FUNCTIONAL CHARACTERIZATION OF LOC115098: A NOVEL GENE CONSERVED IN EUKARYOTIC GENOMES

A THESIS SUBMITTED TO
THE DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS
AND THE INSTITUTE OF ENGINEERING AND SCIENCE OF
BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE

BY
ESRA KARAKÖSE
JANUARY 2007

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assist Prof. Ayşe Elif Erson

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assist Prof. Kamil Can Akçalı

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assist Prof. Uygar Halis Tazebay

Approved for the Institute of Engineering and Science

Director of Institute of Engineering and Science
Prof. Mehmet Baray

ABSTRACT

FUNCTIONAL CHARACTERIZATION OF LOC115098: A NOVEL GENE CONSERVED IN EUKARYOTIC GENOMES

Esra Karaköse
M.Sc. in Molecular Biology and Genetics
Supervisor: Assist Prof. Uygar H. Tazebay
January 2007, 76 Pages

Loc115098 is a newly identified gene in our laboratory. The gene was firstly discovered in an unrelated study in which the cis-acting regulatory elements of a neighboring gene NIS (Sodium-Iodide Symporter) were investigated. It is composed of 4 exons. According to the computational predictions, it has rather interesting features such as having a composition of mostly charged amino acids, carrying a nuclear localization signal and being well conserved throughout the eukaryotic kingdom. Our aim in this study was to understand the function of this new gene. Firstly we confirmed that it is expressed in several different tissues, later we conducted a subcellular localization study whose results led us to hypothesize that the function of loc115098 could be related to cell cycle regulation. According to our hypothesis we decided to utilize a lower eukaryote carrying a human homologue of loc115098 which we selected to be Aspergillus nidulans. It was a very useful model organism for our study not only because it carries a very similar homologue of loc115098 but also it is an easily handled organism which is widely used for genetic studies. Firstly we compared the expression profile of loc115098 in several cell cycle mutants of A. nidulans as well as a wild type strain, the results indicated that the expression of loc115098 decreases in the restrictive temperature of cell cycle mutant strains. We also conducted subcellular localization studies in A. nidulans and the results indicated that it was mostly cytoplasmic and sometimes perinuclear. Our other aim was to utilize A. nidulans as a model for our knock-out experiments. We designed a knock-out system by the help of a sophisticated method called Double-Joint PCR. Our first attempts supported that the knock-out of loc115098 was likely to cause lethality in the organism and therefore we improved our system to conditionally knock out the gene. Our preliminary results indicated that the knock out of loc115098 causes cell death and we suspect that it could be due to a damage in one of the crucial metabolic events for cell viability such as mitosis. Our attempts to functionally characterize the function of loc115098 are currently going on.

ÖZET

ÖKARYOTİK CANLILARDA KORUNMUŞ OLAN LOC115098 GENİNİN İŞLEVSEL TANIMLANDIRILMASI

Esra Karaköse
Moleküler Biyoloji ve Genetik Yüksek Lisans Derecesi
Tez Yöneticisi: Yard. Doç. Dr. Uygar H. Tazebay
Ocak 2007, 76 Sayfa

Loc115098 ilk kez tarafımızdan bulunmuş bir gendir. İlk defa bu genin komşusu olan NIS (Sodyum-Iyot Simporter) geninin anlatımından sorumlu cis kontrol bölgelerinin aydınlatılması amacıyla yapılan bir çalışma sırasında bulunmuştur. Bu gen 4 eksondan meydana gelmektedir. Bioınformatik analizlerden faydalanarak yaptığımız araştırmalar bu genin oldukça ilginç özelliklere sahip olduğunu gösterdi, örneğin proteinin yapısı genel olarak pozitif yüklü amino asitlerden oluşuyor, çekirdek sinyali taşıyor ve ökaryotik organizmalarda oldukça korunmuş bir şekilde bulunuyor. Bizim bu çalışmadaki amacımız bu genin işlevini anlamaktı. Bu nedenle öncelikle farklı dokularda bu genin ifadesi olduğunu teyid ettik. Sonra bu proteinin hücre içersindeki konumunu anlamak için çalışmalar yaptık. Sonuçlar, bu proteinin hücre döngüsü kontrolüyle ilgili bir işleve sahip olabileceği hipotezini kurmamızı sağladı. Bunun üzerine çalışmalarımızı sürdürmek amacıyla loc115098 geninin homoloğunu taşıyan daha düşük bir ökaryot olan Aspergillus nidulans'ı seçtik. Bu organizma bizim çalışmalarımız açısından oldukça kullanışlıydı çünkü hem loc115098 geninin çok benzer bir homoloğunu taşıyordu hem de bir model organizma olarak genetik araştırmalarda sıklıkla kullanılıyordu. Öncelikle loc115098 geninin anlatımını çeşitli hücre döngüsü mutasyonlarına sahip olan A. nidulans ırklarında ve doğal (wild type) ırkta karşılaştırmalı olarak inceledik. Bu çalışmanın sonucu bize mutant olan ırkların yaşamalarına müsaade etmeyen sıcaklık derecelerinde loc115098 ifadesinin azaldığını gösterdi. Aynı zamanda bu organizma ile de LOC115098 proteinin hücre içindeki konumlanışını araştırdık ve sonuç bize bu proteinin çoğunlukla sitoplazmada bazen de çekirdeğin etrafını saracak şekilde bulunduğunu gösterdi. Ayrıca Double-Joint PCR adı verilen oldukça sofistike bir metoddan faydalanarak knock-out çalışmalar yaptık ve ilk elde ettiğimiz sonuçlar bize bu genin, hücrenin yaşamsal olaylarından sorumlu bir işleve sahip olabileceğini gösterdi çünkü bu genin knock-out edildiği durumlarda canlının yaşadığı örneklere hiç rastlamadık. Loc115098 geninin işlevini aydınlatmaya dair çalışmalarımız halen devam ediyor.

All truths are easy to understand once they are discovered; the point is to discover them.

Galileo Galilei

ACKNOWLEDGEMENTS

Fistly, I would like to thank my supervisor Dr Uygar Tazebay for his continious help, support and friendship during my days in Ankara and in Athens.

I would also like to express my gratitude to my second supervisor Dr Georges Diallinas who was not only a great supervisor but also a nice friend whenever necessary.

I would like to thank the members of Molecular Biology and Genetics Department at Bilkent University for the nice working environment especially to Dr Hani Alotaibi for his valuable technical guidance.

I would like to thank Dr Diallinas' group in the University of Athens for the lovely working atmosphere in the lab and their immense technical support. Furthermore I would like to thank Dr Ioannis Serafimidis for both his contributions to the project and his company.

I would also like to express my deep gratitude to my family for their existence and moral support during my entire life.

Lastly I would like to express that I am very grateful to all my friends who never left me alone during this long journey and always stood by me.

This project was supported by a 8-month scholarship for research from the Greek Ministery of Education.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

3D: 3 dimensions

AlcA: alcohol dehydrogenase A

A. nidulans: *Aspergillus nidulans*

BLAST: Basic Local Alignment Search Tool

Bim: blocked in mitosis

BSA: Bovine serum albumin

Bp: base pair

cAMP: cyclic adenosine monophosphate

cDNA: complementary DNA

cGMP: cyclic guanosine monophophate

CM: complete medium

C-terminus: carboxy-terminus

DAPI: 4,6-diamidino-2-phenylindole, dihydrochloride

DEPC: diethylpyrocarbonate

dCTP: deoxycytidine triphosphate

dH_2O: distilled H_2O

dNTP: deoxynucleotide triphosphate

DNA: deoxyribonucleic acid

EDTA: ethylenediaminetetraacetic acid

EtBr: ethidium bromide

EtOH: ethanol

FBS: Fetal Bovine Serum

G1: Gap 1

G2: Gap 2

GAPDH: Glyceraldehyde 3-phosphate dehydrogenase

GFP: Green Fluorescent Protein

HCl: hydrochloric acid

Loc: locus

Kb: kilobase

µg: microgram

µl: microliter

Mb: Megabase

ml: milliliter

MM: minimal medium

mRNA: messenger RNA

N-terminus: amino-terminus

NaCl: sodium chloride

NaOH: sodium hydroxide

NCBI: National Center for Biotechnological Information

Nim: Never in mitosis

NIS: Sodium Iodide Symporter

Nud: nuclear distribution

OD: optic density

ORF: open reading frame

pBKS: pBlueScript vector

PBS: phosphate buffer saline

PCR: Polymerase Chain Reaction

PEG: polyethylene glycol

pGEM-T: pGEM-Teasy vector

RNA: ribonucleic acid

rRNA: ribosomal RNA

rpm: revolutions per minute

RT-PCR: Revers transcriptase-PCR

S: Synthesis

SDS: sodium dodecyl sulfate

SEM: Scanning Electron Microscope

SM: sucrose medium

Sud: suppressor of bim

TAE: tris-acetate-EDTA

TRIS: 2-amino-2-hydroxymethyl-1,3-propanediol

Wt: wild type

UCSC: University of California Santa Cruz

1. INTRODUCTION

1.1. General knowledge about Hypothetical Proteins

The completion of Human Genome Project (Human Genome Sequencing Consortium, 2004), revealed that the expected number of genes in the human genome is neither around 100000 as it was expected in some text books nor around 30000 (Lander et al., 2001) as expected in others. More recent studies have indicated that this number varies between 20000 and 25000. Even though we are able to perform such estimations we are still far from understanding the human genome in its all aspects. Indeed it is well-known that coding sequences, as well as many non-coding regulatory regions tend to keep their sequences conserved during evolution. Although there are certain efforts to understand the function of non-coding regulatory conserved elements at the level of entire genome, the identification, functional annotation and systemic classification of the coding sequences remain incomplete.

Until today the DNA sequences of more than 100 species have been completed and one of the most surprising outcomes of genome research is that roughly 20-40% of genes in newly sequenced genomes do not have statistically significant matches to functionally annotated sequences and they are annotated as 'hypothetical proteins' (Eisenstein et al., 2000).

We believe that one of the most important questions of the today's genome research is to assign annotations to the genes whose biological function is yet to be understood. The most common approaches and several methods will be mentioned in the following sections.

1.2. How to Study the Hypothetical Proteins

Since the annotation of hypothetical proteins encoding genes is crucial for the future of genomic studies, the tools which are generated for this scope is a hot topic in the field, too. There are indeed several different approaches to study the hypothetical proteins and it is possible to classify them as 'the bioinformatics tools' and 'the biochemical and genomic tools'.

1.2.1. The bioinformatics tools

There are several different approaches to functionally identify new genes. However the most accurate and effective annotations which have been performed up to date are based on sequence homology. Among them BLAST and PSI-BLAST are the most widely used tools (Altschul et al., 1997). However more advanced tools which have been used by several groups exist and the number of these tools keep growing everyday. Some of these methods will be named here. One of the most reliable tools is 'The Rosetta Stone Technique' which is based on protein-domain fusion events (Marcotte et al., 1999) (Enright and Spratt, 1999). There are others which are based on chromosomal proximity information. The rational behind this technique is that when two genes appear as a neighboring gene pair in the genomes of several distantly related organisms, meaning that they form a conserved gene cluster. This suggests the possibility that these genes might be functionally related (Overbeek et al., 1999). Another technique is based on the phylogenetic profile method. Basically this method indicates the presence or the absence of the gene of interest in each organism by an entry of 1 or 0. Thus the functional linkage is established when the two genes have similar phylogenetic profiles, they show a correlated pattern of inheritance across the genomes examined (Marcotte, 2000). Some other methods also combine the ones mentioned above (Zheng et al., 2002) in which they explain the genomic functional annotation using co-evolution profiles of gene clusters. Finally, by using specific amino acid frequency and the periodicity at the proteome level, functionally unknown proteins can be characterized (Fujishima et al., 2003).

Among all these methods, we used the most efficient and established one for our

bioinformatics analyses; BLAST. Moreover, we also focused on the biochemical and genomic analyses of our gene of interest.

1.2.2. Biochemical and genomic approaches to study hypothetical proteins

Although there is no certain method to follow in order to conduct biochemical and genomic studies of the hypothetical proteins, the general forward and reverse genetics methods are applicable. Firstly the confirmation of the expression of these genes is essential. After this confirmation, further analyses can be performed depending on the future scope of the study. The overexpression and disruption of the gene of interest are two main approaches in order to observe how crucial the protein is for the organism. As long as there is no domain similarity with previously annotated proteins, any method aiming to identify the protein-protein interaction such as immunoprecipitation or yeast two hybrid methods would be useful to find the targets of the hypothetical proteins. Moreover some localization studies by using immunofluorescence can be conducted to observe the subcellular localization of the hypothetical proteins. All these experiments can also be performed in time or growth dependent manners meaning that different developmental stages of a given organism could be investigated in terms of gene expression or the localization studies can be performed in a time dependent manner and the difference can be observed. Moreover the interaction of the protein of interest with other proteins may also vary according to the developmental or metabolic activities.

On the other hand the disruption of the gene of interest could also be a prominent approach. In case of any phenotypic changes or lethality occurs this would bring new insights to the study. Moreover, according to the bioinformatics analyses, some mutations can be inserted to check the possible outcomes of the mutation of those sequences.

Other more sophisticated approaches such as crystallizing the protein and exploring the 3D structure are possible. However these methods are generally considered to be more complicated and require higher technology and longer periods of commitment.

1.3 General Knowledge about 'Loc115098'

1.3.1. Identification of loc115098 as an expressed gene

Loc115098 was firstly identified in an unrelated study where *cis*-acting transcriptional elements of the NIS (Sodium-Iodide Symporter) gene were being analyzed. The focus of that study was to analyze a putative transcriptional control element localized in the 3' downstream of NIS gene by using a tool called VISTA (Negre et al., 2005). In this study 10 conserved putative regions were identified and we firstly discovered the 5' regulatory region of Loc115098 as one of these putative regions and then the Loc115098 itself was discovered by our group (Hani Alotaibi *et al*, unpublished data). The gene could be found in University of California Santa Cruz database (Kent et al., 2002) with the accession number of LOC115098.

1.4. Model Organisms to Study Gene Function

Model systems have always been useful in basic research. They have contributed to the creation and development of basic biological knowledge for many years. Especially in the last decades due to the rapidly expanding genetic research area revealing the genomes of different species, model organisms have gained even more importance. The common evolutionary heritage makes it possible to use genetically tractable organisms to model important aspects of human medical disorders such as cancer, birth defects, neurological dysfunctions, reproductive failure, malnutrition and aging in systems amenable to rapid and powerful experimentation (Spradling et al., 2006).

Utilizing model organisms will help us identify common genes, proteins and biological processes. This will not only be useful for unrevealing the complications of human genetic disorders but also establish a better understanding of the roles of the common mechanisms which take part in all fundamental life processes, including development, physiology, growth, behavior and aging. Moreover the studies with model organisms provide information which can not practically be obtained in any other way. Mostly due to practical difficulties to utilize and manipulate higher

organisms and sometimes due to the ethical issues, the idea of using a model system has always been very useful.

1.4.1. General Knowledge about Model Organisms

The number of model organisms which are used for genetic, biochemical and molecular studies are growing rapidly. The most commonly used organisms include rat (*Rattus norvegicus*), mouse (*Mus musculus*), zebrafish (*Danio rerio*), the fruit fly (*Drosophila melanogaster*), the budding yeast *Saccharomyces cerevisiae*, the amphibian *Xenopus tropicalis*, the nematode *Caenorhabditis elegans,* cellular slime mold *Dictyostelium discoideum* and finally the ascomycete *Aspergillus nidulans*. They are all useful to study different functions, for instance the *C. elegans* is mostly utilized for neurobiological and developmental studies whereas the *S. cerevisiae* is a good model to study the control of cell division. According to the aim of the study and the laboratory facilities, the most easily handled and the most appropriate organism can be selected from a big variety.

Taking fungi into consideration, it is obvious that they have a lot of features which make them useful as experimental genetic organism. Many of the fungi that are commonly used for genetic studies have most or all of the following features:

• A haploid phase for mutagenesis and screening.
• A dikaryon or heterokaryon phase for complementation and epistasis analysis.
• The ability to grow on defined media for selection and screening of nutritional and drug-resistance markers.
• The ability to grow at wide temperature ranges which facilitates the isolation of conditional lethal mutants.
• Heterothallic strains for out-crossing.
• Homothallic derivatives to allow haploid screens for mutants that are defective in fundamentally diploid processes (for example, meiosis).
• Small genomes (12–50 Mb) to facilitate mutant recovery and gene cloning.
• The ability to carry out tetrad analysis, in which all four meiotic products are held together in an ascus. (Casselton and Zolan, 2002)

1.4.2. Features and Characteristics of Aspergillus nidulans

Aspergillus nidulans is an ascomycetes belonging to the subphylum of Euascomycetes (Ascomycota). However the filamentous ascomycota is called pyrenomycetes. The classification can be made as following according to Micheli ex Link, 1809:

Kingdom: Fungi
Phylum: Ascomycota
Order: Eurotiales
Family: Trichocomaceae
Genus: Aspergillus
Species: nidulans

The yeasts which are unicellular ascomycotan fungi are considered to be the best organisms for the study of basic eukaryotic genetics (Hedges, 2002). The multicellular (filamentous) ascomycotan fungi have apparently larger genomes but they are as useful and practical to study basic eukaryotic genetics.

