www.ebi.ac.uk/Tools/clustalw2/index.html).

Predicted restrictions maps of DNA sequences have been realized with pDraw32 DNA Analyis software (http://www.acaclone.com/).

3.2. Reverse Transcription PCR

Materials:

- Titan One Tube RT-PCR System , Roche, Mannheim, Germany
  Cat No. 11 855 476 01
  Contains: Enzyme mix (Expand High Fidelity, Reverse Transcriptase, AMV), RT-PCR reaction buffer 5x, $MgCl_2$ stock solution, DTT solution.

Primers:

<u>Ect1 (forward):</u>

5'-TCA TAT AGA CTT TTC GTA ATG G-3'

Tm= 58°C

<u>Ect3 (reverse):</u>

5'-ACT TCT CAT CCT CCC TTG G-3'

Tm= 58°C

Methods:

Procedure followed is identical to the reference manual coming with the Roche kit.

3.3 *In-Situ* Hybridization

ISH is performed on embryos at different stages of development on the basis of a three days-long protocol.

Solutions:

PBST:

NaCl	137mM
KCl	2.7mM
Na_2HPO_4	10mM
KH_2PO_4	10mM
Tween 20	0.1% v/v

SSCT:

NaCl	150mM
Sodium Citrate	15mM
Tween 20	0.1% v/v

TBST:

NaCl	150mM
Tris-HCl pH8.0	10mM
Tween 20	0.1% v/v

Blocking Solution:

- TBST
- Serum
- BSA 5%

AP Buffer:

NaCl	150mM
$MgCl_2$	50mM
Tris-HCl pH8.0/9.5	100mM
Tween 20	0.1% v/v

TBSTE:

- Add 20mM of EDTA to TBST solution

Protocol:

- Wash embryos with PBST
- Incubate 45 minutes at 55°C with the hybridization solution for pre-hybridization
- Add tRNA to each sample to saturate non-specific site for the probe
- Denaturate the probe 5 minutes at 55°C
- Incubate overnight the samples with probe at 55°C
- Wash at increasing stringency the samples with hybridization solution, PBST and SSCT at 65°C
- Wash at room temperature with TBST during the time of sedimentation
- Incubate 30 minutes at room temperature with the blocking solution to saturate non-specific sites for the antibody
- Incubate at room temperature with 0.2 µl per sample of alkaline phosphatase-coupled anti desoxygenase
- Wash the samples with TBST, AP buffer pH8.0 and AP buffer pH9.5
- Add NBT/BCIP (substrates for AP) in AP buffer pH9.5. For the samples beyond the gastrula stage, add tetramisole hydrochloride 10x to inhibit endogenous AP
- Incubate at 4°C overnight
- Wash samples three times in TBSTE and one time in Glycerol 25%
- Fix the samples on a glass slide with Glycerol 50% and take pictures

3.4. Cloning

3.4.1. Primer design

Primer design have been performed with the help of the Primer3 (http://biotools.umassmed.edu/bioapps/primer3_www.cgi) and OligoAnalyzer from Integrated DNA Technologies (http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/) online softwares.

Method:

Three sets of primers have been designed in order to amplify different lengths of the upstream region of the coding sequence ("promoter" region). The three amplicons share the same reverse primer positioned on the start codon of the coding sequence in order to achieve the fusion with a GFP reporter gene. The sequences of the primers have been extended to include restriction sites needed for the further cloning procedures (*EcoR*I for the forward primers and *Nhe*I for the reverse primer) and 5 nucleotides have been added in 5' of the restriction sites to ensure correct amplification.

pEct1-*EcoR*I

5'-CC CCG GAA TTC GGG CAG TAA CCG TCA ATT-3'

30 bp, Tm = 60°C

pEct2-*EcoRI*

5'-CC CCC GAA TTC AGA TTT GAA GTA AAG TGC C-3'

29 bp, Tm = 58°C

pEct3-*EcoR*I

5'-CC CCG GAA TTC GGA GCC AAG TAA TAC ATG-3'

29 bp, Tm = 58°C

pEct4-*Nhe*I

5'-CC CCC GCT AGC CAT AAC GAA AAG TCT ATA T-3'

30 bp, Tm=60°C

Restriction sites are underlined; start codon (ATG) on reverse primer is highlighted in red.

3.4.2. Polymerase chain reaction

Materials:

- Elongase Amplification System, Invitrogen, Carlsbad, USA
  Cat No. 10481-018, Lot No. 1200567
  Contains: Elongase Enzyme Mix, Buffer A 5x, Buffer B 5x, dNTPs 10mM
- DyNAzyme II DNA Polymerase, Finnzymes, Espoo, Finland
  Cat. No. F501L, Lot No. 115
- Thermocycler TouchGene, TECHNE Ltd, Cambridge, United Kingdom
  Model FTG 02TD, Serial n°101588-6

Methods:

Detailed protocols of performed PCR are shown in Annexe n°1.

3.4.3. Electrophoresis

DNA electrophoresis

TAE buffer 1x:

- 40mM Tris-acetate
- 5mM EDTA
- pH8.0

0.5% agarose gel (small/big):

- 0.17g/0.5g agarose
- TAE 1x to 35ml/100ml
- 1.4µl/4µl Ethidium Bromide solution

1% agarose gel (small/big):

- 0.35g/1g agarose
- TAE 1x to 35ml/100ml
- 1.4µl/4µl Ethidium Bromide solution

DNA ladder:

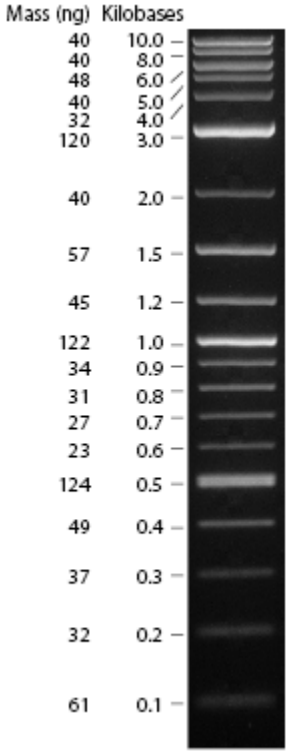


Figure 6 : Biolabs 2log ladder

RNA electrophoresis

MOPS 10x running buffer:

- 0.4 M MOPS, pH 7.0
- 0.1 M sodium acetate
- 0.01 M EDTA

RNA loading buffer:

- 50µl formamide
- 10µl MOPS 10x
- 15µl formaldehyde 37%

Denaturing agarose gel:

- 0.375g agarose
- 3.75ml Formaldehyde 37%
- 2.5ml MOPS 10x running buffer
- distillated water to 25ml

3.4.4. DNA Extraction from agarose gel

Materials:

- Agarose gel DNA extraction kit , Roche, Mannheim, Germany
  Cat No. 1 696 505
  Contains: agarose solubilisation buffer, silica gel matrix suspension, nucleic acid binding buffer, washing buffer

Methods:

Procedure followed is identical to the reference manual coming with the Roche kit.

DNA pellet is dissolved in milliQ water instead of TE buffer.

3.4.5. Phenol-chloroform extraction

Phenol-Chloroform extraction

Minimum working volume required is 50µl to minimize DNA losses. Add adequate volume of milliQ water to the samples under 50µl.

- Add an equal volume of phenol-chloroform pH 4.7 solution to the sample tube
- Vortex to mix the two phases
- Centrifuge shortly at 13'000RPM
- Remove the aqueous phase to a new microcentrifuge tube
- Repeat one more time the previous extraction procedure and remove again the resulting aqueous phase in a new tube
- Add an equal volume of chloroform solution to remove phenol traces
- Vortex to mix the two phases
- Centrifuge shortly at 13'000RPM
- Remove the aqueous phase to a new microcentrifuge tube, ensure that no residual fraction of organic phase subsists
- Proceed to precipitation to recover the purified DNA

DNA precipitation

- Add 1/10 of sodium acetate 3M to the DNA solution
- Complete with 2.5 volumes (sodium acetate volume included) of ethanol 96%
- Precipitate overnight at -20°C or during two hours at -80°C
- Centrifuge during 15 minutes at 13'000RPM and carefully discard supernatant
- Wash DNA with 20-100µl of ethanol 70% depending the pellet size

- Centrifuge during 15 minutes at 13’000RPM
- Carefully discard ethanol (pellet is much more transparent after washing step) and dry the tube in inverted position at room temperature
- When all traces of ethanol vanished, dissolve the pellet in an appropriate volume of milliQ water and proceed to quantification

3.4.6. Digestion

Materials:

Enzymes:

- *Apa*I, 15U/µl, Amersham Pharmacia Biotech, Freiburg, Germany
- *Asp*718, 10U/µl, Boehringer Mannheim, Mannheim, Germany
- *BamH*I, 15U/µl, Amersham Pharmacia Biotech, Freiburg, Germany
- *Bgl*II, 10U/µl, New England Biolabs, Ipswich, USA
- *EcoR*I, 10U/µl, Invitrogen, Carlsbad, USA
- *EcoR*V, 15U/µl, Amersham Pharmacia Biotech, Freiburg, Germany
- *Kpn*I, 10U/µl, Amersham Pharmacia Biotech, Freiburg, Germany
- *Nco*I, 10U/µl, New England Biolabs, Ipswich, USA
- *Nhe*I, 5U/µl, New England Biolabs, Ipswich, USA
- *Not*I, 10U/µl, New England Biolabs, Ipswich, USA
- *Xho*I, 20U/µl, New England Biolabs, Ipswich, USA

Buffers:

- NEB2 Buffer 10x, New England Biolabs, Ipswich, USA
- NEB3 Buffer 10x, New England Biolabs, Ipswich, USA
- NEB BSA 10x, New England Biolabs, Ipswich, USA
- K Buffer 10x, Amersham Pharmacia Biotech, Freiburg, Germany
- L Buffer 10x, Amersham Pharmacia Biotech, Freiburg, Germany
- H Buffer 10x, Amersham Pharmacia Biotech, Freiburg, Germany
- React3 Buffer 10x, Invitrogen, Carlsbad, USA

Methods:

DNA samples are digested with 10U of enzyme in a 20µl volume during 2 hours at 37°C. DNA recovered from midiprep is digested in a 100µl volume.

3.4.7. Ligation

Materials:

- T4 DNA Ligase Kit, Roche, Mannheim, Germany
  Cat No. 481 220, Lot No. 90812824
  Contains: T4 DNA Ligase, Buffer 10x

Methods:

An insert/vector molar ratio of 5:1 is used.

Quantities of insert and vector to be used are calculated with this formula:

((ng vector)x(kb size of insert))/(kb size of vector)) x (molar ratio of (insert/vector)) = (ng insert)

Standard protocol:

Insert+vector	100ng max
T4 Buffer 10x	2µl
T4 Ligase	0.5µl
H_20	Add to reach 20µl final volume

Ligation mix is incubated overnight at 4°C

3.4.8. Transformation of DH10b competent cells

Rubidium Chloride method for transformation competent *E. coli*:

Solutions:

TfbI:

Potassium acetate	30mM
Rubidium chloride	100mM
Calcium chloride	10mM
Manganese chloride	50mM
Glycerol	15% v/v

Adjust pH to 5.8 with dilute acetic acid

TfbII:

MOPS	10mM
Calcium chloride	75mM
Rubidium chloride	10mM
Glycerol	15% v/v

Adjust pH to 6.5 with dilute NaOH

Protocol:

- Inoculate 1ml from overnight into 100ml LB broth (scale up or down as needed). Incubate at 37°C.
- Put on ice 15 minutes
- Pellet cells in appropriate centrifuge tube at 3-5000g for 5 minutes
- Discard supernatant and add 0.4 volume (i.e. of original volume) TfbI, resuspend and put on ice 15 minutes
- Pellet cells
- Discard supernatant and resuspend in 0.04 volume 0.04 volume TfbII and put on ice 15 minutes
- Aliquot 50 to 100µl of cell solution and either use immediately or quick freeze at -70°C for storage

Transformation protocol:

- Take the cells out of the -80°C freezer and put them to thaw on ice
- When cells are thawed, add 2µl of ligation mix to 100µl of cells
- Put cells 20 minutes on ice
- Heat shock during 2 minutes in a thermoblock at 37°C
- Add 400µl of LB Broth and incubate 20 minutes at 37°C
- Inoculate 50-200µl of transformed cells on a LB-Agar Petri dish with ampicilline
- Incubate overnight at 37°C or over the week-end in a drawer at room temperature

3.4.9. Small scale plasmid isolation (miniprep) by lysozyme treatment

Solutions:

STET:

Sucrose	8% w/v
Triton X100	5% v/v
EDTA	50mM
Tris pH8.0	50mM

TEN:

Tris-Hcl	10mM
EDTA	10mM
NaCl	150mM

Preparation:

- Pick one colony to inoculate a 5ml tube of LB-ampicilline medium
- Grow cells at 37°C overnight
- Aliquot 2ml culture samples in microcentrifuge tubes
- Put a large becher of water to boil on a heater
- Prepare a fresh batch of TEN-lysozyme solution ,a tip of scalpel blade of lysozyme powder (135000 U/mg lyo., Roche) dissolved in 300µl TEN, and store on ice

Miniprep protocol:

- Spin down two times the samples tubes at 13'000RPM
- Discard supernatant
- Resuspend cell pellets in 105µl of STET solution
- Add 7.5µl of TEN lysozyme solution
- Incubate 5 minutes at room temperature
- Put tubes 2 minutes in boiling water
- Remove the jelly pellet with a pick

- Precipitate DNA at -20°C during 30 minutes with an equal volume of isopropanol (≈110µl)
- Centrifuge tubes at 13'000RPM during 15 minutes
- Remove carefully supernatant and dry the tubes in inverted position at room temperature
- Dissolve DNA pellet in 70µl milliQ water

3.4.10. Medium scale plasmid isolation (midiprep)

Materials:

- QIAGEN Plasmid Midi Kit, QIAGEN GmbH, Hilden, Germany
  Cat No. 121 43, Lot No. 130164347

Methods:

Procedure followed is identical to the reference manual coming with the QIAGEN kit.

3.4.11. Clones screening by colony PCR

Colony PCR screening is efficient method to analyze a large amount of clones in a reduced time. This method has been used in this project for the preparation of the synthetic mRNA.

Materials:

Please see PCR chapter (3.3.2.).

Methods:

Primers:

Primers were already available from previous projects. The forward primer (CS2+Up) binds prior to the SP6 promoter on the CS2+ vector backbone. The reverse primer (Ect3) is complementary with one conserved domain at the start of the *ectodermin* sequence.

<u>CS2+Up:</u>

5'-GCC ATT CTG CCT GGG GAC G-3'

19bp, Tm=58°C

<u>Ect3:</u>

5'-ACT TCT CAT CCT CCC TTG G-3'

19bp, Tm=64°C

Predicted amplicon sizes are ≈400bp for CS2+ Ecto and ≈600bp for CS+MT Ecto, according to the respective sequence maps.

PCR protocol:

PCR Mix	Vol. [µl]
Cell suspension	5
CS2+Up 10mM	1
Ect3 10mM	1
dNTPs 20mM	1
Buffer 10x	5
DyNAzyme II	0.5
H_2O	37.5

Cycles	Time [s]	Temp. [°C]
Init. Denat.	600	94
30x	15	94
	30	55
	50	72
Final Ext.	120	72

Screening protocol:

- Pick one colony with a sterile 200µl pipette tip
- Resuspend the cells in 5µl milliQ water in PCR tube
- Use 0.5µl of this cell suspension to create a clone bank on a fresh LB-agar ampicilline dish. Use only one dish for all the colonies screened, by disposing them in a grid of dots.
- Grow the cells on the bank and store it at 4°C for further use
- Lyse cells by heat denaturation in the thermocycler (5 minutes at 94°C)
- Add to each tube 45µl of PCR mastermix
- Run the PCR program
- When the reaction is completed, put a 5µl aliquot with 1µl loading dye on 1% agarose gel for analysis.

3.4.12. mRNA transcription

Materials:

- mMESSAGE mMACHINE Kit, Ambion, USA
  Cat No. AM1340

Methods:

Procedure followed is identical to the reference manual coming with the Ambion kit.

3.5. Microinjection of DNA constructs and synthetic mRNA into sea urchin embryos

3.5.1. Microinjection solution

Promoter-GFP constructs:

DNA	1µg
Glycerol 30%	1.5µl
TexasRed dye	0.5µl
H_20	Add to reach 5µl final volume

TexasRed is a dye that permits to discriminate the injected embryos under UV light (red glow).

The microinjection solution has to be centrifuged at 13'000RPM during 10 minutes before the work session to avoid any residual precipitate that could clog the injection capillary. During sessions, the solution is stored on ice. Otherwise store it at -20°C like any DNA sample.

Synthetic mRNA:

mRNA	Up to 10µg
Glycerol 30%	1.5µl
TexasRed dye	0.5µl
H_20	Add to reach 5µl final volume

3.5.2. Embryos preparation

Harvest and treatment of the sea urchin gametes

- Open a fresh urchin taken from the tank by cutting a circular piece of shell around the mouth with clean scissors
- Discard carefully residual sea water and internal organs to the exception of the gonads
- Remove gonads with a clean spoon. If they are male, store them dry in a clean 15ml falcon tube (dilution with water decrease sperm vitality). If they are female, place them in a becher filled with MFSW (Millipore Filtered Sea Water)
- Store the sperm on ice or at 4°C
- Separate eggs from gonads by filtering on a mesh
- Let eggs sediment and change the top water containing immature eggs

Dejellification of the egg's coat

Sea urchin eggs are coated with jelly contained polysaccharides and glycoproteins []. This protection has to be removed to ease penetration of the injection capillary into the egg. As dejellification of the coat is a quite aggressive procedure for the eggs, it's recommended to not use treated eggs for more than one injection session and to use freshly treated eggs as often as possible.

- Add 0.5M citric acid solution to the becher of MFSW contained the eggs (120µl of solution for 60ml of seawater)
- Incubate 1 minute at room temperature. Longer incubation time could irreversibly damage the eggs
- Neutralize acid solution with an equal amount of 1M Tris-HCl solution at pH8.0
- During all the dejellification, stir the becher to keep eggs in suspension
- Let the eggs sediment and change the water
- Store the eggs in a 18°C incubator during all the microinjection session

3.5.3. Microinjection setup

Microinjection procedures are taken and adapted from *Functional studies of regulatory genes in the sea urchin embryo* [9].

Materials:

See the referenced publication.

Methods:

General settings

- Set the nitrogen bottle to deliver a 3bar pressure.
- Switch on injecting pressure generator (IPG), micromanipulator controller and microscope.
- On the IPG: ingoing pressure (P1) should be superior to 3000 HPa, injection pressure (P2) is set to 300 HPa and holding pressure (P3) to 30 HPa. Injection time is 0.2s.

Eggs disposal and fertilization

- Eggs are disposed in a single-cell column with a glass pipette on the microinjection plate filled with MFSW. Two lines are drawn on the bottom of the plate to assist the operator in this task.
- Sperm is diluted in sea water: put 10-20µl of fresh concentrated sperm stored on ice in 10ml MFSW. As a visual indicator, the resulting solution should be slightly opalescent. Too concentrated sperm solutions lead to increasing polyspermia during fertilization and undesired dirts in the plate.
- Eggs are fertilized with a few drops (no more than 3-4) of sperm solution in the microinjection plate.
- Control the phenomenon under the microscope: if fertilization doesn't occur in the 5 minutes following insemination (usually it is completed after 2 minutes), add more sperm, prepare new sperm solution or use fresh eggs.

Needle filling

- Load 0.5 to 1µl of microinjection solution in a sterile disposable microloader tip.
- Insert the loader tip in the rear hole of the needle.
- Low the loader tip until it reaches the tip (constriction) in the needle.
- Gently release the sample on the capillary wall and simultaneously remove the loader from the needle.
- Remove carefully the front protection tip on the needle (by letting it drop to the ground) and attach the needle to the micromanipulator.

Needle positioning in the microinjection field

- Focus the microscope on the embryos attached to the bottom of the plate.
- Position the needle above the liquid surface with a 45° angle to the horizontal.
- Position the needle in the center of the field. The tip should appear under the microscope as a bright unfocused object.
- Low carefully the needle first with the mechanical wheels on the micromanipulator until it touches the liquid and then with the electrical commands.
- When the tip can be seen slightly above the embryos, check if the tip is not clogged by performing test injections in the liquid (a tiny light purple cloud should be visible at the tip each time injection is triggered).

Needle cleaning

How to deal with clogged and dirty needles is probably the most hazardous part of the microinjection process. Usually, after 15-20 injections, the needle is clogged with cell membrane wastes, cytoplasm components and dead sperm, decreasing or nullifying the microinjection efficiency. Simply changing the needle is not an option as they are expensive and it cannot be afforded to lose such a time while working with fertilized embryos.

In function of the nature and the gravity of the clogging, several methods can be used and combined to get rid of the problem:

- Use the cleaning pressure: the IGP has a cleaning function that can be triggered by a simple button press. It injects a high pressure in the capillary and it's usually efficient to eliminate light clogging problem.
- Scratch the tip on the surface of the plate: move the field to an empty area of the plate and focus on the bottom. Low the tip very carefully until it touches the surface. Move it slowly across the surface to remove external wastes attach to the needle. If too much sperm has been used previously, this procedure can amplify the issue, as many dead spermatozoids lay on the ground.
- Create an artificial cavity by scratching one empty area of the plastic surface with a glass blade. Move very gently to the tip of the needle on the edges of this cavity to remove wastes attached to it.

Embryos observation:

For observation purposes, embryos beyond the stage of morula have to be immobilized. For that task, one or two drops of formaldehyde 37% stock solution are put in the microinjection plate. Fixed embryos will die a few hours after this treatment, so fixation is not recommended if observations at a later stage of development are planned.

4. Results

4.1. *In-silico* analysis of *ectodermin*

Two copies of an *X. laevis ectodermin*-like gene have been identified in *P. lividus* genome. To understand the *cis*-regulatory mechanisms driving the expression, the early efforts have been put on the non-coding region preceding the gene. In a second time, the 3' non-coding sequences of the genes have been studied as well.

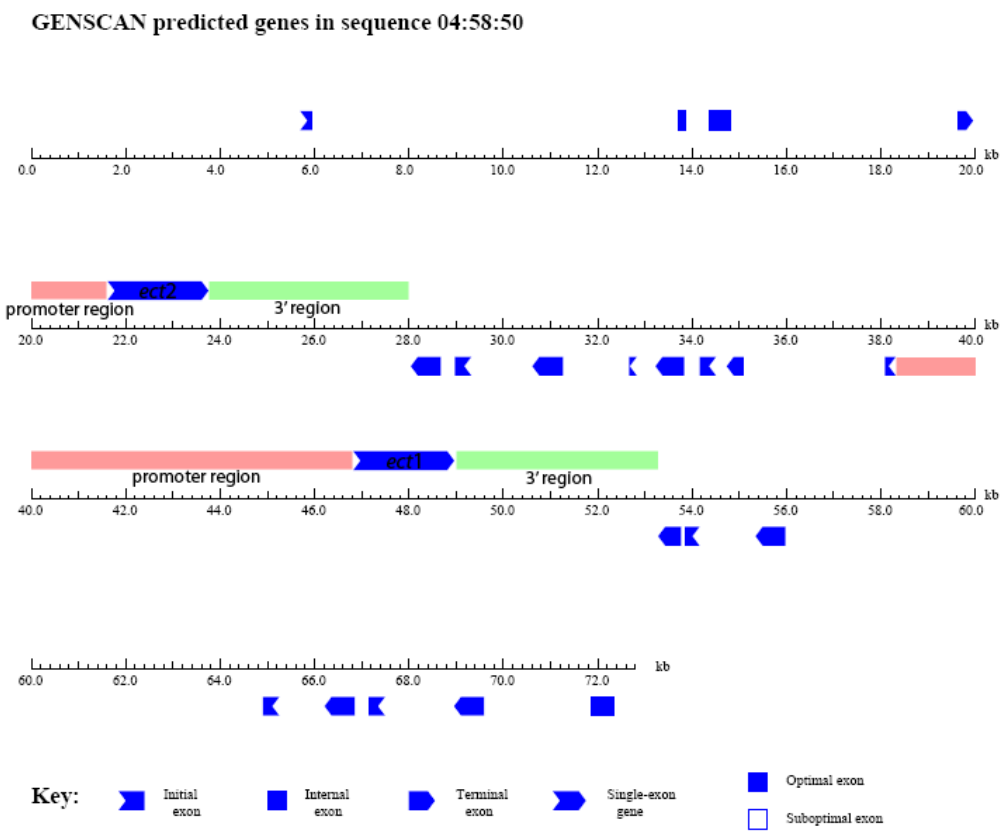


Figure 7 : GENSCAN analysis of putative peptide-coding sequences in *P. lividus* genome.