Aspergillus nidulans first appeared in the literature as an alternative model to *Neurospora crassa.* It was firstly introduced by Guido Pontecorvo in the 1950s (PONTECORVO et al., 1953). Like all other ascomycetes, *A. nidulans* also has a sac formation which is called ascus. The nuclear fusion and meiosis take place in the ascus. This eventually leads to the formation of sexual spores so called ascospores. The ascospores are formed within the ascus by an enveloping membrane system, which packages each nucleus with its adjacent cytoplasm and provides the site for ascospore wall formation. These membranes are derived from the ascus plasma membrane. This larger structure bursts and releases the thick-walled haploid ascospores which are resistant to adverse environments, but when they come across to the right conditions they start germinating and form the new haploid fungus. The hyphea is one filament of A. nidulans carrying several nuclei.

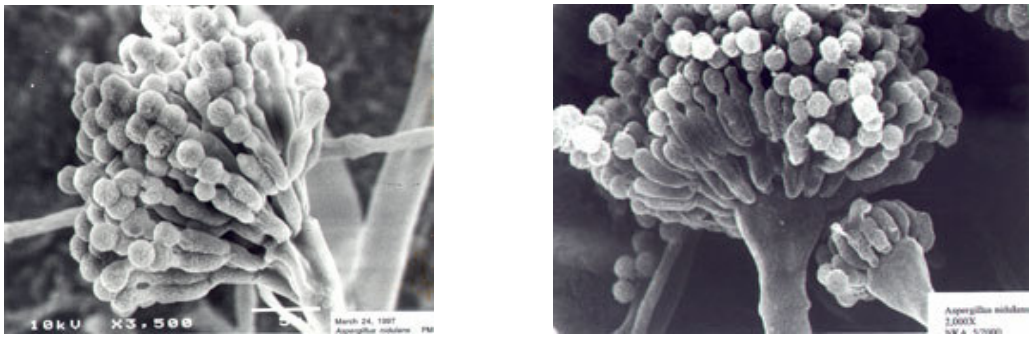

Figure1.1. The pictures of A. nidulans which were taken by SEM (Scanning Electron microscope). Both pictures show how *A. nidulans* look like in its natural environment.

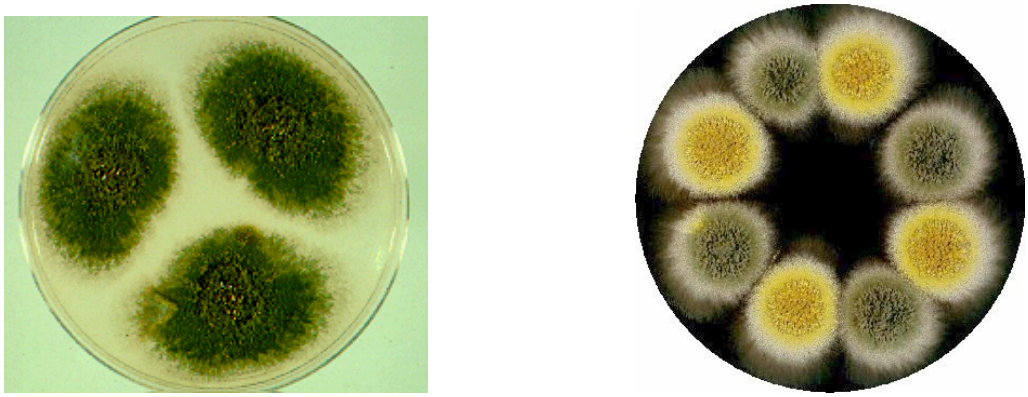

Figure1.2. Different cultures of Aspergillus nidulans. Both pictures show how A. nidulans grows under laboratory conditions. It might have different color mutants as it is indicated in one of the pictures

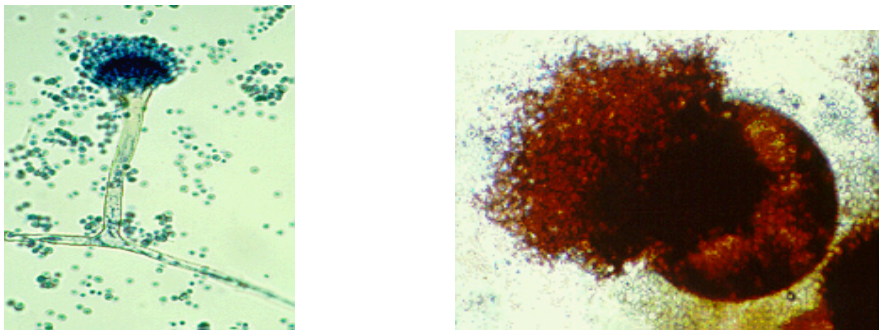

Figure1.3. In the first picture the conidial heads which are short and columnar are shown. Conidiophores are usually short, brownish and smooth-walled. Conidia are globose and rough-walled. In the second picture the cleistothecium of *A. nidulans* is shown pointing numerous reddish-brown ascospores.

1.4.2.1. Life Cycle of Aspergillus nidulans

The life cycle of the fungal mycelium of *A. nidulans* consists of sexual and asexual cycles. *A. nidulans* is a web of branched filaments (hyphae) of connected compartments or cells, each containing several nuclei. This mycelium, or homokaryon, which develops from a single haploid spore, differentiates into many identical asexual spores known as conidia or conidiospores (the asexual cycle). *A. nidulans* is homothallic, which means that it is self fertile, but crosses can be initiated by hyphal fusions between homokaryons with genetically different nuclei. The resulting heterokaryons are not stable, but can be forced to maintain a balanced ratio of the component nuclei by including complementing auxotrophic mutations in the parental nuclei and forcing growth without the corresponding supplements. *A. nidulans* can also reproduce sexually. In the fruiting body, which produces the sexual spores, a pair of nuclei that is destined for meiosis divides in synchrony to form a mass of cells known as the ascogenous hypha. These hyphae are highly branched and each tip cell becomes an ascus (a specialized cell) in which the two haploid nuclei fuse. The diploid nucleus undergoes meiosis followed by a post-meiotic mitosis, which results in the formation of eight haploid ascospores. The fruiting body, called the cleistothecium, can hold tens of thousands of ascospores, which are released into the environment when the cleistothecium bursts open. In addition to an asexual cycle and sexual cycle, a parasexual cycle offers the genetic benefits of meiosis achieved through a mitotic route (PONTECORVO and KAFER, 1958). The parasexual cycle is initiated when haploid nuclei fuse in the vegetative cells of a heterokaryon and continue to divide mitotically. Crossing over might occur between homologues and random chromosome loss restores the haploid chromosome number, which is eight in *A. nidulans*. These events can be used to map gene orders and assign new genes to the eight linkage groups. (See Figure 1.4)

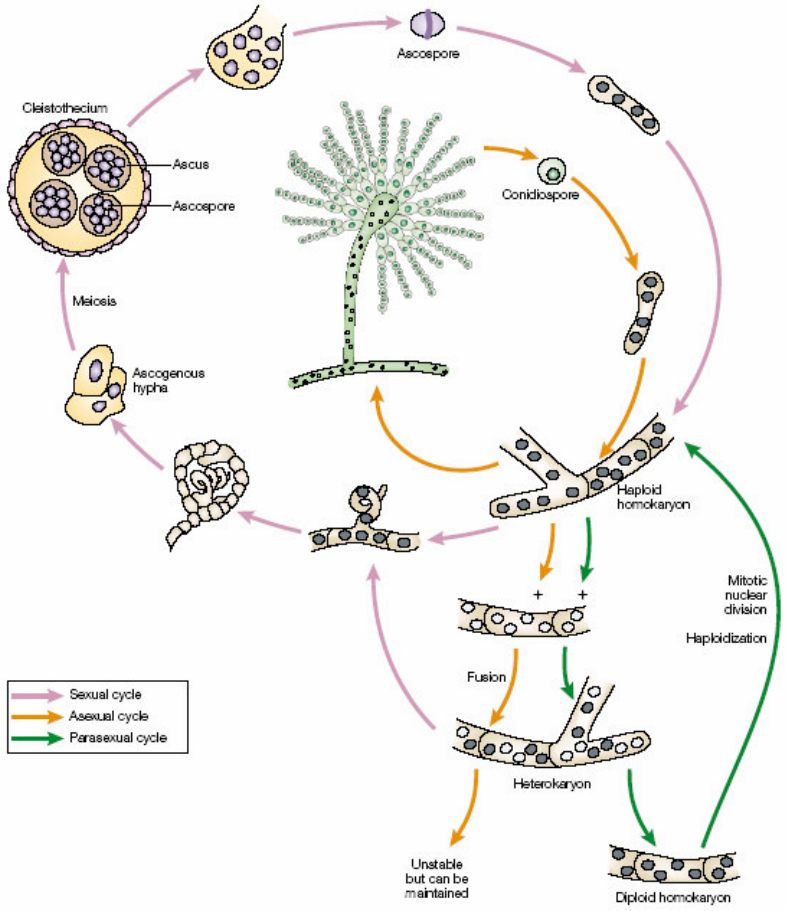

Figure1.4. The life cycle of *A. nidulans* shown schematically. The sexual cycle is shown in purple, the asexual cycle in orange and the parasexual cycle in green. (Casselton and Zolan, 2002)

1.4.2.2. Cell cycle Mutant Strains of Aspergillus nidulans

The *A. nidulans* cell cycle mutants have long been a good model to study gene function. The first mitotic mutants were screened by Ron Morris in 1975. He carried out the classical microscopic screen in which the morphology of the germinating conidiospore was used as the identifying phenotype. As the conidiospore germinates, the germ tube elongates to generate the first hyphal compartment, into which the conidiospore's nucleus moves and divides. The mitosis mutants were those in which the nuclear division (but not cell elongation) was affected at the restrictive temperature. In this screen the idea was to isolate the mutants as temperature-sensitive conditional-lethals because otherwise the cells which cannot go through

mitosis would immediately die. The permissive temperature was 32ºC and the restrictive temperature was 42ºC. Basically 3 classes of mutants were isolated in this study: nim (never in mitosis), bim (blocked in mitosis), nud (nuclear distribution). (See Figure 1.5)

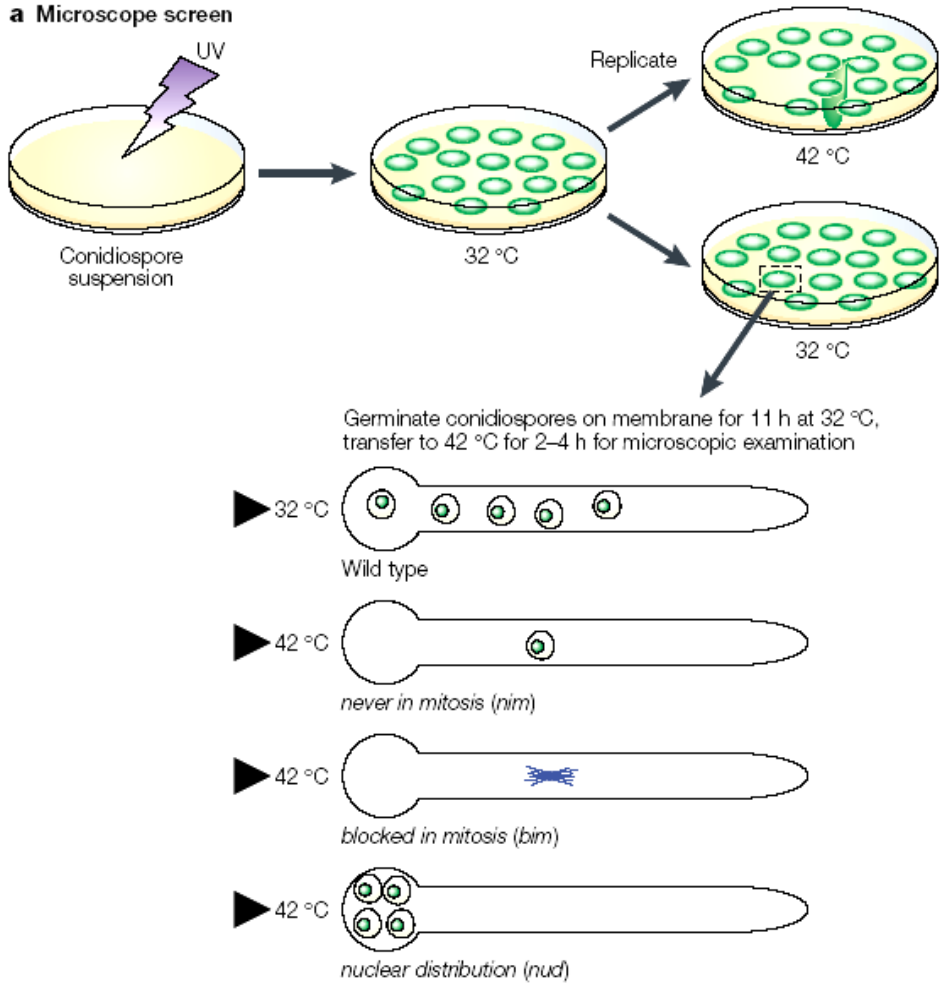

Figure1.5. The schematic representation of cell cycle mutants and how they were obtaine. The permissive temperature is 32ºC and the restrictive temperature is 42ºC. (Casselton and Zolan, 2002)

The mutant strains we use in this study are *nim*, *bim* and *sud* mutants. The NimA protein is a kinase that controls the entry into mitosis (Osmani *et al*., 1987). Thus the *nimA* mutants are arrested in late in G2 with duplicated spindle pole bodies, the fungal equivalent of the centrosome (Oakley *et al*., 1983). Further genetic analyses

showed that the interphase arrest of *nimA5* mutants was bypassed by *bimE7* mutation, with double mutants entering an aberrant mitotic state characterized by abnormalities of the nuclear envelope, which remains intact during mitosis in fungi, and the failure to form bipolar spindle (Osmani et al., 1991). In other words the *bimE7* mutation causes nuclei to accumulate and persist in a pre-anaphase mitotic state with tightly condensed chromatin and a normal-appearing spindle apparatus at restrictive temperature(James et al., 1995). (See Figure 1.6)

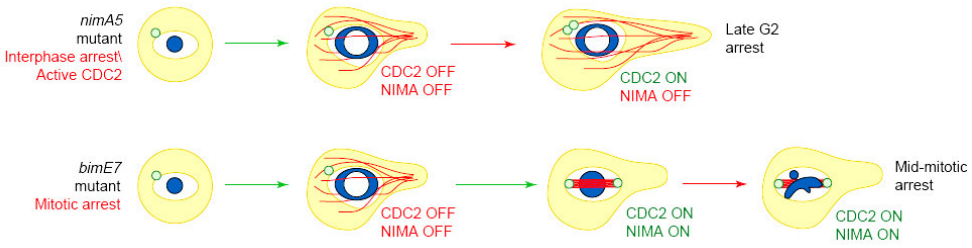

Figure1.6. The *nimA5* and the *bimE7* are shown in the picture. In the *nimA5* mutant the cell cycle arrest occurs at late G2 whereas in the *bimE7* mutant mid-mitotic arrest occurs. (O'Connell et al., 2003)

The *sud* mutant has also the background of *bimD6* mutation. The protein which is encoded by *bimD* shares sequence motifs with known DNA binding transcription factors (May et al., 1992a). The *bimD6* mutant germlings at restrictive temperature block in anaphase and eventually undergo nuclear fragmentation. In other words the *bimD6* mutation confers a heat-sensitive mitotic defect characterized by a failure of chromosomes to correctly attach to the spindle microtubules, resulting in an increase in chromosome loss. The extragenic suppressors of *bimD6* (sud) were isolated and characterized before (Holt and May, 1996b). Moreover it was discovered that in addition to their ability to suppress *bimD6*, *sud* genes are themselves required for successful nuclear division in *A.nidulans*. The germlings harboring *sud* mutations with the *bimD6* mutation undergo normal rounds of interphase and mitosis at 37ºC. After a shift to 25ºC for 8 hours, the approximate length of one cell cycle, nuclear morphology of all *sud* strains strikingly alters. The nuclei are not seen as delimited organelles but stretch about the hyphal interior as disorganized masses of chromatin. (Holt and May, 1996a).

1.4.3. Why we selected Aspergillus nidulans

There were several reasons which led us to choose the *Aspergillus nidulans* as our experimental model. First of all, the homologues counterpart of *loc115098* in the Aspergillus genome had 34.4% identity with the human homologue. (From now on the Aspergillus homologue will be called AN0879.3.) This high conservation rate would make our results more reliable. Furthermore *Aspergillus nidulans* is a non pathogenic and easily handled model organism for genetic, molecular and biochemical studies. It grows on very cheap and simple growth media. *Saccharomyces cerevisiae* could have been another alternative for this study since this organism is equally easy and cheap to handle. Interestingly this organism does not have a homologue of the protein of interest, therefore the idea was dismissed. In addition to these advantages of *A. nidulans* mentioned above, it was also important that all the methods such as genetic transformation and targeted gene replacement were described in the 80's and 90's which obviously made all manipulations easier because the establishment of such studies and protocols is also quite time consuming and hard to deal with. Moreover the *A. nidulans* has always played an important role in the genetic dissection of the cell cycle. And last, but not the least, the experience and expertise of Dr. Diallina's laboratory was an extra stimulus for us to conduct our experiments on *A. nidulans*.

1.4.5. Aim

Our aim in this study was firstly to characterize the *loc115098* gene. For this purpose we designed experimental approaches for both the human cell lines and the filamentous fungus *A. nidulans*. Our initial observations which will be explained in the 'Results' pointed that Loc115098 could be involved in mitosis and therefore we concentrated on the cell cycle regulation and the possible participation of Loc115098 into this process. Since *loc115098* is a newly identified gene, its features, function and any related data were still elusive. Our aims in this study were focusing on the identification of such data by utilizing as many tools and approaches as possible. Therefore we concentrated on a) localization, b) over expression, c) knock-out studies as well as the bioinformatics analyses of this new gene. The idea

behind all these approaches was firstly to understand the characteristics of Loc115098 and to check its expression profile and correlate it with the data coming from the bioinformatics analyses. In addition to this, we sought to identify the subcellular localization of this protein and to observe any phenotypical changes caused by the overexpression and/or the knock-out of the gene.

The subcellular localization studies were mainly focusing on the determination of the localization of Loc115098 protein in the cell. It might be crucial to understand where this protein localizes in the cell because it might give us a clue about the possible Loc115098 interacting proteins. In the overexpression studies; our aim was to observe the possible differences in the phenotype of the cells which were caused by the increase in *loc115098* or AN0879.3 expression. The results coming from this experiment would lead us to conduct some microscopic or biochemical analysis to observe and understand these changes in a better way. Lastly, our knock-out strategy was set in order to be able to observe the possible outcomes of the deletion of AN0879.3.

Certainly, all the results obtained from our already designed approaches would bring new aspects for the current project. This study was a part of our long term attempt to functionally characterize loc115098.

2. MATERIALS AND METHODS

2.1. Bioinformatics

2.1.1. Data bases

The accession number of human loc115098 in the NCBI database (http://www.ncbi.nlm.nih.gov) is NM138442. The same sequence can also be obtained from Genome Browser database at the University of California Santa Cruz (Kent *et al.*, 2002) with the accession number of LOC115098. The Zebrafish *loc115098* can also be found in NCBI with the accession number of BC066532. The *A.nidulans* copy of *loc115098* which we call AN0879.3 can be found in the *A.nidulans* database prepared by Broad Institute: (http://www.broad.mit.edu/annotation/genome/aspergillus_nidulans/Home.html).
The accession number for Loc115098 protein is AN0879.3; the mRNA data can also be reached and the exons and the introns of the gene can optionally be viewed. In order to find the homologues of loc115098 in different organisms generally the NCBI Blast (Altschul et al., 1990; Altschul et al., 1997) search was utilized.

2.1.2. DNA analysis

Once the sequences of different organisms were obtained, they were not only used for further studies such as amplification and cloning but also for performing some bioinformatical sequence comparisons. For these comparisons mainly the ClustalW (Higgins, 1994) (http://www.ebi.ac.uk/clustalw/#) was used.

When *loc115098* was firstly identified the annotation for the human sequence was obtained from the Genome Vista database (Couronne et al., 2003), and the sequences were aligned with the mVista too (Frazer et al., 2004) (Hani Alotaibi unpublished).

2.2. Plasmids and Cloning

The plasmids which were used for localization studies in human cell line MCF-7 were p3XFLAG-CMV-10 (Sigma) and p3XFLAG-CMV-14 (Sigma). The plasmids which were used for *A. nidulans* studies were pBKS (Stratagene) and pGEM-T (Promega). All plasmids had ampicillin resistance.

The standard protocol for cloning was to double digest both the plasmids and the inserts which have already been amplified with the appropriate enzyme and reaction conditions. This was followed by a ligation step and later the ligated plasmid was transformed into *E.coli* DH5α strain. After that the plasmids were rescued from the bacteria. Then several further digestions and PCR analyses were carried out in order to obtain the plasmid carrying the gene of interest.

The enzymes used for this purpose and their reagents were as following:

Table2.1. The restriction enzymes which were used for the cloning and plasmid confirmation purposes.

RESTRICTION ENZYMES		
Name	**Temperature**	**Buffers**
BamHI	37ºC	10mM Tris-HCl (pH 8.0 at 37°C), 5mM MgCl2, 100mM KCl, 0.02% Triton X-100, 1mM 2-mercapthoethanol, 0.1mg/ml BSA
BglI SalI	37ºC	50mM Tris-HCl (pH 7.5 at 37°C), 10mM MgCl2, 100mM NaCl, 0.1mg/ml BSA
EcoRI	37ºC	50mM Tris-HCl (pH 7.5 at 37°C), 10mM MgCl2, 100mM NaCl, 0.02% Triton X-100, 0.1mg/ml BSA
HindIII	37ºC	10mM Tris-HCl (pH 8.5 at 37°C), 10mM MgCl2, 100mM KCl, 0.1mg/ml BSA
KpnI	37ºC	10mM Tris-HCl (pH 7.5 at 37°C), 10mM MgCl2, 0.02% Triton X-100, 0.1mg/ml BSA
NcoI	37ºC	33mM Tris-acetate (pH 7.9 at 37°C), 10mM magnesium acetate, 66mM potassium acetate, 0.1mg/ml BSA
DpnI	37ºC	
SmaI	30ºC	
XbaI	37ºC	

The p3XFLAG-CMV-10 was basically used for N-terminus tagging of Loc115098. It was firstly double digested by *Xba*I and *Hind*III enzymes. The Loc115098 was amplified by the primers called HALNG-F1 and HALNG-R1 and these two fragments were ligated in appropriate conditions and the transformation was carried out. The p3XFLAG-CMV-14 was digested by *Bgl*I and *Hind*III enzymes and the

PCR product of Loc115098 was obtained by using the primers called HALNG-F1 and LOC-HN-Ctrev. Later the same ligation and transformation protocol was carried out. Several enzyme digestions were done to confirm the correct plasmid.

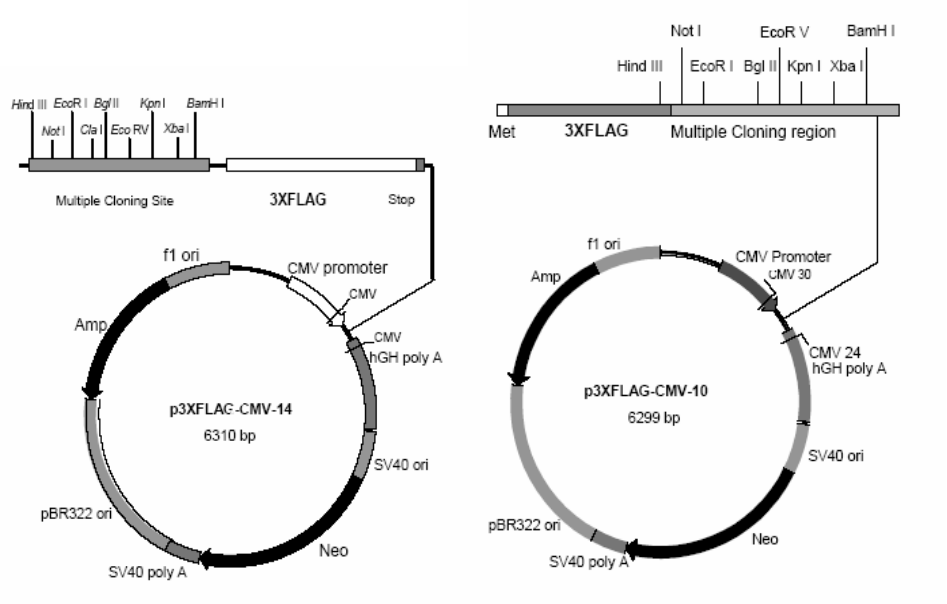

Figure2.1. The vector maps of plasmids which were used to construct the N and C-termini FLAG epitope tagged Loc115098. The p3XFLAG-CMV-14 was double digested with *Bgl*I and *Hind*III and Loc115098 was inserted between these 2 sites with the same linkers. The p3XFLAG-CMV-10 was double digested with *Xba*I and *Hind*III and Loc115098 was cloned in between these two sites with the same linkers.

The pALC plasmid was kindly provided by Dr V. Sophianopoulou's group. The plasmid was already carrying the alcA promoter cloned in the polyT site of the plasmid. Thus we inserted the AN0879.3 carrying *Nco*I linker on both sides into *Nco*I digested pALC. After the ligation and E. coli transformation steps the orientation of AN0879.3 in the plasmid was further confirmed by PCR. Later the *argB* selection gene was cloned into the same plasmid with *Eco*RI sites. After all the cloning steps this plasmid was named pALCLOC. This plasmid was transformed into each strain by the method which is explained in detail in the methods section. Following the transformation the transformants were grown on CM in Petri dishes and they were checked whether or not they were carrying our gene of interest together with alcA promoter in the correct orientation by PCR. The correct

orientation meant that the AN0879.3 was intact with the promoter, confirming that it is under the control of this promoter. Later the copy number of these transformants was detected by Southern Blot analyses and the ones carrying multi copy were further studied.

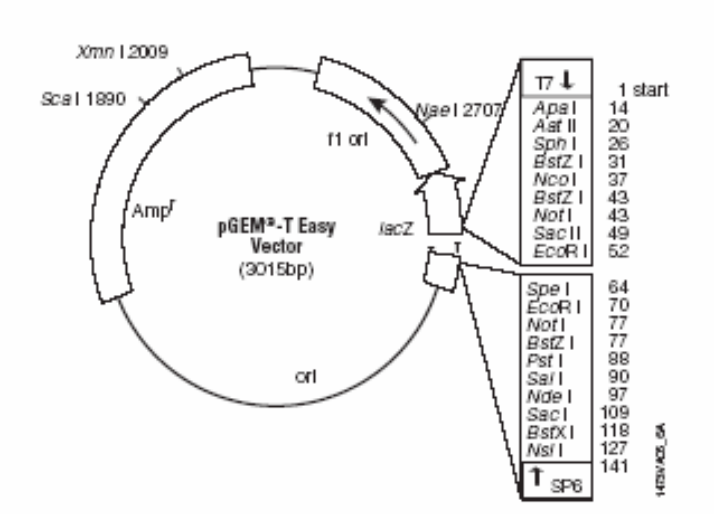

Figure2.2. The vector map of pGEM-T Easy plasmid which was used for overexpression studies.

2.2.1. Site directed mutagenesis

After cloning *A. nidulans* homologue of *loc115098* into the pBSK, the stop codon of AN0879.3 was destroyed by site directed mutagenesis by using the primers called 'McaP no-stop NcoI FOR' and 'McaP no-stop NcoI REV'. For this reaction Phusion High Fidelity DNA polymerase (Finnzymes) was used according to its manual. After the PCR amplification this product was digested by *Dpn*I enzyme for 1,5 hours. After that it was transformed into *E.coli* DH5α.

2.3. Primers

Table2.2. The primers which have been used concerning the current study.

Name	Sequence 5' →3'	Expected Size
	Locus Expression Primers	
EKYG F2	GAGGAGGAGGACTCCAAGCTC	318bp
EKYG R2	CTGGGCTTCCTCAAAGGC	
F1-ZFYGxp	ACCAAAGGCTTTTGGAGG	314bp
R1-ZFYGxp	GTTGGCTTCTTCAAACGC	
GAPDH-F	GGCTGAGAACGGGAAGCTTGTCAT	143bp
GAPDH-R	CAGCCTTCTCCATGGTGGTGAAGA	
	Locus Cloning Primers	
HALNG-F1	GAATTCAAGCTTATGCCCAAGAAGTTCCAG	672bp
HALNG-R1	GGATCCTCTAGAGCTCACTTGGGGGCATTG	
HALNG-F1	GAATTCAAGCTTATGCCCAAGAAGTTCCAG	667bp
LOC-HN-Ctrev	GATATCAGATCTGCCTTGGGGGCATTGAAGGGC	
pALCLOC Conctruction Primers		
McaP Forward NcoI	CGCC**ATG**GGCGGCAAAAAGGGC	1337bp
McaP Reverse NcoI	CGCCATGGTGCTGAGCGCGAGCTATG	
Knock-out Cloning Primers		
5' FOR KpnI	GCGGTACC-TATCGAATGGCCAATTCC	2034bp
5' REV EcoRI	GCGAATTC-AGGTGAATGAGGCTTGTC	
3' FOR EcoRI	CGGAATTC-TGAGGATTGGAAGGACTG	2149bp
3' REV XbaI	CGTCTAGA-GTGGATGATAGTGCACGA	
Pyr4 FOR EcoRI	CGGAATTC-AAGCTTAGACTTCAACAA	1859bp
Pyr4 REV EcoRI	CGGAATTC-GGATCTTCATCATTCGTC	
5 ΔMcaP Fo	GATTCGCACAGACAGCCATAG	4087bp
5 ΔMcaP Re	GTCTTCACCTCGTAGGGGATG	
3 ΔMcaP Fo	CATCCCCTACGAGGTCAAGAC	3699bp
3 ΔMcaP Re	CGCTGCATCATAGCTGAGAC	
Double-Joint PCR Primers		
CPFP4Fo	CACAAGAAACGGACAAGAAGGTCTGAAAAGCTTAGACTTCAACAACCCCACCATC	1859bp
CPFP4Re	TAAAACACAATCATTATGTAAACAAATCATGGATCTTCATCATTCGTCGCTTTCGGG	
CPF5Fo	GCAAGTCTTAGCGAGTG	3013bp
CPF5Re	GATGGTGGGGTTGTTGAAGTCTAAGCTTTTTCAGACCTTCTTGTCCGTTTCTTGTG	
CPF3Fo	CCCGAAAGCGACGAATGATGAAGATCCATGATTTGTTTACATAATGATTGTGTTTTA	3160bp
CPF3Re	GCCAGTCTGTGTGAAGT	
PNPCRFo	CGTGGAACACCTTACTC	6200bp
PNPCRRe	GAAACTCTCGCTGGACA	
pLOCGFP Construction Primers		
McaP upstream BamHI	GCGGATCCTGATCGGGGAGGGTAGCA	1985bp
McaP downstream EcoRI	CCGGAATTCTGCTGAGCGCGAGCTATG	
McaP no-stop NcoI FOR	GCGAGACTCACGAGGTCCATGGTTTGTTTACATAATGATTG	site directed mutagenesis
McaP no-stop NcoI REV	CAATCATTATGTAAACAAACCATGGACCTCGTGAGTCTCGC	
GFP NcoI linker FOR	GGCCATGGTAATGGTGAGCAAGGGCGA	745bp
GFP NcoI linker REV	CGCCATGGTTACTTGTACAGCTCGTCC	
	Conditional Knock-out Primers	
ALCMcaP For+	GACAAGCCTCATTCACCTGAATTCGCGAGCTACCATCCAATAACCCCCAGCTG	1371bp
ALCMcaP Rev+	GTTGAAGTCTAAGCTTGAATTCCGGGATGAACCGACATACAGAATAAGTGAC	

2.4. Cell Culture

2.4.1. Cell lines

In order to check the expression profile of Loc115098 several hepatic and mammary gland cell lines were used. The hepatic cell lines were HepG2, Hep40, SNU182, SNU387, SNU398, SNU449, Huh7, and Focus. The mammary gland cell lines are which were used were BT474, BT20, T47D, MCF7, 453, 468, HME, HCC, 231. The cDNAs of these cell lines were all kindly provided by Prof Mehmet Ozturk's group.

2.4.2. Growth Media

MCF-7 cell line were maintained in high glucose Dulbecco's modified Eagle's medium (Gibco) [supplemented with 10% fetal bovine serum (FBS), 1% penicillin/streptomycin (P/S) and 1% L-glutamine (Biochrom)], abbreviated as DMEM, at 37°C in a 5% CO_2 incubator.

2.4.3. Transfection

The transfection of MCF-7 cells was carried out with plasmids by using FuGENE reagent (Roche). Cells which were 90%-100% confluence in 100 mm dishes were harvested by trypsinization and washed once with DMEM (not containing the FBS, penicillin/streptomycin and L-glutamine) then resuspended in 10 ml DMEM, and diluted 1:7 in culture medium (3 ml cells in 20 ml medium), diluted cells were then seeded in 24-well plates for transfection; 400 µl cells per well.