Analysis of the promoter region

Both copies of *ectodermin* are preceded by long non-coding regions (dubbed promoter regions, highlighted in red on figure 7) including the promoter module. For *ect*1, the size of the promoter region reaches 10kb and for *ect*2, it's around 2.5kb. Alignment of the two promoter regions has been accomplished (ClustalW output shown in Annexe n°2). It revealed that the fragment of 2.5kb directly

preceding the start codon of *ect*1 shares some similarities with the promoter region of *ect*2, but homology is not complete.

In-silico analysis of the putative transcriptional binding sites has been performed with the TESS online software to unravel the mechanisms driving the expression of Ectodermin. As there is not a complete consensus sequence shared by the two promoter region analyzed, putative binding sites differ from one to another.

Some putative factors have been related to known GRNs in sea urchin embryos.

Putative sites for Twist, an activator involved in skeletogenesis and mesenchymes differentiation GRNs [10], have been found through the two sequences.

TCF binding sites have been as well spotted in both sequences. TCF is involved in the Wnt8 regulation pathway responsible for the patterning of the animal-vegetal axis in sea urchins [11]. It mediates the β-Catenin signal to induce vegetalization of the embryo.

GATA is a transcriptional repressor that acts conjointly with Goosecoid, an oral ectoderm-restricted transcriptional repressor [12]. As Goosecoid expression is triggered by Nodal signals, GATA could be an interesting candidate in regard of the start hypothesis of this project.

Various binding sites for homeodomain-related factors such as Hox[13], Cdx[14], Nkx[15] or HNF[16] have been as well identified.

Analysis of the 3' region

Both copies of *ectodermin* are followed by a 4.5kb non-coding region (highlighted in green on figure 9). Alignment of these two sequences has been performed to discover if they have more similarities than the size. It resulted that they share a conserved sequence of 2.5kb, starting from the stop codon of the gene. TESS queries have been performed on this conserved sequence and Twist, Cdx, Nkx and HNF binding sites have been found.

4.2. In-Situ Hybridization of *ectodermin* mRNA

In-Situ Hybridization of the complementary sequence of a conserved domain of *ectodermin* has been performed at different stage of the development to obtain insights on the spatial and temporal expression of Ectodermin during embryogenesis:

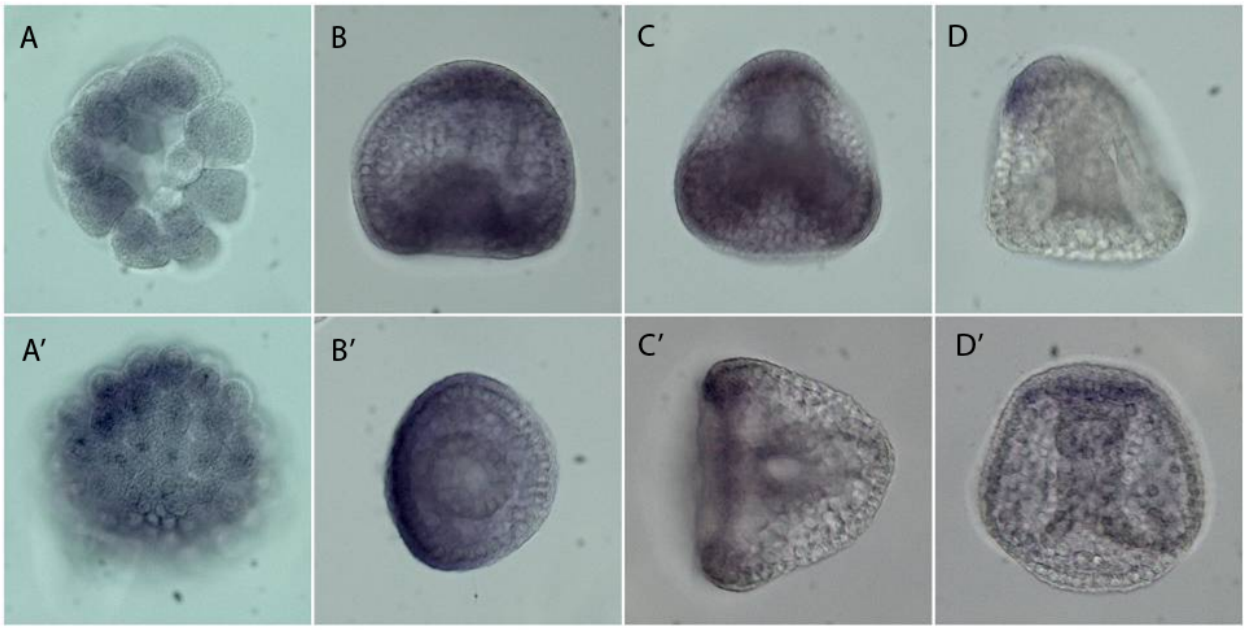


Figure 8 : ISH in embryos: morula stage (A, A'), early gastrula (B oral side view, B' top view), prism (C oral side view, C' top view), and late gastrula (D lateral view, D' oral side view).

At the morula stage (A, A'), it's not yet possible to discriminate the future oral region from the aboral, but it can be noticed that expression of Ectodermin is clearly limited to one domain on the side of the embryo. In the stages occurring later, the expression is still clearly restricted to one specific region that appears to be the oral ectoderm. This is especially visible in the late gastrula stage (D, D'), where expression is only visible in a thin band of cells near the apical plate, at the exact location where the mouth will open in the prism stage.

4.3. Reverse Trancription PCR (RT-PCR)

RT-PCR has been performed in preparation of the Q-PCR, to assess level of expression of Ectodermin in function of the development time. Successive stages of the development have been analyzed through this method. Data obtained are briefly summarized by the following graphic:

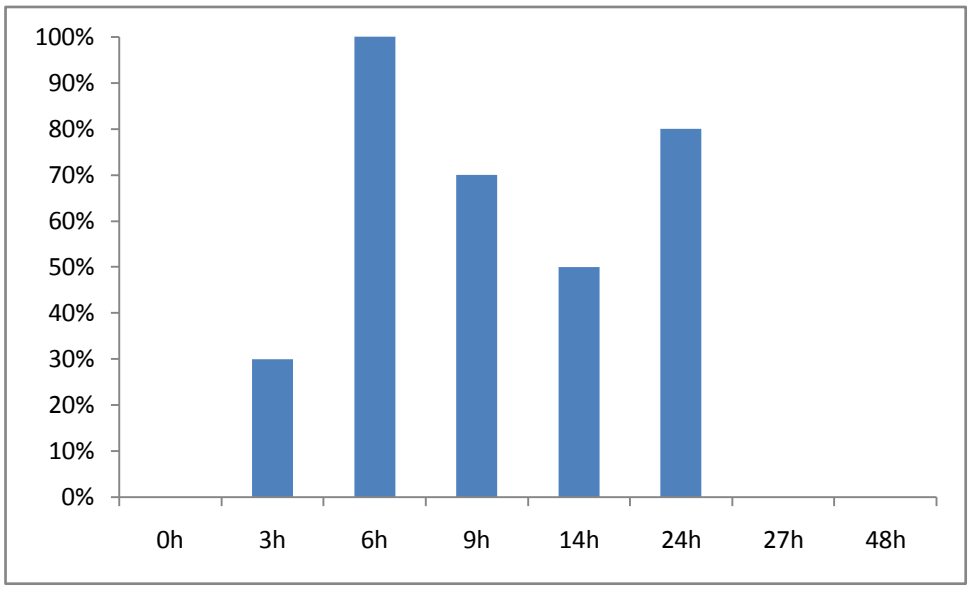


Figure 9 : Relative levels of expression of Ectodermin at different stages of the embryogenesis. 0h: unfertilized egg. 3h: 16-cells morula. 6h: morula. 9h: blastula. 14h: mesenchyme blastula. 24h: gastrula. 27h: late gastrula. 45h: pluteus.

Levels of expression are represented relatively to the highest recorded (at 6h). In the fertilized egg, no *ectodermin* mRNA is detectable, meaning that expression of Ectodermin is driven internally in the zygote and is not dependant of maternal factors. mRNA starts to be detected after 3h which is quite an early event and levels of sequences transcripted reach a peak at 6h before slowly decreasing in the following stages. At the gastrula stage, the rate of transcription increases again. Unfortunately, after this stage no additional transcripted mRNAs are detectable, but an error during the extraction procedure of these samples is not to be excluded.

4.4. Cloning

4.4.1. Promoter-GFP transgene contructs

As described in the objectives, it was intended to design and clone three GFP transgenes containing promoter region-fragments of *ectodermin* (Figure 10).

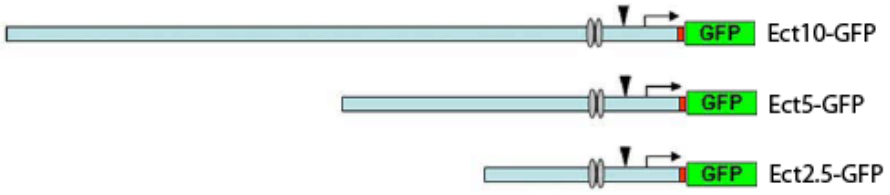


Figure 10 : Schematic drawing of the GFP transgene constructs driven by ectodermin cis-regulatory apparatus. A black arrowhead indicates the TATA-box sequence, while the two light grey ovals indicate CAAT-boxes. The bent arrow denotes the putative transcription start site. The name of each construct is listed on the right.

For this purpose, fragments have been cloned in pHE-GFP, a modified pGL3 vector (Promega) that already contained a 2.9kb insert (Figure 11).

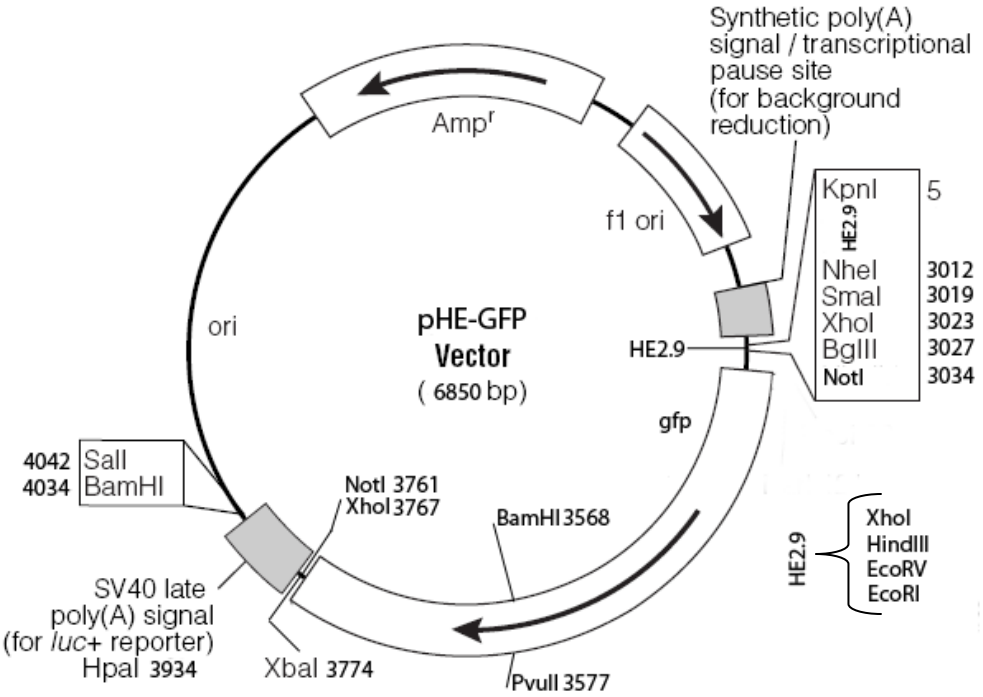


Figure 11 : map of pHE-GFP, the modified pGL3 plasmid used for the cloning of the GFP transgenes. In the bottom-right corner are indicated restriction sites of HE2.9 that will be conserved after removal of the insert.

Amplification of Ectodermin 10kb, 5kb and 2.5kb promoter region

Long-distance PCR has been performed overnight to amplify three different-sized fragments of the promoter region, using the Invitrogen Elongase Amplification System. The PCR products have been analyzed by electrophoresis on a 0.5% agarose gel.

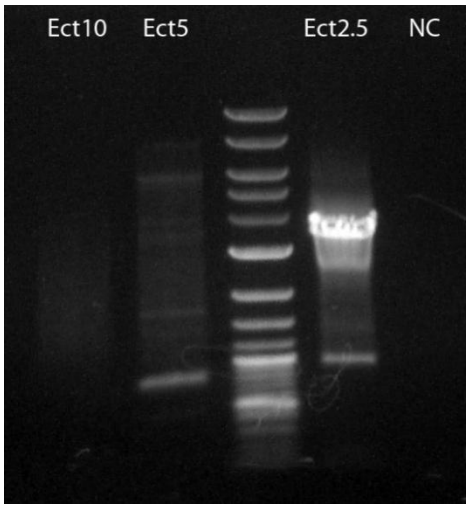


Figure 12 : PCR products of the promoter region: Ect10: 10kb fragment; Ect5: 5kb fragment; Ect2.5: 2.5kb fragment; marker: 2log ladder.

The Ect10 lane shows no amplified product, only the trail of the genomic template used.

Several amplicons are visible in the Ect5 lane with the most marked one around 800bp. One band is as well slightly visible at 5kb.

On the Ect2.5 lane, a vast amount of amplified products are visible around 3.5kb as well as a minor band at 1kb.

Considering these results, it appears that no satisfying amplicon have been obtained with this reaction. No amplification at all occurs with the 10kb primers set, a lot of byproducts with the desired amplicon in minority for the 5kb set and one amplicon 1kb longer than the desired size for the 2.5kb set. However it has been decided to focus the following optimizations of the PCR protocol on the 3.5kb fragment obtained for Ect2.5. Indeed, given the specificity of the primers set, it cannot be excluded that this amplicon is actually a fragment of the promoter region but 1kb longer than expected due to some polymorphism phenomenon.

Optimizations of the PCR protocol for Ect2.5

The first optimization attempted consists to increase the annealing temperature from 55°C to 58°C for the ten first cycles and from 60°C to 62°C for the thirty next. In the same time the $MgCl_2$ concentration in the PCR mix is lowered from 1.6 mM to 1.5 mM. The elongation time is adjusted to 3 minutes as this time only the 2.5kb is being amplified. In the same time the same optimized protocol is performed with another PCR kit, DyNAzyme II from Finnzymes for traditional PCR program.

Due to the loss of some of the early data of this project, it's not impossible to show any photographies of the resulting gels. Nevertheless, this optimization gave worst results than the original protocol with no visible sign of amplification on the electrophoresis gels.

In order to amplify the 3.5kb band obtained by the first PCR, a nested PCR has been performed using the previous product as template. The resulting product has been isolated on the gel and purified with the Roche Agarose gel DNA kit.

Cloning of Ect2.5 into pHE-GFP

No data of this cloning will be shown as it results in a complete failure, but the strategy will be exposed in the following lines for a better understanding of the further experiments:

The purified amplicon and pHE-GFP have been sequentially digested with *EcoR*I and *Nhe*I enzymes. The resulting linear vector has been put on agarose gel and extracted and purified with the same kit (Roche) than the amplicon. They have been ligated together following a molar ratio of 5:1 (insert:vector). The plasmid obtained has been transformed into DH10B *E. coli* competent cells by heat shock. The transformation has been repeated three times with positive control and fresh batch of competent cells. Recombinant colonies were obtained only with the control. Therefore, it has been chosen to restart the cloning of the 2.5kb fragment from scratch with a new strategy.

Cloning of Ect2.5 by shotgun ligation

After the failure of the previous attempt, it has been decided to minimize the intermediate steps between amplification and ligation. The agarose extraction kit, used in the precedent procedure, has been suspected to introduce contaminants inhibiting the ligase used. Therefore, the purification steps have been reduced to the minimum with only phenol-chloroform extractions after the digestions to remove proteins. The ligation mix containing vector, insert and undesired sequences, it is necessary to proceed afterwards to an extensive step of screening of the recombinant colonies obtained.

A new PCR reaction has been set, using the first protocol, as no samples of the first PCR subsisted. In parallel fresh plasmid pHE-GFP has been digested with *EcoR*I and *Nhe*I. Both samples have been put on 0.5% agarose gel with negative controls for each:

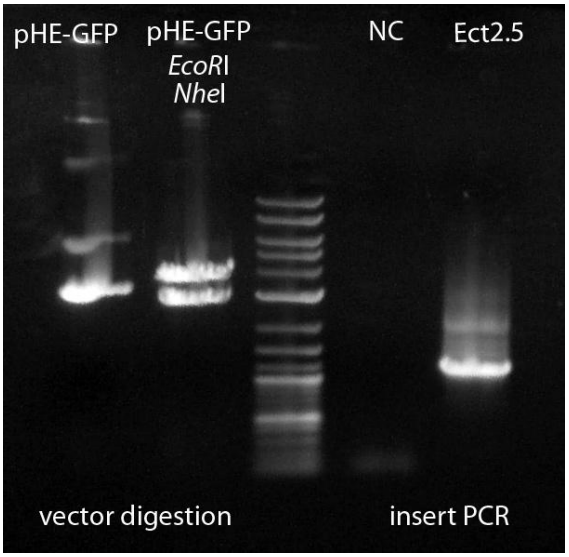


Figure 13 : On the left side, pHE-GFP vector used for the cloning: undigested and digested with *EcoR*I and *Nhe*I; on the right side, results of the PCR: negative control (NC) and 2.5kb amplicon. Marker: 2log ladder.

The control lane for the plasmid is all right: the various (open-circle, coiled, supercoiled) circular forms are visible. The digestion results in two bands, one at 4kb and another at 3kb whose sizes correspond respectively those of the backbone and the inserting removed.

Concerning the PCR, the negative control is OK. The PCR with the 2.5kb primers set shows multiple bands with a major one around 1.2kb. But interestingly, another band is clearly visible at 2.5kb (the size of the desired amplicon), which was not the case on the first reaction performed.

The PCR product has then been purified by phenol-chloroform extraction and the resulting DNA precipitated during two hours at -80°C. The pellet obtained has been resuspended in 10 µl of milliQ water.

Quantifications of the purified PCR product and the plasmid digestion have been performed by spectrophotometry and on an agarose gel for confirmation:

- PCR product: 146.1 ng/µl (A_{260}/A_{280} = 1.42)
- Plasmid digestion: 457.2 ng/µl (A_{260}/A_{280} = 1.66)

Then, the PCR product has been digested with the same set of enzymes than the plasmid, before proceeding to the shotgun ligation.

After overnight ligation, DH10b competent cells have been transformed with the resulting mix. Two Petri dishes have been inoculated, one with 200 µl of culture and the other with 50 µl. The dishes have incubated overnight at 37°C in the incubator. On both laid several hundreds of colonies.

Screening of recombinant clones

Screening has been performed by analysis of the restriction map of the putative constructs. Screening has been planned in two steps, a first assay with a single enzyme (*Xho*I) to identify a pool of putative positives and a second with a combination of several enzymes for confirmation.

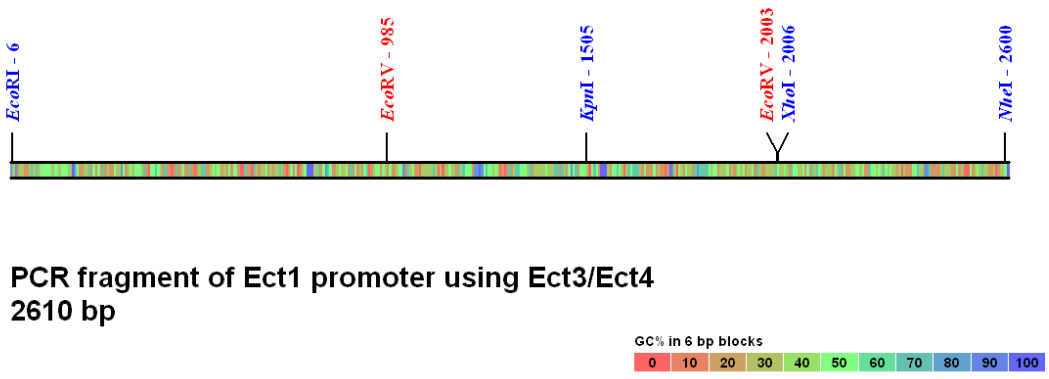
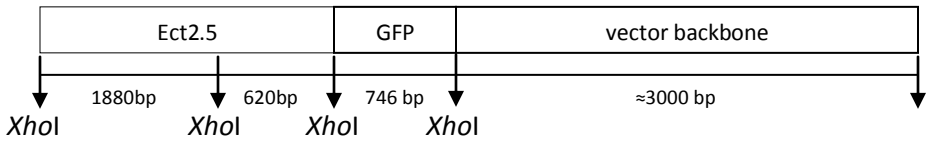


Figure 14 : Predicted restriction map of Ect2.5 insert.

Twelve colonies from the 50 µl dish and six from the 200 µl dish have been picked for screening. Plasmid DNA has been extracted by miniprep and then digested with *Xho*I for analysis. Culture tubes of each clone have been stored at 4°C for later uses. The positive pEct2.5-GFP clone should have four sites for that endonuclease (schematic not drawn to the scale):



After digestion, the samples have been analyzed by electrophoresis on a 1% agarose gel:

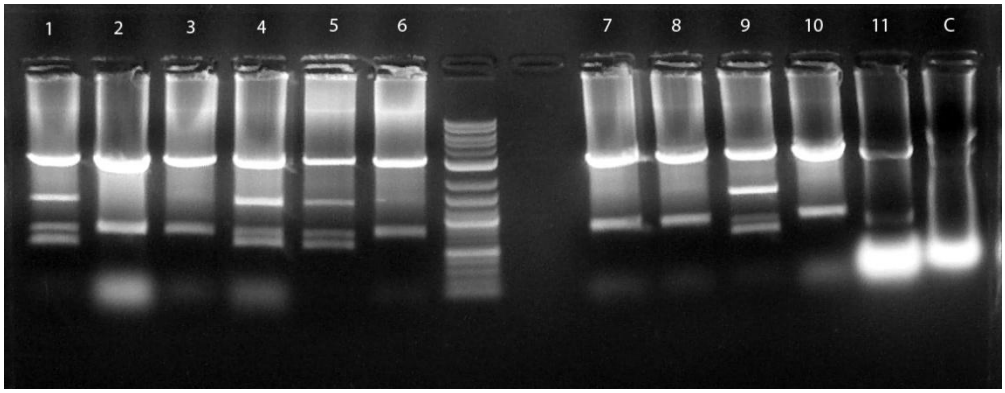


Figure 15 : Screening of candidates for pEct2.5-GFP by *Xho*I digestion. Lanes 1 to 11: samples. Lane C: undigested plasmid DNA as control. Marker: 2log ladder.

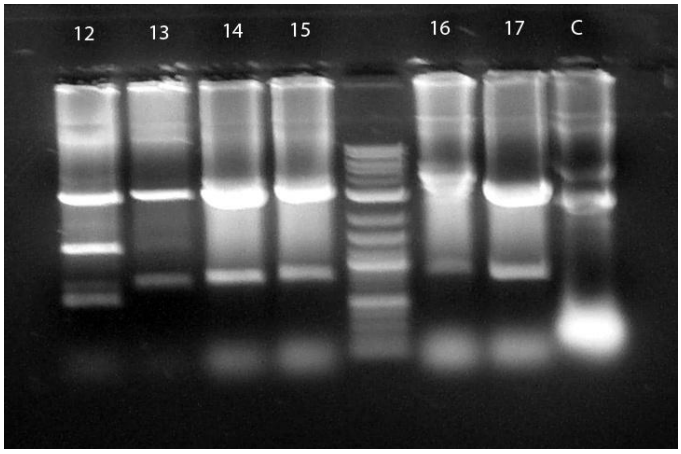


Figure 16 : Screening of candidates for pEct2.5-GFP by *Xho*I digestion. Lanes 12 to 17: samples. Lane C: undigested plasmid DNA as control (C). Marker: 2log ladder.

Given the poor resolution of these gels, it's difficult to get definitive answers. Nevertheless, two patterns of restriction can be observed: the first has four distinctive bands (samples 1, 4, 5, 9 and 12), one high at approximately 3kb, followed by another around 1.5kb and finally two small of very similar sizes between 0.5 and 1kb. The second pattern (samples 2, 3, 6, 7, 8, 10, 11 and 13 to 17), has only two visible bands

that it shares with the first pattern: the high band at 3kb and the low between 0.5 and 1kb. Possible incomplete digestion can be observed in the sample 11.

The first pattern could correspond to a positive pEct2.5-GFP clone but better resolution is needed to confirm. The second pattern is most certainly related to a recombination of the original pHE-GFP plasmid during the shotgun ligation as the bands sizes are similar to those of the backbone and GFP fragments.

In addition of this screening, a second have been performed under the same conditions with fifteen additional colonies from the 50 µl dish. The digested samples have been analyzed on a 1% agarose gel as well as the putative positives of the first screening:

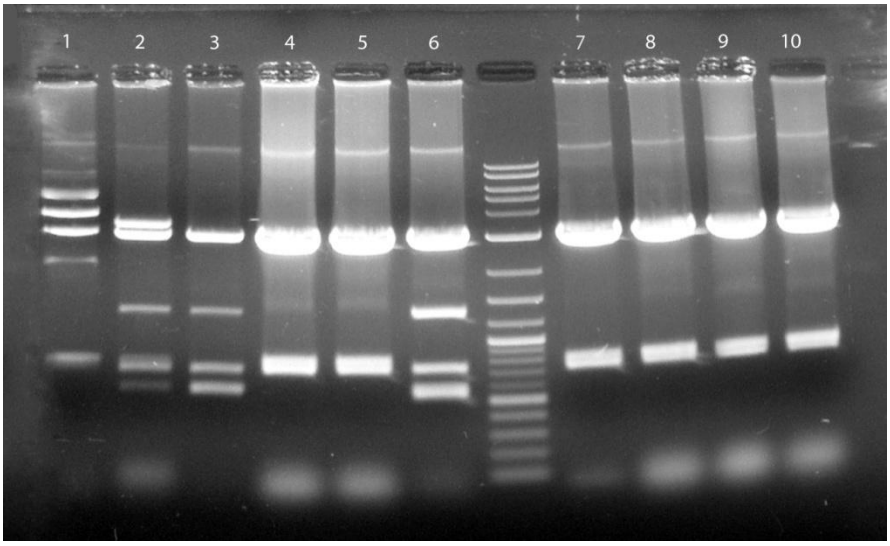


Figure 17 : Second screening for pEct2.5-GFP candidates. Samples 1 to 10. Marker: 2log ladder.

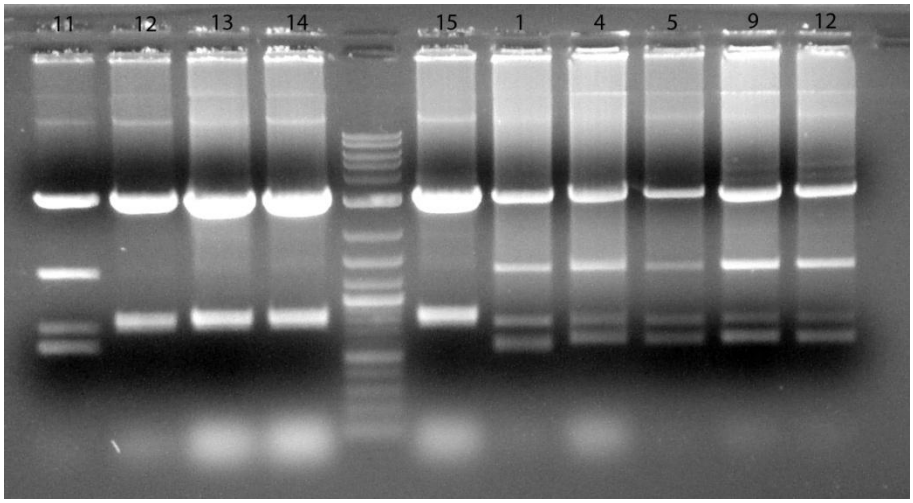


Figure 18 : Second screening for pEct2.5-GFP candidates. Samples 11 to 15. After sample 15 are shown putative positives of the first screening (1, 4, 5, 9, and 12). Marker: 2log ladder.

The same two patterns revealed previously can be observed among the samples of the second screening. The putative positive is present in sample 2, 3, 6 and 11. It's possible now to distinguish with more precision the sizes of the bands: the highest at 3kb, the next around 1.3-1.4kb, and the two lowest at 750 and 600bp. The better resolution of these gels allows seeing some slight variations in this pattern like in samples 6 (Figure 17) and 1 (figure 18) where the last band is slightly lower than in the others samples.

In the end, the putative positive pattern differs only by one band from the predicted one. To the exception of the 1.4kb band, all the others match with the restriction map of the plasmid. Before emitting a final statement on these clones, it's necessary to perform additional restriction analysis with other endonucleases.

Clones 5 and 9 (Figure 18) have been used for the following digestions. Four enzymes have been combined in five reactions to have a complete understanding of the restriction map of the clones. The bands visible (except the high backbone band) should be the followings:

- *Kpn*I: 1504bp
- *EcoR*V: 985bp & 1018bp
- *EcoR*V + *Bgl*II: 985bp, 1018bp & 609bp
- *Bgl*II: linearized plasmid
- *Bgl*II+*EcoR*I: 2609bp

The results are shown on these 1% agarose gel electrophoresis pictures:

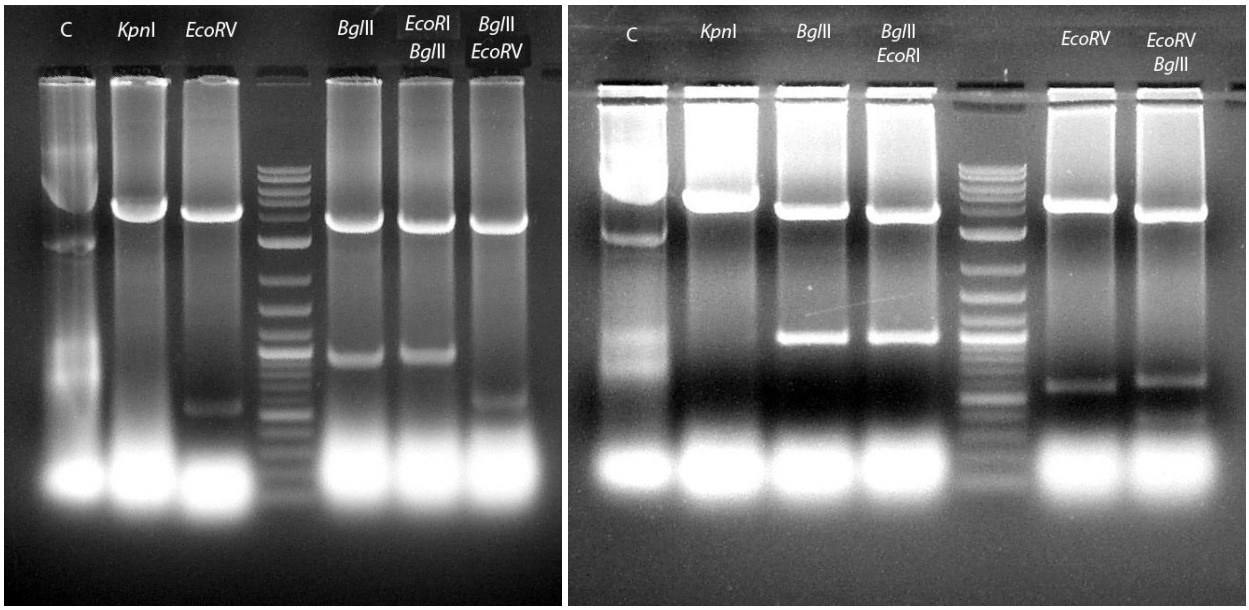


Figure 19 : Restriction maps of clones 5 (left) and 9 (right). Undigested plasmid DNA as control (C). Marker: 2log ladder.

First, it can be observed that the restriction patterns on the two clones are identical. Concerning *Kpn*I, a single high molecular weight band is visible above 4kb. On *EcoR*V lane, two bands are visible, one at 4kb and another around 550 bp. The restriction with *Bgl*II and *Bgl*II combined to *EcoR*I shows the same pattern: one band slightly under 4kb and another around 1kb. *Bgl*II combined to *EcoR*V reveals three bands at 4kb, 550bp and 350bp. These results don't correlate at all the predicted restriction map. In addition the sum of the bands for each digestion doesn't match with the sum of the bands seen with the *Xho*I restriction performed earlier.

A sequencing of 300bp from *EcoR*I site just before the insert confirmed that clone 5 was positive. Apparent lacks in the restriction patterns observed could only be explained by admitting the occurrence of polymorphism among the specimens. It have to be remembered that all the primer design and the restriction predictions have been done on the basis of a single individual genomic sequence, as no others were available.

Preparation of the plasmid for microinjection

200µl of plasmidic DNA have been obtained by midiprep. Original tube of clone 5 has been used as inoculum for the midiprep culture. *Xho*I digestion has been performed on midiprep DNA to ensure no genetic loss occurred during storage:

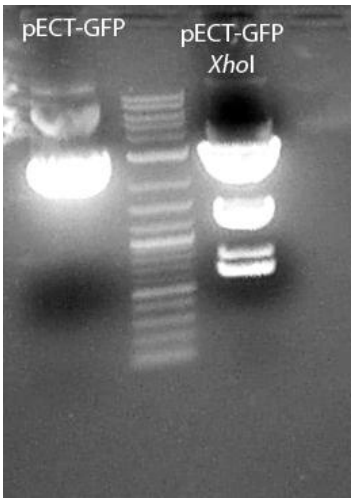


Figure 20 : 1% agarose gel of *Xho*I control digestion. To the left, undigested pEct2.5-GFP plasmid as control. To the right, pEct2.5-GFP digested with *Xho*I. Marker: 2log ladder.

The restriction pattern on figure 8 is identical to the one shown on figure 6 for clone 5. No loss of the plasmid occurred.

Quantification of plasmidic DNA obtained by midiprep has been performed by spectrophotometry. The DNA concentration reached 1176.9ng/µl (A_{260}/A_{280}=1.52).

The plasmid has been successfully linearized with EcoRI, purified by phenol-chloroform extraction (as shown on figure 20), precipitated and dissolved in 50µl milliQ water.

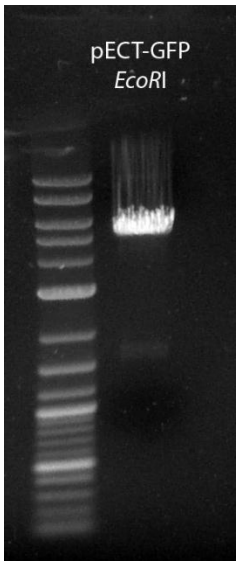


Figure 21 : 1% agarose gel of purified pEct2.5-GFP linearized with EcoRI. Marker: 2log ladder.

The final concentration of pEct2.5-GFP solution after linearization and purification steps is 459.8ng/µl (A_{260}/A_{280}=1.77).

The plasmid is then stored at -20°C, before to be used for the microinjection experiments.

Additional promoter-GFP designs

Shortened pEct2.5-GFP by *Kpn*I restriction (pECT1.1-GFP)

In order to optimize expression of the pEct2.5-GFP construct in embryos, a fragment of 1500bp have been removed by digestion from the plasmid, leaving a 1100bp long promoter region. Digested plasmid has been purified by phenol-chloroform extraction, precipitated and dissolved in 10µl milliQ water, for a final concentration of 398.9ng/µl (A_{260}/A_{280}=1.91).

pEct1.45-GFP construct

In the late stage of this project, a new construct including a 1.45kb long promoter region has been realized by another member of the team.

The cloning strategy was the same that was used for Ect2.5 construct, so this report will not detail these experiments. Nevertheless, it could be interesting to look at the PCR step: amplifications (under identical conditions) have been performed with a set of specific primers on genomic templates of four different specimens of *P. lividus*. The amplicons obtained are shown on the following 1% agarose gel:

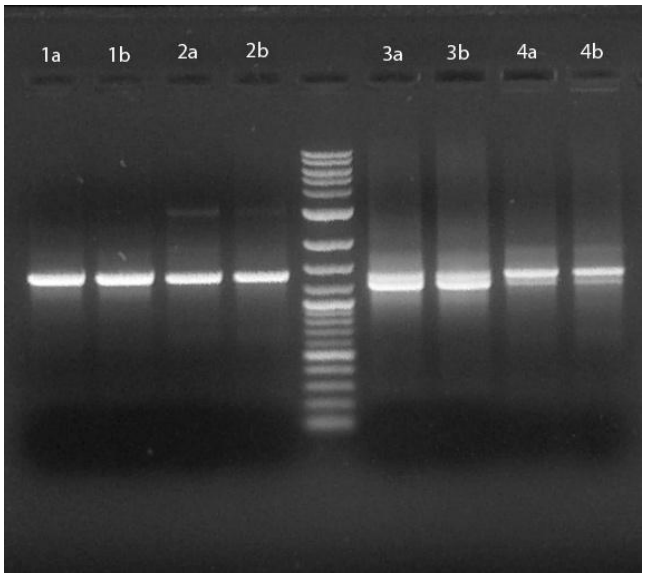


Figure 22 : PCR products of Ect1.45. Genomic templates 1 to 4 (in duplicates)

It can be noticed significant variations among the genomic material used. In all the specimens, the 1.4kb band is clearly visible but, to the exception of the template 1, it's accompanied by secondary products whose positions differ among the samples. It's a proof of the strong occurrence of polymorphism in *P. lividus* genome.

Ect1.45 n°1 has then been successfully cloned into pHbox12-GFP vector and prepared for microinjection following the same procedure as Ect2.5 construct.

4.4.2. Synthetic mRNA

In addition to the promoter-GFP constructs, mRNA of the coding sequence of Ectodermin has been synthesized. The injection of this RNA in the zygote will induce ectopic expression of the gene and possibly allow better understanding of its function in sea urchin embryos.

As full coding sequence of Ectodermin was already available on a pGEM vector, no PCR was needed for this cloning.

Two vectors were used: CS2+ and CS2+MT (documentation of these plasmids is available in Annexe 3).

Shotgun ligation strategy has been followed like for the promoter-GFP constructs to reduce steps before transformation.

*EcoR*I has been used for cut the Ectodermin gene from the pGEM vector and for linearize the CS2+ and CS2+MT vector. Two ligation mixes have been set (one for each plasmid).

DH10b competent cells have been transformed with the ligation products (two plates for each vector) and several hundreds of recombinant colonies were available on each plate after incubation.

As primers were available for the CS2+ backbone and the Ectodermin gene, screening has been achieved by colony PCR method.

188 colonies have been screened, divided in four sets of 47. The PCR products resulting have been analyzed by electrophoresis on a 1% agarose gel. As only one positive result has been found among the 188 clones screened, only the concerned gel is shown in this report (figure 23):

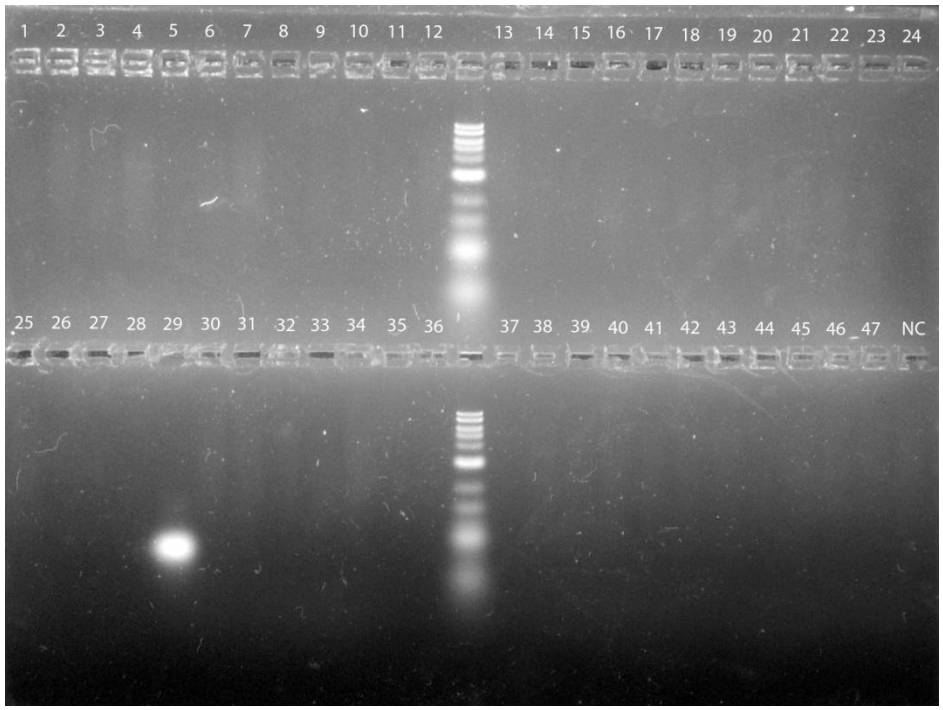
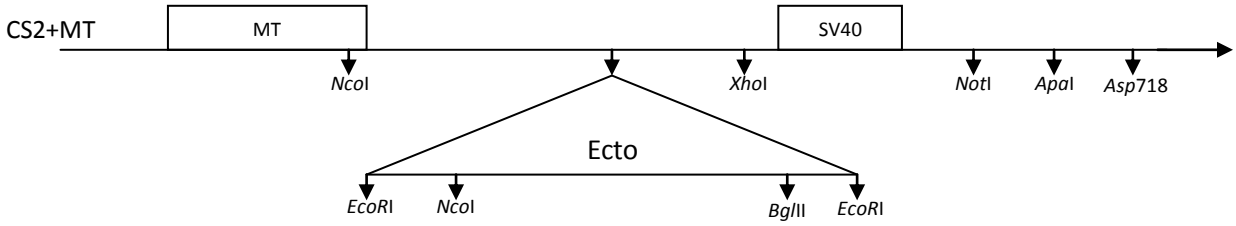


Figure 23 : PCR colony screening. Sample 1 to 24: CS2+ Ecto. Sample 25 to 47: CS2+MT Ecto. NC: negative control. Marker: 2log ladder.

In order to confirm the positive CS2+MT clone, its restriction map has been analyzed by three enzymatic digestions. Three additional digestions have been set for choose the most adapted endonuclease for linearization. The clone colony have been picked from the clone bank plate and inoculated in 5ml tube culture for DNA extraction by miniprep.

The restriction map established is the following (not drawn to the scale, restriction map of Ecto insert shown in Annexe 4):



- *Nco*I : 634bp
- *EcoR*I : 1.85kb (Ecto insert)
- *Bgl*II + *Xho*I : 463bp
- *Apa*I, *Asp*718 and *Not*I : linearized vector ≈6.2kb

Digested samples have put on 1% agarose gel for analysis :

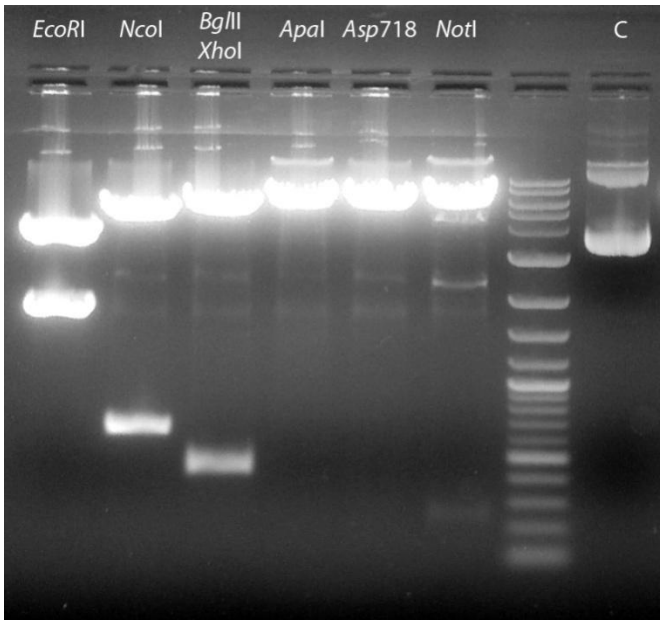


Figure 24 : Restrictions performed on the putative positive CS2+MT Ecto clone with undigested plasmid DNA as control (C). Marker: 2log ladder.

The trails visible on the samples lanes indicate partial digestion. Nevertheless, digestion is complete enough to study the restriction map. For *EcoR*I, two bands are visible, above 4kb and at 2kb. *Nco*I lane shows again two bands, the first between 5 and 6kb and the second around 600-700bp.Same situation for *Bgl*II combined to *Xho*I, with bands at 5-6kb and 400-500bp. Linearization digestions (*Apa*I, *Asp*718 and *Not*I) show unique bands above 6kb. According to the restriction map, this clone is positive.

Large amount of plasmidic DNA have been extracted by midiprep from a fresh culture of the positive clone. DNA pellet has been dissolved in 250µl milliQ water for a final concentration of 613.5ng/µl (A_{260}/A_{280}=1.77).

10 µg of CS2+MT Ecto have been digested with *NotI* for linearization before transcription of the mRNA.

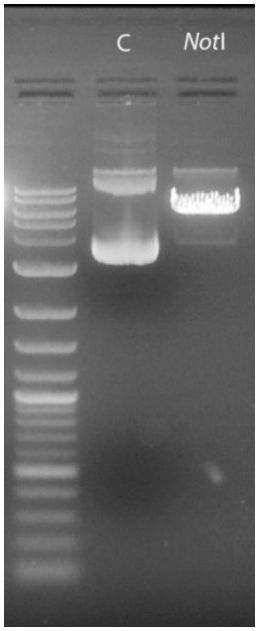


Figure 25 : Linearization of CS2+MT Ecto with undigested plasmid as control (C). Marker: 2log ladder.

Linearization of the vector is almost fully completed. Residual undigested plasmids should not interfere significantly with the transcription reaction.Linearized plasmid has then been purified by phenol-chloroform extraction before proceeding to transcription in mRNA.mRNA has been then purified by phenol extraction and quantified. 3.5µg/µl was obtained.

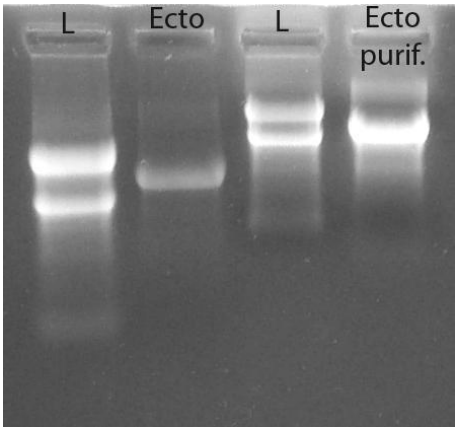


Figure 26 : Denaturing agarose gel for RNA with total RNA as marker (L). Ecto: mRNA obtained after transcription. Ecto purif.: mRNA purified by phenol extraction.

4.5. Microinjection

4.5.1. Promoter-GFP transgenes

Embryos have been injected with the three promoter-GFP transgene produced (Ect2.5-GFP, Ect1.1-GFP and Ect1.45-GFP) along with an Hbox12-GFP construct as positive control. Development has been followed until pluteus stage (48h) to allow the accumulation of sufficient quantity of GFP to be detected.

Positive control showed results comparable to those described in the related study [6]. Faint fluorescence have been detected in the oral territory in embryos injected with Ect2.5-GFP, but it only concerned two over several hundred injected. This result cannot be taken as statistically significant.