2.4.4. cDNA synthesis and RT-PCR

The cDNAs of the cell lines were kindly provided by Dr Mehmet Ozturk's group whereas the cDNAs of *Danio rerio* (Zebrafish) were kindly provided by Dr Ozlen Konu's group and the cDNAs from human brain tissues were again kindly provided by Dr Cengiz Yakicier's group. In order to perform cDNA synthesis 2µg of RNA was used and the reaction was done according to the Revert-Aid First Strand cDNA Synthesis Kit (Fermentas) manual. The rat tissue RNAs were isolated by Dr Hani Alotaibi and the cDNA synthesis was performed by using 2µg of these RNAs. After

the PCR amplifications the PCR products were resolved on 2% agarose gels stained with ethidium bromide and visualized using the Gel Doc-2000 supported with the Multi-Analyst Ver.1.1 image analysis software (Bio-Rad).

Table2.3. The PCR Polymerase enzymes which have been used for this study.

PCR ENZYMES		
Company	**Name**	**Reagents**
Stratagene	Platinum *Pfx* DNA Polymerase	1xbuffer,0.3mM dNTPs each,1mM $MgSO_4$,0,3µM each primer,1Uenzyme
Finnzymes	Phusion	1xbuffer,200µM dNTPs each,0,3µM each primer,1Uenzyme
Finnzymes	DyNAzyme EXT DNA Polymerase	1xbuffer,360µM dNTPs each,0.3µM each primer,3Uenzyme
HyTest	HyTest Taq Polymerase	1xbuffer,0.3mM dNTPs each,1mM $MgSO_4$,0,3µM each primer,1Uenzyme
Roche	Taq DNA Polymerase	1xbuffer,0.3mM dNTPs each,1mM MgCl2,0,3µM each primer,1Uenzyme

2.5. *Aspergillus nidulans* handling

2.5.1. Strains

Several different strains were used in this study. There were some wild-type strains (*pyrG89 paba panto, pyrG89 argB2 paba panto*) as well as some cell cycle mutant strains SO6 (*nimA5, wA2, pyrG89, cnxE16, choA1, yA2, chaA1, sC12*), DBE4 (*bimE7, riboA2, pyrG89*) which were kindly provided by Dr. S. Osmani and 8-16(1-3) (*riboA1, wA3, bimD6, pyroA4, sudD7*), 8-16(3-6) (*pyrG89, [bimD6?], pyroA4, sudD7*) which were kindly provided by Dr G May and Dr S Osmani.

2.5.2. Growth Media

The complete medium (CM) and minimal medium (MM) for *A. nidulans* have been described previously by Tilburn *et al*., 1983; Cove, 1966.

2.5.2.1. Complete Medium (CM)

Vitamin solution	10 ml
Salt solution	20 ml
Casamino acids	10 ml
D-glucose	10 gr
Bacto-peptone	2 gr
Yeast Extract	4 gr
Agar	30gr (3%)

pH adjusted to 6.8 and filled with dH_2O to 1 liter

Vitamin Solution

20 mM p-amino benzoic acid, 1mM Biotin, 0,1M Nicotinic acid, 50 mM Calcium-D-pantothenate, 0,1M Riboflavin, 50mM Pyridoxine, 50mM Thiamine filled with dH_2O to 1 liter

Salt Solution

26g KCl, 26g $MgSO_4$7.dH_2O, 76g KH_2PO_4, 50ml trace elements, 2ml chloroform filled with dH_2O to 1 liter

2.5.2.2. Minimal Medium (MM)

20ml salt solution, 10g D-glucose, pH adjusted to 6.8 and filled with dH_2O to 1 liter

2.5.2.3. Fructose Medium (FM)

0,1g fructose, 2ml salt solution, 2g agar, pH adjusted to 6.8 and filled with dH_2O to 1 liter

2.5.2.4. Sucrose Medium (SM)

20ml salt solution, 342,3g sucrose, 10g glucose, 2% agar pH adjusted to 6.8 and filled with dH_2O to 1 liter

2.5.3. Growth tests

After each transformation and when the phenotypic observations were needed to be made, we performed the growth tests. Basically different strains or different transformants were inoculated on either CM or MM which contained the appropriate oxotrophies and usually grown at different temperatures in parallel to observe the changes.

2.5.4. Transformation

Transformations were made with protoplasts of mycelia from the appropriate strains. The transformation was essentially as described by Tilburn *et al*., 1983 (Tilburn et al., 1983). The routine protocol was adapted from M.Koukaki *et al*., 2003 (Koukaki et al., 2003). *Aspergillus* strains were grown on CM for 3-4 days at various temperatures depending on the strain. From these Petri dishes the conidiospores were taken by carefully scraping the dish into 5 ml dH_2O containing falcon and they were separated from mycelia by filtration through sterile blutex. Afterwards, $10^{10\text{-}11}$ conidiospores were inoculated in MM, supplemented with 5mM urea as a nitrogen source, necessary vitamins and either 5mM arginine (*argB2* strains) or 10mM uracil and uridine (*pyrG89*) or both (*argB2 and pyrG89*) for 3,5-4 (7 hours for *nimA5*) hours at the appropriate temperature at 130rpm. Upon the initiation of the germination, the conidiospores were collected with centrifugation at 4000rpm for 10 minutes and resuspended in 10 ml of the standard isosmotic buffer (Tilburn *et al*., 1983). In order to start the protoplastation, Glucanex (Novozymes) was added in the concentration of %90 (m/v) and it was incubated at 30ºC for 1,5 hour to let it act. After the completion of protoplastation conidiospores protoplasts were washed twice in 1M Sorbitol, 10mM Tris-HCl, and 10mM $CaCl_2$ pH 7.4, resuspended in the same buffer at a concentration 10^8 conidiospores/200µl and distributed into sterile eppendorfs. Each eppendorf was used for a single transformation experiment with 0.5µg of plasmids. After the mixing the appropriate plasmid with the conidia protoplasts, a drop of PEG was added on top and it was chilled on ice for 15 minutes before adding the 500µl PEG. They were incubated at room temperature for 15 minutes and then centrifuged for 10 minutes at 6000rpm. After the centrifugation they were washed with the above mentioned solution, they were again centrifuged at 5000rpm for 4 minutes and later they were plated on SM and allowed to grow for 3-4 days at various temperatures depending on the strain.

2.5.5. DNA Extraction

The DNA isolation protocol was previously described by Lockington *et al.*, 1985 (Lockington et al., 1985). The strains were inoculated and grown overnight in 50ml MM containing the appropriate vitamins and urea as a nitrogen source at distinct temperatures depending on the strain at 130rpm. The next day mycelia were collected by filtering through blutex and they were firstly kept at -80ºC for 10 minutes. Later they were homogenized by mortar and pestle in liquid nitrogen. An aliquot of 100-200mg was kept in an eppendorf and 800µl breaking solution (Tris-HCl pH8 0,2M; SDS 1%; EDTA pH8, 1mM) was added to further break the cells. Then it was vortexed vigorously and kept on ice for 20 minutes. Later phenol chloroform extraction was carried out.

2.5.6. RNA Extraction

The strains were inoculated and grown overnight in MM containing the appropriate vitamins and urea as a nitrogen source at distinct temperatures depending on the strain at 130rpm. All the materials that were used for RNA extraction were RNase free. The next day mycelia were collected by filtering through blutex and they were firstly kept at -80ºC for 10 minutes. Later they were homogenized by mortar and pestle in liquid nitrogen. An aliquot of 100mg was kept in an eppendorf and 1ml of trizole was added together with 10mg of glass beads. It was vortexed vigorously and kept at room temperature for 10 minutes. Then it was centrifuged at 4ºC for 15 minutes at 12000rpm. The supernatant was taken into a clean eppendorf and 200µl of chloroform was added, it was shaken by hand and incubated at room temperature for 2 minutes. The centrifugation was repeated with the same conditions. After the centrifugation the upper phase was transformed into a new clean eppendorf. 500µl isopropanol was added on top of it and it was chilled on ice for 10 minutes after shaking the eppendorf well. After the centrifugation at 4ºC for 10 minutes at 12000rpm the supernatant was discarded and the pellet was washed by 1 ml of 75% Ethanol. After the washing step the pellet was dried and resuspended in DEPC treated dH_2O.

2.5.7. Southern Blotting

The nucleic acid manipulations were done according to the previously prepared protocols by Sambrook *et al.* After the DNA extraction step, the concentration was measured by Optic Density (OD) measurements. 2,5-3µg of DNA was digested by EcoRI for 2 hours at 37ºC and then more enzyme was added and the digestion was carried out overnight. The next day 1% agarose gel was prepared to run the digested fragments in 1xTAE (242g Tris base, 57.1ml Glacial Acetic Acid, 18.6 g EDTA in 1 lt dH_2O). The electrophoresis was done at 58V for 4 hours. After the completion of the electrophoresis the gel picture was taken. The gel was baked at 100% Ultraviolet light (254nm) for 5 minutes and the irrelevant parts of the gel was cut and discarded. Then the baked gel was incubated in denaturing buffer (0,5M NaOH and 1,5M NaCl) for half an hour. Then the gel was incubated in the neutralizing solution (1M Tris-HCl pH 8 and 1,5M NaCl) for half an hour. After these steps the overnight capillary transfer was carried out in alkaline conditions using the sandwich method. It is called sandwich because the gel and the membrane are kept together between two layers of Whatman paper. It is necessary for the capillary movement. 20XSSC (0,3 M $Na_3C_6H_5O_7.2H_2O$ and 3M NaCl pH 7) was used for the transfer. In the sandwich method, firstly the long Whatman No:2 papers were placed in such a manner that both ends of the paper could be deep in the transfer buffer. The gel was placed on top of the Whatman papers. The edges of the gel were wrapped by plastic wrap so as to let the capillary movement continue through the gel and the membrane. On top of the gel the nylon membrane was placed and it was made sure that there were no air bubbles left in between the gel and the membrane. 2 pieces of Whatman paper were cut in the same size as the membrane and wetted in 2X SSC then they were placed on top of the membrane. The air bubbles were smoothened out by a glass rod. A stack of paper towels (5-8 cm) were placed on the Whatman papers and lastly a 500-g weight was placed on top of all. The tissue papers were changed as they became wet. The transfer was carried out for at least 20 hours. Following the transfer the membrane was cross linked by UV light. Then it was ready for hybridization.

For the hybridization step Church buffer (0,15M BSA; 0,5M orthophosphate; 1mM

EDTA; %20 SDS and dH_2O) was used. After placing the membrane in a cylinder with 100ml of Church buffer the pre-hybridization step took place where the membrane was incubated at 65ºC for 2 hours. Meanwhile the [^{32}P]dCTP labeled probe was prepared using a random hexanucleotide-primer kit following the supplier's instructions (Promega). After the prehybridization, the probe was added in the cylinder and the hybridization was initiated by leaving the membrane for incubation at 65ºC overnight. The next day probe containing church buffer was removed and the washing steps were carried out. The washings were as following: 30 minutes at 65ºC in 2xSSC, 0.1% SDS twice and in the cylinder. After that 30 minutes at 65ºC in 1xSSC, 0.1% SDS in a plastic container. Once the washing was completed the membrane was placed in a light-proof cassette with a film on and it was kept at -80ºC for a period ranging from 3hours to 7 days (depending on how strong the probe is). The development was made manually by using developer and fixer solutions.

2.5.8. Northern Blotting

The nucleic acid manipulations were done according to the previously described protocols by Sambrook *et al.* Once the RNA extraction was completed the OD measurement was carried out. All the materials were either RNase free or made RNase free manually by adding DEPC treated water or NaOH washings when necessary. Firstly the running buffer (10mM orthophosphate) was prepared which was used both for electrophoresis and the preparation of the agarose gel. 10µg RNA was mixed with the following mixture; for x volume of RNA, 3x volume of 11,5µl DMSO; 2,25µl (0,1M pH7) orthophosphate; 3,3µl glyoxal. This mixture was incubated at 50ºC for one hour and then on ice 5 minutes. After the 1% agarose gel was ready, and the incubation on ice was over, the samples were ready to be loaded after adding the dye (10mM orthophosphate, 2%Bromophenol blue, 50% glycerol) to a final ratio of 1/8 of the total volume. After the loading of the samples, the electrophoresis was carried out. The voltage was adjusted according to the distance between the two electrodes; it was carefully calculated to 3V per centimeter. After letting the samples run for 5 minutes, the peristaltic pump was started to keep the currency stable (maximum 60mA). The samples were allowed to run for 5 hours and

once the electrophoresis was over the transfer was ready to be started. The sandwich model was prepared as described in the 'Southern Blotting' section. The transfer was carried out in 20X SSC overnight. Next day the membrane was baked under UV for 5 minutes and then it was stained with methylene blue dye (0,5M NaOAc pH: 5.2 with 0,03% methylene blue) for 2 minutes and a picture was taken. After that the membrane was soaked into the destaining buffer (1% SDS; 0,2X SSC) until there was no stain left on the membrane. After the destaining, the membrane was ready to be hybridized. The hybridization was the same as it was described in 'Southern Blotting' section. The washings were as following: 10 minutes at 65ºC in 100 ml of 2xSSC, 0.1% SDS twice and in the cylinder. After that, 20 minutes at 65ºC in 200 ml of 1xSSC, 0.1% SDS in a plastic container. Once the washing was completed the membrane was placed in a light-proof cassette with a film on and it was kept at -80ºC for a period ranging from 3hours to 7 days (depending on how strong the probe is). The development was made manually by using developer and fixer solutions.

2.6. Immunofluorescence (IP)

The localization studies which were performed with MCF-7 cell line were carried out as following: Firstly the cells were grown on cover slips in 8-well plates and when they became confluent the cover slips were removed from the cell culture dish and they were treated with 4% paraformaldehyde at room temperature for 15 minutes. After that the cells were treated with 1:1000 TritonX-100 in PBS for 6 minutes at room temperature. The blocking was performed by blocking solution (5% BSA in PBS) for 1 hour at room temperature. Later ANTI-FLAG M2 monoclonal antibody (Sigma) was applied to these cover slips. They were kept in this antibody diluted in blocking solution (20µg/ml) for one hour in a dark place. Afterwards they were washed 3 times in 1xPBS. Then they were kept in anti-Texas Red antibody which was diluted in blocking solution (1:200) for one hour at room temperature in a dark place. The washing step was repeated and then the counter stain HOESCHT (1:1000) was applied for 5 minutes. Following the washing step the cover slips were mounted by glycerol and they were ready for visualization.

2.7. Microscopy

In the subcellular localization studies which were performed by using *A.nidulans* the samples were prepared as following. Initially the cover slips were washed in nitric acid, water and aceton and later poly-L-lysine hydrobromide (Sigma) was applied on each cover slip. Later two of these clean and poly-L-lysine hydrobromide treated cover slips were placed in sterile Petri dishes. Following the addition of the medium into the dish, the appropriate strains were inoculated and they were allowed to grow at either their permissive or restrictive temperature. When the strains were grown according to their growth rate, the cover slips were removed from the petri dish and the fixation and staining step took place according to Harris *et al*., 1994 (Harris et al., 1994) and May *et al*., 1992 (May et al., 1992b). After removing the cover slips from the petri dish firstly the fixation buffer (3.7% formaldehyde, 50mM phosphate buffer pH 7, 0.2% TritonX-100) was applied to the cover slips. They were incubated at room temperature in a dark for 20 minutes and then they were then washed in dH_2O, later the stains were applied. For the staining of the nucleus DAPI (4,6-diamidino-2-phenylindole, dihydrochloride) and for the staining of the cell wall calcofluor were used. They were incubated in these dyes for 10 minutes and then rinsed in dH_2O. The cover slips were then mounted on slides which had 20µl mounting buffer (10% phosphate buffer, 50%glycerol and dH_2O) and they were ready for microscopic analyses. The photographs were taken by Axionplan Zeiss phase-contrast fluorescent microscope and appropriate filters.

The photographs of MCF-7 cells were taken by AxioCamMRm Zeiss fluorescent microscope.

3. RESULTS

3.1. Computational Data about Loc115098

3.1.1. The identification of loc115098

As it was explained in the introduction section, Loc115098 was firstly identified in an unrelated study where *cis*-acting transcriptional elements of the NIS (Sodium-Iodide Symporter) gene were being analyzed. The focus of that study was to analyze a putative transcriptional control element localized in the 3' downstream of NIS gene by using a tool called VISTA (Negre et al., 2005). Loc115098 was one of the 10 conserved putative regions identified. (Alotaibi *et al*, unpublished data). (See Figure 3.1)

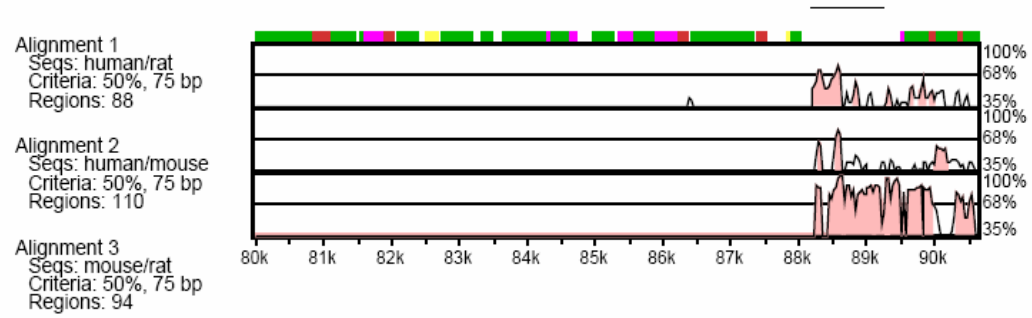

Figure 3.1. VISTA plot of conservation levels in a 90 kbp genomic DNA in human, mouse and rat. Only the 10th region is shown here. Percent nucleotide identities between human, mouse and rat DNA sequences are plotted as a function of position along the human sequence. Peaks of evolutionary conservation in overlapping exonic sequences are shaded blue. Aligned regions of more than 50% identity over 75 bases are shaded pink. (Alotaibi *et al*., unpublished)

3.1.2. Features

According to our computational analyses, the human Loc115098 is located in chromosome 19p22. The gene consists of 4 exons and the predicted protein is 223 amino acid-long. It has rather an unusual amino acid composition and it is very hydrophilic. 42% of the protein consists of charged amino acids such as Arginine, Lysine, Aspartate and Glutamate. Therefore it is a very basic protein (pI=10). It bears no domain similarities with any previously known proteins. The conservation of Loc115098 is remarkably high. Our *in silico* analysis through the databases has proved that it is conserved in the eukaryotic kingdom with the exception of the budding yeast *Saccharomyces cerevisiae*. Although the similarity ratios differ, there is more than 75% identity in mammals, 50% identity with insects and nematodes and 30-40% identity with plants and ascomycetes.