4.5.2. Ectopic expression induced by mRNA

Synthetic mRNA produced by transcription of CS2+MT Ecto has been injected at a concentration of 2ng/nl in fertilized eggs. Development has been followed for each batch until the pluteus larva-stage.

No noticeable phenotype change has been observed until the transformation in pluteus. At larval stage, several highly-occuring abnormal skeleton phenotypes have been reported in injected embryos. A collection of them is exposed in the following panel:

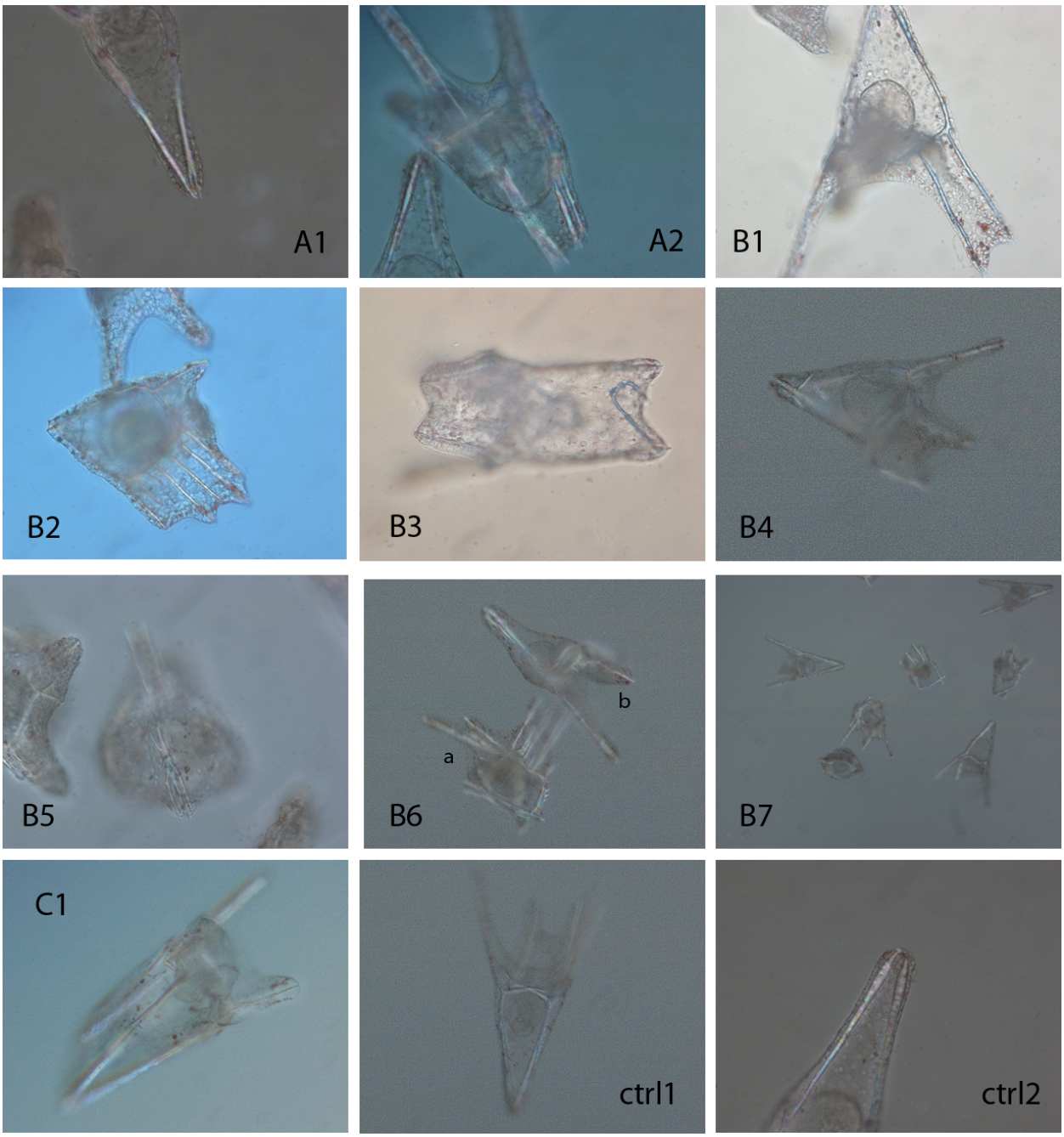


Figure 27 : Ectopic expression of Ectodermin in pluteus larvae. Eggs originated from three different females (A, B and C). Pictures of A taken after 72h, of B after 72h and of C after 90h. Uninjected embryos after 72h as control: full view of the larva (ctrl1), focus on the spicules vertex (ctrl2).

The abnormal phenotypes observed can be categorized in three families: forking at the spicule tip (A1, A2, B5, and B6b), supernumerary spicules (B1, B2, B6a, and C1) and aberrant mineralization at the spicule tip (B3, B4).

In the larvae from female A, only the fork phenotype has been reported, with various degree of severity among the individuals, ranging from very slight mutation (A1) to obvious (A2).

The batch issued from female B contained all the mutant phenotypes registered, with an extreme level of mutation. Multiple supernumerary arms, deforming the oral side of the pluteus, can be observed in specimens B2 and B6a. An extreme version of the fork phenotype can be seen in specimen B5, where a complex triple branching occurred at the tip of one spicule, forming a trident-like structure. Larvae from female B are as well the only one to show evidence of new mineralization at the type of spicules (B3 and B4), a phenomenon that normally never occurs in this localization.

In the specimen C1, it can be observed an interesting variation of the supernumerary spicules phenotype: a third spicule radiating from the oral side in direction of the vertex, and not of the arms as it can be observed in B1.

5. Discussion

In Situ Hybridization

ISH results show that expression of Ectodermin is clearly restricted to the oral territory; that confirms the start hypothesis that Ectodermin is in some way involved in the development of this region of the embryo. More advanced embryos show no expression of Ectodermin outside of this particular region.

Reverse Transcription PCR

Result of the RT-PCR shows that no mRNA of Ectodermin is present in the unfertilized egg. Unlike in *X.laevis*[2], maternal mRNA of Ectodermin are not present in the sea urchin egg, meaning that expression is certainly triggered by zygotic factors. Expression of Ectodermin rises up during the cleavage stage to reach a peak at the morula stage. During blastulation, expression levels decrease before re-increasing when the gastrulation starts. No significant data has been collected for the following stages, but it can be proposed that, similar to what happened during the cleavage, the level of Ectodermin continues its increase during gastrulation.

From these data, it can be concluded that expression of Ectodermin seems to occur in two waves in the embryo: the first during the cleavage process and the second during the gastrulation. Probably Ectodermin assumes different function in each of these events but more experiments are needed to answer these questions.

***In-silico* analysis of putative *cis*-regulatory elements**

Database queries on the known sequence of *ectodermin* and its surroundings provided some hints on the putative regulatory mechanisms involved, especially in the light of the ectopic expression results. Connections with Twist signaling should be explored as it's a key player in triggering skeletogenesis [10]. No clear implication of Nodal-related factors has been established so far, but given the fact that sequence databases concerning sea urchin regulome are far from being complete, this hypothesis shouldn't be dropped yet.

It should as well be reminded that only one full genome sequence issued from one individual is available for *Paracentrotus lividus*, therefore polymorphism and sequencing errors can seriously alter the significance of the *in-silico* results gathered.

Promoter-GFP transgenes

Three promoter-GFP transgenes (with promoter fragments of 2.5, 1.45 and 1.1kb) have been successfully cloned during the period of this project. Correct amplification of the desired fragment was probably the major issue. The occurrence of polymorphism makes difficult to design highly specific primers set for the desired amplicon. As it can be seen in figure 10, different genomic templates, issued from the sperm of four specimens, gives variable PCR products. Predicted restriction maps are as well disturbed by polymorphism phenomenon and therefore it's difficult to rely only on them to screen the

clones. Extensive use of a Roche agarose gel extraction kit in the early attempts has been suspected of hindering the ligase activity, as none of the ligation performed ever gave rise to a single recombinant colony after transformation. As soon as the kit was dismissed in favor of the phenol-chloroform extraction combined to a shotgun ligation strategy, recombinant colonies were seen at each transformation performed.

As explained in the results section, none of these transgenes worked as expected in the sea urchin embryo. Ect2.5-GFP has been the first construct to be injected. Given the issues encountered to realize this construct and the fact that the confirmation of the result came only from a partial sequencing of the upstream region of the promoter region, it's possible that a point mutation, induced by the PCR, occurred in one of the supposed regulatory modules or in the promoter module. Another hypothesis came after closer examination of the sequence: repeated simple DNA patterns have been spotted in the upstream region, that could lead to heterochromatin formation and thus preventing expression of the GFP.

To overcome this issue, two strategies have been deployed: the first approach was to remove the anterior part of the sequence by enzymatic digestion (conducted with *Kpn*I) to cope with the supposed heterochromatin-forming sequences, leading to a new transgene, dubbed Ect1.1-GFP. The second strategy was to design an entirely new construct containing a 1.45kb sequence of *ect*1 promoter region (pEct1.45-GFP).

As neither of this new transgenes induced expression of the GFP in the injected embryos, another hypothesis has been envisaged. Examination of the genome organization of *P. lividus* has shown that both *ectodermin* genes are followed by similar-sized non-coding sequences of 4.5kb. Alignment of these sequences revealed a 2.5kb conserved sequence starting from the stop codon of *ectodermin*. Such conserved non-coding sequence at the 3' extremity of both copies is most probably involved in the cis-regulation of the gene (and in-silico analysis showed that many putative binding sites are present). The new design strategy deployed currently for the transgenes is to clone this conserved region at 3' extremity of the GFP sequence along with the 1.45 kb promoter region in 5'.

Ectopic expression of Ectodermin

Injection of synthetic mRNA to induce ectopic expression of Ectodermin gave unexpected results, in respect to the start hypothesis, stating that Ectodermin should act as a repressor of Nodal pathway and block the differentiation of the oral ectoderm. The effects of ectopic expression do not concern directly the oral ectoderm; actually the oral region is normally developed in all the injected embryos observed. Ectopic expression effects occur in a total different tissue: the skeleton of the pluteus larva. Skeletogenesis had suffered severe disturbances leading to the mineralization of abnormal structures such as forked spicules, supernumerary spicules radiating from existing and new mineralization in regions that usually don't know this phenomenon. The severity and the type of the mutations differ from one batch of eggs to another, as well as among the specimens of the same batch. The variation between batches can be simply explained by polymorphism (different batches, different females). The

variation among individuals is due to the fact that, along the microinjection session, injected volume of mRNA is not constant, so some embryos will receive more than others and show stronger effects.

Alterations in the skeletogenesis are not a new-discovered phenomenon in sea urchin. Nickel treatment is known to produce radicalized embryos with disorganized spicules growing along the radial axis. But the mutations induced by the Ectodermin do not alter the basic skeletal structure of the pluteus. In most cases, it's possible to recognize the genuine features of a normal embryo like the two spicules radiating from the oral side to form the vertex and the four arms extending around the mouth area.

How does Ectodermin interfere with the PMCs responsible of the biomineralization is yet still unknown. It can be proposed (by analogy to the studies in CRC cells [1]), the enzyme promotes the proliferation of the PMCs that will start new biomineralizations out of the predicted grid. That could explain the supernumerary phenotype where new spicules radiate in a parallel way to the native spicules.

6. Conclusion and Perspectives

A few steps in the comprehension of Ectodermin role in the embryogenesis of *P. lividus* have been achieved. In Situ Hybridization showed this enzyme is restricted to the oral territory of the embryo and RT-PCR suggested that it's probably expressed at two different times. No significant results have been obtained with the promoter-GFP transgenes but ectopic expression experiments showed unexpected effects on the skeletogenesis of the embryos. Polymorphism proved to be a challenging issue, as all the sequence predictions have been made on the single genomic sequence available.

However, a lot a work still needs to be done to put the *ectodermin* gene in its right place in the ectoderm network as well as in the oral ectoderm/PMC inductive signaling.

First, designing a functional promoter-GFP-3' region transgene has to be completed in order to understand which non-coding sequences are involved in the *cis*-regulation of the expression. Morpholino knock-down of putative binding sites, unraveled by *in silico* analysis, must be performed to identify which factors are involved. Ectopic expression of an mRNA coding for an inactive Ectodermin (mutated in the catalytic domain of the enzyme) could complete the data obtained with the ectopic expression and help to assess the exact function of this protein during the development of the embryo.

7. Acknowledgements

First I would like to thank Prof. Sergio Schmid and Prof. Letizia Vittorelli who gave me the possibility to come in the University of Palermo to realize my diploma project. Then, I thank Prof. Giovanni Spinelli for being my advisor and welcoming me in his research group and Dr. Enzo Cavalieri for driving me through this project and teaching me day after day more than what I could have ever expected to know about sea urchins and biology of the development. Last but not least, I thank Rosa Guarcello, who provided me assistance whenever I needed it and will now carry on the project on Ectodermin.

8. Bibliography

[1]Ashby Meredith, *The Sea Urchin Regulome in Development*, Thesis for the Degree of Doctor in Philosophy, Californian Institute of Technology, Pasadena (2007)

[2]Dupont S, *et al.*, Germ-*Layer Specification and Control of Cell Growth by Ectodermin, a Smad4 Ubiquitin Ligase*, Cell Vol.121 (2005), 87-99

[3] Sasai Y, *Developmental biology: a blank canvas no more*, Nature 435(2005), 433-434

[4]Whitman M, *Nodal signaling in early vertebrate embryos: themes and variations*, Dev. Cell 1 (2001), 605-617

[5]Duboc V, *et al.*, *Nodal and BMP2/4 Signaling Organizes the Oral-Aboral Axis of the Sea Urchin Embryo*, Dev. Cell Vol. 6 (2004), 397-410

[6]Sea Urchin Development Basic to Understand Genomic Circuits Chapter, consulted the 01/03/2009 http://www.bio.davidson.edu/courses/GENOMICS/method/UrchDev.html

[7]Cavalieri V, Di Bernardo M, *et al.*, *cis-Regulatory sequences driving the expression of the Hbox12 homeobox-containing gene in the presumptive aboral ectoderm territory of the Paracentrotus lividus sea urchin embryo*, Dev. Biol 321(2008), 455-469

[8]Sambrook J, Fritsch EF and Maniatis T, *Molecular Cloning – A Laboratory Manual 2nd Edition*, Cold Springer Harbor Press (1989)

[9]Cavalieri V, Di Bernardo M, Spinelli G, *Functionnal Studies in the sea urchin embryo*, Microinjection, Methods in Molecular Biology Vol.518 (2008), 1-14

[10]Wu SY, Yang YP, McClay DR, *Twist is an essential regulator in the skeletogenic gene regulatory network in the sea urchin embryo*, Dev. Biol 319 (2008), 406-415

[11]Vonica A, *et al.*, *TCF Is the Nuclear Effector of b-Catenin Signal That Patterns the Sea Urchin Animal-Vegetal Axis*, Dev. Biol 217 (2000), 230-243

[12]Kiyama T, et al., *Strongylocentrus purpuratus transcription factor GATA-E binds to and represses at an Otx-Goosecoid cis-regulatory element within the aboral ectoderm-specific spec2a enhancer*, Dev. Biol 280 (2005), 436-447

[13]Popodi E, *et al.*, *Sea urchin Hox genes : insights into the ancestral Hox cluster*, Molecular Biology and Evolution Vol. 13 (1996), 1078-1086

[14]Arnone MI, *et al.*, *Genetic organization and embryonic expression of the ParaHox genes in the sea urchin S. purpuratus: Insights into the relationship between clustering and colinearity*, Dev. Biol 300 (2006), 63-73

[15]Smolenicka Z, *et al.*, *Sequence of a conserved region of a new sea urchin homeobox gene from the NKfamily*, Cell Biology International Vol. 27 (2003), 81-87

[16] Poustka AJ, *et al.*, *On the origin of the chordate central nervous system: expression of onecut in the sea urchin embryo*, Evolution & Development 6:4 (2004), 227-236

9. Annexes

9.1. PCR protocols

Amplification of Ectodermin 10kb, 5kb and 2.5kb promoter region

Invitrogen	Vol. [µl]
genomic DNA	4
F-primer 10mM	1
R-primer 10mM	1
dNTPs 10mM	1
Buffer A	5
Buffer B	5
Enzyme	2
H_2O	31
Total vol.	50

DyNAzyme	Vol. [µl]
genomic DNA	4
F-primer 10mM	1
R-primer 10mM	1
dNTPs 10mM	1
Buffer 10x	5
Enzyme	0.5
H_2O	37.5
Total vol.	50

Cycles	Time [s]	Temp. [°C]
Init. Denat.	60	94
10	30	94
	30	58
	180	68
30	30	94
	30	62
	180	68
Final Hold	∞	4

Optimizations of the PCR protocol for Ect2.5

Invitrogen	Vol. [µl]
genomic DNA	4
F-primer 10mM	1
R-primer 10mM	1
dNTPs 10mM	1
Buffer A	4
Buffer B	6
Enzyme	2
H_2O	31
Total vol.	50

Cycles	Time [s]	Temp. [°C]
Init. Denat.	60	94
10	30	94
	30	55
	300	68
30	30	94
	30	60
	300	68
Final Hold	∞	4

Cloning of Ect2.5 by shotgun ligation

Invitrogen	Vol. [µl]
genomic DNA	4
F-primer 10mM	1
R-primer 10mM	1
dNTPs 10mM	1
Buffer A	4
Buffer B	6
Enzyme	2
H_2O	31
Total vol.	50

Cycles	Time [s]	Temp. [°C]
Init. Denat.	60	94
10	30	94
	30	55
	300	68
30	30	94
	30	60
	300	68
Final Hold	∞	4

9.2. Sequence alignments

Alignment of *ect*1 and *ect*2 promoter regions:

```
CLUSTAL 2.0.10 multiple sequence alignment


ect2            ------------------------------------------------------------
ect1            GGGTGGTGATGTTTGTTACTCTCATGGTTCGTTAGTCCGAAAATGGAAAAAAGGTTCGTT 60


ect2            ------------------------------------------------------------
ect1            AGTCCGAAAATGGAAAAAAGGTTCGTTAGTCCGACAATAGAAAAGGTTCGTTAGTCCGAA 120


ect2            ------------------------------------------------------------
ect1            AATAGAAAAAAGTTTCGTTAGTCCGAAAATGGAAAAAAGGTTCGTTAGTCCGACAATAGA 180


ect2            ------------------------------------------------------------
ect1            AAAGGTTCGTTAGTCCGAAAATGGAAAAAAGGGTCGTTAGTCCGAAAATAGAAAAGGTTC 240


ect2            ------------------------------------------------------------
ect1            GTTAGTCCGAAAATAGAAAAAAGGTTCGTTAGTCCGAAAATGGAAAAAAGTTTGTTAGTC 300


ect2            ------------------------------------------------------------
ect1            CGAATATGGAACAAAGGGTTCGTTAGTTCGAAAATGGAAAAAAGGTTCGTTAGTCCGAAA 360


ect2            ------------------------------------------------------------
ect1            ATGGAAAAAAGGTTCGTTAGTCCGAAAATAGAAAAGGTTCGTTATTCCGAATTTTTGTTA 420


ect2            ------------------------------------------------------------
ect1            TATCCAAAAATAGAAAAGGTTTGCTATTCCGAAAATAGAAAAGGTTCGTTATTCCGAAAA 480


ect2            ------------------------------------------------------------
ect1            TAGAAAAAGGTTTGTTAGTCCGAAAATAGAAAAGGTTCGTTATTCCGAAAATAGAATGTG 540


ect2            ------------------------------------------------------------
ect1            GTTCGTTATTCCAAATCGTTAGTCCAAACGTAGAAAACGTTTGTTATTCCGAAAATATTA 600


ect2            ------------------------------------------------------------
ect1            AAAGGCTCGTTAGTCCAAAATTAGAAAAGGTTTGTTATTCCTAATTTTCGTTATTTTTCA 660


ect2            ------------------------------------------------------------
ect1            AAAGTTTAACTCGAATTTATATAGAGAGGAAAAGAGAACTTAAATTTTGGCCCTCTTTTT 720


ect2            ------------------------------------------------------------
ect1            CTTGCTCTCGCATGTTTTTTTTTCCACCCGTATATTGAGATTATTATAAAATTGAAAAGC 780


ect2            ------------------------------------------------------------
ect1            ATATAATTATAGAATATATCCAAGCGATCATATAATGTCAACAGTTTTCTCTCCTTTTTT 840


ect2            ------------------------------------------------------------
ect1            GCCTTGCTTTCAAATCATTAACGAATAACGAACTTTATTGTATTTTGGGAATAATTAACC 900
```

```
ect2            ------------------------------------------------------------
ect1            TTTTTGTATTTTTGGATATAACAAAAATTCGGACTAACGAACCTTTTCTATTTTCCATTT 960

ect2            ------------------------------------------------------------
ect1            TCTGAATAATGAACCTTTTATATTTTCGGACTAACGAACCTTTTCTATTATCGGACTAAC 1020

ect2            ------------------------------------------------------------
ect1            GAACCTTTTCTATTTTCGGACTAACGAACCTTTTTTCCATTTTCGGACTAACGAACCTTT 1080

ect2            ------------------------------------------------------------
ect1            TTTTCCATTTTCGGACAAACGAACCTTTTTTCTATTTTCGGACTAACGAACCTTTTCTAT 1140

ect2            ------------------------------------------------------------
ect1            TTTTGGACTCATCCATTTTCGGAATAACGAACCATTTACTACTTTCAGAATAACGAACCT 1200

ect2            ------------------------------------------------------------
ect1            TCAGAGAAACAAACATCACCCCCCCCCCCCCCCCGCGGACTATCGAACCCGCGGACTATT 1260

ect2            ------------------------------------------------------------
ect1            AACGGACCCGCGGACTATCGGGCAGTAACCGTCAATTTTTGTAAAATTTGATGACGTCAT 1320

ect2            ------------------------------------------------------------
ect1            CGAAAAGTTATGGACAAAAAATAGCGAATTTAAAATGCATTACATAACAAGGGATTACAT 1380

ect2            ------------------------------------------------------------
ect1            CATTGTTTCCTTATGAAAATGACATCACAATGACTGCATTGAAATTTGCCTTTCCATTGA 1440

ect2            ------------------------------------------------------------
ect1            TACCATATAAAATATAGGTTGTCACAGAAAACTTTTGAGTAAATCACGTTTTTGTAGGAG 1500

ect2            ------------------------------------------------------------
ect1            CCATTATGCCTGCTTACCTAATAATTGCAAATTTGACAGAGCCATGTGCACGCCGGATAA 1560

ect2            ------------------------------------------------------------
ect1            TATGGAAAGAAAGCGAAGTTCTCCAAATTGCACAGCCAATACACAAACCGTATTCGTTTT 1620

ect2            ------------------------------------------------------------
ect1            ACATACAGATCTTGTATAATTATGTAAAGCCTACTCCCGTTTCCCAATATACTTCAAAAT 1680

ect2            ------------------------------------------------------------
ect1            ATACGTTTACATTTTTATAATCCTCGATAGGCCTATATTTTAGTCAAAGAGTCAATGCAC 1740

ect2            ------------------------------------------------------------
ect1            GCACATCGTTCATGTCAACGATTGCAGTTCTAGTCATTATTATTGTTATTATTGTTATTT 1800

ect2            ------------------------------------------------------------
ect1            TTATCATTACTATCATCTTTACTAATATTGATTAATACAACTCTATTCATGTTGAACGAA 1860

ect2            ------------------------------------------------------------
ect1            ATTACATGCATGCTCTGCCGATCTATAGAGGGTAAATTTTTATCATGCTGAAGAGGACGA 1920

ect2            ------------------------------------------------------------
ect1            TTTCCATATATTCTGTATGCTCAAAGTGTCAAAGGTTGATGGGGAAAACCCAGTGTTATA 1980
```

```
ect2            ------------------------------------------------------------
ect1            CTTGCACGTCCAACGATCGTTCCAACACACGTAACTTGGTCTACGTTGTCCGAATATGAT 2040

ect2            ------------------------------------------------------------
ect1            TAACTTTAACCTGACTACGTGACGTCAAATAATTAAGTTTCTTGAACGTACCAGATTGTT 2100

ect2            ------------------------------------------------------------
ect1            ATTGTTAAAATTCCCTGTTTGGTTGGTTGGATGGATGGATGAATGGAGAACAGTTGGATG 2160

ect2            ------------------------------------------------------------
ect1            GTGAATGATATGACTCTACATGCCGTGTATTTTTGAAAGCTCTATGGCATTAAATCATAT 2220

ect2            ------------------------------------------------------------
ect1            GAACCGTTAATACCACGTTTATATTTTGCGAAAAGGTGCTCATGAATTAGGTAAGTTCGT 2280

ect2            ------------------------------------------------------------
ect1            ATTACCAAATAGGCCTAGAAATTGTTTCGGTGTGTAAGCTTAAGCAGATTATGCTCATAA 2340

ect2            ------------------------------------------------------------
ect1            CATTAACTTCCATATTATGTTTTAACTAAACCATTATACGTCTAAGGCATAATATAATAT 2400

ect2            ------------------------------------------------------------
ect1            ATATTGCAGGGGCCGCTGAGCGTTTTTGCAAGTGGAGGGGCTGAGCGGAAAGTGGGGGGG 2460

ect2            ------------------------------------------------------------
ect1            GGGGGGGATCTGAACAGGTTAACATTAAAAATCTATATATGGTCATTTTCGCGATTTTAT 2520

ect2            ------------------------------------------------------------
ect1            ACACGTTTTGGCTGAGAAAAGTGGGAGGGGGGGGGGGGCTCCCCCCGGCTCCGCGGCCC 2580

ect2            ------------------------------------------------------------
ect1            TTGTATTGGAACATTCTTTGCCTTTTCGGTAATTTTGTTTTTGGAATAGGTGTACCATGT 2640

ect2            ------------------------------------------------------------
ect1            GGATGAGTTTTAAATGTAATGGGAGCTTCAATTACTTTTCAGTAATCATCTTTTTACTAC 2700

ect2            ------------------------------------------------------------
ect1            TGAAGGGCCCACATTAAATTTTATCTTTATGATAGTGTTTGGATTGAGAGAGCTAAAAAG 2760

ect2            ------------------------------------------------------------
ect1            AGGTTAGCATGGTTGTGAGGGGATATGGGTGAGGCAGGAAGTGCATGAGAGGTGGGGGGG 2820

ect2            ------------------------------------------------------------
ect1            GGGGTGTTGGGAGAGAAAGAGAGAAAAGAAAAAGAAACGAAGAAAGATAGAGAAGCAGAT 2880

ect2            ------------------------------------------------------------
ect1            AAATTAAGAGAAGGGAGAGATAAAGAGAAAGAGGGGAATATAATAGAAGGCGGGAAATAA 2940

ect2            ------------------------------------------------------------
ect1            GAAGAAATTAAAACAAAGTTAACATGGGCGATCACCTCGCTTCTGTGTAACTATTCTTTG 3000

ect2            ------------------------------------------------------------
ect1            CTCTCTTTTCTTTTTCTATCTCTTTCTATCTCATCTCTCTCTCTCCCCCAATCTCTCTCT 3060
```

```
ect2            ------------------------------------------------------------
ect1            CTCAATTTTCTCGCTATCTGTATTGTTTAGAGAGAGAGAGGGGGGAGGAGAGAGGGGGAG 3120

ect2            ------------------------------------------------------------
ect1            GAGAGAGAGGGGAGGAGAGAGAGAGAGAGAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGA 3180

ect2            ------------------------------------------------------------
ect1            GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAAAGAGGCTACATCGGTTGTGGAGCAATTGC 3240

ect2            ------------------------------------------------------------
ect1            GATATTTTTATTTTAATTATTATTTATTATTTTTATTATCATCATCATTATCATTACTAT 3300

ect2            ------------------------------------------------------------
ect1            TATTAATTAATTCAACTCTATTCATGTCGAACGAAATTACGAGGGTAAATTTTTATCATG 3360

ect2            ------------------------------------------------------------
ect1            CTGAGGATGATTTCCATAATTATATTCTGTATGCTCAAAGTGTCAAAGGTTGATGGGGAA 3420

ect2            ------------------------------------------------------------
ect1            AACCCAGTGTTATACTTGCACGTCCAACGATCGCTCCAATACACGTAACTTTGTCTTTAT 3480

ect2            ------------------------------------------------------------
ect1            ATATATTGCCAAGTATGATTGCTGTTAGCTTGACATAATAACGTCAAATAATTATGTTTT 3540

ect2            ------------------------------------------------------------
ect1            CTTGAATGTGTCATGTTGTTATAGTTAAAATTCCCTGTTCGGTTGGATGTATGTATGAAT 3600

ect2            ------------------------------------------------------------
ect1            GGAGAGCAATTGGATGGTGAATGATGATGCAGCCGTGTAATGTTGGAAAGCTCTAAAACA 3660

ect2            ------------------------------------------------------------
ect1            TATCAAACTGTTCACAACACTGTTAGTTTTTTGTTTTGCGAAAAGGTGCTCATAAATTTT 3720

ect2            ------------------------------------------------------------
ect1            TTTGAAGTTTTTTGAAAACCAGATAGGCTTATACATGTATATGGTCGATTCCAGCGAAAA 3780

ect2            ------------------------------------------------------------
ect1            GTGTACACACTGTTTTTGATCATGTTCAGGCCAATAAAAAAATCCAAAAAAATGTAGGAA 3840

ect2            ------------------------------------------------------------
ect1            CCTCAATTTTGCGTTTTATGTATCCGCACCGGTGAACTTATTCGTGCCGCATTTGTTCCA 3900

ect2            ------------------------------------------------------------
ect1            ACTTCATTTGAAGTGAACTTTTTTTTTAACGATTTAAAAACATGCCATTTTTGGCCATTT 3960

ect2            ------------------------------------------------------------
ect1            TTGGAAAAAAAAAAAAATTAATTTTTTTTTTCGCGACCAAACTTTTTATAGAAAATCATCA 4020

ect2            ------------------------------------------------------------
ect1            AATAATTTAGATTCAAGTTGTATTTTAGGATGGCACACATCTCAGGAGAAATTGTACCAA 4080

ect2            ------------------------------------------------------------
```

```
ect1            ATTTGAATAATATAGGGCAATCCAAACAATCTCTGGAATTTTTTTTCGAGTGGTAAATTT 4140

ect2            ------------------------------------------------------------
ect1            GGATGTTACCTACGAATACAGCTTTTTTTAAAAGCTTATAAACCATACTAGATATAACGA 4200

ect2            ------------------------------------------------------------
ect1            AGTACGGCTCCTAATTGCCGTACAGAAGCGCCTCACCATTCCCTATCGACATGCATTGTT 4260

ect2            ------------------------------------------------------------
ect1            CCCGGTGACATTGCTTTCTTTGTTAAACTGGCAGAAGAGGTCAAAGGTCCGAAATTTTCG 4320

ect2            ------------------------------------------------------------
ect1            AGTGGTCAAAAAATTCAGTTACCTACGAATACACCTATAAACCACTGCCATTTCTATACC 4380

ect2            ------------------------------------------------------------
ect1            TTAAGAGATTAGTGACAAATATTGTAGGGAAAGTTGACAAACGATGTCTTGCTGACGTGT 4440

ect2            ------------------------------------------------------------
ect1            GTAAACTTTTATCATCCTTGAACAATAGGTTTTTGAGATATTGCCTTATAAGTGCTGTAT 4500

ect2            ------------------------------------------------------------
ect1            TCGTAGGTAACATTTGACATTCCGAATGTTACCTACGAATACAAGTATGCATAGCGTGTT 4560

ect2            ------------------------------------------------------------
ect1            TCTGGCATATATTTTAGTAATTTGTCTTGAGACTAAAGTTACAAACTTTTGTAGTTTCAT 4620

ect2            ------------------------------------------------------------
ect1            TTGTCATATAAATAGGATATTGTGTGACAATTTTCTTGTTCAGAATCTTATGTTTTCATT 4680

ect2            ------------------------------------------------------------
ect1            ATGTTTCGACATTTTCCTTATTAATGTAAAATAATTTATATTTTTTAATATTATTGTTGA 4740

ect2            ------------------------------------------------------------
ect1            GAAGTAATTTTCCTCTGAAACAATGTTATTATCAAGTTTAATTTTTTTCGTAGCGCTTAT 4800

ect2            ------------------------------------------------------------
ect1            CAAGTGGGGAAACTCTTTCAAATTTAATTACCAAAAAAATTATAGCTTTTCAGGAATAAA 4860

ect2            ------------------------------------------------------------
ect1            TAAAAACATGGGATAGCATTTCCTGCAGTGGTGTCACGTGTATTCGTAGGTAACAATGTG 4920

ect2            ------------------------------------------------------------
ect1            TATTCGTAAGTAACAATGTAACGTAAACTGCCATTTTCGGCTTTACAATTAACCTCGCAT 4980

ect2            ------------------------------------------------------------
ect1            TGCTTTTTCTTTATTGGAAAGGGAGGAATAGTCCCAATAGATATTGGGGACATTGTTTGG 5040

ect2            ------------------------------------------------------------
ect1            ACCCAATAATATAAGTTATTTTATTGTATTAGTGATTATCAGATTTGAAGTAAAGTGCCC 5100

ect2            ------------------------------------------------------------
ect1            CAAATTATAGCTTATTGCATTATTCGGGATTAACAAATCCTTTTTTCCTATTTCATTCAA 5160
```

```
ect2            ------------------------------------------------------------
ect1            GGCACGTTTGCCGATTGACACAATCTGTACATAACAGATGTGAATTAACCTTCTCTAATT 5220

ect2            ------------------------------------------------------------
ect1            TTATTTCAAAATACGAAATGAATAATATTTACTGAGATGGATTTTCATTACTTATAAAAA 5280

ect2            ------------------------------------------------------------
ect1            AAATGAAAACAAAACTAAACGTTGATTTTTTTTTCCGATTTATCCTTTTTTTAGATTTCA 5340

ect2            ------------------------------------------------------------
ect1            TTTATCAGGTGTTTATGTTTCTTATTAATGAATTGCAATCTGGAGTTCAATACAAAGCAA 5400

ect2            ------------------------------------------------------------
ect1            ACTTGCCCCATTCTTTAAGTTATCTTTAATTTCCAAATTGGGATTTCATGCGTATTTACT 5460

ect2            ------------------------------------------------------------
ect1            CAGACCTCATAATTTCATAAGTTATCGAAACCATGTTGATTAAAGTAAAGTTGACTTCAT 5520

ect2            ------------------------------------------------------------
ect1            TTTCATTTTTATTTTTGATGAAAATGAAAATGATGTAAAAGTGACAGATGCAGACGTATA 5580

ect2            ------------------------------------------------------------
ect1            TCGCTCTCTTTAATCATACAGTGCAGCAGAAATTCAATTTGGTTGGGATATTCAGCCTGG 5640

ect2            ------------------------------------------------------------
ect1            ATGTGTGTCATTGACGTATTATATTCTCTCAAAATACTTTGATTTTGGTTACGTTCGAGC 5700

ect2            ------------------------------------------------------------
ect1            AAGAGGTCTTTAATTCTTTCCAATATTCATATCATATCAAGAAACAGTCAGTTATGAATA 5760

ect2            ------------------------------------------------------------
ect1            TTTTAACTCATAAGCAGTTCAAGACAGTCACTGTTACCCACAAATAGTATCATTGGAATA 5820

ect2            ------------------------------------------------------------
ect1            AGATAACAGAATGGGATACTCATATGTGCATGTAAGTGCCTAGATATATACGACGATACA 5880

ect2            ------------------------------------------------------------
ect1            ATACAACATTTTTTTTTATTAGTTAAAGGAAGGGTATAATGAATATTTAAATTTTAAAAT 5940

ect2            ------------------------------------------------------------
ect1            GATATATAGCTAATATTTTGTTCGTTTATGGTTTCACGCTTATGTAAATATGATTTCAGA 6000

ect2            ------------------------------------------------------------
ect1            ATAAACATGAGGAAAGAATGCCTATACCAGTTAGTGAATCATTAAAAAAAAGTGCTCTAC 6060

ect2            ------------------------------------------------------------
ect1            ATCTATAATGTAATGTTCGTCGAAAGACATTTGATTAACTTTTTTTTCTTCGAATTATGT 6120

ect2            ------------------------------------------------------------
ect1            TACAGTATTTATGTTAGACAGAAGATATCGTATTCATTTTTTATTGTTGGCTAGGGGCTG 6180

ect2            ------------------------------------------------------------
ect1            ATGCTATTAATAAATATTCGCAATAAATTATTCTCATGTGTGAATCTTTTAGTTTAGACG 6240
```

```
ect2            ------------------------------------------------------------
ect1            TTTATCATTTTAATCGGTATTTATGCTTGTTATCATGCACAATGTTGAGAAAAATACCTT 6300

ect2            ------------------------------------------------------------
ect1            ACACTTATTAGTACCAATAATCTTGATTAAAAGATATAAACATACGATAAACCGGACAAA 6360

ect2            ------------------------------------------------------------
ect1            TATTGACGTTTTTGGTTAAAAGTTAGTGTTTCCCCCATACTTGGTAGTATGTTCCCATAT 6420

ect2            ------------------------------------------------------------
ect1            ATTGGCTGAGAAAGATATTTTTCACTAGAAACAATTAGTTTGATACCCCATCACGATTAT 6480

ect2            ------------------------------------------------------------
ect1            TATATCAATTGGATCCAAACAAATAGTTTAAAATATACATGTAAGCCAGTTATACCTATG 6540

ect2            ------------------------------------------------------------
ect1            TTACCTACGAATACATTAGCTGTTACCTACGAATACATATGTTTCTACCTCTTCTATAAA 6600

ect2            ------------------------------------------------------------
ect1            CAAAAGTAATGGGCGGATTCCAATTTTTTTAGGGTTTTAATGAGGAGTCATTCTCTTTTC 6660

ect2            ------------------------------------------------------------
ect1            TTTCAGTAAGATATGACATATCTTAATTAGCGTCGTGAAGATTAAACAATGCAAACTGAC 6720

ect2            ------------------------------------------------------------
ect1            ACAACCCCATGTTTTTAAAACAGATTGTCTTAAAACTATAACTTTAATGAGATATGGCAA 6780

ect2            ------------------------------------------------------------
ect1            ATACAAGGAAATTAGGAGCACAAGCCCTGCTCTTACCGATTTTAACTATTGTTAGAATCT 6840

ect2            ------------------------------------------------------------
ect1            TATTGATGGAATGGAATGACATATTTGGAAGGAAAATGTATATTAGGAGGCAAAATTCTG 6900

ect2            ------------------------------------------------------------
ect1            ACTGTATCACATTAGGGTACGTGTATTCGTAGGTAACATGCAAAAAATATGTTTTATTTT 6960

ect2            ------------------------------------------------------------
ect1            ACCCCTCTGCCACAAAGTCAATTTGGGAGATTAAGCGTGGCAATTTATCACACGAATAGG 7020

ect2            ------------------------------------------------------------
ect1            TTGATATGTTGACTTTTAGTGGGTCAAAGGACTAAATCATCAAAGAGGTTCTAAATTTGG 7080

ect2            ---------------------------------------GGGAGGTGTTAACAAATTTTC 21
ect1            ACTCAACTTTTTCGCCGTGTACACTTTTCGCTGGAATCGACCATATATATATATATTTTT 7140
                                                        *  * *  * * *****

ect2            G---GCCTTTGGCCGTTCACAAGTTCAGCAAATATTACCTATAGGGGTAGGGTCTGGCCG 78
ect1            GTTTGTTTTTGTTTGTTTTTTGTTTCTTTGTTTCTTTGTTTTTTGTTGGGGGTCTGTGAG 7200
                *   *  ****   ***      ***      * **   * *  *    *******   *

ect2            CTCGATCGGCCCCGAAGAAATTGACTAGCAAAAAAAAACAAAACAAAAACAA-ACAAACC 137
ect1            CATAATCAGATTATGCACATTAAACAACTATACCTCTAAGACACATATATATTGGAAATT 7260
                *   *** *         * *  ** *  * *     *  * *** * * *    ***

ect2            AAAAATGTTATTTAAATTTAACGGGGGGAGGGGGGGGGGCTCACTAAATAAAAACAAGG 197
ect1            ATTCGTCTTTTTGGTAATTTCTTTTTTGGGAATAGATAGGCCTATACCATGTGGATAAGA 7320
```

```
                 *    * ** **   * **        * *    *   ***  *    **    * ***

ect2             ATCATGTACAAAAGGGATGCAGTTTTCCGATCTTACGGGGTTCATCCGCACCCTCCGCCC 257
ect1             GATAAG-ATTATGAGAGTGAAACCCAGAAACAGTGCAATTAATGATCAGGGACTTTATTG 7379
                    * * *  *   *  ** *        *   * *          *     **

ect2             CCTGGATCCGCTCCTGTATTTGGTTCGTTGTCGCTTTATCTTCAGAACTAGTTGAGATGT 317
ect1             TT---ACGCTTTATAATACAAACTAAGTAATAGTATTATTGTCATTATTAATGATGATCG 7436
                      *  *  *    **     *  **  * *  ****  ***  * ** *   ***

ect2             GAGCTTTA-AACTCATGAGTAGTATGTATCACC-----ATAGGAAGATGACTTGATTAGA 371
ect1             GAGCCAAGTAATACATGTATAGCAAGTATCATTTTGATACAAGGGGATGAAAGGGTGAGA 7496
                 ****     **  ****  *** * ******       * * *  *****   * * ***

ect2             TTTTGTAGTATGTAG-GTCAATGTTCAAGGGTCATATGGGTTAAAGATGCACGTTCCTCA 430
ect1             GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAAAGAGAGAGAGAGAGAGTAA 7556
                     *    *   ** *  *  *    ** *  * *       *    *   *    * *

ect2             TAGTGT---AGAGGGAGTTCAAAATACGTACGCGGTAACGA---------------ACAT 472
ect1             GAGAGAGAGAGAGAGAGAGTAAGAGAGATGGGGGGGGGCAATGTTCCTAGAAATCTAAAT 7616
                  ** *    **** ***   ** * *  *  * **   * *               * **

ect2             TGCAAGATGCCCATTCA---------TACCAACGATGTGCGACTCTTACGCA-----GTG 518
ect1             TGTGATACGTACAGCCAAAAATCAAGCACAAAACAAAGGGTACTTTTGCGTACTATTGTG 7676
                 **  * * *  **  **          ** **  *   *  *** ** ** *     ***

ect2             GTCCAAAAACCGAAAGCTAGAACGCC---CGAACGATTGCCGA-AGTCATGAGCCAATTT 574
ect1             TTTCTATCATATAAAGTTGGCAAATCAAGTGTACATCCGCCGTCACTTCTATAATGACAT 7736
                  * * *  *   **** * * *   *    * **    ****  * *  *      *  *

ect2             GTTATCGGCATTGATTCGGCTGCTC--GCATGTTTGGATGCACAAGAACACTGAATTGAT 632
ect1             CGCATGATCATAGACGTAACGTCTATGGCATGATCAAGGGCAAATTGAAAAAAGACTGAC 7796
                    **   *** **     *  **   ***** *     *** *   * *    * ***

ect2             GCAGTGCATG----------------------CTCTACACACACACACGCTT-------- 662
ect1             GTTGAGCATGGGCCAATATACTCGTAATATCGCTCAAGATTTATACATAATTTCTTTTTT 7856
                 *  * *****                      *** * *   * ***   **

ect2             -GTACGCTA---ATGCAGCGCGATCGCGATTGGCTGGTGTGTGTTTGCATAAAACCGTAT 718
ect1             CGCACACTATAGATACTTCGCTATT-TGTTTTGTTTTGCTGGGTTAAAAAAAAAATGTAT 7915
                  * ** ***   ** *  *** **   * ** * *    ** ***   * ****  ****

ect2             GCAAATTAAGTCATGCTCTAAGCT-TCGTGAACAATA--TGGCTGCGA----CCACACAG 771
ect1             TCT--TTCAGTCGCGCCTCAAACTATAATTGACAATCACTGACAGTAAAGAACCACATGA 7973
                  *   ** ****  **   ** ** *  *  *****   ** * *  *    *****

ect2             CGCATCTTGAATTCTCGTTTTAGTTTAGCGAT--GCTCGTGAATGACCACCTTTAAATTT 829
ect1             GAAATAGAGAAGATATTAGACGAATGAAAGATCAGTCTGTGACAAATC-TATATAATTAT 8032
                    **   ***             * *  ***  *   ****   * *   * *** * *

ect2             TGAAATTCC------ATGTGTTGAGT--TCGTTATCGCAATATCTCGAGAATC------- 874
ect1             TATGTTGCCTGAATTATGATTTAAATAATCGCCCTTACATTGTTTTGGTAATTTGAGGTG 8092
                 *    * **      ***  ** * *  ***   *  ** * * * *  ***

ect2             -----AAATGACGTATTAGCTTCAA-----ACTAAGGGGTAGTATGC----ATCATCATA 920
ect1             TACATAGATTGCATATCA-TTACAACATGTACAAGGCAGAAGCATATCTTTATTATTTTA 8151
                      * **  * *** *  * ***     ** * *  * ** **      ** **  **

ect2             TGAAGA-TGGTCTGATTAGATTGTGCTTTATGTATGTCGAAGTCAAGGGTCAGAGAGGTC 979
ect1             TCAACCCTAATCTAATTTGGCTATTTTTCATA---ATTTGAGTCGATGAGGGGGGGGGGG 8208
                 * **   *  *** *** *  * *  ** **     *   **** * *    * * **

ect2             AAAATGTGAAATTCCATGTACTGAGTTCATTA--TCTAGATATTTCAAGAGCCAATTCAA 1037
ect1             GGGGGGCTTGTTCCATTTCATCGGGTTAATCGCGTTTGGTCGGGCTACCCGATCTTTCAA 8268
                      *     * *  *  *  * *** **    * * *       *   *    *****

ect2             GCTTACACTGTAAAAAATATTTGTATTTTTACAGCAAATCTGCTGTCAGCAGAGTCGCCA 1097
ect1             ACTT---TTTTGGAAAGTTGGTCTCTTATGATGAGGAAACTGAT-TTACCCGAATGTTTG 8324
                  ***    * *  *** *   * * ** * *     ** *** * * * * ** *

ect2             TAAAAAA-CGAATTATAAAAAAAC--AGTCTCGTTGTAAATCTAGATATTTCAAGAGCCA 1154
```

```
ect1            TGGAAAAGTTACCTTTAAAATGACGAAGAATCGG-GTAATATTATCTATCCAATATGGCT 8383
                *  ****   *  * *****  **  **  ***  ****   **  ***   *   * *

ect2            ATTGAGATGTAAAAT-TCAATCTTACACTGTAAAAAATATTTGTATTT--TTACAGCAAA 1211
ect1            GTAGGTCTATAACGTGTTTGTTTGATATCATTGCCAATTCATGTTTTAAATTGGAGAGAG 8443
                 * *   * ***  * *   * * * *   *    ***   *** **   **  **  *

ect2            TCTGCTGTCAGCAGAGTCGCCATAAAAAA----ACGAATTATAAAAA--AACAGTTT-CG 1264
ect1            AGGGAAAGGGGCAAAATCACACCGTAAACCTCTGCAAATTGTGTATATTATTATTGTATG 8503
                   *      *** * ** *     ***      * **** *  * *  *  * * *  *

ect2            TTGTAAATTTCCCCTGCAAAT---------AATTATTTGAGAAATACAAAATATAC---- 1311
ect1            TTACAGAGAATCAAACCAAAGCTAGTGGCAAGTTGTTTTAATAGAACAAAAAATATTGAG 8563
                **  * *    *    ****          * ** *** *  *  ****** ***

ect2            AT---CAGG-AGTAACACATTTGTTCCGTTTTAAACAACTGTGTAAGCTGTAGAACTCTA 1367
ect1            ATATGCAGACAGGAGAAAGTTTGAACAAAACTGGACAAACG-ATGTGCGGTGGAATGGTT 8622
                **   ***  ** *  *  ****  *     *  ****  *  *  ** ** ***   *

ect2            ATGATTTTTTTTTTTT-----------------CTTCTCTTTCA---TTAATTATTAGAG 1407
ect1            GCGGAAATTTTCTTCGGGGAGGGGGGGGGGGGCTGGAGGGTAAAGCCCACTCCTCTGAT 8682
                  *    **** **                    **      * *     * *  *  **

ect2            GAGA----AATTTACAGCAAAACTGTTATTCTAAATTGCATTTTTGTATGGCGACTC-TG 1462
ect1            GGGATTAGGGTTTATTGGAAATAAAATATTGTAAAATTAATGTGGGTGTTACTTTTTATG 8742
                * **      ****  * ***     **** **** *  ** *  ** *  *   *  **

ect2            CTGACATCACTTT--GCTGTATTATTTTTTACAGTGTAGGTGTAGTATGTATCGCCATAG 1520
ect1            CCTCCACCACCGAAGGTGATGCCATCATGTGTT-CGTACCTAC-GAATGTATCACTATAG 8800
                *   ** ***     *   *   **  * *     ***  *   * ******* * ****

ect2            GAAGATGGTCTGATTAGATTATGTTGTGCCTAGGTCAAAGGTCATGGGTCACAGAGGTTT 1580
ect1            GAAAACGATCTGATTAGACTATCGAGTATGTAGGTCAAAGATCAAGGGTGACAGAGGTCA 8860
                *** * * ********** ***   **   ********** *** **** ********

ect2            GGGATCGCAAGGTCAATGT--------CAAGGAT--AAGGATGTCTTTGTATAAAG---G 1627
ect1            AATGGGTGAAATTCCATGTATTTCAGGCGCGGATCCAGAGACGGACTTTTACAGTGAGCG 8920
                        **  ** ****        *  ****  *  ** *   ** ** *  *   *

ect2            AAAATAAATAGATGAATATGTCATATATCTCTGAT-------CTGGCATTGGTGGAGG-- 1678
ect1            AAAAGGTACCTCTATTTTTTTCACATGACGCTGGTGGAATGATAAGCGCGGGGGGGGGGG 8980
                ****   *    *   * * *** **  * *** *          **   ** ** **

ect2            ----------CATATATTTCGACCGCGCGCGGTCGAAGATTCATTTTGTTTCGTTTGTTA 1728
ect1            GGGGTGTTATCAAATTTTCGGCCTTTGGCCGTTCACAAGTTCAACAAATATTACCTATAG 9040
                          ** ** **  * *   *  ** **  *  ****     * *    * *

ect2            TATTA----CTATACGTACCAGGATCATCACTTCATTTGCTAGAAAATAGTAATGCACA- 1783
ect1            GGGTAGGGTCTAGCCGCTCGATTGGCCCGAATAAATTGTCAAGCAAAAAACAAAACAAAA 9100
                   **    ***  **  * *    *   * *  ***  * ** *** *  **  ** *

ect2            TAAAGTTT-------------------GTTGTGTGTTTTTTTATGC---------CTCCG 1815
ect1            CAAAGCGTAAACAAACAAACAAAAAAGGTTATTTAAATTTTAATGGGGGGGGGGGGCTCAC 9160
                 ****  *                   *** * *   **** ***          ***

ect2            CCACCGAAGGTG--------GTGCCGGAGGCATCATGTTTTCGGGTT-------GTCCGT 1860
ect1            TAAATAAAAACAAGGTTCATGTACAAAAGGGATGTAGTTTTCCAATCTTATGGGGTTCAC 9220
                  *   **            ** *   *** **   ******   *        ** *

ect2            CCGT-CCGTCCG---------TCCGTCCCG----TTTGCTCGTTATCGCGATATCTCAAG 1906
ect1            CCGCACCCTCCGAACCCTGGATCCGCTCCTGTATTTGGTTCGTTGTCGCTTTATCTTTAG 9280
                ***  ** ****         ****  **     ** * ***** ****  *****  **

ect2            AACCACTTCAGATATCAACTTCAAACTTATGGATGGTATTTATTACCATAGGAAGATGAT 1966
ect1            AACTAGTTGAGATGTGAGCTTTAAACTTATGAGTAGTATGTATCACCATAGGAAGTTGAC 9340
                *** * ** **** * * *** *********  * **** *** *********** ***

ect2            CTGATTAGATTTTGGAGTATGTAGGTCAAAGGTCAAGGGTCACAGAGGTCAAATTGTGAA 2026
ect1            TTGATTAGATTTTGGAGTATGTAGGTCAATGTTCAAGGGTAATGG--GTTAAATTTTGAA 9398
                ***************************** * ******** *  *  ** ***** ****
```

```
ect2            ATTCCATGTATTAAGCTCGTTATCGCGTTTTCTCAAGAACCACTTGAGATATCAACTTCA 2086
ect1            ATTCCATGTGTTGAGTTCGTTATCGCGATATCTCGAGAATCGAATGACGTATGATCTTCA 9458
                ********* ** ** *********** * **** **** *   ***  *** * *****

ect2            AACTTGGGGATGGTATGTATCACCATAGGAAGATAATCTGATTAGATTTTGGAGTATGTA 2146
ect1            AACTAAGGGGTAGCATGCATCACCATATGAAGATGGTCTGATTAGATTGTGCTTTATGTA 9518
                ****  *** * * *** ********* ******  ************ **   ******

ect2            GGTCAAAAGTCAAGGGTCACAGAGGTCAAATTGTGAAATTTCATGTATTGTGTTAGTTAT 2206
ect1            GGTCGAA-GTCAAGGGTCAGAGAGGTCAAAATTTGAAATTCCATGTACTGAGTTCATTAT 9577
                **** ** *********** ********** * ******* ****** ** ***  ****

ect2            CGCCATAT----------------------GAAGACGATCTTAT-TATATTTTGGAGTGT 2243
ect1            CGAGATATTTCGAGAGCCAATTGAGATGTGAAATTCAATCTTAAGTGTAGTATGTATCGC 9637
                **  ****                       **  * ******  * ** * ** *  *

ect2            -GTAGGTCAAAAATTTAAT----------------------AGGGACTATGATCTAATAA 2280
ect1            CATAGGAAGATGGTCTGATTAGATTATGTAGTGCCTAGGTCAAAGGTCAAGAGTCACAGA 9697
                  ****   *   * * **                      *  *   * **   *   *

ect2            GGTTTGGGGTCACAAGGTCAAAGGTCAAGGTCAAAGATGTCTTTGTATAAAGGAAAATCA 2340
ect1            GGTTTTGGATCGCAAGGTCAATG-TCAAGGTTAAGGATGTCTTTGTATAAAGGAAAATAA 9756
                ***** ** ** ********* * ******* ** ********************** *

ect2            ATAGATGATTGTGCTGTATATTTCTGATCTGGCAATGGCGGAGGCATAAATTTCGACCGT 2400
ect1            ATAGATGAATATGTCATATATTTCTGATCTGGCAATGGTGGAGGCATATATTTCGACCGC 9816
                ******** * **   ********************** ********* **********

ect2            GCGCGGTCGAGGATTCATCTTGTTTTTA-------------------------------- 2428
ect1            GCGCGGTCGAGGATTCATTTTGTTTGTAATATTACTATTCGTACCAGGATCATCACTTCA 9876
                ****************** ****** **

ect2            ----------------------TATAA--------------------------TTGTGTG 2440
ect1            TTTGCTAGAAAATAGTAATGGACATAAAGTTTGTTGTGTTTTTTTTTTAAATATTGTGTG 9936
                                      ****                           *******

ect2            AGTAAATACCAAAGCATGATTGATTGTCCTCATGATTCAAGGTAGATTTTCATTTCGTTT 2500
ect1            AGTAAATACCAAAGCATGATTGATCGTCCTCACGATTCAAGATAGATTTTCATTTCGTTT 9996
                ************************ ******* ******** ******************

ect2            TATATTTCCTATAGACTTTTCGTTATG 2527
ect1            TATATTTCATATAGACTTTTCGTTATG 10023
                ******** ******************
```