In the simplest possible model organism *A.nidulans*, the gene is located on the 8th chromosome, Contig 14: 46513-47339 +. It has the ID number of AN0879.3 and can be reached at the Broad Institute web page with the same accession number. (http://www.broad.mit.edu/annotation/genome/aspergillus_nidulans/Home.html).
The Aspergillus and human homologue of this protein had a 34.4% identity. However the Aspergillus homologue was 222-amino acid long but the composition of the amino acids showed similar features. (See Figure 3.9)

3.1.3. Post Translational Modifications

Aiming to better understand this unusual protein carrying no domain similarities with other proteins, we also searched for the post translational modifications through Expasy/Prosite webpage (http://www.expasy.org/prosite/). The results of this study indicated that both the human and the Aspergillus homologues had several predicted post translational modifications. Both of them had predicted Tyrosine kinase, protein kinase C, casein kinase II, cAMP and cGMP dependent protein kinase phosphorilation sites, N-myristoylation sites. In addition, the Aspergillus copy has an amidation site and an alanine rich region in the C-terminus and the human copy has

an extra tyrosine sulfation site. Moreover, they both carry bipartite nuclear targeting sequence in the C-terminus of the protein. All of these post translational modification predictions might be very important and we intend to further investigate these modifications in near future.

3.1.4. Biological Significance of Loc115098

Our bioinformatics analyses have indicated that *loc115098* is an evolutionarily conserved gene suggesting that it exists in all eukaryotic kingdom. We believe that, this high percentage of conservation might be pointing at a biological significance. All genes which are very well conserved during evolution have important functions in crucial biological events. Among the most conserved genes, the histones are best known examples. 16S RNA which is a subunit of ribosomal RNA is another example of the conserved genes which also has an essential role for the organism. Along with this proof we hypothesize that *loc115098* could be playing an important role in one of the well conserved biological mechanisms such as mitosis.

3.2. Analysis of Loc115098 Expression in Different Tissues

After the identification of *loc115098*, the aim of the study was initially to confirm that *loc115098* is expressed in different tissues of different organisms. For this purpose several tissues were investigated whether or not they expressed the *loc115098* gene.

3.2.1. Breast cell lines

Firstly, the expression pattern of cDNAs of hepatic cell lines was analyzed. Here we confirm that the appropriate primers are able to amplify the *loc115098.*

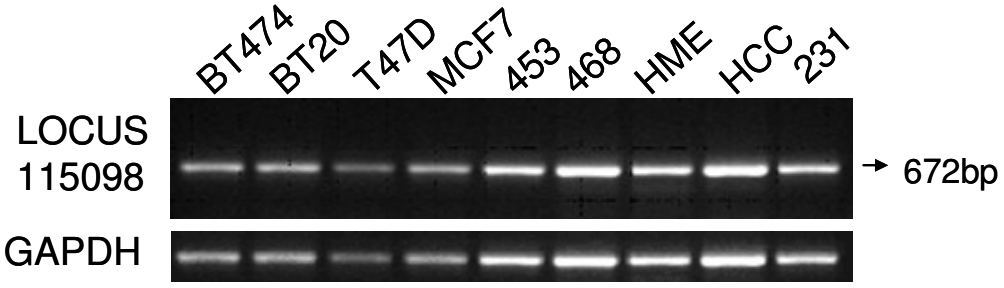

Figure3.2. The mammary gland cell lines were grown in DMEM with the appropriate supplements and RNA was prepared. Then the expression of *loc115098* was evaluated by RT-PCR. GAPDH was used as an internal control to verify the equal loading of wells. It was confirmed that Loc115098 is expressed by all mammary gland cell lines analysed.

3.2.2. HCC cell lines

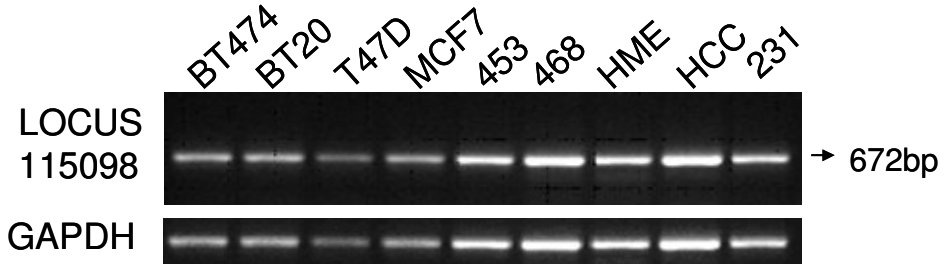

Figure3.3. The hepatic cell lines were grown in DMEM with the appropriate supplements and RNA was prepared. Then the expression of *loc115098* was evaluated by RT-PCR. GAPDH was used as an internal control to verify the equal loading of the synthesized cDNA. It was confirmed that Loc115098 is expressed by all hepatic cell lines analysed.

3.2.3. Human brain tissues and Mouse tissues

After the confirmation of Loc115098 expression in several breast and HCC cell lines, we also analyzed the expression profile of Loc115098 in human brain tissues, kindly provided by Dr Cengiz Yakicier's group. Meanwhile with the same purpose we synthesized cDNAs of the already existing RNAs which were extracted from several rat tissues by our group. The results indicated that these tissue types also had a persisting Loc115098 expression. (See Figure 3.4)

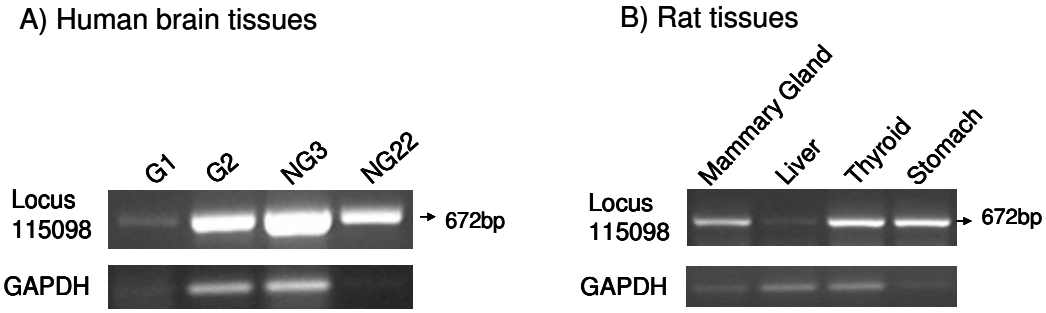

Figure3.4 A) The cDNAs from human brain tissues indicate the persistent expression of Loc115098. B) RNA was isolated from the mammary gland, liver, thyroid and stomach tissues were isolated of rat and the expression of *loc115098* was analysed by RT-PCR. GAPDH was used as an internal control to verify the quality of the synthesized cDNA. It was confirmed that Loc115098 is expressed in all tissues. (MG: Mammary gland).

3.2.4. Zebrafish tissues

To further extend our expression studies on Loc115098, we also designed primers to confirm Loc115098 expression in zebrafish tissues. The cDNAs and Mipep primers were kindly provided by Dr Ozlen Konu's group. Mipep is a gene which is used as an internal control in this study. cDNAs were synthesized from different growth stages as well as from several tissues. Our results indicated that Loc115098 is expressed in pectoral fin, dorsal fin, ovary, but not in heart, gills, muscle and liver. Following zebrafish through embryonic development, we observed the first Loc115098 expression at the 24^{th} hour post fertilization whereas later the expression could not be observed neither at the 48^{th} hour nor on the 3^{rd} day post fertilization however it could again be observed on the 4^{th}, 5^{th} and 6^{th} day post fertilization.

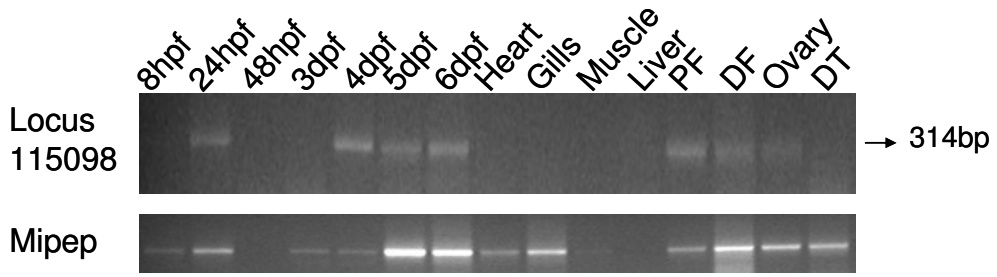

Figure3.5. The expression of *loc115098* was evaluated by RT-PCR. Mipep was used as an internal control to verify the quality of the synthesized cDNA. (8hpf:8th hour post fertilization, 4dpf: 4th day post fertilization, PF: pectoral fin, DF: Dorsal fin, DT: Ductal track)

3.3. Subcellular Localization of Loc115098 in MCF-7 Cell Line

Once we confirmed that Loc115098 has a persistent expression profile in different cell lines and tissues, we set a new goal to identify the subcellular localization of Loc115098 protein. For this reason we firstly prepared two constructs, one carrying a C-terminus and the other N-terminus flag epitope tagged Loc115098. In these two different plasmids, Loc115098 was cloned either in the C-terminus or in N-terminus of 3XFLAG protein expressing sequence. Then we transfected these constructs into MCF-7 cell line, separately.

3.3.1. Construction of pFLAGLOC and pLOCFLAG plasmids

The p3XFLAG-CMV-10 was used for N-terminus tagging of Loc115098. It was firstly double digested by *Xba*I and *Hind*III enzymes. The Loc115098 was amplified by the primers called HALNG-F1 and HALNG-R1 and these two fragments were ligated in appropriate conditions and the transformation was carried out. The p3XFLAG-CMV-14 was digested by *Bgl*I and *Hind*III enzymes and the PCR product of Loc115098 was obtained by using the primers called HALNG-F1 and LOC-HN-Ctrev. Later the same ligation and transformation protocol was carried out. Several enzyme digestions were done to confirm the correct plasmid. (See Figure 3.6)

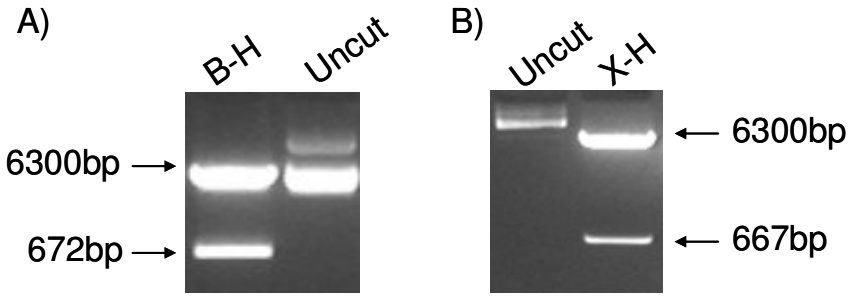

Figure3.6. The plasmids which were used to construct the N and C-termini FLAG epitope tagged Loc115098. In figure A, the p3XFLAG-CMV-14 was double digested with *Bgl*I and *Hind*III and Loc115098 was inserted between these 2 sites with the same linkers. In figure B, the p3XFLAG-CMV-10 was double digested with *Xba*I and *Hind*III and Loc115098 was cloned in between these two sites with the same linkers.

3.3.2. Ectopically expressed loc115098 is likely to involve in mitosis

When N and C-termini FLAG tagged Loc115098 was transfected into the MCF-7 cell lines, together with the HOESCHT-33252 staining we observed that Loc115098 protein is very likely to involve in cell cycle regulation, because as shown in the figure 3.7 when cells are not dividing Loc115098 is localized in the cytoplasm whereas it is relocated to the nucleus upon the initiation of mitosis. In this figure it is clearly observed that the cells which are pointed with arrows in A' representing those that express N-terminus tagged *locus115098*, the localization is mostly cytoplasmic with a much diffused pattern. Arrows in panel B' indicate cells in mitosis and unlike non-dividing cells, Loc115098 is observed in the nucleus. (See Figure 3.7)

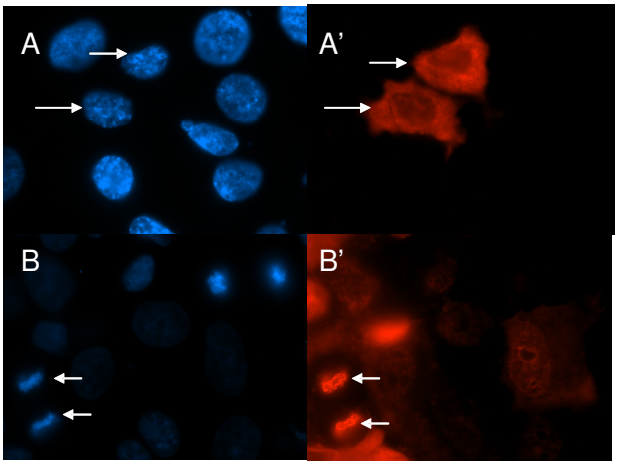

Figure3.7. Plasmids carrying N-terminus flag-epitope tagged Loc115098 were transfected to MCF-7 human mammary cells. Nuclear structures were monitored by epifluorescence microscopy using DNA intercalating Hoescht-33258 (Sigma) blue fluorescing dye (Panels A, B). Transfected cells were analyzed by immunofluorescence by using texas-red (red fluorescing dye) labelled flag-specific antibodies (Panels A' and B').

3.3.3. Localization studies with anti-Loc115098

Another construct was prepared by inserting *loc115098* with *BamHI* and *HindIII* into a plasmid called pcDNA3.1/*myc*-His C (Invitrogen) and we made further localization analyses but this time with anti-Loc115098 antibody which was designed for the first 24 amino acids in the whole sequence. In this study our results indicated that Loc115098 is almost always cytoplasmic and even though there were cells going through mitosis, the localization was persistent and we never observed it in the nucleus. Instead in the cells which were going through mitosis the localization was around the nucleus. Interestingly enough, the staining of Loc115098 was observed in the centrioles as well. This data does not support the previous one, however it is likely to be more reliable because the staining was carried out by an antibody specific to Loc115098 itself. (See Figure 3.8)

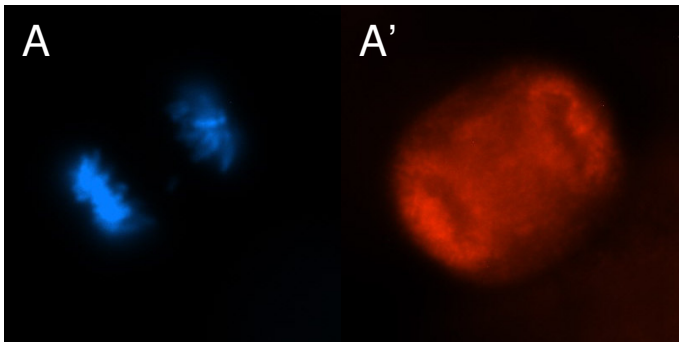

Figure3.8 Plasmid carrying Loc115098 was transfected to MCF-7 human mammary cells. Nuclear structures were monitored by epifluorescence microscopy using DNA intercalating Hoescht-33258 (Sigma) blue fluorescing dye (Panel A). Transfected cells were analyzed by immunofluorescence by using texas-red (red fluorescing dye) labelled Loc115098-specific antibody (Panel A’).

3.4. *Aspergillus nidulans* Expression Data

3.4.1. Human and Aspergillus protein alignment confirms the conservation

After we observed the localization of Loc115098 in the nucleus we set new goals for our project. Since A. nidulans is an easily handled organisms we decided to carry on our studies with this lower eukaryote to better understand the function of Loc115098 in cells. Firstly the conservation between these two proteins was checked and 34.4% identity was discovered, indicating a good conservation. (See Figure 3.9)

```
Aspergillus  MGGKKGGGEN SKKAAGNARK AEAAANKKAI EDQKRAAEED KQWAKGA.KS
      Human  .MPKKFQGEN TKSAAARARR AEAKAAADAK KQKELEDAYW KDDDKHVMRK
  Consensus  .ggKKgqGEN sKkAAarARr AEAaAaadAi e#qeraaaed K#daKga.rk

Aspergillus  SSKKEEAEAK KAEAARKKAE RDALLAAEEA SQPS..KPKN NKSAAKKNAP
      Human  EQRKEEKEKR RLDQLERKKE TQRLLEEEDS KLKGGKAPRV ATSSKVTRAQ
  Consensus  eqrKEEaEar ra#aarrKaE r#aLLaaE#a kqkg..aPrn akSaakkrAq

Aspergillus  SRGTLNLD.Q LDDAPSSRAS A......... ..LNASGIDN AL.......D
      Human  IEDTLRRDHQ LREAPDTAEK AKSHLEVPLE ENVNRRVLEE GSVEARTIED
  Consensus  irdTLrrD.Q Lr#APdsaak A......... ..lNargi## al.......D

Aspergillus  ALSLTSKDTS KVDRHPERRY KAAYAAFEAR RLPEIEAENP GLRRQQRIEL
      Human  AIAVLSVAEE AADRHPERRM RAAFTAFEEA QLPRLKQENP NMRLSQLKQL
  Consensus  AiallSkaee aaDRHPERRm rAA%aAFEaa rLPrieaENP n$RrqQri#L

Aspergillus  VKKEFDKSPE NPFNQVHVAF DASKEEIAAV REAERKKTEA RLTR
      Human  LKKEWLRSPD NPMNQRAVPF NAPK...... .......... ....
  Consensus  lKKEfdrSP# NPmNQraVaF #ApK...... .......... ....
```

Figure3.9 The human Loc115098 was aligned with its homologue in *A.nidulans* by a commonly used alignment tool ClustalW. Here we observe the high conservation between these two organisms.

3.4.2. Different strains of Aspergillus nidulans indicate different expression patterns

Several *A.nidulans* strains were used in this study. These strains had different growth requirements and expression levels.

3.4.2.1. The growth difference between the wild type and cell cycle mutant strains

The wild-type strains grow normally at 37ºC whereas the cell cycle mutant strains have restrictive and permissive temperatures. At their permissive temperatures they grow normally, however when they are grown at restrictive temperatures the cell cycle gets arrested at different stages depending on the strain. Before starting our studies we wanted to confirm this data.

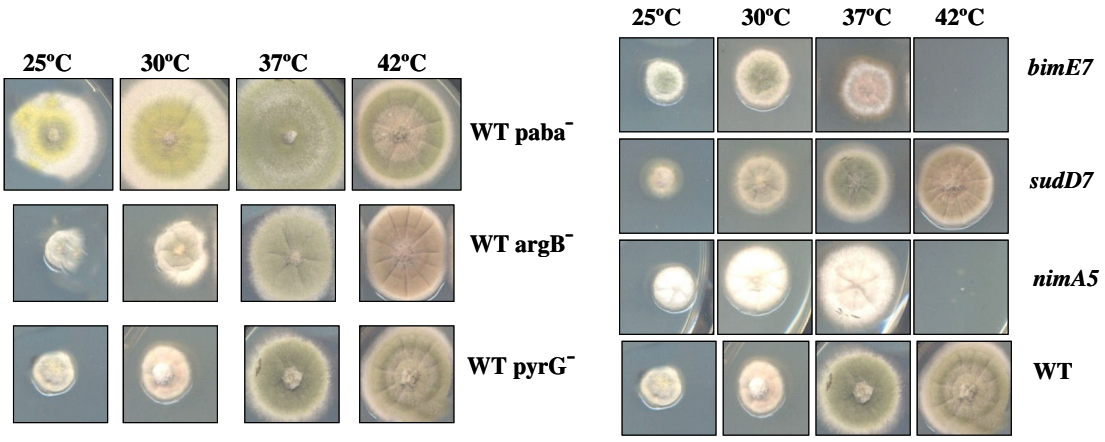

Figure3.10. All the *A.nidulans* strains were grown on complete media. The *bimE7* and *NimA5* strains grow normally at 30ºC and are not viable at 42ºC which is their restrictive temperature. The *sudD7* grows normally at 37º and very poorly at 25ºC. (WT: wild-type)

3.4.2.2. The Northern Blot Assays confirm the different expression pattern at different temperatures

After observing the growth difference of several different strains at various temperatures, the AN0879.3 expression was further analyzed. In order to understand its expression pattern, the Northern blot analyses were performed (see Figure 3.11A). Before the hybridization, the methylene blue staining was performed in order to confirm the equal loading of the samples (see Figure 3.11B). The AN0879.3 probe was utilized in order to hybridize the isolated RNAs. While preparing the probe, the ORF (open reading frame) of the gene was taken into consideration. Our results consistently indicate that the cell cycle mutant strains have a higher expression profile at their restrictive temperatures. The restrictive temperatures of these strains represent a mitotic arrest at different stages of the cell cycle. This result has also been confirmed by densitometric analyses of the bands (See Table 3.1). In this analysis, firstly the density of the bands corresponding to each strain was measured by using the software Dolphin-1D, Wealtec Corp., and in the table it is represented as 'Raw Data'. The ratio of each strain in its permissive and restrictive temperature was

indicated as 'Fold Change.' For the measurement of equal loading, the rRNAs were stained by methylene blue; with the same method the density of each band corresponding to the 18S rRNA was measured. Later the 'Fold Change' was corrected by finding the fold difference between the density of the Northern Blot bands and 18S rRNA bands. This was called 'Corrected Fold Change.' Together with the densitometric analyses it is very likely that the AN0879.3 expression increases depending on the cell cycle arrest at various stages whereas in the wild type strain the corrected fold difference is less than one. In other words, when the cells are arrested at different cell cycle stages depending on the mutations they carry the AN0879.3 expression increases accordingly.

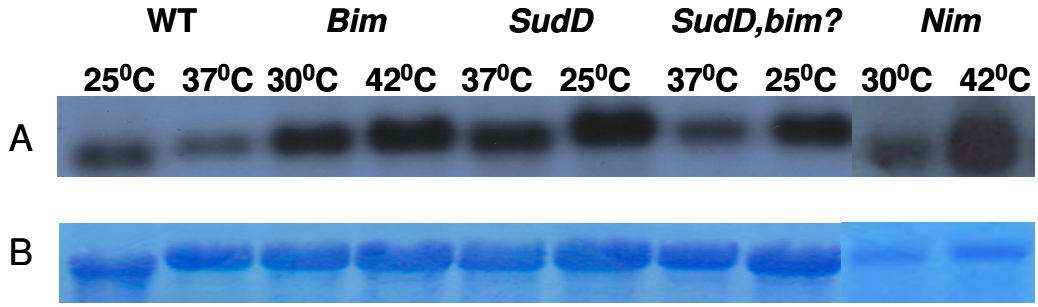

Figure3.11. In this Northern blot analysis RNA isolated from different *A. nidulans* strains are probed by AN0879.3 and their expression profiles are compared. A: Probed samples are indicated. B: Methylene Blue Staining was done for the confirmation of equal loading. The expression profile of the cell cycle mutants differ dramatically upon change of temperature when compared to wild-type strain. It is speculated that mitotic arrest causes this overexpression of AN0879.3, indicating a possible link between this protein and the process of mitosis.

Table 3.1 The densitometric analyses of Northern Blot results.

Strains	Raw Data	Fold Change	Corrected Fold Change
WT 25	24,875		
WT37	13,397	0,538572864	0,439132729
BimP	35,464		
BimR	47,717	1,345505301	1,139749686
SudDP	46,335		
SudDR	64,436	1,390655012	1,145929928
SudD,bim?P	26,207		
SudD,bim?R	52,841	2,016293357	1,262465523
NimP	0,352		
NimR	10.102	28698,86364	13205,4811

3.5. Subcellular Localization of AN0879.3 in *Aspergillus nidulans* mycelia

Our next goal was to analyze the subcellular localization of AN0879.3 in *A.nidulans* strains. For this purpose we used *NimA*5, *BimE*7, *SudD*7 and wild type control strains, the cell cycle mutants were specifically important for the study because a change in the localization of AN0879.3 during different stages of cell cycle could have been very important and crucial for our studies because our hypothesis was that AN0879.3 was somehow involved in mitosis. On the other hand this would be a good control of our previous localization studies as we would be able to compare the subcellular localization of the protein in lower and higher eukaryotes.

3.5.1. The subcellular localization of AN0879.3 in several strains of A. nidulans

For this experiment, the plasmid which was planned to be used in this study was constructed firstly by inserting a AN0879.3 fragment with *Bam*HI and *Hind*III poly linkers into pBSK which was already digested with the same enzymes. Later the stop codon of AN0879.3 was destroyed and an *Nco*I site was inserted into the sequence as described in the methods. The GFP sequence was amplified and inserted into this plasmid with an *Nco*I linker. Eventually we obtained the plasmid called pLOCGFP which carried GFP in the C-terminus of AN0879.3. This plasmid was transformed into each strain by the method which is explained in detail in the methods section. Following the transformation, the strains were grown on CM in Petri dishes and they were checked whether or not they were carrying our gene of interest together with GFP. Later the copy number of these transformants was detected by Southern Blot analyses and the ones carrying single and multi copy were further studied. The results were the same for both and they were both promising. (See Figure 3.15)

For microscopic analyses the conidiospores were inoculated in MM growing on poly-L-lysine coated cover slips. When they were forming mycelia the growth was terminated and the cover slips were removed from the Petri dishes to perform fixation and further staining. The results we obtained from epifluorescent studies

indicated that AN0879.3 was almost always cytoplasmic independently of the cell cycle regulation. It was mostly located in the cytoplasm however rarely observed around the nucleus of cell cycle arrested cells. There was an obvious change in the localization characteristics of Loc115098 in two different organisms and possible reasons causing this outcome will be discussed in the discussion part.

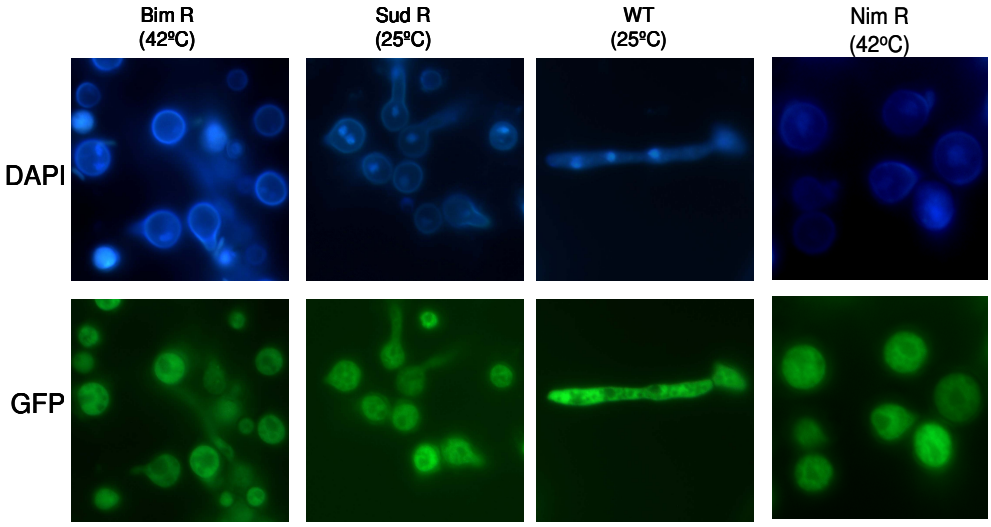

Figure3.12. The pLOCGFP transformed strains are grown in minimal media and on cover slips. The pictures which were taken by epifluorescent microscope indicate the cytoplasmic localization of AN0879.3. The temperatures are the restrictive temperatures of the mutants and 25ºC is the optimal condition for GFP observation in wild-type strains.

3.6. Overexpression Studies in *Aspergillus nidulans*

We also aimed to check the possible effects of AN0879.3 overexpression in wild-type strains. For this purpose we constructed a plasmid in which the AN0879.3 was under a controllable alcA (alcohol dehydrogenase) promoter. This method was designed according to Felenbok *et al*., 2001 (Felenbok et al., 2001). Upon the addition of ethanol in the medium, the alcA promoter gets activated and the AN0879.3 expression increases whereas the presence of glucose in the medium represses the promoter and the AN0879.3 expression is turned down. Thus, our aim was to observe any phenotypical changes due to overexpression of AN0879.3.

3.6.1. The outcomes of overexpression of AN0879.3

Firstly we designed a construct to transform the strains. The pALC plasmid was kindly provided by Dr V. Sophianopoulou's group. After all the cloning steps this plasmid was named pALCLOC (see Figure 3.13). This plasmid was transformed into each strain by the method which is explained in detail in the methods section. Following the transformation the transformants were grown on CM in Petri dishes and they were checked whether or not they were carrying our gene of interest together with alcA promoter in the correct orientation by PCR. The correct orientation meant that the AN0879.3 was intact with the promoter, confirming that it is under the control of this promoter. Later the copy number of these transformants was detected by Southern Blot analyses and the ones carrying multi copy were further studied. The results represent the multi copy transformants. We chose the multi copy transformants on purpose because our aim was to observe the possible outcomes of the overexpression of AN0879.3. Thus the more copy numbers meant more expression.

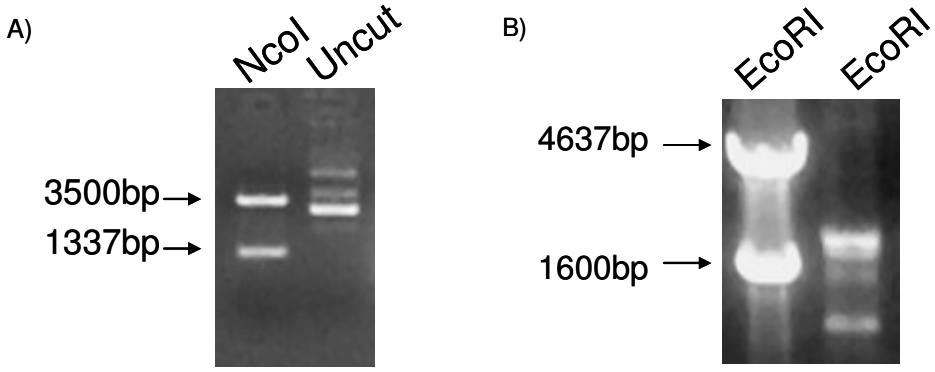

Figure3.13. The confirmation of pALCLOC construct. In figure A, the 1337bp long fragment represents the insert. The uncut plasmid is loaded into the gel together with the digested plasmid. In figure B, the plasmid containing 1600bp long *argB* selection gene is digested by *EcoRI* indicates the insert. An empty vector which is digested with the same enzyme is present as a control.

There were two wild-type strains which were used for this study; their genetic background was as following: *pyrG89, paba, panto* and *pyrG89, argB, paba, panto*. Following the transformation and isolation of *pyrG89*-positive and *argB*-positive

transformants, the growth tests were carried out in order to see the possible effects of overexpressed AN0879.3 in these strains. The overexpression results indicated no phenotypic change of these strains. The possible reasons and future perspectives about this result will further be discussed in the discussion part.

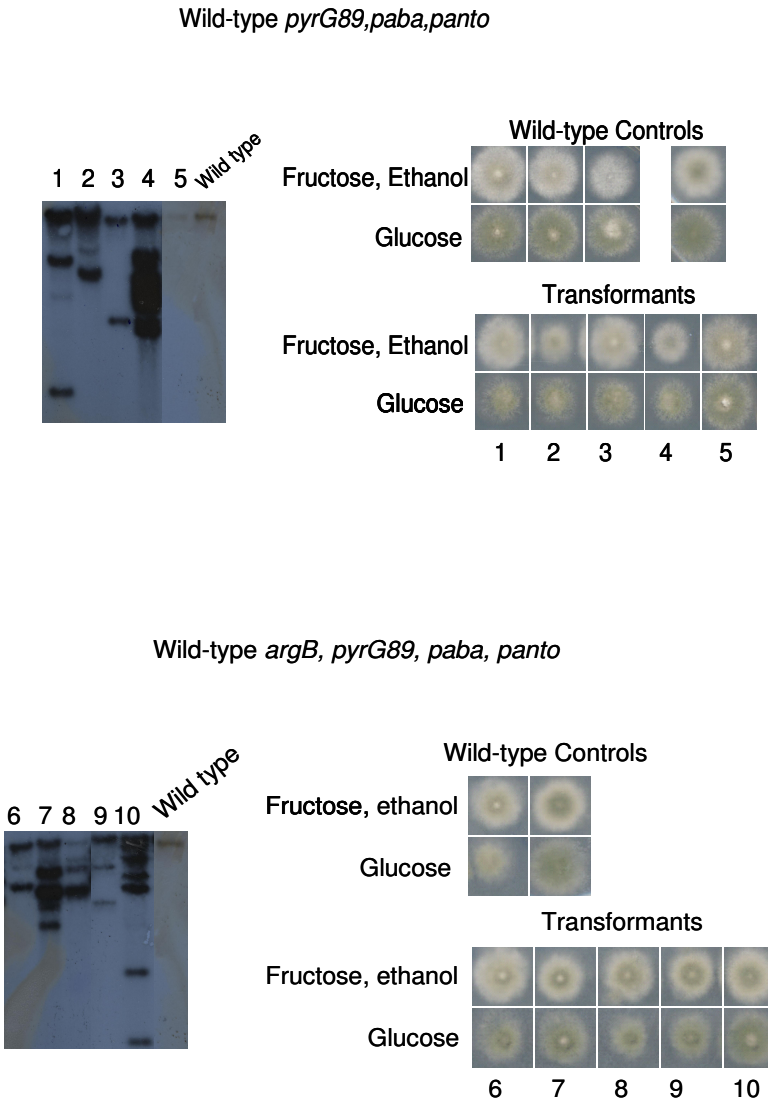

Figure.3.14. The wild-type, *pyrG89, paba, panto* and wild-type *argB, pyrG89, paba, panto* strains were transformed with pALCLOC plasmid and the transformants were confirmed by Southern blot analyses. The growth tests were performed in order to observe the possible outcomes of AN0879.3 overexpression. No dramatic phenotype change was observed according to the growth tests. (Growth tests were conducted in the same petri dish).

We also conducted some experiments to measure the growth rate of these strains together with a wild-type control. We initiated the experiment with the same conidia number for each strain by counting the conidia in the heamocytometer. After that, every 2 hours we checked the conidia number under the microscope for each strain. This has not revealed a significant growth rate difference between the AN0879.3 overexpressing strains and the strains carrying a single copy of AN0879.3.

3.7. Knock-out Studies of AN0879.3

One of the most important goals for this project, and the main reason why *A. nidulans* was used as a model system in this study, was the generation and characterization of mutants lacking functional AN0879.3. A sophisticated method called double-joint PCR was initially utilized for the preparation of linear AN0879.3 knock out construct. Later on, a cloning strategy was followed in order to make the constructs necessary for the investigation of strains where AN0879.3 is conditionally knocked-out.

3.7.1. Strategy for the knock-out construct by Double-Joint PCR

The technique was adapted from Jae-Hyuk Yu *et al.*, 2004 (Yu et al., 2004). It is schematically explained in the Figure 3.15. In order to knock out the AN0879.3, Firstly, different primers (A, B, C, D, X, Y) were designed to amplify the 5' AN0879.3, 3' AN0879.3 and *pyr4* gene separately. This 1st round PCR was performed according to the protocol which was as follows: 1xPhusion HF buffer, 200µM of each dNTP, 0,5µM of each primer, 1U of Phusion High-Fidelity DNA Polymerase (Finnzymes), the reaction was performed in 50µl (98ºC 30seconds, 35 cycles of [98ºC 15 seconds, 54ºC 30seconds, 72ºC 2,5 minutes] and 72ºC 10 minutes). In this case the *pyr4* is the selection gene which was obtained from *Neurospora crassa* genome. It is selected for the complementation of *pryG89;* these two genes have different DNA sequences however the proteins are capable of complementing each other. Some of these primers (C, D, X, Y) had flanking

sequences which are shown in different colors in the figure. Once these three fragments were obtained separately, a new PCR amplification which we called 2nd round PCR was set to obtain the knock out construct that bears 5' AN0879.3, 3' AN0879.3 and *pyr4* gene**.** In this second round PCR, the flanking regions of the previously amplified fragments served as primers which maintained the annealing of these 3 fragments and following the annealing step the amplification was carried out in the same reaction with the primers called CPF5Fo shown as A and CPF3Re shown as B in the schematic explanation. The condition for this reaction was as following: 1xEXT buffer, 360µM of each dNTP, 1µM of each primer, 3 U of DyDNAzyme EXT DNA Polymerase (Finnzymes), the reaction was performed in 50µl (4 minutes at 94ºC, 9 cycles of [30 seconds 94ºC, 30 seconds at 54ºC, 6 minutes and 40 seconds at 70ºC], 19 cycles of [30 seconds at 94ºC, 30 seconds at 54ºC, 6 minutes and 40 seconds at 70ºC -in every cycle 20 seconds were added onto the extension time], 10 minutes at 70ºC). The final product was gel extracted by using the Qiagen gel extraction kit and this fragment was confirmed by several restriction enzyme digestions. (See Figures 3.15 and 3.16)

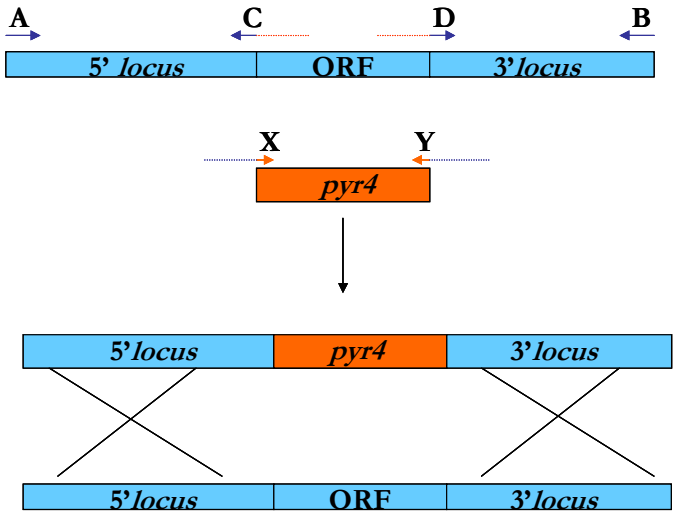

Figure3.15. The rationale of the Double-Joint PCR is explained in a schematic manner. The primers A and C were used to amplify 5' end of AN0879.3, the primers called B and D were used to amplify 3' AN0879.3 and the primers called X and Y were used to amplify *pyr4* gene. The C, D, X and Y had flanking regions to serve as primers for the 2nd round PCR. (ORF: Open Reading Frame)

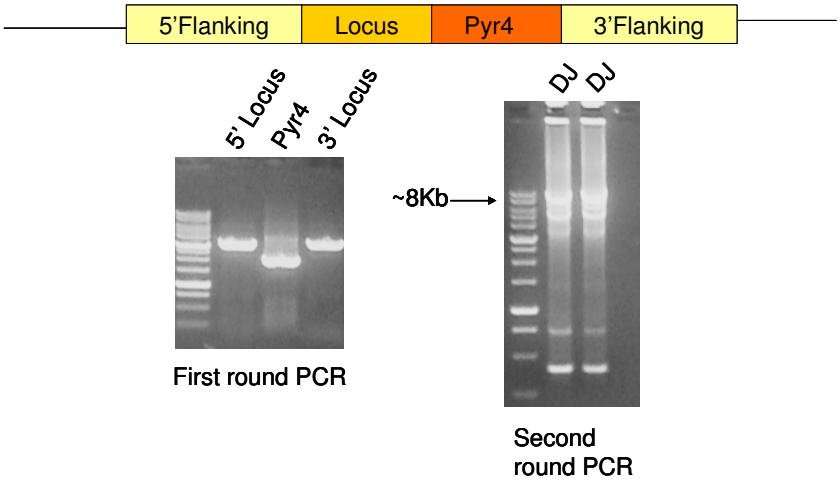

Figure3.16. In the first round PCR, 3 fragments were amplified separately, in the second round these 3 fragments were annealed and amplified with the primers called A and B. The final product was pointed with an arrow in the figure.

Later this construct was utilized to transform *A.nidulans* strains. After the transformation, the construct we introduced created a knock-out allele through the homologues recombination at the 5' AN0879.3, 3' AN0879.3**.** Our preliminary data has indicated that this gene is essential for life because we haven't observed any living cells carrying the knock out construct so far. All the transformants we obtained were checked for the existence of the inserted construct which in all cases was found to be ectopic.

The primers called '5ΔMcaP Fo', '5 ΔMcaP Re', '3 ΔMcaP Fo', '3 ΔMcaP Re' were specially designed to check the existence of the inserted construct. The primer called '5ΔMcaP Fo' was designed 357bp upstream of the 5'flanking start region and the '5ΔMcaP Re' was designed in the 717bp downstream of 5' flanking end region. The '3ΔMcaP Fo' primer was 222bp upstream of the 3'flanking start region and the '3ΔMcaP Re' primer was 334bp downstream of the 3'flanking end region. The rationale was to observe the amplification of this fragment in case the linear ΔAN0879.3 was integrated in the genome. Such amplification was never achieved (data not shown) and it proved us that knocking out the AN0879.3 from the A. nidulans genome is likely to cause lethality.

3.7.2. A new strategy to conditionally knock out the AN0879.3

Our new strategy was to develop a new technique to be able to conditionally knock out AN0879.3. For this reason we decided to prepare a knock-in where by homologous recombination, the AN0879.3 promoter will be replaced by the alcA promoter (see Figure 3.17). This way we can isolate a transformant where AN0879.3 expression is under the control of an inducible promoter, allowing us to turn it on and off just by manipulating the growth medium. For instance, initially different strains can be grown and at different stages during development the gene can be turned off and the morphology, growth rates or nuclear conditions can be investigated, moreover this experiment can be compared with the cell cycle mutants.

For this reason we designed new oligos and we continued our work by cloning these fragments into pBKS. Firstly the 5' flanking region of AN0879.3 was amplified by 5' FOR KpnI and 5' REV EcoRI and cloned into pBKS which was double digested with the same enzymes. After the ligation and transformation into *E. coli* the plasmid called p5'LOC was obtained. Later this plasmid was digested by *EcoR*I and *Xba*I and the *pyr4-3'* AN0879.3 fragment was amplified with the primers called "Pyr4 FOR EcoRI" and "3' REV XbaI" and inserted into this plasmid. After that pΔLOC plasmid was obtained. Later the primers called 'ALCMcaP For+' and ALCMcaP Rev+ were utilized to amplify the AlcLOC fragment from the pALCLOC which we prepared earlier. And later both this fragment and the pΔLOC were digested with *EcoR*I and they were ligated to obtain pCONKO, there were two correct plasmids which were named as 1 and 13. Afterwards with these plasmids several different restriction enzyme digestions were performed to confirm the orientation of the 5' AN0879.3-AlcA-AN0879.3-Pyr4-3' AN0879.3 fragment in the plasmid. The results we obtained from the *Hind*III, *EcoR*I and *Sma*I enzymes not surprisingly proved us that both plasmids gave the same restriction pattern. Later when we digested them with *Sal*I we observed that these two plasmids had different orientations.

According to our cloning strategy we knew exactly where these enzymes had their recognition sites therefore we named the number 1 as pCONKO(+) and the number 13 was named as pCONKO(-). The rationale to keep both of these plasmids

was to utilize them both for further studies, because after the transformation of the linear fragments into *A.nidulans*, it causes homologues recombination through the recognition sites. For the (+) oriented plasmid pCONKO(+), there were two possibilities of homologues recombination, it would either occur through the flanking region recognition sites or each site together with AlcA promoter meaning that it would give rise to unwanted integrations and unwanted knock-out sequences. However, the second possibility was not very likely to occur due to the size of the flanking regions which were big enough to provide appropriate conditions for the homologous recombination to occur, we still decided to keep the (-) oriented plasmid pCONKO (-) which had a higher possibility of the desired recombination to occur. The desired recombination occurs when the long flanking regions are used to give rise to homologues recombination.

After obtaining these two plasmids, both were double digested by *Kpn*I and *Xba*I to obtain the linear fragment which was ready to be transformed into *A. nidulans*. The transformation was performed once and the preliminary results indicate that when the AN0879.3 expression is allowed in the medium containing ethanol there was no obvious phenotypic change and the transformants grew normally whereas when the AN0879.3 expression was not allowed by growing the transformants in the glucose containing medium the growth was obviously affected in comparison with the internal control.

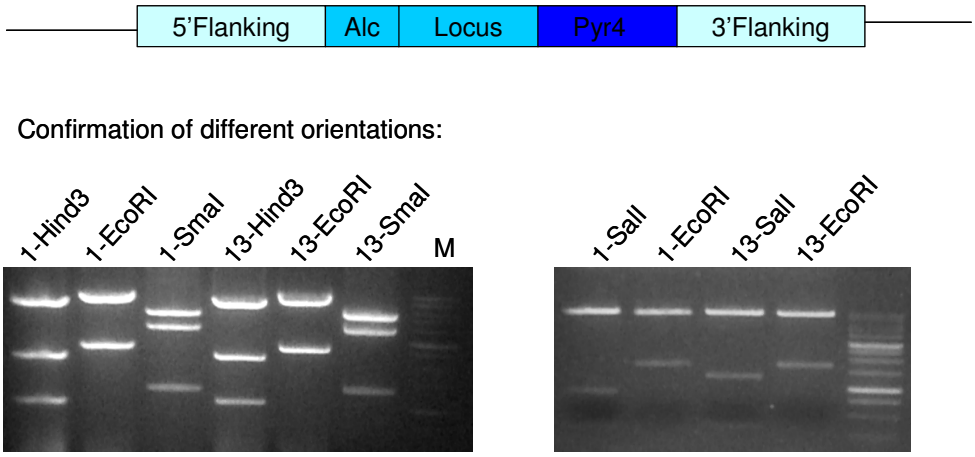

Figure3.17. The 2 pCONKO plasmids which were obtained after our cloning procedure were further analyzed by several restriction enzyme digestions and the different orientations were confirmed. These 2 plasmids were both utilized for the transformation experiments separately.

4. DISCUSSION

AN0879.3 is a newly identified gene by our group. The bioinformatics analyses revealed very interesting results about the features of this gene. The high conservation rate indicates that this is rather an important gene which is conserved during evolution. Therefore any attempt to functionally characterize the gene should be considered as a very crucial step and needs to be taken into consideration with special rigour. However prior to the initiation of the project we were aware of some certain difficulties. Firstly the gene of interest does not fall into any category with the previously known genes. Thus it was considered to be a 'hypothetical protein' and this obliged us to firstly confirm that the transcript of the gene exists in different tissues and organisms.

We could have initiated our study from totally a different point of view. As it is explained in the scientific work which was carried out by Eisenstein *et al*., 2000 the structural approach could have been the best and the most reliable way to study this new protein. In the paper the importance of studying the structure of a given hypothetical protein is clearly explained and suggested as a promising method. However there are certain difficulties of following this suggested method such as the bottlenecks of crystallization, X-ray diffraction and NMR (Nuclear Magnetic Resonance) spectroscopy. All of these techniques require specialization and very sophisticated equipments. Thus we started our study from a more biochemical point of view.

4.1. The subcellular localization studies

Our localization studies have indicated that in MCF-7 human mammary gland cell lines the localization of Loc115098 is mostly cytoplasmic but also in the nucleus in mitosis whereas in *A. nidulans* the localization of AN0879.3 was always observed in cytoplasm and never in nucleus. This was a very consistent result. There might be several explanations to this result.

Firstly in human cell line MCF-7 we performed our experiments by tagging Loc115098 either with N-terminus FLAG epitope or C-teminus FLAG epitope. In *A.nidulans* studies the AN0879.3 was tagged with GFP only in the C-terminus of the protein. We suspect that the more charged amino acid composition of the C terminus of Loc115098 may play an important role in directing the subcellular localization of this protein. However there is another evidence that the mitosis occurs differently in these two organisms. It is a well known fact that *A. nidulans* has a closed mitosis whereas in higher eukaryotes like humans the open mitosis occurs. As it is explained by S.Sazer (Sazer, 2005), the nuclear envelope never breaks down to reform itself again. In other words in higher eukaryotes the nuclear envelope disassociates during the mitosis whereas in lower eukaryotes it does not change structure (See Figure 4.1). This might be another explanation for the discrepancy of the result we obtained. After observing AN0879.3 never localizing in the nucleus in *A. nidulans* it is possible to speculate that AN0879.3 is a dynamic component of the nuclear envelope with a short turn over period. It is mostly observed in the cytoplasm although it should be taken into consideration that both the human and the *A. nidulans* copy of the gene carry nuclear localization signals predicted by *in silico* methods. Therefore we might expect to observe the localization of AN0879.3 in the nucleus.

Our preliminary results from the experiments with anti-Loc115098 has indicated that even the cells are dividing Loc115098 is localized around the nucleus as if it is surrounding the nuclear material. This seems as a supporting data confirming the *A. nidulans* experiment results. However it is still elusive where exactly Loc115098 protein localizes during mitosis. This preliminary data suggesting the perinuclear localization may be due to the different stages of mitosis. Because we conducted

no synchronization experiment we are unable observe the different stages of mitosis in one go. Therefore it is possible to speculate that in dividing cells the localization of Loc115098 might alter due to different mitotic stages. In other words we can speculate that the localization of Loc115098 differs while the cells proceed into mitosis. For instance it might be speculated that while cells are going through anaphase and telophase it is observed as a surrounding matter around the nucleus whereas during metaphase it is located in the nucleus. Even though there is still some uncertainty it is likely to be a component of the nuclear envelope. However further investigations should be done concerning the possible interactions between the Loc115098 and other proteins.

Another observation coming from anti-Loc115098 staining is that, this protein also localizes in the centrioles. Thus centriole staining together with Loc115098 might be a good approach to better understand this issue. As well as centriole staining another good approach might be to stain microtubules to distinguish the localization of different cell components during mitosis.

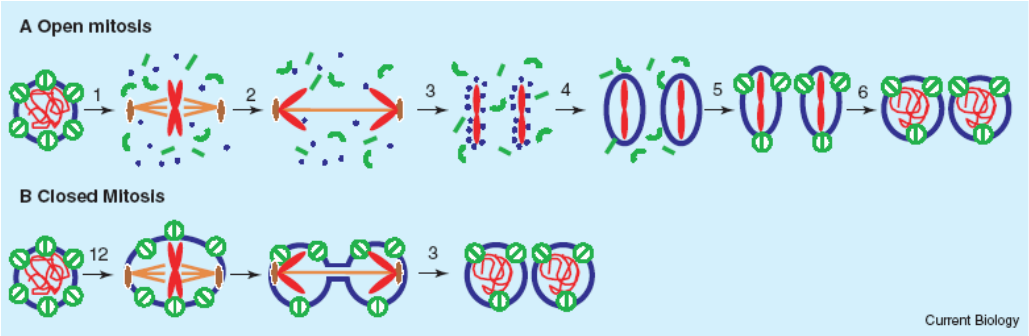

Figure4.1 The schematic explanation of the mitosis in higher and lower eukaryotes. The higher eukaryotes have open mitosis whereas the lower eukaryotes go through closed mitosis.

Considering where Loc115098 is localized in the cell, there is another interesting study carried out by Thisse *et al*, 2004 (Thisse et al., 2004). In this study they performed a high throughput localization screen of some expressed genes in the pectoral fin stage of the development by in-situ hybridization. Their results indicated that Loc115098 has rather a diffused localization in the embryos meaning that it localizes in the nucleus as well as the cytoplasm. (see Figure 4.2)

Figure4.2. The in-situ hybridization to identify Loc115098 localization in the zebrafish embryos. All embryos are in pectoral fin stage during the development. The stained regions represent the embryos and unstained regions are the yolks.

4.2. The overexpression study

Our overexpression study in *A.nidulans* aimed to observe any phenotypical changes due to the overexpression of AN0879.3 in the cell. In our results we have not observed such a phenotypic change. There might again be several explanations for this outcome. Firstly the repressible AlcA promoter might have a higher background than it is thought to have and the expression level might not be as strong as it was expected to be. However this is probably not the case because in some studies it is accepted to be one of the most reliable controllable promoters (Pachlinger et al., 2005). The overexpression of AN0879.3 caused no phenotypical change most probably due to a possible tolerance mechanism in the cell. It might be assumed that only a certain amount of AN0879.3 is needed in the cell and therefore the excess is not utilized by the cell. Concerning the experiment we have a serious drawback because no Northern Blot analyses have been performed in our experiments to check the expression levels. Instead we assumed that strains carrying multiple copies of Alc-AN0879.3 were overexpressing the protein. More experiments should be performed to further analyze and confirm this result.

4.3. The knock-out studies

Our results have indicated that knocking out AN0879.3 from the genome is very likely to cause lethality in the organism. In other words it proves that AN0879.3 is a very crucial gene for the survival of the organism. We firstly designed a linear construct according to Yu *et al*., 2004 to replace the ORF of the gene with the marker gene *pyr4*. Theoretically, after the transformation of *A. nidulans* with this causes a homologous recombination event which eventually gives rise to a knock-out strain. We obtained and checked several transformants by PCR with specially designed oligos however we never came across a true knock-out strain. After that we aimed to confirm this through another system which we generated and called 'Conditional Knock-Out Strategy'. Even though this effort is currently in progress, we are confident that the knock-out of the gene of interest causes lethality. Therefore both the results and the conditional knock-out aim are crucial for this study.

4.4. Future perspectives

Concerning Loc115098, there are still plenty of questions to be answered. The immediate goal of this project is to carry out a knock-out study in human mammary gland cell line MCF7 by using RNAi methodology. The aim in this study is to silence the *loc115098* gene in the genome by designing 21 nucleotide long oligos and observe the obvious changes in MCF7 cell line. This is an ongoing project and the results from this study will reveal new perspectives and will bring new insights to the project.

In addition to that we have already designed an anti-Loc115098 antibody and confirmed that it is working properly by testing it with Western Blot assays. Since we have the anti-Loc115098 antibody we are currently aiming to perform immunoprecipication experiments to find out the putative Loc115098 interacting proteins in the cells. Moreover the expression of Loc115098 is also being checked by immunohisto chemistry to further confirm the in situ labeling of the gene in different mouse tissues.

As mentioned before there are a lot of questions remained to be addressed in order to learn and understand the function of Loc115098 more. Therefore all different approaches aiming to functionally characterize this new protein will indeed be valuable. We are confident that when the function and the pathways in which Loc115098 participates are revealed, this will answer at least some of the questions which still remain unanswered. Hopefully it will be an important scientific contribution to the understanding of certain metabolic events.

REFERENCES

Altschul,S.F., Gish,W., Miller,W., Myers,E.W., and Lipman,D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403-410.

Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W., and Lipman,D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389-3402.

Casselton,L. and Zolan,M. (2002). The art and design of genetic screens: filamentous fungi. Nat. Rev. Genet. *3*, 683-697.

Couronne,O., Poliakov,A., Bray,N., Ishkhanov,T., Ryaboy,D., Rubin,E., Pachter,L., and Dubchak,I. (2003). Strategies and tools for whole-genome alignments. Genome Res. *13*, 73-80.

Eisenstein,E., Gilliland,G.L., Herzberg,O., Moult,J., Orban,J., Poljak,R.J., Banerjei,L., Richardson,D., and Howard,A.J. (2000). Biological function made crystal clear - annotation of hypothetical proteins via structural genomics. Curr. Opin. Biotechnol. *11*, 25-30.

Enright,M.C. and Spratt,B.G. (1999). Multilocus sequence typing. Trends Microbiol. *7*, 482-487.

Felenbok,B., Flipphi,M., and Nikolaev,I. (2001). Ethanol catabolism in Aspergillus nidulans: a model system for studying gene regulation. Prog. Nucleic Acid Res. Mol. Biol. *69*, 149-204.

Frazer,K.A., Pachter,L., Poliakov,A., Rubin,E.M., and Dubchak,I. (2004). VISTA: computational tools for comparative genomics. Nucleic Acids Res. *32*, W273-W279.

Harris,S.D., Morrell,J.L., and Hamer,J.E. (1994). Identification and characterization of Aspergillus nidulans mutants defective in cytokinesis. Genetics *136*, 517-532.

Hedges,S.B. (2002). The origin and evolution of model organisms. Nat. Rev. Genet. *3*, 838-849.

Higgins,D.G. (1994). CLUSTAL V: multiple alignment of DNA and protein sequences. Methods Mol. Biol. *25*, 307-318.

Holt,C.L. and May,G.S. (1996a). An extragenic suppressor of the mitosis-defective bimD6 mutation of Aspergillus nidulans codes for a chromosome scaffold protein. Genetics *142*, 777-787.

Holt,C.L. and May,G.S. (1996b). An extragenic suppressor of the mitosis-defective bimD6 mutation of Aspergillus nidulans codes for a chromosome scaffold protein.

Genetics *142*, 777-787.

James,S.W., Mirabito,P.M., Scacheri,P.C., and Morris,N.R. (1995). The Aspergillus nidulans bimE (blocked-in-mitosis) gene encodes multiple cell cycle functions involved in mitotic checkpoint control and mitosis. J. Cell Sci. *108 (Pt 11)*, 3485-3499.

Kent,W.J., Sugnet,C.W., Furey,T.S., Roskin,K.M., Pringle,T.H., Zahler,A.M., and Haussler,D. (2002). The human genome browser at UCSC. Genome Res. *12*, 996-1006.

Koukaki,M., Giannoutsou,E., Karagouni,A., and Diallinas,G. (2003). A novel improved method for Aspergillus nidulans transformation. J. Microbiol. Methods *55*, 687-695.

Lander,E.S., Linton,L.M., Birren,B., Nusbaum,C., Zody,M.C., Baldwin,J., Devon,K., Dewar,K., Doyle,M., FitzHugh,W., Funke,R., Gage,D., Harris,K., Heaford,A., Howland,J., Kann,L., Lehoczky,J., LeVine,R., McEwan,P., McKernan,K., Meldrim,J., Mesirov,J.P., Miranda,C., Morris,W., Naylor,J., Raymond,C., Rosetti,M., Santos,R., Sheridan,A., Sougnez,C., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Wyman,D., Rogers,J., Sulston,J., Ainscough,R., Beck,S., Bentley,D., Burton,J., Clee,C., Carter,N., Coulson,A., Deadman,R., Deloukas,P., Dunham,A., Dunham,I., Durbin,R., French,L., Grafham,D., Gregory,S., Hubbard,T., Humphray,S., Hunt,A., Jones,M., Lloyd,C., McMurray,A., Matthews,L., Mercer,S., Milne,S., Mullikin,J.C., Mungall,A., Plumb,R., Ross,M., Shownkeen,R., Sims,S., Waterston,R.H., Wilson,R.K., Hillier,L.W., McPherson,J.D., Marra,M.A., Mardis,E.R., Fulton,L.A., Chinwalla,A.T., Pepin,K.H., Gish,W.R., Chissoe,S.L., Wendl,M.C., Delehaunty,K.D., Miner,T.L., Delehaunty,A., Kramer,J.B., Cook,L.L., Fulton,R.S., Johnson,D.L., Minx,P.J., Clifton,S.W., Hawkins,T., Branscomb,E., Predki,P., Richardson,P., Wenning,S., Slezak,T., Doggett,N., Cheng,J.F., Olsen,A., Lucas,S., Elkin,C., Uberbacher,E., Frazier,M., Gibbs,R.A., Muzny,D.M., Scherer,S.E., Bouck,J.B., Sodergren,E.J., Worley,K.C., Rives,C.M., Gorrell,J.H., Metzker,M.L., Naylor,S.L., Kucherlapati,R.S., Nelson,D.L., Weinstock,G.M., Sakaki,Y., Fujiyama,A., Hattori,M., Yada,T., Toyoda,A., Itoh,T., Kawagoe,C., Watanabe,H., Totoki,Y., Taylor,T., Weissenbach,J., Heilig,R., Saurin,W., Artiguenave,F., Brottier,P., Bruls,T., Pelletier,E., Robert,C., Wincker,P., Smith,D.R., Doucette-Stamm,L., Rubenfield,M., Weinstock,K., Lee,H.M., Dubois,J., Rosenthal,A., Platzer,M., Nyakatura,G., Taudien,S., Rump,A., Yang,H., Yu,J., Wang,J., Huang,G., Gu,J., Hood,L., Rowen,L., Madan,A., Qin,S., Davis,R.W., Federspiel,N.A., Abola,A.P., Proctor,M.J., Myers,R.M., Schmutz,J., Dickson,M., Grimwood,J., Cox,D.R., Olson,M.V., Kaul,R., Raymond,C., Shimizu,N., Kawasaki,K., Minoshima,S., Evans,G.A., Athanasiou,M., Schultz,R., Roe,B.A., Chen,F., Pan,H., Ramser,J., Lehrach,H., Reinhardt,R., McCombie,W.R., de la,B.M., Dedhia,N., Blocker,H., Hornischer,K., Nordsiek,G., Agarwala,R., Aravind,L., Bailey,J.A., Bateman,A., Batzoglou,S., Birney,E., Bork,P., Brown,D.G., Burge,C.B., Cerutti,L., Chen,H.C., Church,D., Clamp,M., Copley,R.R., Doerks,T., Eddy,S.R., Eichler,E.E., Furey,T.S., Galagan,J., Gilbert,J.G., Harmon,C., Hayashizaki,Y., Haussler,D., Hermjakob,H., Hokamp,K., Jang,W., Johnson,L.S., Jones,T.A.,

Kasif,S., Kaspryzk,A., Kennedy,S., Kent,W.J., Kitts,P., Koonin,E.V., Korf,I., Kulp,D., Lancet,D., Lowe,T.M., McLysaght,A., Mikkelsen,T., Moran,J.V., Mulder,N., Pollara,V.J., Ponting,C.P., Schuler,G., Schultz,J., Slater,G., Smit,A.F., Stupka,E., Szustakowski,J., Thierry-Mieg,D., Thierry-Mieg,J., Wagner,L., Wallis,J., Wheeler,R., Williams,A., Wolf,Y.I., Wolfe,K.H., Yang,S.P., Yeh,R.F., Collins,F., Guyer,M.S., Peterson,J., Felsenfeld,A., Wetterstrand,K.A., Patrinos,A., Morgan,M.J., de,J.P., Catanese,J.J., Osoegawa,K., Shizuya,H., Choi,S., and Chen,Y.J. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Lockington,R.A., Sealy-Lewis,H.M., Scazzocchio,C., and Davies,R.W. (1985). Cloning and characterization of the ethanol utilization regulon in Aspergillus nidulans. Gene *33*, 137-149.

Marcotte,E.M. (2000). Computational genetics: finding protein function by nonhomology methods. Curr. Opin. Struct. Biol. *10*, 359-365.

Marcotte,E.M., Pellegrini,M., Ng,H.L., Rice,D.W., Yeates,T.O., and Eisenberg,D. (1999). Detecting protein function and protein-protein interactions from genome sequences. Science *285*, 751-753.

May,G.S., McGoldrick,C.A., Holt,C.L., and Denison,S.H. (1992a). The bimB3 mutation of Aspergillus nidulans uncouples DNA replication from the completion of mitosis. J. Biol. Chem. *267*, 15737-15743.

May,G.S., McGoldrick,C.A., Holt,C.L., and Denison,S.H. (1992b). The bimB3 mutation of Aspergillus nidulans uncouples DNA replication from the completion of mitosis. J. Biol. Chem. *267*, 15737-15743.

Negre,B., Casillas,S., Suzanne,M., Sanchez-Herrero,E., Akam,M., Nefedov,M., Barbadilla,A., de,J.P., and Ruiz,A. (2005). Conservation of regulatory sequences and gene expression patterns in the disintegrating Drosophila Hox gene complex. Genome Res. *15*, 692-700.

O'Connell,M.J., Krien,M.J., and Hunter,T. (2003). Never say never. The NIMA-related protein kinases in mitotic control. Trends Cell Biol. *13*, 221-228.

Osmani,A.H., McGuire,S.L., O'Donnell,K.L., Pu,R.T., and Osmani,S.A. (1991). Role of the cell-cycle-regulated NIMA protein kinase during G2 and mitosis: evidence for two pathways of mitotic regulation. Cold Spring Harb. Symp. Quant. Biol. *56*, 549-555.

Overbeek,R., Fonstein,M., D'Souza,M., Pusch,G.D., and Maltsev,N. (1999). The use of gene clusters to infer functional coupling. Proc. Natl. Acad. Sci. U. S. A *96*, 2896-2901.

Pachlinger,R., Mitterbauer,R., Adam,G., and Strauss,J. (2005). Metabolically independent and accurately adjustable Aspergillus sp. expression system. Appl. Environ. Microbiol. *71*, 672-678.

PONTECORVO,G. and KAFER,E. (1958). Genetic analysis based on mitotic recombination. Adv. Genet. *9*, 71-104.

PONTECORVO,G., ROPER,J.A., HEMMONS,L.M., MACDONALD,K.D., and BUFTON,A.W. (1953). The genetics of Aspergillus nidulans. Adv. Genet. *5*, 141-238.

Sazer,S. (2005). Nuclear envelope: nuclear pore complexity. Curr. Biol. *15*, R23-R26.

Spradling,A., Ganetsky,B., Hieter,P., Johnston,M., Olson,M., Orr-Weaver,T., Rossant,J., Sanchez,A., and Waterston,R. (2006). New roles for model genetic organisms in understanding and treating human disease: report from the 2006 Genetics Society of America meeting. Genetics *172*, 2025-2032.

Thisse,B., Heyer,V., Lux,A., Alunni,V., Degrave,A., Seiliez,I., Kirchner,J., Parkhill,J.P., and Thisse,C. (2004). Spatial and temporal expression of the zebrafish genome by large-scale in situ hybridization screening. Methods Cell Biol. *77*, 505-519.

Tilburn,J., Scazzocchio,C., Taylor,G.G., Zabicky-Zissman,J.H., Lockington,R.A., and Davies,R.W. (1983). Transformation by integration in Aspergillus nidulans. Gene *26*, 205-221.

Yu,J.H., Hamari,Z., Han,K.H., Seo,J.A., Reyes-Dominguez,Y., and Scazzocchio,C. (2004). Double-joint PCR: a PCR-based molecular tool for gene manipulations in filamentous fungi. Fungal. Genet. Biol. *41*, 973-981.

Zheng,Y., Roberts,R.J., and Kasif,S. (2002). Genomic functional annotation using co-evolution profiles of gene clusters. Genome Biol. *3*, RESEARCH0060.