Alignment of ect1 and ect2 3’ regions:

```
CLUSTAL 2.0.10 multiple sequence alignment


ect1            ------ATAACTGGGTGATAGCTTTAGTCGGTTGTTAATTACCTGTACATTATTGAAACA 54
ect2            AGATATATAATTGGGTGATAGCTTTAGTCGGTTGTTAATTATCTGTACATTATTGAAACA 60
                      **** ****************************** ******************

ect1            ATAATCCTGAATGAAAATTATTTCCGACATCTTTGGAAATGAAATTCTCTCTTGAATTAA 114
ect2            ATAATCCTGAATGAAAATTATTTCCGACATCTTTGAAAATGAAATTCTCTCTTGAATTAT 120
                ************************************ ************************

ect1            TTGACAGATGATTAAAAAACATGAAGGGGGCTTATACAGAATACGAGAGTAGTAGCTTAC 174
ect2            TTGACAAATGATTAAAAAACATAAATGGGGCTAATACAGAATAAGAGAGTGGTAGCTTAC 180
                ****** *************** ** ****** ********** ****** *********

ect1            AGTGTCAAAAAAAAAACTAAAAAAACAAAAACAGTGAAAATAAACCGTATACAAATTATT 234
ect2            AGTGTCAACACAAAAACCAGTGAAA-ATACACCGTATTA------CGTATACAAATTATT 233
                ******** * ****** *   *** * * ** **   *      ***************

ect1            CAAGAATGATAAAACCGTCTAACATTATTCCATCATAATAATGATTCTGGCATTGGCGTA 294
ect2            CAAGAATGATAAAATCGTCTAACATTATTCCATCATAATAATGATTCTGGCATTGGCGTA 293
```

```
                 ************** ********************************************

ect1             TACAATCGACTAATATCTAATATCTCAACTTCCTATAGATGTAGATTTGAAATAAATTTT 354
ect2             TACAATCGACTAATATCTAATATCTCAACTTCCTATAGATTCC---------TAAAAATT 344
                 ****************************************            ****  **

ect1             GTGATAAAATATTTAAATACTTACATGATTACAATAGTTCACATGCTACCAGACATTACA 414
ect2             GTGATAAAATATTTAAAAATTTACATGATTACAATAGTTAACATGCTATCAGACATTACA 404
                 ***************** * ******************* ******** ***********

ect1             TAGTAATAAGTATTTTCTACCACTTTCCGTGATTGTTAGTTCTGTCAATAATATTTTATC 474
ect2             GAGTAATAAGTATTTTCTACA---TTTCGTGATTGTTACTTCTGTCAATAATATTTTATC 461
                  *******************    ** *********** *********************

ect1             CTCAGTACATTATGAGGTTTATTTGTGTATTATTACTATAATTTTTCTATGTCTGCAACA 534
ect2             CTCAGTACATTATGAGGTTTATTTGTGTATTATTACTATTATGTTTCTATATCTGCAACG 521
                 ************************************** ** ******* ********

ect1             AATTGGGAGAAATGTTTCCATTTAATTTAAAACAAA-----AGTAAGATGGTTCAATATG 589
ect2             AATTGGGAGAAATGTTTTTATTTCATTTACAACAAGCCAAGAGTATTATGGTTCAATATG 581
                 *****************  **** ***** *****      ****  *************

ect1             TAACAATACCATTAATGATATTAATTGATTAATAGTATCAAACCCGTTTGCGAGTAATAG 649
ect2             TAGCAATACCATTAATGATATT----------CAGTATCAAACTCGTTCGCGAGTAATAG 631
                 ** *******************           ********** **** ***********

ect1             CTAGACGATAAATATTAGATATCTTAACTTCCTAGATGCAGTCTTGAAGTTAACTTTATA 709
ect2             CTATACGATAAATATTAGATATCTTAACTTTGTTGATGCAGTTTTAAAGTTAATTATATA 691
                 *** ***************************  * ******** ** ******* * ****

ect1             AATTAAGTACATGTATATCGTCTTCCTATTGTATGACCACGC------------GAGTTG 757
ect2             ACTGAAGT------ATATCGTCTTCCTATTGTATGACCATGATGACCACGCGGGGAGTTG 745
                 * * ****      ************************* *             ******

ect1             TAAGTAATTTCTAACACTTTTACTGATTGTTACTTTTTCTTCAAATGGGCAGCATTACAA 817
ect2             TAAGTAATTTCTAACACTTTTACTGATTGCTACTTCTTCTTCAAATGGGCAACATTACAA 805
                 ****************************** ***** *************** ********

ect1             AAAAAATTTTGAGTGGGATATTCTCTCTCTTCACATATGACTTACATATCTATACCTTGC 877
ect2             TTT---TTTTCAGTGGGATATTCTCACTCTTCACATATAACTTACATATCTATACCTTGC 862
                       **** ************** ************ *********************

ect1             ATTTATAAAAATCACATTTCATCTTACTGAATAAAAATACGTTTCTATTTCAATTTTCAT 937
ect2             ATTTATAAAAATCACATTTCATCTTACTGAATAAAAACACGTTTCTATTGTAATTTTCAT 922
                 ************************************** ***********  *********

ect1             TATATGATTCAAATTATTTATATGGGTAATAATTTTATTAGGGAAGAGTAGAACCCAGTG 997
ect2             TATATGATTCA------TTATATGGTTAATAATTTTATTAGAGAAAAGTAGAACCCAGTA 976
                 ***********      ******** *************** *** *************

ect1             TACATGTATAGGTTTATCTCAATTTGGTCTAACATCAGTTGGCCAAATTCACTGTTTTTT 1057
ect2             TACTAG----GTTTTATATCAATTTGGTCTAACATCAGTTGGCCAAATTCACTGTTTTTT 1032
                 ***  *    * ***** ******************************************

ect1             CTAATACCACATAGGCTACACCCATTTGGTCTTCTAAATTTTTTGTCAAATTTCCATTCG 1117
ect2             CTA-TACCACATAGGCTAAACCCATTTGGTCCTCTAAAATTTTGGTCAAATTTCCATTCG 1091
                 *** ************** ************ ****** **** ****************

ect1             GTTTATAATTAGCATAATTCTCACTTGGTATAATGGTCAACTCGTCTAATAGCCAAATTG 1177
ect2             GTTTATAATTAGCATAATTCTCACTTGGTATAATGGTCAACTCGTCTAATAGCCAAATTG 1151
                 ************************************************************

ect1             TCTAAATACATTTTCGTCTATTAATTCCCCAAAAGTCTATTTGGCTTATTACCCGTCTGT 1237
ect2             TCTAACTACATTTTCGTCTATTAATTCCCCAAAAGACTATTTGGCTTATTACCCGTCTGT 1211
                 ***** ****************************** ************************

ect1             TACAGAAATTGGCCAAGTGATGTTAGACCAAGTGGACATTAGACGAAATGGGGATTGGAC 1297
ect2             TACAGAAATTGGCCAAGTCATGTTAGACCAAGTGGACATTAGACGAAATGTGGATTGGAC 1271
                 ****************** ******************************** *********

ect1             AAAGTAGTAAATAGGCAAAGTGTCAATTAGACCAAAATGTAATAAGACAAACTGGTTGTA 1357
```

```
ect2            CAAATAGTAAATAGGCAAAGTGTCAATTAGACCAAAATGTAACAAGACAAACTGGTAGTA 1331
                 ** ************************************** ************** ***

ect1            GACGAAATGGAATTGGACAATGGAAGGTAAACATGGTAAAAGCGGGTTTAGGCTACTATT 1417
ect2            GACGAAATAGAACCATGTGA--------------GGTGAAAGCGGGTTTAGGCTACTATT 1377
                ******** ***       *              *** **********************

ect1            AAGTACCAATAAATTGAACAACAATAAACACCATAGCCAATAAGGTCAATTTCATGCTAA 1477
ect2            AAGTACCAATAAATTGAATAAC---AAATACCATAGCTAACAAGGTCAATTTCATGCTAA 1434
                ****************** ***   *** ******** ** *******************

ect1            TTCACATTTTACAAAAGGAAACAAAAAGCTAAAAAGCAAACAAACAAACA----AATGTA 1533
ect2            TTCACATTT-ACAAAAAGAAACAAAAACAAAAAAAGCAAACAAATAAACATAATAATGTA 1493
                ********* ****** **********   ************** *****    ******

ect1            ACGTCGTTTACCATGATTACCGACCAAATTTTGCTTGCCGTTTCTAAGGAAATGGAGTCT 1593
ect2            ACGTCGTTTACCATGATTGCCGACCAAATTTTGCTTGCCGTTTCTAAGGAAATTGAGTCT 1553
                ****************** ********************************** ******

ect1            GCATGAGTGGAATTTTTTTGCACATTTCTTTCCCTATCACTAACATATTTAGCTGTATTA 1653
ect2            GCATGAGTGGAATTTTTC-GCACACTTCTTTCCCTATCACTAACATATTTAGCTGTATTA 1612
                *****************  ***** ************************************

ect1            ATTTTAAATAATTCTAAGTCTTATGTGTATTATTGTGTATTCCAATACAAACAACCTAGT 1713
ect2            ATTTTAAATAATTCTAATTCTTATGTGTATTTTTGTATATTGTA-TACTAACAACCTAGT 1671
                ***************** ************* **** ****  * *** ***********

ect1            AATTTATGTAAATTTGTAACGTGCAGTTACCGTGGAATTGC-CATAAGTGTTAGGTATAC 1772
ect2            AATTTATGTAAATTTGTAATGTGCAGTTACCGTGGAATTGCACATAAGTGTTAGGTATAC 1731
                ******************* ******************** ******************

ect1            AATATTGATCAGCAATATCAATAAG------CATTTGCTTATAAATCATCAAATTTTGGT 1826
ect2            AATATTGATCAGCAATATCAACAGGTTTATGCATTTGCTTATAAATCATCAA---TTGGT 1788
                ********************* * *      *********************   *****

ect1            AGGTATGAATGATAACACATTGGACATATTATACTATTGTGTATTAACCAATTTGCCTAG 1886
ect2            AGGTATGAATGATAACACATTTGACATATTATACTATTGTGTATTAACCAATTTGTCTAG 1848
                ********************* ********************************* ****

ect1            TTGAGATGCAGGGCGATTGCGACATTATCAAAGA----ATCGGGTGTATTAAGAAATGCC 1942
ect2            TTGGGATGCAAGGCGATTGCGACATTAACAAAGACTCAATCGGGTATATTAAGAAATACC 1908
                *** ****** **************** ******    ******* *********** **

ect1            CAGAAAACCAGTGTCGATACGTACTTATATTAAAATGATGTACAGTTGAGTGGTTTCACT 2002
ect2            CAGAAAACCAGTGTCGATACGTACTTACCTTGCA-TGATCTACAGTTGAATGGTTTCACT 1967
                ***************************  **  * **** ********* **********

ect1            TAGTGTTACAGTATACTTTATGTGAATATCACAATTGGAATTTTGAAATATTTGTATTGA 2062
ect2            TAGT------GTATACTGTATGTGAATATCACAATTTGAATTTTGAAATATTTTTATTGA 2021
                ****      ******* ****************** **************** ******

ect1            TGTTATACAACGCCATTACGTATATTCGGTGTGACATGGTTGAGGAAAATTAAGATATAG 2122
ect2            TGTAATACAACGCCATTACGTATATTCGGTGC-ATATGGTTGAGGAAAATTAAGATATAG 2080
                *** ***************************  * *************************

ect1            AAGACAAATATTTACCTTCTTTTACCTTGTTGAACTTTATTGGCACATTGTATATGAATT 2182
ect2            AAGACAAATATTTACCTTCTTTTACCTTGTTGAACTTTATTGGCACACTGTATATGAATT 2140
                ********************************************** ************

ect1            AAACATCACAATACTATTTACAGACCTTTTACGCAATCAAATCAATTAAGATCAAAGATA 2242
ect2            AAACATCACAATGT--------------------------------------AAAAATA 2161
                ************                                       *** ***

ect1            TAGGTGCACCTTGCCAAATTTCCTCGGATCATTGCGCTTACGAAAAGATTTACAACTGAA 2302
ect2            CAGGTGCACCTTGCCAAATTTCCCCGGATCACTGCGCTTACGAAAAGATTTTCAACTGAA 2221
                ********************** ******* ****************** ********

ect1            ATATTGTAGTTATTCTGATTAAAAATATATCCATGACATTTAGGTAATTT-TGTACACTT 2361
ect2            ATATTGTAGTTATTCTGATTCAAAATATATCCATGACCTTTTGGTAATTTGTGTACACTT 2281
                ******************** **************** *** ******** *********
```

```
ect1            GTTAATAGTAACTACCACAGACTGTTGTATATTCATAATGACGCTTATTGTTTGCTTGTA 2421
ect2            ATTAATAGCAACCACCACAGACTGTTGTATATTCATAATGACGCTTATTGTTTGCTTGAA 2341
                 ******* *** ********************************************** *

ect1            TATTCTTATTTATATGTGTTTTTAATCGGTGTACTAAGGAGTGGCACAATTAACAATATT 2481
ect2            TATTCTTATTTATATGTGTTTTTAATCGGTGTATTAAGGAGTGGTACAATTAACAATATT 2401
                ********************************** ********** ***************

ect1            CATAATGTAAACAGTTAACCTTATATATTTCATAATCGGTGTAGAAGGTTTGGTAAATTA 2541
ect2            CATAATGTAAACAGTTAACCTTATATATTTCATAATCGGTGTAGAAGGTTGGGTAAATTA 2461
                ************************************************* *********

ect1            CAACTGCAATATTGATG---TAATTCTGGTTTTGGAATGTACCATCAGCTTTTTGCATTT 2598
ect2            CAACTGCAATATTGATGATGTAATTCTGGTTTTACACTCT-CAGAAAAATTTTAACCAAC 2520
                *****************   *************  * * * *    *  ****  *

ect1            TTGTTTTTGTTTTTACTCTTTATACTATTCCCATAAACATTTAACATGTGTTTGATAATT 2658
ect2            CGGGCAAAATTTTAACCAACCGT-TGGTTCAAATATGAACTTACCAACTGTTGGTTAAAC 2579
                  *      **** **      *    ***  ***   * *** **  **** * ***

ect1            GTACAAAGT--TCGCAAT-----GAAATTTATCCTCCGTTTTGTTTT--CTTTTCTTCGA 2709
ect2            GCGTTAACCAACCCCAGTTGGTTAAAATTTAACCAATTATGTATTAAAACTTTAATAAAA 2639
                *    **     * ** *      ******* **     * * **    ****  *   *

ect1            ATGAGATTTAAA------AATCAGATTTGTCAAAGAAAATGC----ATGTGTAGTCGTTA 2759
ect2            AGTTGGTTAAAATATTTTAACCAACCCTAGTTGGTAAAATTTTTTAATACTTAATTGGTT 2699
                *   * ** ***      ** **    *       *****      **   ** * * *

ect1            AAATACCTTTGGGGCCAT----TGGTTTAAATTCAACCAAGGCTGATGATTAAAAAGTGA 2815
ect2            AATTTCTTTTTTTAATACATCATGGTTGGTTAAAAAAAAAATCAAGCGATAGACCGGCCT 2759
                ** * * ***      *     *****       **  **  *    ***  *   *

ect1            AGTCTCTCCAATTCAACGTTTAATTAAAATGTTTTTGAAGAAAATATTCTAAAAATGTTC 2875
ect2            GGACTGGCCTTTTGACAATTTGATCAAAGTATTGGTACAAATAACATGA-AGAGTTATGC 2818
                 * **  **  ** *   *** ** *** * **  *  * * ** **   * *  * * *

ect1            CTAAATCACTTTACTGAAAAAAAAAAT----AACATAGACATTA-AGACCTACTTTGTGT 2930
ect2            CTGTATTATTAGACTACATGCATTATTTGTTGGCCTAAATAGGGTAGGCCTACAGTAGCC 2878
                **  ** * *  ***  *   *  * *      * ** * *    ** *****  *

ect1            AATGGGATGACGTAATTCTGAATTGTATGAATGGTTTATATGGTTTATCTTTAAAAAAA- 2989
ect2            TAAGTATTGGTCCAGTACAATATCATTACTTCGGTTCAGACGTGAGGTCTACAATCTAGT 2938
                 * *   **    * * *   **  *      **** * * *     ***  **   *

ect1            --AGTAAC--AGGTGCACTATATTAATTTGATAAAGCTAT--TCTTAAATTGAAATGTTA 3043
ect2            CCAGTAGCTGAGATCTACTGGATCTAATT-ATAGATCTAGAATATAAAAATGAGATGTTG 2997
                  **** *  ** *  ***  **  * ** *** * ***   * * *** *** *****

ect1            AA-TAAAT----------AATTTCATGGTCTT---TTGAAAATCATT-------GTTTTC 3082
ect2            GCCTAGATTCTACGCTTAGATTCCAAGGTCATCACTCTGAAACCAGTCAAGTAGGCTTAG 3057
                   ** **           *** ** **** *   *   *** ** *       * **

ect1            TCTCAAAATGCATTATTCATGTTATTGAATATGAAGT------GAATATATTT-GACACA 3135
ect2            ATCTATAGGACTATACCCACATCGAGAGGTCTAAAAATATTTAGGCTAAATTTAGATGCA 3117
                    * *   *  **  **  *       * * **        *  ** **** **  **

ect1            AA-AATGCGTTTGCCTTATTCTTATAAATAAA---AAAGGAGCATATAATGTAACAAGCT 3191
ect2            AGTGATGGTTCTAACGTAGTTCTAAATGTTTAGCCAACTAAATCTATGACCTCA-AGACT 3176
                *   ***  * *  * ** *  ** *  *  *   **   *   *** *  * * *  **

ect1            TAGCTTTTAACTTAGGCTCCTTCATATTTCTTTCTTATACTACTGATGTCATTCTTCTTA 3251
ect2            TAGTCCAAGACCAAGCTGGTTGGAGAGCGAAGTATTTAGATTCTAACGTTAGACCGGAGC 3236
                ***      **  **     *  * *      * **    * ** * ** *  *

ect1            TAACTTTGTCAACGTATTGTTTTGATATAACATATCATTTTTTTTTTAAATGTTAATATT 3311
ect2            TGCATAAATTCAAGCAAGTTTAATACCTAACGTTAGGTCTAACGTCTCAACCGAAATAAT 3296
                *   *   *  * * *   **   *  **** *    * *    * * **    **** *

ect1            GTTTGATTGGAAATATGGAATAAACTAAACTGAAACTGAAAT-ATTTTGACATTCTGAGA 3370
ect2            TAAGAACCGTACTGACCATGTAGTCCAAAACATAATGCATGTTAATCCTATATCACAGGC 3356
                     *  * *   *     **  * ***    **   *  * * *   * **     *
```

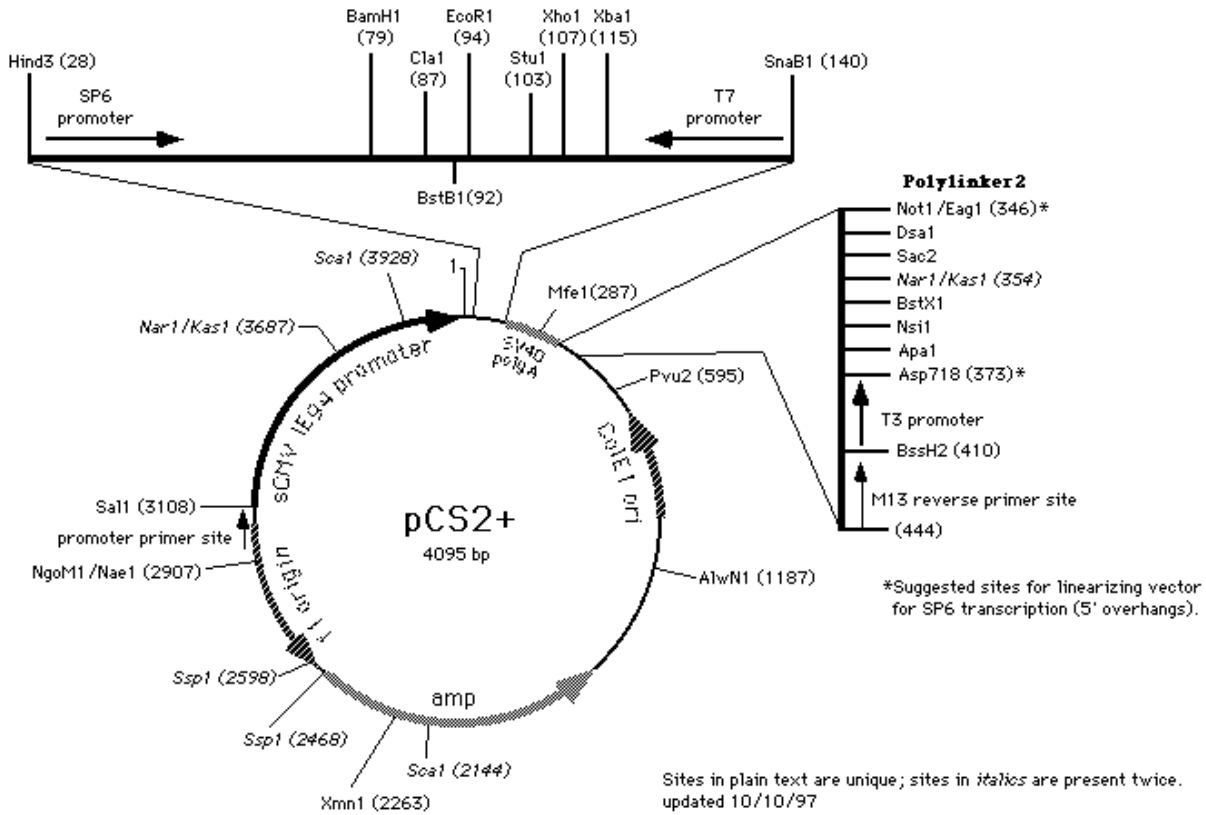
```
ect1            A-ATACTCTTAAATAAAATGTTGAATATAAACC--CTTTTATGGTTATGGGAAAATGATT 3427
ect2            CTATACTCTTCAAGTCATTTCTACCTACCAATATACTTTGATCAGATTGTCAAGATAGCT 3416
                  ******** **   * *  *   **  **    **** **     **  ** **   *

ect1            GTTTTTCTCAACAACAAAAAATAAC--GTTAATTTCCATTAACTTAAAAAAAAAATTACA 3485
ect2            AGGTCCAGCCAGACCAGCAGTCAGCTTGTGTAACCCTGTTGTCGTCGTGAGCAAAGTACG 3476
                   *    * * * **  *   * *  **  *   *  **  * *    *  *** ***

ect1            TCCG--------GGTTTTAAATACTTT-----------TTAGTACTTCATCTTTCTGACT 3526
ect2            TGCCTAATGAATAGGCGCATATATCTTACGTGTATGAATCGAAATCGCAAGATTTTTAGC 3536
                * *          *    * ***  **           *    *   **   ** * *

ect1            ACTAGTATATAAAATCTTATCTTCGTGA---TCATGTTTGA-ATTGAGAAAGCTGGAAAG 3582
ect2            GCGAGGAGCGGAGTGCGTATCTTTGTTAAATCCCTGTTTCCCATTGGTTGTTTTTCTTCC 3596
                 * ** *    *   * ****** ** *    * *****   ****       *

ect1            ATGTAAACATGGTCGAGAGGGAATT--TGGGTGAAGCAGGAA--GAGCTAGAGAGAGAGA 3638
ect2            ACGTGATCAATTTTTTTAACCAACTATTGGTTAAAATTGGAATTGGGCTTTTTTAACCAA 3656
                * ** * **   *    *   ** *  *** * **   ****  * ***      *   *

ect1            GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA------GAGAGAGA 3692
ect2            CCCATGCTTTTTCACGTTTAACCAACTTGTTTTTTAACCAACTTTTCTTTTTAATACATT 3716
                   *         *     *   *    *      *   *             * * *

ect1            CTTGGGAGAGAAAGAGAGTGAGAGGAAGAGAAAAGG--TTAAATTTAAGC------CTGA 3744
ect2            ATTGGTTAATATTTTTAACCAACTGGTTTTAATATGCATTTTTTTTAACCAGCGTTCTGA 3776
                 ****   * *     *   *   *     ** * *  **   ***** *      ****

ect1            GAGGG--AGATAAATTAAAAGGGCCCAT--ATTACCGAGTATTATTTGTAGAAAGAAAGC 3800
ect2            GAGTGTAGAATGTACCCATCAGCTTTTTTCATTTTTGA-TTTTGTTTTTACTCTTTATAC 3835
                *** *    **  *   *   *     *  ***   ** * ** *** **      *  *

ect1            ATACATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGGA 3860
ect2            TTTTCCCATAAACATTTAACATGTGTTTACATAGTTCGCAATGAAATTTATCCTCAATTT 3895
                 *    *   *       * *       * * **   *  *   *           *

ect1            AGAGATAT--------AGAGAGA---AAGAGAGAGA---GAGAGAGAGAGGGAAAACTAA 3906
ect2            TCTTTTCTTTTCTTCGAATGAGATTTAAAAATCAGATTTGTCAAAGAAAATGCATGTGTT 3955
                     * *        *  ****   ** *   ***   *  * *** *  * *

ect1            ATTTATAGAAGAGGGAAAATAAGAAAGAGTTGAAAAAATGAGAATCTATTCGTTTCTAGT 3966
ect2            GTCGTTAAAATACCTTTGGGGACATTGGTTTAAAGTCAACCAAAGCTGATGATTAAAAAA 4015
                 *   ** ** *         * *  *  ** **   *    ** **  *  **   *

ect1            TATTTTTTCTTCTTTTCTCTATTATACTGACATGACATTGAGTTAAAAATGG-GAAGATG 4025
ect2            TAAAGTCTCTCCAATTCAACGTTTAATTAAAATGTTTTTGAAGAAAATATTCTAAAAATG 4075
                **   * *** *  ***    **  * * * ***   ****   *** **    ** ***

ect1            TATGTTAACATATATAGGCGATCACCTCTCTTTTGTGTAACTATTCTTTGCTCTTTTTCT 4085
ect2            TTCCTAAATCACTTTACTGAAAAAAAAAATAATAACATAGACATTAAGACCTACTTTGTG 4135
                *   * **    * **    *  *        *    **   ***     **  ***

ect1            CCTT----CTATCTCTTTTTATCTCCTCTCTCTCTCTCTCTCTCTCTCTTGCTCTCAACC 4141
ect2            TAATGGGATGACGTAATTCTGAATTGTATGAATGGTTTATATGGTTTATCTTTAAAAAAA 4195
                   *      *  *  ** *   *  * *   *   *       * * *   *   **

ect1            TTTTTTCTTTTTTCTAATTCTGGTTTTTCTTTTTTTTTTTAGAGAGAGAGAGAGAATCAG 4201
ect2            AATAACAGGTGCACTATATTAATTTGATAAAGCTATTCTTAAATTGAAATGTTGAATAAA 4255
                  *      *   ***  *    **  *     * ** *** *  ** *    **** *

ect1            TAGTCTCGTTTTCAATTTTAAAGTA--ACCCACATCCAAAGCTATACATTCATTAAAAT- 4258
ect2            TAATTTCATGGTCTTTTGAAAATCATTGTTTTCTCTCAAAATGCATTATTCATGTTATTG 4315
                ** * ** *  **  **  ***  *       *   ****       ******   * *

ect1            --CATTAATCAGATATCTCCACAACAGTATTGTAATCTTATTTTCTCAAATGATTATTGC 4316
ect2            AATATGAAGTGAATATATTTGACACAAAAATGCG-TTTGCCTTAATATTCTCATAAATAA 4374
                   ** **    **** *     ***  * **   * *   **  *    * ** * *

ect1            ATAAACAACATATATTAACACAACACAACACAACAAAATATACTA---TTTATATAAAAT 4373
ect2            AAAAGGAGCATATAATGTAACAACCTAAGGGGGCCGT-CACACTGAGGTCGTTCTATAAC 4433
```

```
                 * **  * ****** *   *****  **     *     * ***    *   * ** **

ect1             GTCATTTCATAATACATGTTTATTCAATTGCAAAATATTAACATTTCACTGATTCCGACA 4433
ect2             GTTCTTGCCCG-TTTTTGCCCGTTTTAAAAAAAAATGCTGAACGGGCTTTGGTCGGGTTG 4492
                 **  ** *    *   **    **  *    *****  * *     *  ** *   *

ect1             AAAC-----ATTCCAAAAAGTATCA--------ACAAAGTCAATGAGTTCACCAAACAAT 4480
ect2             GAACTTATAACTTTGAAAAATAATTTCTTGAATACGATGCTAGGCAGTTCAT-AAATAGT 4551
                  ***     * *   **** **           ** * *  *   ******  *** * *

ect1             GTAACAGTACATTATTTATAT-TCAGTAGAGTTTCAGAGTCS------------------ 4521
ect2             -TCTCGTTACGTCAATTCCGTCTTTATCGCGTCCGCAATACGTTGTTATCCAGTTCAGAA 4610
                  *  *  *** * * **   * *   * * **     *  *
```

9.3. CS2+ & CS2+MT Plasmids

Documentation available on the website of the University of Michigan: CS2 Vector Resource Homepage (http://sitemaker.umich.edu/dlturner.vectors/home)

9.3.1. CS2+



CS2+ and CS2- polylinker I:

```
                 (Aat2)                                                                                  BstB1          Xho1
                    |      Hind3                                                   BamH1    Cla1   | EcoR1    Stu1 |       Xba1                      SnaB1
                    |        |                                                      |        |     | |         |   |        |                              |
1                  20        |          40                   60                    80       |     | |   100   |   |        |  120                    140         150
*        *          *        | *         *          *          *          *         |*       |  *  | |     *   |   |  *     |    *          *          *           *
CGCCATTCTG CCTGGGGACG TCGGAGCAAG CTTGATTTAG GTGACACTAT AGAATACAAG CTACTTGTTC TTTTTGCAGG ATCCCATCGA TTCGAATTCA AGGCCTCTCG AGCCTCTAGA ACTATAGTGA GTCGTATTAC GTAGATCCAGA
GCGGTAAGAC GGACCCCTGC AGCCTCGTTC GAACTAAATC CACTGTGATA TCTTATGTTC GATGAACAAG AAAAACGTCC TAGGGTAGCT AAGCTTAAGT TCCGGAGAGC TCGGAGATCT TGATATCACT CAGCATAATG CATCTAGGTCT
                                  |-SP6 Promoter---->|---------------------------| G   S  H  R   F  E  F   K  A  S  R   A  S  R   T  I  V   S  R  I  T   *
                                                   5' leader sequence          D   P  I  D   S  N  S   R  P  L   E  P  L  E   L  *  *
                        5' CCCAAG CTTGATTTAG GTGAC-> 3'          (27 nt.)          I  P  S  I   R  I  Q   G  L  S   S  L  *    <--T7 Promoter----|   |------->
                              SP6 primer                                                                                                                  SV40 polyA
                                                        |---------|                                                                  3' <-GATATCACT CAGCATAA 5'
                                                       CS2 insertion                                                                         T7 primer
```

Notes: Base number 1 is the first nucleotide of the sCMV IE94 mRNA (this is true in all CS2 derivatives as well). Restriction sites shown are unique, except for Aat2. The SP6/T7 polylinker cassette can be removed as a Hind3 to SnaB1 fragment. The polylinker contains no ATG start codons, and no stop codons before the Xba1 site. Stop codons are present in all three frames at or after the Xba1. The translation of the polylinker is shown in all three reading frames after the first cloning site (BamH1). All sites have 5' overhangs except Stu1 and SnaB1 (both blunt), and Aat2 (3').

CS2 versus CS vectors: The CS2 series of vectors contain an insertion of 10 nucleotides between the SP6 promoter and the polylinker of the original CS series vectors. This insertion was introduced to improve the efficiency of SP6 RNA synthesis (CS2+/- generates about 5 times as much SP6 RNA as CS+/-). The inserted nucleotides do not introduce or alter any restriction sites, nor do they appear to alter the function of the CMV IE94 promoter. When used as DNA expression vectors in frog embryos or mammlian cells, CS and CS2 appear to function identically. All CS series vectors are obsolete, as all of the basic vectors are available as CS2 versions. Note that there are several CS- derivatives, but there are no CS2- derivatives.

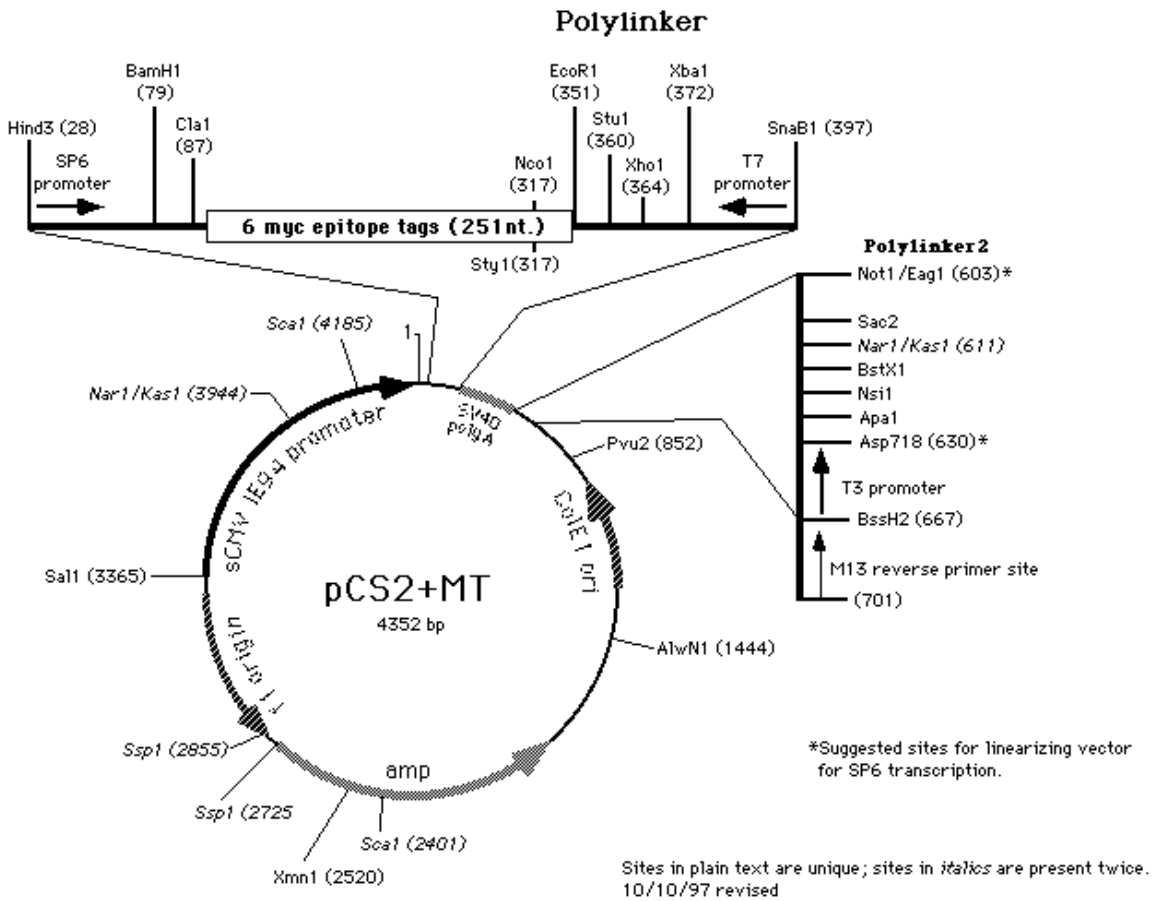
```
LOCUS       PCS2+        4095 BP SS-DNA   CIRCULAR  SYN       15-JAN-2001
DEFINITION  -
ACCESSION   -
KEYWORDS    -
SOURCE      -
FEATURES             Location/Qualifiers
     mRNA            1..1
                     /note="base1 = cap site for sCMV IE94 promoter"
     frag            35..51
                     /note="SP6 promoter"
```

```
     frag            54..>63
                     /note="CS2 insertion"
     frag            <64..78
                     /note="B-globin 5' mRNA sequence"
     frag            79..120
                     /note="polylinker"
     frag            123..139
                     /note="T7 promoter (reverse orientation)"
     frag            144..339
                     /note="SV40 late polyadenylation site (Bam to Dra1)"
     frag            340..372
                     /note="donwstream polylinker (from pVZ1)"
     frag            373..3108
                     /note="Bluescript II KS+ from Kpn1(759) to old Pvu2(52 ;
                     destroyed)"
     frag            393..409
                     /note="T3 promoter (reverse orientation)"
     frag            426..441
                     /note="M13 reverse primer binding site"
     frag            2580..3040
                     /note="f1 ori"
     frag            3109..>4095
                     /note="simian CMV IE94 enhancer/promoter "
     frag            3900..4095
                     /note="based on DT sequences"
     frag            3109..3899
                     /note="based on database sequences"
BASE COUNT     1046 A    993 C   1006 G   1050 T      0 OTHER
ORIGIN      -
        1 CGCCATTCTG CCTGGGGACG TCGGAGCAAG CTTGATTTAG GTGACACTAT AGAATACAAG
       61 CTACTTGTTC TTTTTGCAGG ATCCCATCGA TTCGAATTCA AGGCCTCTCG AGCCTCTAGA
      121 ACTATAGTGA GTCGTATTAC GTAGATCCAG ACATGATAAG ATACATTGAT GAGTTTGGAC
      181 AAACCACAAC TAGAATGCAG TGAAAAAAAT GCTTTATTTG TGAAATTTGT GATGCTATTG
      241 CTTTATTTGT AACCATTATA AGCTGCAATA AACAAGTTAA CAACAACAAT TGCATTCATT
      301 TTATGTTTCA GGTTCAGGGG GAGGTGTGGG AGGTTTTTTA ATTCGCGGCC GCGGCGCCAA
      361 TGCATTGGGC CCGGTACCCA GCTTTTGTTC CCTTTAGTGA GGGTTAATTG CGCGCTTGGC
      421 GTAATCATGG TCATAGCTGT TTCCTGTGTG AAATTGTTAT CCGCTCACAA TTCCACACAA
      481 CATACGAGCC GGAAGCATAA AGTGTAAAGC CTGGGGTGCC TAATGAGTGA GCTAACTCAC
      541 ATTAATTGCG TTGCGCTCAC TGCCCGCTTT CCAGTCGGGA AACCTGTCGT GCCAGCTGCA
      601 TTAATGAATC GGCCAACGCG CGGGGAGAGG CGGTTTGCGT ATTGGGCGCT CTTCCGCTTC
      661 CTCGCTCACT GACTCGCTGC GCTCGGTCGT TCGGCTGCGG CGAGCGGTAT CAGCTCACTC
      721 AAAGGCGGTA ATACGGTTAT CCACAGAATC AGGGGATAAC GCAGGAAAGA ACATGTGAGC
      781 AAAAGGCCAG CAAAAGGCCA GGAACCGTAA AAAGGCCGCG TTGCTGGCGT TTTTCCATAG
      841 GCTCCGCCCC CCTGACGAGC ATCACAAAAA TCGACGCTCA AGTCAGAGGT GGCGAAACCC
      901 GACAGGACTA TAAAGATACC AGGCGTTTCC CCCTGGAAGC TCCCTCGTGC GCTCTCCTGT
      961 TCCGACCCTG CCGCTTACCG GATACCTGTC CGCCTTTCTC CCTTCGGGAA GCGTGGCGCT
     1021 TTCTCATAGC TCACGCTGTA GGTATCTCAG TTCGGTGTAG GTCGTTCGCT CCAAGCTGGG
     1081 CTGTGTGCAC GAACCCCCCG TTCAGCCCGA CCGCTGCGCC TTATCCGGTA ACTATCGTCT
     1141 TGAGTCCAAC CCGGTAAGAC ACGACTTATC GCCACTGGCA GCAGCCACTG GTAACAGGAT
     1201 TAGCAGAGCG AGGTATGTAG GCGGTGCTAC AGAGTTCTTG AAGTGGTGGC CTAACTACGG
     1261 CTACACTAGA AGGACAGTAT TTGGTATCTG CGCTCTGCTG AAGCCAGTTA CCTTCGGAAA
```

```
1321 AAGAGTTGGT AGCTCTTGAT CCGGCAAACA AACCACCGCT GGTAGCGGTG GTTTTTTTGT
1381 TTGCAAGCAG CAGATTACGC GCAGAAAAAA AGGATCTCAA GAAGATCCTT TGATCTTTTC
1441 TACGGGGTCT GACGCTCAGT GGAACGAAAA CTCACGTTAA GGGATTTTGG TCATGAGATT
1501 ATCAAAAAGG ATCTTCACCT AGATCCTTTT AAATTAAAAA TGAAGTTTTA AATCAATCTA
1561 AAGTATATAT GAGTAAACTT GGTCTGACAG TTACCAATGC TTAATCAGTG AGGCACCTAT
1621 CTCAGCGATC TGTCTATTTC GTTCATCCAT AGTTGCCTGA CTCCCCGTCG TGTAGATAAC
1681 TACGATACGG GAGGGCTTAC CATCTGGCCC CAGTGCTGCA ATGATACCGC GAGACCCACG
1741 CTCACCGGCT CCAGATTTAT CAGCAATAAA CCAGCCAGCC GGAAGGGCCG AGCGCAGAAG
1801 TGGTCCTGCA ACTTTATCCG CCTCCATCCA GTCTATTAAT TGTTGCCGGG AAGCTAGAGT
1861 AAGTAGTTCG CCAGTTAATA GTTTGCGCAA CGTTGTTGCC ATTGCTACAG GCATCGTGGT
1921 GTCACGCTCG TCGTTTGGTA TGGCTTCATT CAGCTCCGGT TCCCAACGAT CAAGGCGAGT
1981 TACATGATCC CCCATGTTGT GCAAAAAAGC GGTTAGCTCC TTCGGTCCTC CGATCGTTGT
2041 CAGAAGTAAG TTGGCCGCAG TGTTATCACT CATGGTTATG GCAGCACTGC ATAATTCTCT
2101 TACTGTCATG CCATCCGTAA GATGCTTTTC TGTGACTGGT GAGTACTCAA CCAAGTCATT
2161 CTGAGAATAG TGTATGCGGC GACCGAGTTG CTCTTGCCCG GCGTCAATAC GGGATAATAC
2221 CGCGCCACAT AGCAGAACTT TAAAAGTGCT CATCATTGGA AAACGTTCTT CGGGGCGAAA
2281 ACTCTCAAGG ATCTTACCGC TGTTGAGATC CAGTTCGATG TAACCCACTC GTGCACCCAA
2341 CTGATCTTCA GCATCTTTTA CTTTCACCAG CGTTTCTGGG TGAGCAAAAA CAGGAAGGCA
2401 AAATGCCGCA AAAAAGGGAA TAAGGGCGAC ACGGAAATGT TGAATACTCA TACTCTTCCT
2461 TTTTCAATAT TATTGAAGCA TTTATCAGGG TTATTGTCTC ATGAGCGGAT ACATATTTGA
2521 ATGTATTTAG AAAAATAAAC AAATAGGGGT TCCGCGCACA TTTCCCCGAA AAGTGCCACC
2581 TAAATTGTAA GCGTTAATAT TTTGTTAAAA TTCGCGTTAA ATTTTTGTTA AATCAGCTCA
2641 TTTTTTAACC AATAGGCCGA AATCGGCAAA ATCCCTTATA AATCAAAAGA ATAGACCGAG
2701 ATAGGGTTGA GTGTTGTTCC AGTTTGGAAC AAGAGTCCAC TATTAAAGAA CGTGGACTCC
2761 AACGTCAAAG GGCGAAAAAC CGTCTATCAG GGCGATGGCC CACTACGTGA ACCATCACCC
2821 TAATCAAGTT TTTTGGGGTC GAGGTGCCGT AAAGCACTAA ATCGGAACCC TAAAGGGAGC
2881 CCCCGATTTA GAGCTTGACG GGGAAAGCCG GCGAACGTGG CGAGAAAGGA AGGGAAGAAA
2941 GCGAAAGGAG CGGGCGCTAG GGCGCTGGCA AGTGTAGCGG TCACGCTGCG CGTAACCACC
3001 ACACCCGCCG CGCTTAATGC GCCGCTACAG GGCGCGTCCC ATTCGCCATT CAGGCTGCGC
3061 AACTGTTGGG AAGGGCGATC GGTGCGGGCC TCTTCGCTAT TACGCCAGTC GACCATAGCC
3121 AATTCAATAT GGCGTATATG GACTCATGCC AATTCAATAT GGTGGATCTG GACCTGTGCC
3181 AATTCAATAT GGCGTATATG GACTCGTGCC AATTCAATAT GGTGGATCTG GACCCCAGCC
3241 AATTCAATAT GGCGGACTTG GCACCATGCC AATTCAATAT GGCGGACTTG GCACTGTGCC
3301 AACTGGGGAG GGGTCTACTT GGCACGGTGC CAAGTTTGAG GAGGGGTCTT GGCCCTGTGC
3361 CAAGTCCGCC ATATTGAATT GGCATGGTGC CAATAATGGC GGCCATATTG GCTATATGCC
3421 AGGATCAATA TATAGGCAAT ATCCAATATG GCCCTATGCC AATATGGCTA TTGGCCAGGT
3481 TCAATACTAT GTATTGGCCC TATGCCATAT AGTATTCCAT ATATGGGTTT TCCTATTGAC
3541 GTAGATAGCC CCTCCCAATG GGCGGTCCCA TATACCATAT ATGGGGCTTC CTAATACCGC
3601 CCATAGCCAC TCCCCCATTG ACGTCAATGG TCTCTATATA TGGTCTTTCC TATTGACGTC
3661 ATATGGGCGG TCCTATTGAC GTATATGGCG CCTCCCCCAT TGACGTCAAT TACGGTAAAT
3721 GGCCCGCCTG GCTCAATGCC CATTGACGTC AATAGGACCA CCCACCATTG ACGTCAATGG
3781 GATGGCTCAT TGCCCATTCA TATCCGTTCT CACGCCCCCT ATTGACGTCA ATGACGGTAA
3841 ATGGCCCACT TGGCAGTACA TCAATATCTA TTAATAGTAA CTTGGCAAGT ACATTACTAT
3901 TGGAAGGACG CCAGGGTACA TTGGCAGTAC TCCCATTGAC GTCAATGGCG GTAAATGGCC
3961 CGCGATGGCT GCCAAGTACA TCCCCATTGA CGTCAATGGG GAGGGGCAAT GACGCAAATG
4021 GGCGTTCCAT TGACGTAAAT GGGCGGTAGG CGTGCCTAAT GGGAGGTCTA TATAAGCAAT
4081 GCTCGTTTAG GGAAC
```

//

9.3.2. CS2+MT



CS2+MT(myc tag)

Designed for making epitope-tagged fusion proteins. Includes 6 copies of the myc epitope recognized by the 9e10 monoclonal antibody cloned between Cla1 and EcoR1 (see Roth, M.B., Zahler, A.M., & Stolk, J.A. [1991] J. Cell Biol. 115: 587-596.). Functional CS2+MT (or CS3+MT, see next page) fusions have been made for many

transcription factors (MyoD, XASH3, others), several signal transducing proteins (ras, mCAP, others), and at least one transmembrane protein (Notch). Most fusion have been made to the N-terminus, but C-terminal fusions also work. A few internal fusions have been made, but these are riskier. In a number of cases, subcellular localization of a protein has been determined using fusions to the multimerized myc-epitope tag (6MT). We know of no example where the tag has altered the subcellular localization of the fusion protein, but obviously this is a possibility that should be kept in mind. At very high levels of expression (e.g. after injection of large amounts of RNA into Xenopus embryos), nuclear proteins with the 6MT can be detected in both the nucleus and cytoplasm. Presumably this is an artifact arising from the large amount of protein present. In the absence of an insert, the 6MT protein expressed from CS2+MT (or CS3+MT) is cytoplasmic. It seems quite stable in Xenopus embryos.

```
                                                        BamH1    Cla1  (Dra1)
Hind3                                                    |        |      |
|          40                   60                      80       |      |     100                120                 140                160
| *         *          *         *          *          |*       |  *   |       *         *         *         *          *         *         *         *
AAG CTTGATTTAG GTGACACTAT AGAATACAAG CTACTTGTTC TTTTTGCAGG ATCCCATCGA TTTAAAGCT ATGGAGCAAAAGCTCATTTCTGAAGAGGACTTGAATGAA
ATGGAGCAAAAGCTCATTTCTGAAGAGGACTTGAATGAA
       |-SP6 Promoter---->|--5' leader sequence-------|--polylinker--|         M  E  Q  K  L  I  S  E  E  D  L  N  E   M  E  Q  K  L  I  S  E  E  D  L
N  E
                                                                                |----1st myc epitope tag----39nt------| |----2nd myc epitope----tag39nt---
----|

  180                200                 220                240                 260                280
    *         *         *         *         *         *         *         *         *         *         *         *         *         *
  ATGGAGCAAAAGCTCATTTCTGAAGAGGACTTGAATGAA ATGGAGCAAAAGCTCATTTCTGAAGAGGACTTGAATGAA ATGGAGCAAAAGCTCATTTCTGAAGAGGACTTGAATGAA
   M  E  Q  K  L  I  S  E  E  D  L  N  E   M  E  Q  K  L  I  S  E  E  D  L  N  E   M  E  Q  K  L  I  S  E  E  D  L  N  E
  |----3rd myc epitope tag----39nt------| |----4th myc epitope tag----39nt------| |----5th myc epitope tag----39nt------|
                                                                                                                                  L   N   S   R
                                                                  Xho1
                     Nco1                                EcoR1     Stu1 |       Xba1                      SnaB1
                      |                                   |         |   |        |                          |
    300               | 320                 340            |       360   |        |      380                 | 400
      *         *     |   *         *         *         *|         |   |      * |        *          *      |   *
  ATGGAGAGCTTGGGCGACCTCACC ATGGAGCAAAAGCTCATTTCTGAAGAGGACTTGAATTCAA GGCCTCTCGA GCCTCTAGAA CTATAGTGAG TCGTATTACG TA
   M  E  S  L  G  D  L  T   M  E  Q  K  L  I  S  E  E  D  L  N  S   R  P  L  E   P  L  E   L  *  *
  |----unique sequence---| |-------6th myc epitope tag--------|                           <--T7 Promoter----|
                                                     |-----polylinker------------|
                                                                                   3' <-GATATCACTC AGCATAA 5'
    5' GAGCTTGGGCGACCTCA-> 3'                                                              T7 primer
        MT primer
```

Note that blunt end fusions to an inserted protein can be made in all three reading frames:

```
  Stu1 (blunt):
        TTG AAT TCA AGG
         L   N   S   R
```

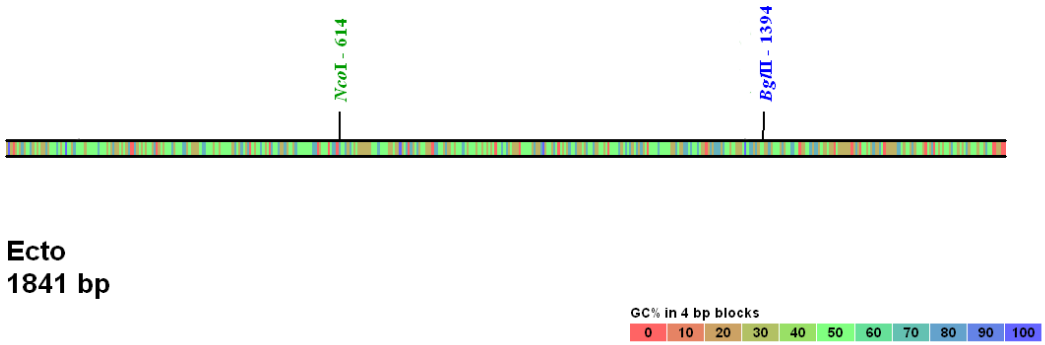
```
EcoR1 (filled in):
      TTG AAT T
       L   N

Xho1  (filled in):
      TTG AAT TCA AGG CCT CTC GA
               L   N   S   R   P   L
```

```
LOCUS       PCS2+MT      4352 BP SS-DNA   CIRCULAR  SYN       15-JAN-2001
DEFINITION  -
ACCESSION   -
KEYWORDS    -
SOURCE      -
FEATURES             Location/Qualifiers
     mRNA            1..1
                     /note="base1 = cap site for sCMV IE94 promoter"
     frag            35..51
                     /note="SP6 promoter"
     frag            54..>63
                     /note="CS2 insertion"
     frag            <64..78
                     /note="B-globin 5' mRNA sequence"
     frag            join(79..>92,<351..377)
                     /note="polylinker [Split]"
     frag            100..350
                     /note="6 myc tag repeats"
     frag            380..396
                     /note="T7 promoter (reverse orientation)"
     CDS             100..387
     frag            401..596
                     /note="SV40 late polyadenylation site (Bam to Dra1)"
     frag            597..629
                     /note="donwstream polylinker (from pVZ1)"
     frag            630..3365
                     /note="Bluescript II KS+ from Kpn1(759) to old Pvu2(529;
                     destroyed)"
     frag            650..666
                     /note="T3 promoter (reverse orientation)"
     frag            683..698
                     /note="M13 reverse primer binding site"
     frag            2837..3297
                     /note="f1 ori"
     frag            100..350
                     /note="1 to 251 of CS+MT XA10-1/MT S+L"
     frag            3366..4352
                     /note="3109 to 4095 of pCS2+"
     frag            3366..>4352
                     /note="simian CMV IE94 enhancer/promoter "
```

```
     frag            4157..4352
                     /note="based on DT sequences"
     frag            3366..4156
                     /note="based on database sequences"
BASE COUNT     1140 A   1030 C   1073 G   1109 T      0 OTHER
ORIGIN      -
        1 CGCCATTCTG CCTGGGGACG TCGGAGCAAG CTTGATTTAG GTGACACTAT AGAATACAAG
       61 CTACTTGTTC TTTTTGCAGG ATCCCATCGA TTTAAAGCTA TGGAGCAAAA GCTCATTTCT
      121 GAAGAGGACT TGAATGAAAT GGAGCAAAAG CTCATTTCTG AAGAGGACTT GAATGAAATG
      181 GAGCAAAAGC TCATTTCTGA AGAGGACTTG AATGAAATGG AGCAAAAGCT CATTTCTGAA
      241 GAGGACTTGA ATGAAATGGA GCAAAAGCTC ATTTCTGAAG AGGACTTGAA TGAAATGGAG
      301 AGCTTGGGCG ACCTCACCAT GGAGCAAAAG CTCATTTCTG AAGAGGACTT GAATTCAAGG
      361 CCTCTCGAGC CTCTAGAACT ATAGTGAGTC GTATTACGTA GATCCAGACA TGATAAGATA
      421 CATTGATGAG TTTGGACAAA CCACAACTAG AATGCAGTGA AAAAAATGCT TTATTTGTGA
      481 AATTTGTGAT GCTATTGCTT TATTTGTAAC CATTATAAGC TGCAATAAAC AAGTTAACAA
      541 CAACAATTGC ATTCATTTTA TGTTTCAGGT TCAGGGGGAG GTGTGGGAGG TTTTTTAATT
      601 CGCGGCCGCG GCGCCAATGC ATTGGGCCCG GTACCCAGCT TTTGTTCCCT TTAGTGAGGG
      661 TTAATTGCGC GCTTGGCGTA ATCATGGTCA TAGCTGTTTC CTGTGTGAAA TTGTTATCCG
      721 CTCACAATTC CACACAACAT ACGAGCCGGA AGCATAAAGT GTAAAGCCTG GGGTGCCTAA
      781 TGAGTGAGCT AACTCACATT AATTGCGTTG CGCTCACTGC CCGCTTTCCA GTCGGGAAAC
      841 CTGTCGTGCC AGCTGCATTA ATGAATCGGC CAACGCGCGG GGAGAGGCGG TTTGCGTATT
      901 GGGCGCTCTT CCGCTTCCTC GCTCACTGAC TCGCTGCGCT CGGTCGTTCG GCTGCGGCGA
      961 GCGGTATCAG CTCACTCAAA GGCGGTAATA CGGTTATCCA CAGAATCAGG GGATAACGCA
     1021 GGAAAGAACA TGTGAGCAAA AGGCCAGCAA AAGGCCAGGA ACCGTAAAAA GGCCGCGTTG
     1081 CTGGCGTTTT TCCATAGGCT CCGCCCCCCT GACGAGCATC ACAAAAATCG ACGCTCAAGT
     1141 CAGAGGTGGC GAAACCCGAC AGGACTATAA AGATACCAGG CGTTTCCCCC TGGAAGCTCC
     1201 CTCGTGCGCT CTCCTGTTCC GACCCTGCCG CTTACCGGAT ACCTGTCCGC CTTTCTCCCT
     1261 TCGGGAAGCG TGGCGCTTTC TCATAGCTCA CGCTGTAGGT ATCTCAGTTC GGTGTAGGTC
     1321 GTTCGCTCCA AGCTGGGCTG TGTGCACGAA CCCCCCGTTC AGCCCGACCG CTGCGCCTTA
     1381 TCCGGTAACT ATCGTCTTGA GTCCAACCCG GTAAGACACG ACTTATCGCC ACTGGCAGCA
     1441 GCCACTGGTA ACAGGATTAG CAGAGCGAGG TATGTAGGCG GTGCTACAGA GTTCTTGAAG
     1501 TGGTGGCCTA ACTACGGCTA CACTAGAAGG ACAGTATTTG GTATCTGCGC TCTGCTGAAG
     1561 CCAGTTACCT TCGGAAAAAG AGTTGGTAGC TCTTGATCCG GCAAACAAAC CACCGCTGGT
     1621 AGCGGTGGTT TTTTTGTTTG CAAGCAGCAG ATTACGCGCA GAAAAAAAGG ATCTCAAGAA
     1681 GATCCTTTGA TCTTTTCTAC GGGGTCTGAC GCTCAGTGGA ACGAAAACTC ACGTTAAGGG
     1741 ATTTTGGTCA TGAGATTATC AAAAAGGATC TTCACCTAGA TCCTTTTAAA TTAAAAATGA
     1801 AGTTTTAAAT CAATCTAAAG TATATATGAG TAAACTTGGT CTGACAGTTA CCAATGCTTA
     1861 ATCAGTGAGG CACCTATCTC AGCGATCTGT CTATTTCGTT CATCCATAGT TGCCTGACTC
     1921 CCCGTCGTGT AGATAACTAC GATACGGGAG GGCTTACCAT CTGGCCCCAG TGCTGCAATG
     1981 ATACCGCGAG ACCCACGCTC ACCGGCTCCA GATTTATCAG CAATAAACCA GCCAGCCGGA
     2041 AGGGCCGAGC GCAGAAGTGG TCCTGCAACT TTATCCGCCT CCATCCAGTC TATTAATTGT
     2101 TGCCGGGAAG CTAGAGTAAG TAGTTCGCCA GTTAATAGTT TGCGCAACGT TGTTGCCATT
     2161 GCTACAGGCA TCGTGGTGTC ACGCTCGTCG TTTGGTATGG CTTCATTCAG CTCCGGTTCC
     2221 CAACGATCAA GGCGAGTTAC ATGATCCCCC ATGTTGTGCA AAAAAGCGGT TAGCTCCTTC
     2281 GGTCCTCCGA TCGTTGTCAG AAGTAAGTTG GCCGCAGTGT TATCACTCAT GGTTATGGCA
     2341 GCACTGCATA ATTCTCTTAC TGTCATGCCA TCCGTAAGAT GCTTTTCTGT GACTGGTGAG
     2401 TACTCAACCA AGTCATTCTG AGAATAGTGT ATGCGGCGAC CGAGTTGCTC TTGCCCGGCG
     2461 TCAATACGGG ATAATACCGC GCCACATAGC AGAACTTTAA AAGTGCTCAT CATTGGAAAA
     2521 CGTTCTTCGG GGCGAAAACT CTCAAGGATC TTACCGCTGT TGAGATCCAG TTCGATGTAA
     2581 CCCACTCGTG CACCCAACTG ATCTTCAGCA TCTTTTACTT TCACCAGCGT TTCTGGGTGA
     2641 GCAAAAACAG GAAGGCAAAA TGCCGCAAAA AAGGGAATAA GGGCGACACG GAAATGTTGA
```

```
2701 ATACTCATAC TCTTCCTTTT TCAATATTAT TGAAGCATTT ATCAGGGTTA TTGTCTCATG
2761 AGCGGATACA TATTTGAATG TATTTAGAAA AATAAACAAA TAGGGGTTCC GCGCACATTT
2821 CCCCGAAAAG TGCCACCTAA ATTGTAAGCG TTAATATTTT GTTAAAATTC GCGTTAAATT
2881 TTTGTTAAAT CAGCTCATTT TTTAACCAAT AGGCCGAAAT CGGCAAAATC CCTTATAAAT
2941 CAAAAGAATA GACCGAGATA GGGTTGAGTG TTGTTCCAGT TTGGAACAAG AGTCCACTAT
3001 TAAAGAACGT GGACTCCAAC GTCAAAGGGC GAAAAACCGT CTATCAGGGC GATGGCCCAC
3061 TACGTGAACC ATCACCCTAA TCAAGTTTTT TGGGGTCGAG GTGCCGTAAA GCACTAAATC
3121 GGAACCCTAA AGGGAGCCCC CGATTTAGAG CTTGACGGGG AAAGCCGGCG AACGTGGCGA
3181 GAAAGGAAGG GAAGAAAGCG AAAGGAGCGG GCGCTAGGGC GCTGGCAAGT GTAGCGGTCA
3241 CGCTGCGCGT AACCACCACA CCCGCCGCGC TTAATGCGCC GCTACAGGGC GCGTCCCATT
3301 CGCCATTCAG GCTGCGCAAC TGTTGGGAAG GGCGATCGGT GCGGGCCTCT TCGCTATTAC
3361 GCCAGTCGAC CATAGCCAAT TCAATATGGC GTATATGGAC TCATGCCAAT TCAATATGGT
3421 GGATCTGGAC CTGTGCCAAT TCAATATGGC GTATATGGAC TCGTGCCAAT TCAATATGGT
3481 GGATCTGGAC CCCAGCCAAT TCAATATGGC GGACTTGGCA CCATGCCAAT TCAATATGGC
3541 GGACTTGGCA CTGTGCCAAC TGGGGAGGGG TCTACTTGGC ACGGTGCCAA GTTTGAGGAG
3601 GGGTCTTGGC CCTGTGCCAA GTCCGCCATA TTGAATTGGC ATGGTGCCAA TAATGGCGGC
3661 CATATTGGCT ATATGCCAGG ATCAATATAT AGGCAATATC CAATATGGCC CTATGCCAAT
3721 ATGGCTATTG GCCAGGTTCA ATACTATGTA TTGGCCCTAT GCCATATAGT ATTCCATATA
3781 TGGGTTTTCC TATTGACGTA GATAGCCCCT CCCAATGGGC GGTCCCATAT ACCATATATG
3841 GGGCTTCCTA ATACCGCCCA TAGCCACTCC CCCATTGACG TCAATGGTCT CTATATATGG
3901 TCTTTCCTAT TGACGTCATA TGGGCGGTCC TATTGACGTA TATGGCGCCT CCCCCATTGA
3961 CGTCAATTAC GGTAAATGGC CCGCCTGGCT CAATGCCCAT TGACGTCAAT AGGACCACCC
4021 ACCATTGACG TCAATGGGAT GGCTCATTGC CCATTCATAT CCGTTCTCAC GCCCCCTATT
4081 GACGTCAATG ACGGTAAATG GCCCACTTGG CAGTACATCA ATATCTATTA ATAGTAACTT
4141 GGCAAGTACA TTACTATTGG AAGGACGCCA GGGTACATTG GCAGTACTCC CATTGACGTC
4201 AATGGCGGTA AATGGCCCGC GATGGCTGCC AAGTACATCC CCATTGACGT CAATGGGGAG
4261 GGGCAATGAC GCAAATGGGC GTTCCATTGA CGTAAATGGG CGGTAGGCGT GCCTAATGGG
4321 AGGTCTATAT AAGCAATGCT CGTTTAGGGA AC
//
```

9.4. Restriction map of ectodermin for colony PCR screening assay



9.5. Embryo picture

An enlarged view of the picture B7 from the figure 24 is available here: