Faculty of Science and Technology

MASTER THESIS

Study program/ Specialization: MSc in Biological Chemistry	Spring semester, 2014 Open access
Author: Natalia Hansen	…… (Author's signature)

Faculty supervisor: Professor Peter Ruoff

External supervisor(s):

Title of thesis:

The Influence of Inorganic Nitrogen Compounds on Gene Expression Profiles in *Neurospora crassa:* A Microarray Analysis

Credits (ECTS): 60

Key words: Nitrate metabolism Circadian rhythm *Neurospora crassa* Microarray analysis	Pages: 117 + enclosure: CD Stavanger, 03.02.14

University of Stavanger

Department of Mathematics and Natural Science

MASTER THESIS

Biological Chemistry
Natalia Hansen

Acknowledgements

Thank you!

This project would not have been possible without the goodwill of many people throughout the entire process. In particular, I would like to give special thanks to some.

***Professor Peter Ruoff**, f*or favouring me with excellent instructions and advice. Without his support, it would have been impossible for me to complete my thesis. His incredible understanding of research, his rigorous way of thinking as well as his personal qualities have benefited me a great deal.

PhD student Oleg Agafonov, who kindly provided me with experimental results which were invaluable to the completion of this work. His technical support was one of the most leading points in my thesis.

Manager at SLab Arvid Børretzen. for giving me the opportunity to combine job and education, to develop myself, and complete this thesis.

Family and friends, especially my husband for his support and encouragement.

I also would like to thank our teachers through the two years at University of Stavanger. Each and every one have contributed to give me the necessary knowledge to complete this project.

Natalia Hansen

Abstract

The purpose of this study was to investigate expression profiles of *Neurospora crassa* genes related to the nitrogen metabolism. Two nitrogen sources (ammonium and nitrate) and responses to light were meant to alter patterns of gene expression. The advent of full-genome microarrays enabled the measurement of mRNA levels in response to the environmental alterations.

During the course of this experiment, and for all of the data sets discussed here, a total of 10526 ORFs based on MIPS or Broad Institute predictions had expression data. Further investigations involved recognizing the statistically different expressed genes. Two approaches were used for that purpose– filtering software of BRB-array tool, and arbitrary cut-off function of MatLab-script.

The next step in the analysis was to find the patterns in gene expression by two clustering methods – hierarchical clustering and *k*-means clustering. The six clusters revealed by hierarchical clustering showed a great deal of overlap with the clusters recognized by *k*-means clustering. The analysis of clusters involved visual investigation and functional prediction, which resulted in heat-maps, where one can connect the functions to a specific environmental condition. However, the majority of the statistically different expressed genes were not identified by functional prediction software.

The genes related to a specific functional category were also investigated and results were compared to other studies.

Abbreviations

con	*conidiation*
DD	constant darkness
FLO	*frq*-less oscillator
FRH	*frq*-interacting RNA helicase
frq	*frequency*
glm	*glutamine*
GOGAT	glutamate synthetase
GS	glutamine synthetase
LL	constant light
luc	*luciferase*
NiR	nitrite reductase
nit	*nitrate non-utiliser*
nmr	*nitrogen metabolite repressor*
NR	nitrate reductase
NV	Normal Vogel
wc	*white collar*
WCC	white collar complex

Table of Contents

1. Introduction

1.1 Microarray analysis

DNA microarray analysis is today one of the most promising methods in functional genomics. It is a high throughput technology used in molecular biology and biotechnology to simultaneously access the gene expression profile of thousands of genes. Changes in gene expression as a response to changing environment conditions, diseases, drug treatment or chemotherapy medications, can be detected. This gives the possibility to screen hundreds or thousands of gene fragments in one assay and determine thousands of expression values in hundreds of different conditions. Microarray analysis, therefore, seems to be an important tool for diagnosis of diseases at a molecular level. Applications are, for example, the improvement of diagnosis and treatment of cancer and the improvement of the effectiveness of drug treatment (Clarke et al., 2001).

Different microarray methods have been developed and there are more that are in the process of being developed. However, each array system seems to follow the same basic method. Typically, a microarray chip consists of an arrayed series of thousands of microscopic spots of DNA oligonucleotides, each containing a small amount of a specific DNA sequence of a gene.

In standard microarrays, the probes are attached to a solid surface made of glass, plastic or silicon by a covalent bond to a chemical matrix via epoxy-silane, amino-silane, lysine and polyacrylamide. On the substrate one places single stranded DNAs (ssDNA) with various sequences, arranged in a specific array. This solid substrate with printed ssDNA is called a probe. The purpose and technology of the array are the basic factors that determine the deposit of the probe. Cytoplasmic concentrations of mRNA from the investigated organisms in different conditions/states of life cycle, et cetera are used as indicators for gene expression. Enzyme reverse transcriptase produces a DNA sequence out of the harvested mRNA (complementary DNA or cDNA represents the coding sequence of a gene and its bordering regions). The nucleotides, which are used for the synthesis of the cDNA are called targets and include different fluorescent tags, commonly a green Cyanine fluorescent dye (called Cy3) and a red dye (called Cy5). The targets reflect the amount of mRNA produced from a sample obtained under a particular influence factor. The sets of cDNA from different

samples are mixed together and introduced to the prepared probe at which point hybridization, or base pairing, occurs between the DNA immobilized on the probe and the fluorescently tagged DNA. The probe is then washed and stained if necessary, depending on the tag being used.

After an incubation time, the cDNAs that did not bind to any spots are washed away. Then, the location and intensities of the fluorescent dyes are recorded with a scanner. The scanner consists of lasers with different wavelengths and a sensor. The scanning process results in two digital monochrome images of the microarray, one for the green dyes and one for the red dyes. The amount of fluorescence emitted by each spot will be proportional to the amount of mRNA produced by the gene with the corresponding DNA sequence. In general, the ratio of the relative intensities, (Cy5/Cy3), is used for the analysis of gene expression. To visualize the relative gene expression, the two images are pseudo-colored and then merged to a ratio image of the microarray (Figure 1). There are many different methods which can be used to analyze the received image and quantify the signals into numerical values (Wiltgen and Tilz, 2007).

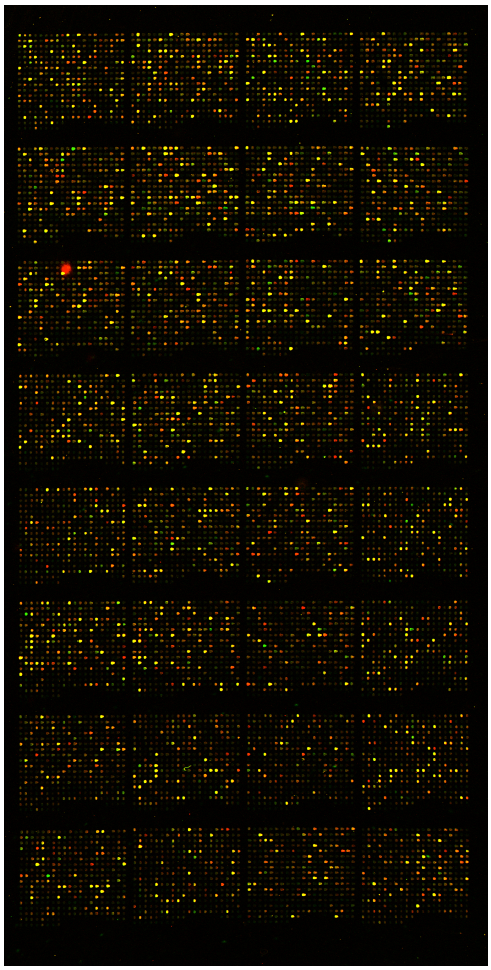


Figure 1. Scanned image of a DNA Microarray chip containing the genome of *N.crassa.* Scanned image represents microarray generated in Jay Dunlap's lab by Carol Kringelberg, Dartmouth Medical School, USA. The data from this microarray are used in this thesis. Spots with only green or red appear in positions where only one of the tagged DNA bases paired to the probe. In general, spots that are yellow are positions where both targets paired to the probe to some degree. The black spots visualize the places where no pairing occurred.

Further data analysis includes the identification and clustering of common patterns of gene expression. This can determine which genes are induced or repressed in response to a step of the cell cycle, a development phase, or changes in the environment, such as drug treatment. Some of the most popular procedures to extract the valuable data from the microarray images are clustering methods like: hierarchical clustering, self organizing maps and support vector machines.

Clustering is the grouping of objects based on their similarity. It can also be described as the partitioning of a data set into subsets, so that the data in each subset share some common trait. The measurement for a common trait is defined before the clustering is performed. Such a trait is often a distance metric defining the relative similarity between the two objects. Clustering is commonly used for statistical data analysis, and has applications to many fields, including machine learning, data mining, pattern recognition, image analysis and bioinformatics.

Clustering of gene expression data helps in identifying genes of similar function (Eisen et al., 1998).

A distance metric is primarily used to define the similarity or difference between two objects. Some of the most common distance metrics used are Euclidean distance, Manhattan distance and Correlation distance. Euclidean distance is the distance between two points that can be measured with a simple ruler, or can be calculated by repeated application of the Pythagorean Theorem. Manhattan distance can be defined as the distance between two points expressed as the sum of the absolute differences of their coordinates. Correlation distance is mostly used for clustering of gene expression data and measures the similarity between two points expressed as the correlation between the two objects (Babu, 2004).

Hierarchical clustering generates a hierarchy among objects based on their similarity or differences. Hierarchical clustering may be constructed using an agglomerative or divisive approach. Hierarchical clustering results in a tree also known as a dendrogram, with individual elements at one end and a single cluster containing every element at the other. Agglomerative algorithms begin at the leaves of the tree, whereas divisive algorithms begin at the trunk. Agglomerative clustering can include single linkage clustering, complete linkage clustering or average linkage clustering.

In single linkage clustering and complete linkage clustering, pairs consisting of one object from each group are considered. In single linkage clustering, the distance between the groups is defined as the distance between the closest pair of objects. Complete linkage clustering, also called farthest neighbor clustering, is the opposite of single linkage. The distance between groups is defined as the distance between the most distant pair of objects, one from each group. Average linkage clustering characterizes the distance between two groups as the average of distances between all pairs of objects, where each pair is made up of one object from each group (Babu, 2004; Do and Choi, 2008).

The k-means algorithm helps to cluster objects into k partitions using the similarity between the objects. The constant k indicates the number of partitions/clusters and is provided by the user. The algorithm starts by partitioning the input points into k initial sets and then calculates the centroid (mean point), of each set. Thereafter, it constructs a new partition by associating each object with the closest centroid. Afterward, the centroids are recalculated for the new clusters, and the process is repeated by alternate application of these two steps until

convergence, which occurs when the objects no longer switch clusters or the centroids no longer change. K-means is one of the most commonly used clustering methods and has a wide application in microarray studies (Do and Choi, 2008).

Received fundamental patterns allow conclusions about the common functional behaviour of the genome, and provide insights into the dynamics of a genome or genomic shift in metabolis. (Draghici, 2011 ; Wiltgen and Tilz, 2007).

1.2 *Neurospora crassa* as a model organism

1.2.1 History review

If in the middle of the nineteenth century, someone told a baker that the orange fungus would become one of the most favoured research organisms for biologists in a couple of centuries, he would never believe it. This fungus, which infected almost all bakeries in France, also spreads fast on burned areas and broken volcanoes.

In 1927 microbiologists C.L. Shear and B.O. Dodge assigned this organism to the new genus *Neurospora*, which belongs to the family of ascomycetes fungi, and described its life cycle. This family consists of 10 species, where *N.crassa* is one of the most studied (Perkins, 2001). In 1941, Beadle and Tatum received the first biochemical mutants. Investigating *N.crassa*, they made a hypothesis "one gene – one enzyme" (that one gene codes synthesis of one specific protein - ferment), which can be found in all books on biochemistry and genetics nowadays. In 1958, the authors were awarded the Nobel prize, and their work is considered to be the bridge between biochemistry and genetics. From that moment, the era of biochemical genetics and molecular biology begins. Filamentous fungus *Neurospora* as well as *Drosophila* became an extremely convenient model of the eukaryotic organism for experiments in these branches of science. The popularity of *Neurospora* can be explained by its quite simple structure and its easy and fast reproduction cycle in the laboratory, which allows investigation of a couple of generations in a short period of time.

In the beginning of the ‘60s, the Fungal Genetic Stock Center (FGSC) was established in order to systemize the job of cataloging the continuously growing amount of *Neurospora´*s mutants. The center collects and keeps all genetic materials for *Neurospora*. The catalog of the more than 7,500 fungal mutants is available on the Center’s website - www.fgsc.net. The

collection includes descriptions of the genetics of the wild type *Neurospora* and its mutants, as well as other strains of the fungus and its various mutants, which were found in more than 4,000 populations in nature.

The increased popularity of the scientific investigations of the red bread mould also resulted in many thematic conferences, scientific articles and even an annual paper, the "*Neurospora* Newsletter". Because of the huge amount of papers in fungal genetics, the word “genetic” was added to the name of the fungal conferences and newsletter: "Fungal Genetics Conference" and "Fungal Genetics Newsletter" from 1986.

The big discoveries in genetics were achieved by use of different *Neurospora* mutants. Repeat-induced point mutations, which happen to genes in double DNA strains before meiosis is one such discovery. They were described in 1987 by E. Selker and seem to be an important defense mechanism for the eukaryotic genome (Selker et al., 2003).

The peak of the genetic investigations of the *Neurospora crassa* is, of course, reporting its genome sequence, which occurred in 2003 with the cooperation of more than 70 scientists (Galagan et al., 2003). It was found that the seven fungal chromosomes contain about 10,000 genes. That is twice as much as *Schizosaccharomyces pombe* and just a quarter less than the fruit fly *Drosophila*. The analysis of the genetic sequences allowed the discovery of some previously unknown features of the *Neurospora´*s biology, concerning the genes from secondary metabolism as well as differences in Ca^{2+} signaling systems between plants and animals. In addition, it was found that the RIP-mutations are able to affect gene evolution, dramatically slowing the formation of new genes and leading to the creation of a genome with an extremely low content of closely related genes.

In that way, *Neurospora crassa*, one of the simplest eukaryotic organisms, became a popular experimental model organism. Physiology, molecular genetics, biochemistry, molecular cell biology, plant development, photobiology, circadian rhythms and evolution – are just a few of the many scientific branches which are taking advantage of *Neurospora´*s complex, yet genetically and biochemically tractable life cycle (Borkovich et al., 2004).

1.2.2 Life cycle of *Neurospora crassa*

Like all ascomycetes, *Neurospora* can undergo reproduction by two different ways – asexual (through spores, formed from one organism) and sexual (which requires two organisms). For scientists it is very important that the phases of N.*crassa´*s development are identifiable by its morphology, and the transition from the one phase to another is caused by certain environmental factors (e.g. changes in water, light, starvation). The fungus is able to grow in an environment composed of inorganic salts, vitamin biotin and a source of carbon. By adding or excluding certain nutrients it is possible to track changes in the ascomycetes morphology.

The fungus is multicellular and most of its life cycle is spent in the haploid state. Haploid spores appear immediately from the diploid zygote as the result of meiosis. The general life cycles of *Neurospora* are shown in Figure 2 (The life cycle of *Neurospora crassa*) and described in sections 2.2.1 and 2.2.2, respectively.

1.2.2.1 Asexual cycle

The asexual cycle includes the formation of a haploid asexual spore, the multinucleate macroconidia and uninucliate microconidia, formed on aerial hyphae. Germination and growth of macroconidia results in a mass of branched tubular filaments – vegetative hyphae. Vegetative hyphae are tip-growing, regularly branching cellular elements, which constitute a colony called micelium. Hyphae consist of one or more cells with internal cross walls, called "septa". Septum is perforated by pores, which are large enough to allow ribosomes, mitochondria and nuclei to flow between cells. In such a way, a single hypha contains many haploid nuclei.

If growing on a Petri dish, the branched hyphae of the mould cover the growth medium with even white pellicle. In optimal conditions, the mould is able to grow at a rate of 10 centimeters per day. In response to some environmental factors, like stress, desiccation or nutrition deprivation, specialized aerial hyphae can be differentiated from vegetative hyphae. Aerial hyphae form chains of round asexual spores – millions of macroconidia which disperse and repeat the asexual cycle if they land on a suitable substrate. In the light, the macroconidia develop an intense orange, carotenoid pigment, which gives *Neurospora* its typical color (Davis, 2000; Seale, 1973; Springer and Yanofsky, 1989).

The uninucleate microconidia is produced by microconidiophores or directly from the vegetative hyphae and can function either as male gametes in fertilization of protoperithecia or as asexual reproductive structures or both (Maheshwari, 1999).

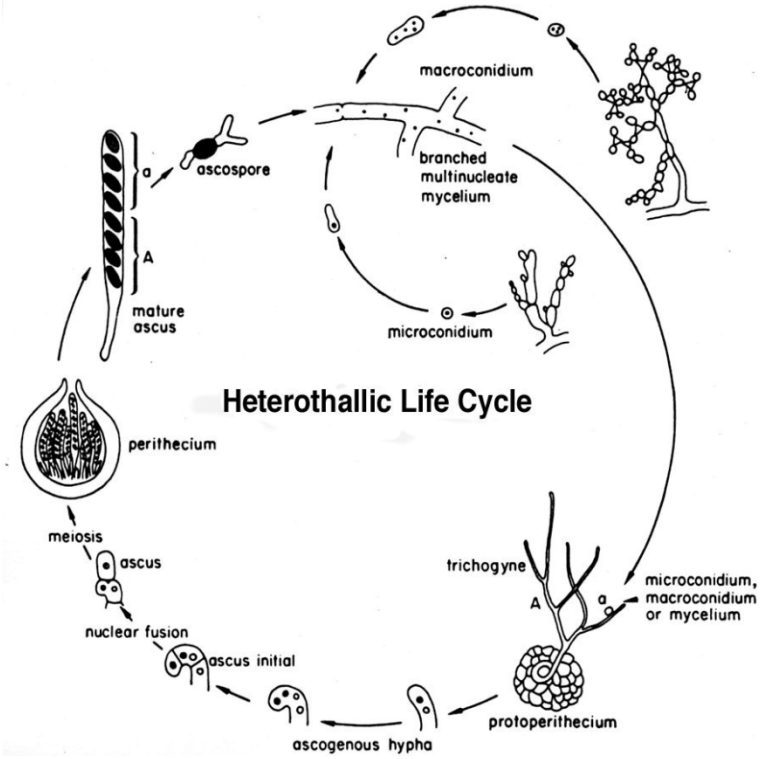


Figure 2. The life cycle of *Neurospora crassa.* The asexual cycle is shown by the smaller sequence. Aerial hyphae form macroconidia or microconidia, which germinate and form a new mycelium. The larger sequence represents the sexual cycle and starts with the formation of the protoperithecium, which is fertilized via its trichogyne by a conidium of the opposite mating type, and ends in the production of meiotically-derived ascospore (http://www.fgsc.net/*Neurospora*/photos/lgHeterothallicLifecycle.jpg).

1.2.2.2 Sexual cycle

In limiting nutrition conditions *Neurospora* forms multicellular female sexual organs – protoperithecia - a small knot of hyphae that surrounds a few special cells. From protoperithecium differentiates ascogonium – one cell of which acts as a female gamete while the other forms a dense protective layer. From the female gamete come branched filamentous trichogynes. Trichogynes grow toward the conidia of the opposite mating type in response to its pheromone emission. Finally, when fusion occurs a nucleus of the conidium travels through the trichogyne to the ascogonial cell in the protoperithecium. Nuclei of both mating types divide many times and then fusion takes place. After two meiotic divisions of the

diploid zygote and mitotic division of the products, the hard melanized black sexual spores – ascospores - appear. Spores are activated by heat and grow thereafter in the presence of suitable nutrients (Davis, 2000).

1.3 Nitrogen metabolism in *Neurospora crassa*

Nitrogen is a necessary element for all living organisms as it is an integral part of proteins and nucleic acids. All organisms require nitrogen, but only plants, bacteria and fungi are able to utilize its mineral forms, including nitrate and nitrite salts. Nitrogen is often a limiting factor in the environment and many organisms therefore developed a complex regulatory system for its metabolism (Fu and Marzluf, 1987).

*Neurospora´*s nitrogen control circuit has been studied extensively and is comprised of a set of genes that encode enzymes that enable the fungus to use various secondary nitrogen sources, e.g., nitrate, nitrite, purines, or amino acids, when preferred nitrogen sources such as ammonia or glutamine are unavailable. If ammonium is present in the environment, it is transported into the cell by a highly specific system. In the cell, ammonium is further assimilated into organic form by two systems. The first one is common with a high level of ammonium concentration in the substrate. Using NADPH as a reducer, glutamate dehydrogenase (NADP-GDH) converts α-ketoglutarate and ammonium to glutamate:

$$NH_4^+ + \alpha - ketoglutarate \xrightarrow{NADP-GDH} Glutamate$$

NADP-GDH is a constitutive enzyme encoded by the *am* locus.

Low ammonium concentrations are suitable for the second assimilating system. It consists of a cycle of two enzymes. One is glutamine synthetase (GS) and the other is glutamate synthase. GS synthesises glutamine from glutamate,

$$Glutamate + ATP + NH_4^+ \xrightarrow{GS} Glutamine$$

whereas glutamate synthase transfers the amide nitrogen of glutamine to α-ketoglutarate, forming one extra glutamate molecule,

$$Glutamine + NADPH + \alpha - ketoglutarate \xrightarrow{GOGAT} 2 * Glutamate$$

Glutamate synthase is also known as glutamine amide: 2-oxoglutarate amino transferase, or GOGAT. *N. crassa* is known to possess two glutamine synthetase polypeptides that are encoded by different genes- *glm-1* (glutamine-1), encoding the β-glutamine synthetase polypeptide that forms an octameric enzyme required for growth in the absence of glutamine, and *glm-2*, encoding the α-glutamine synthetase polypeptide. The purified α-glutamine synthetase polypeptide forms a homotetramer with lower activity than the homooctameric β-glutamine synthetase. When both proteins are present they form a heteromeric enzyme and the overall activity of the enzyme is thought to be determined by the ratio of the polypeptides (Greenwald, 2010).

Glutamine may be the end product of the process whether it is primary donor of amid nitrogen in biosynthetic reactions, and the part of the signal of nitrogen catabolite repression.

In the absence of preferable sources, nitrate repression is lifted, and nitrate is transported into the cell. Nitrate transport is performed by the NIT-10 – inducible high-affinity transporter protein, encoded by the nitrate nonutiliser-10 (*nit-10*) structural gene. NIT-10 is a nitrate permease of 541 amino acids and is used for active nitrate transport through the cell membrane (Borkovich et al., 2004; Gao-Rubinelli and Marzluf, 2004).

In the cell, nitrate converts to ammonium by the use of two enzymes, nitrate reductase (NR) and nitrite reductase (NiR), encoded by structural genes *nit-3* and *nit-6* respectively. In general terms, the pathway is an enzymatic stepwise reduction of nitrate to nitrite and glutamine. A simplified nitrate assimilation pathway is shown on Figure 3.

Figure 3. Nitrate assimilation pathway. Nitrate assimilation pathway. Nitrate is reduced to nitrite by nitrate reductase NR, and nitrite reductase NiR catalyses the reduction of nitrite to ammonium. Ammonium in turn is further converted to glutamine by glutamine synthetase GS.

In *Neurospora*, members of the nitrate assimilatory pathway make up the transcriptional/translational negative feedback loop regulating assimilation of nitrate (see Figure 4).

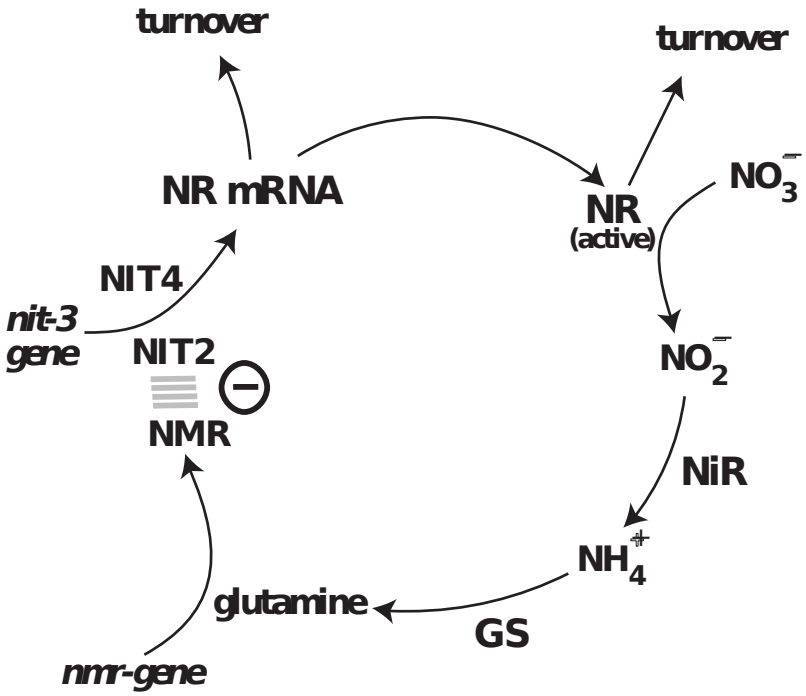


Figure 4. A model for the NR negative feedback loop. Heterodimer formed by NIT-2 and NIT-4 binds to the *nit-3* promoter. The *nit-3* gene product, nitrate reductase, reduces nitrate to nitrite and leads to accumulation of glutamine. Sufficient levels of glutamine trigger binding of NMR to NIT-2, resulting in inhibition of the *nit-3* transcription. The whole circle results in a transcriptional/translational negative feedback loop (Christensen et al., 2004).

When nitrate is present in the cell, positively acting transcriptional factors, NIT-2 and NIT-4 cooperate and activate the transcription of the *nit-3* gene, which codes the structural protein of nitrate reductase (NR). The expression of *nit-10* has also been found to require the presence of both NIT-2 and NIT-4 (Gao-Rubinelli and Marzluf, 2004).

NIT-2 is a global Cys2/Cys2-type zinc finger DNA-binding protein and a member of the GATA family of transcriptional activators along with WC-1, WC-2, SRE and ASD4 in *Neurospora*. It is involved in the activation of transcription of many structural genes of the nitrate assimilatory pathway, including *nit-3*, *nit-6*, and *nit-10*. It has been shown that the

transcription of the *nit-2* increases 2 to 3 times under nitrogen repressed conditions, i.e., lack of a primary nitrogen source (Fu and Marzluf, 1987).

NIT-4 protein is a pathway-specific transcriptional activator, which contains a putative DNA-binding domain consisting of an N-terminal Cys6/Zn2-type zinc finger motif, and a C-terminal leucine-rich activation domain. The NIT-4 protein is a member of the GAL4 family of fungal transcription factors. The *nit-4* protein is constitutively expressed at low levels. The action of NIT-4 is required for expression of both the *nit-3* and *nit-6* genes (Feng and Marzluf, 1996).

The *nit-3* promoter region contains binding sites for both NIT-2 and NIT-4. NIT-2 has been shown to have three binding sites on the *nit-3* promoter, while NIT-4 has two. All binding sites have been shown to be required for full *nit-3* expression.

A protein-protein interaction between NIT-2 and NIT-4 is also required for transcriptional activation of *nit-3*. NIT-2 and NIT-4 form a heterodimer, and protein analysis indicates that the zinc finger DNA-binding domain of NIT-2 as well as both the DNA-binding domain and the C-terminal coiled-coiled activation domain of NIT-4, are sufficient and essential for the interaction (Feng and Marzluf, 1996; Gao-Rubinelli and Marzluf, 2004).

The negative acting element in the *Neurospora´*s transcriptional/translational negative feedback loop regulating assimilation of nitrate has been identified as the nitrogen metabolite repression (*nmr*) gene. The *nmr*-gene is not regulated, but is expressed constitutively at a very low level. It has no distinctive characteristics such as DNA binding or protein kinase motifs (Young et al., 1990). Direct binding of NMR to two α-helices of NIT-2 hinders its binding to the *nit-3* promoter, resulting in the inhibition of *nit-3* transcription (Pan et al., 1997).

NR is a large homodimeric multi-redox protein. It contains three domains separated by short hinge regions. Each domain contains a cofactor essential for the enzyme's catalytic activity. *Neurospora* uses NADPH as a reducing agent and electrons from NADPH are transferred to a C-terminal flavin domain, which contains a FAD cofactor, then to a central heme-containing domain, and finally to an N- terminal molybdopterin-containing domain where the reduction of nitrate to nitrite takes place (Campbell and Kinghorn, 1990; Davis, 2000; Marzluf, 1997).

Nitrite reductase also appears to be a large homodimer with a novel prosthetic group, siroheme. As was shown above, synthesis of the NIT-3 requires the action of many other genes, like *nit-2* and *nit-4*. It has also been shown that molybden cofactor of NR requires the action of the *nit-1*, *nit-7*, *nit*-8 and *nit-9* genes. *nit-4* and *nit-5* genes are shown to be alleles from the same locus and regulate the induction of the pathway by nitrate and nitrite (Tomsett and Garrett, 1980).

Expression of NR sets off a series of reactions whereby nitrate is reduced to nitrite, then ammonium and finally glutamine. Presence of ammonium leads to strong nitrogen repression but is not itself active, since it does not cause repression in mutants lacking glutamine synthetase. Accumulation of glutamine leads to an inactivation of transcription of *nit-3*. Glutamine appears to be the critical metabolite, which exerts nitrogen catabolite repression. The signal pathway that senses the presence of repressing levels of glutamine is still unknown (Marzluf, 1997).

1.4 Circadian rhythms

1.4.1 General review

Circadian and biological rhythms are periodically repeated changes in the intensity and nature of biological processes. A circadian network (or biological clock) confers a competitive advantage to an organism, probably by enabling it to anticipate cyclic changes in the environment. Circadian rhythms with very similar properties characterize, apparently, almost all organisms, controlling processes from cyanobacterial cell division to human sleep-wake cycles (Dunlap, 2004).

An essential function of circadian rhythms is to provide an internal estimate of the external local time, thereby allowing an organism to program its activity at a specific time of the day with adaption for the predictable changes and coordination of biological processes with exogenous cycles of the environment.

Three characteristics of circadian rhythms have been emphasized. First, circadian rhythms are controlled by self-sustained oscillators, which continue to oscillate in constant conditions. Second, these oscillators are temperature compensated so that they run with approximately the same period at different constant ambient temperatures. Third, entrainment

is the most important property for determining the phase relationship of the clock, i.e. external factors can shift the phase of circadian rhythms and change their amplitude (Bell-Pedersen, 1998; Dunlap, 2004).

An oscillator, the mechanism that drives the rhythm, can be defined as a set of interacting variables that according to the amount of available energy show sustained oscillations in at least one variable. The environmental cues are "input" variables that can affect the oscillator. Light, temperature and pH are considered to be the most important environmental cues for circadian rhythms.

Most circadian oscillators are comprised of components whose main purpose is the function of the clock, either in instigating transcription of necessary genes based on environmental indicators or in posttranslational modification of clock proteins. It has been commonly assumed that this core clock is responsible for driving physiological systems, which rely on timekeeping for the function of the organism. The general system driving this rhythm has been considered to be a transcriptional/translational negative feedback oscillator containing both positively and negatively acting elements - multiple interlocked feedback loops controlling the rhythmic expression of key genes (Lillo et al., 2001).

The study of circadian rhythms relies on organisms that display an easily assayed output which can vary greatly from organism to organism. *Neurospora crassa*, a model eukaryotic organism, has played a major role in investigating the key molecular processes involved in maintaining a circadian clock. It shows an easily observable 22-hour rhythm in asexual spore development (conidiation) when cultures are grown in constant darkness, as well as rhythms in gene expression, metabolism, pheromone production, stress response, and other processes.

The *Neurospora* conidiation rhythm is in practice expressed in nearly all strains, but one mutant, band (*bd*) exhibits the rhythm much more clearly and under more varied conditions than the wild-type (*wt*) strains. Since its discovery, the *bd* strain is being used in nearly all circadian clock experiments (de Paula et al., 2007).

1.4.2 The circadian system of *Neurospora crassa*

Recently the circadian system of *Neurospora* has been widely discussed. However, it is clear that the mechanism that drives rhythmic conidiation in *Neurospora* is more complex than a single, simple feedback loop and may be composed of more than one self-sustained oscillator. In addition to the well-known functional feedback loop FRQ/WCC, there has been recently reported regulation and rhythmicity in its absence (called the *frq*-less oscillator, or FLO). This fact allows suggestion that the clock might be organized as a network of coupled oscillators, which on their own might be less robust. The existence has also been predicted of genetically distinct circadian oscillators that function as core oscillators to regulate specific outputs (de Paula et al., 2007).

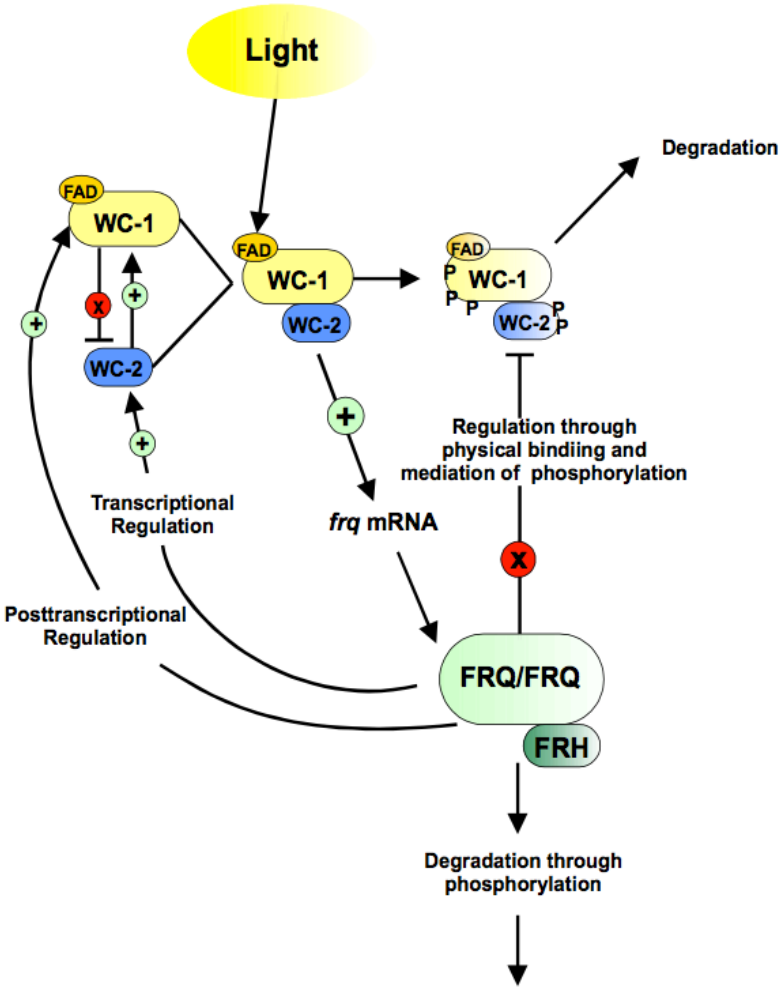


Figure 5. Model of the *Neurospora crassa´s* circadian oscillator. The positive transcriptors WC-1 and WC-2 form a heterodimer which activates the *frq* gene. The FRQ protein inhibits the WCC thus inhibiting its own transcription (Christensen, 2007).

However, the Frequency/White Collar Complex (FRQ/WCC) is considered to be the core circadian oscillator necessary for generating many of the observed circadian rhythms, including the developmental rhythm. The FRQ/WCC oscillator is known to be composed of an autoregulatory, transcriptional/translational negative feedback loop where two positive elements function as activators of the transcription of a negative element, and the negative element acts to repress its own transcription by inhibiting the activity of the positive elements. It involves the frequency (*frq*) and white collar (*wc-1*, *wc-2*) genes and their protein products (Liu and Bell-Pedersen, 2006). Figure 5 shows a current model of the *Neurospora crassa´*s circadian oscillator.

The two positive elements are the WC-1 and WC-2 transcription factors. Both proteins are GATA-type zinc finger DNA-binding proteins. WC-1 and WC-2 form the heterodimeric white collar complex (WCC) via their Per-Arnt-Sim (PAS) domains. In addition to its essential role in the *Neurospora* feedback loops, WC-1 is also known as the primary photoreceptor for the *Neurospora* clock and light-induces the formation of a larger WCC complex that may be responsible for the regulation of many light-controlled genes (*lcgs*). The WC-1 senses light through its PAS/LOV domain (light, oxygen, voltage domain), which counteracts the WCC formation at dawn, and enhances downregulation of *frq* transcript levels at dusk, subsequently preventing or promoting the resetting of the circuit (Wijnen and Young, 2006). The WCC binds to the cis-acting sequence called clock-box (C-box) in the *frq* promoter and directly activates transcription of the *frq* gene. *frq* mRNA is translated to the FRQ protein which exists in two isoforms. The relative levels of these isoforms change with temperature as a result of thermosensitive splicing, yielding a bifurcated, temperature-dependent protein pathway.

FRQ is the key negative element of the oscillator, which dimerizes through the coiled-coil region near its N-terminus and forms a complex with FRH (a FRQ-interacting RNA helicase). The FRQ-FRH complex (FFC) enters the nucleus and dissociates WCC from the *frq* promoter by promoting phosphorylation of WCC by several kinases, including casein kinase 1(CK-1a) and CK2, as well as protein kinase A and C (PKA and PKC). The process happens in the morning, when the relative concentrations of the hypophosphorylated nuclear FRQ protein are relatively low. Once hyperphosphorylated, DNA-binding activity of the WCC is inhibited such that transcription of *frq* is suspended and the level of FRQ protein is

decreased. The inhibition of WCC transcriptional activity by FRQ-dependent phosphorylation closes the circadian negative feedback loop (de Paula et al., 2007).

Accumulation of hyperphosphorylated cytoplasmic FRQ protein later in the day and its subsequent phosphorylation-induced decay in conjunction with dephosphorylation of the WCC by protein phosphatase 2A, promotes the assembly of cytoplasmic WCC complex, thereby stimulating the reactivation of *frq* transcription. Five kinases (CK-1a, CK2, calmodulin kinase 1 (CAMK-1), PKA, and checkpoint kinase 2 (chk2/PRD-4)) have been shown to phosphorylate FRQ, where CK-1a and CK2 seem to be the main kinases in FRQ phosphorylation. Such cyclic activation, repression and initiation of *frq* expression generate the endogenous circadian rhythmicity, which controls expression of clock-controlled genes (ccgs)(Jinhu and Yi, 2010).

It is also considered that the FRQ and FRH can act in a positive feedback loop, maintaining levels of *wc-2* mRNA and WC-1 and WC-2 proteins. The role of the positive loop is suggested to confer stability and robustness to the FRQ/WCC oscillator (Cheng et al., 2005; de Paula et al., 2007).

The conflicting roles of FRQ in positive and negative feedback are explained in terms of spatial location and posttranslational modification. Negative feedback through WCC includes nuclear FRQ that is hypophosphorylated and thus begins early after FRQ is expressed. Progressive phosphorylation of FRQ seems to turn the nuclear repressor FRQ into a cytoplasmic activator of WC-1 protein accumulation. This regulation may be through FRQ-mediated phophorylation of cytosolic WC-1 or WC-2, which could enhance their assembly and stability. The mechanism by which FRQ regulates the levels of *wc-2* mRNA is currently unknown (de Paula et al., 2007).

It is also noted that WC-1 and WC-2 regulate their own expression to form an additional loop in the FRQ/WCC oscillator. WC-2 stabilizes WC-1 protein by forming the WCC, and WC-1 negatively regulates transcription of *wc-2* through interaction with a putative repressor. The mechanism of this inhibition is yet to be discovered (Cheng et al., 2002; Cheng et al., 2003).

1.4.3 Nitrogen metabolism and the circadian clock

Genes involved in assimilation of nitrogen in *Neurospora crassa* are suggested to be circadian regulated. Independent metabolic rhythm - a FRQ/WC oscillator (FWO) has been observed for nitrogen reductase activity and diacylglycerol accumulation (Christensen et al., 2004). The enzymatic activity of nitrate reductase in *Neurospora crassa* grown on media containing nitrate as the only nitrogen source constitutes a metabolic rhythm with an approximately circadian period length that persists when the canonical circadian clock (FWO) is dysfunctional owing to prolonged exposure to constant light or to a genetic defect at the *frq* locus. The fact that the oscillations in NR activity have also been found in a *wc-1* mutant, but with lower amplitude, suggests this metabolic rhythm to be coupled to the circadian clock. Based on the observation that the nitrogen metabolite regulation protein (NMR), a transcriptional inhibitor of NR, is activated by glutamine - a metabolite of assimilated nitrate, a model for the NR metabolic feedback loop has been proposed (Lakin-Thomas and Brody, 2004).

2. Materials and methods

The expression profiles of N. crassa genes during its growth under a variety of constant conditions, including constant darkness (DD) and constant light (LL) on different substrates were represented in a text-file. The values were normalized and represented as log2. The data from this file was used for this study.

The analysis was based on the recognition and investigation of genes which showed statistically different expression profiles. For that purpose, two different tools were used.

The first one, the filtering function of the online software, BRB-array tool, was uploaded from http://linus.nci.nih.gov/BRB-ArrayTools.html. The recommendations from the manual to the BRB-array tool were used for choosing the parameters to analyze (Dr. Richard Simon and Corporation, November, 2012). The chosen parameters include lowess normalization, Log expression variation filter, $p = 0.01$, deletion of the subsets of genes with missing values (< 20%) and class comparison, $p=0.001$.

The second way to recognize statistically significantly different expression values was the use of arbitrary cut-off function of MatLab-script (Moler, 2004). This function allowed the cut-off of genes which showed expression values higher than 2 and lower than -2. The functions of MatLab "and" and "or" recognized groups of genes expressed similarly in particular conditions.

Indicated statistically significantly expressed genes were subjected to clustering analysis. Two clustering methods – hierarchical clustering and *k*-means clustering – were performed using Cluster 3.0 (Eisen et al., 1998).

Unsupervised hierarchical clustering was done with the method of average linkage after centering with the mean value. *k*-means clustering was performed with the following parameters: $k = 6$ clusters, number of runs – 100,000, and Similarity Metric – Euclidean distance. All named parameters were chosen after the recommendations from the manual to the Cluster 3.0 (http://bonsai.hgc.jp/~mdehoon/software/cluster/manual/) and Babu (Babu, 2004). Graphical representations of clustering results were generated using Java TreeView 1.1.6r4 (http://jtreeview.sourceforge.net). The analysis of clusters involved visual investigation and functional prediction by online software Functional Catalogue Database (Ruepp et al., 2004).

Particular genes of interest were extracted from the text-file with the expression data with usage of the Perl script written by Ruoff (Tisdall, 2001). The genes, related to a specific functional category were also investigated by use of Functional Catalogue Database, Broad Institute database (http://www.broadinstitute.org) and by Blast (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?CMD=Web&PAGE_TYPE=BlastHome).

3. Results

3.1 Groups analysis

To determine genes associated with nitrogen metabolism circuit, the full genome microarrays were used to monitor gene expression profiles during growth of *N.crassa* strains under a variety of constant conditions, including constant darkness (DD) and constant light (LL) on different substrates. RNA was harvested from mycelium of *Neurospora* that had first been grown in light on a Normal Vogel medium and then transferred to dark or light conditions, on a low sucrose medium containing just nitrate or just ammonium as the only nitrogen source for 16 hours. Nitrate and ammonium media were prepared on the basis of Vogel's medium with some changes. For the medium with the nitrate as the solo nitrogen source the normally used NH_4NO_3 was replaced by $NaNO_3$. In the medium with ammonium as the solo nitrogen source, NH_4NO_3 was replaced by an equimolar amount of ammonium chloride (Christensen et al., 2004). A kind of "standard sample" – DDNV - was grown on a Normal Vogel medium in darkness.

The microarray data were generated in Jay Dunlap's lab by Carol Kringelberg, Dartmouth Medical School, USA. Data were normalized and represented by a text file with gene annotations and expression values. The data from this file were used for this analysis. Two approaches were used in order to identify differentially expressed genes: cutoff at the arbitrary level and class comparison.

3.1.1 Recognition of differentially expressed genes by arbitrary cut-off

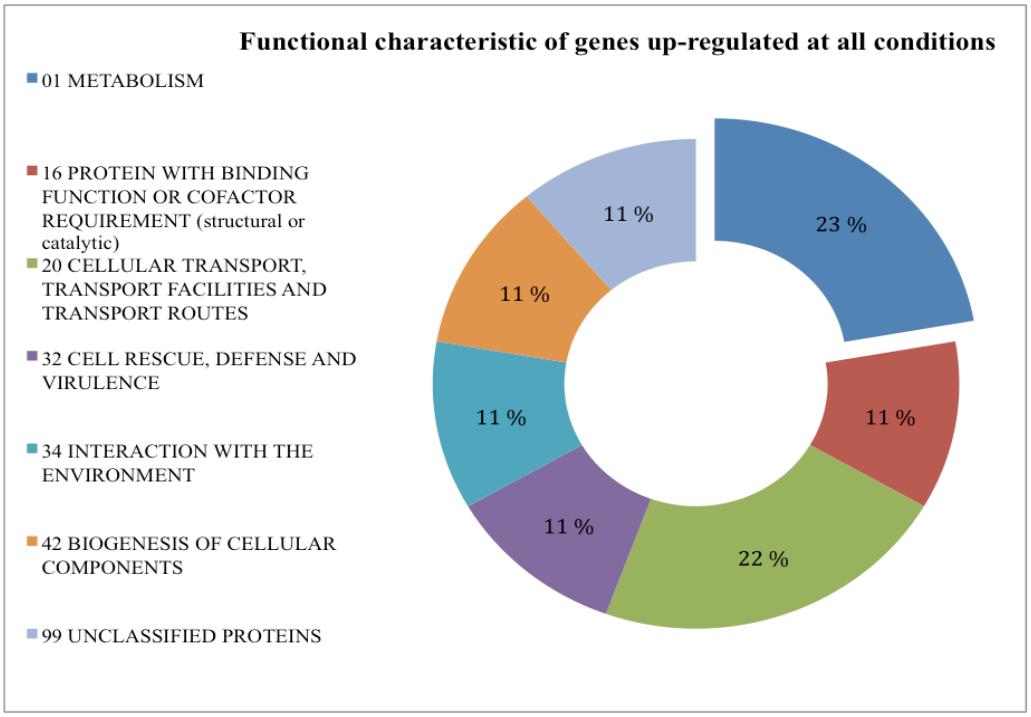
The first way to recognize genes showing differentially expressed values was done by arbitrary cut-off function of MatLab. For further analysis, genes were only used which showed expression values higher than 2 (arbitrary cut-off more then two) and lower than -2 (arbitrary cut-off less then minus two) (Appendix 1). It was also noted which genes showed similar behavior at all conditions (Table 1). The functions which were used for this purpose were "or" and "and".

	Expression values				
Gene ID	$DDNO_3$	$DDNH_4$	DDNV	$LLNH_4$	$LLNO_3$
NCU05897	2,546	3,693	3,149	3,716	3,396

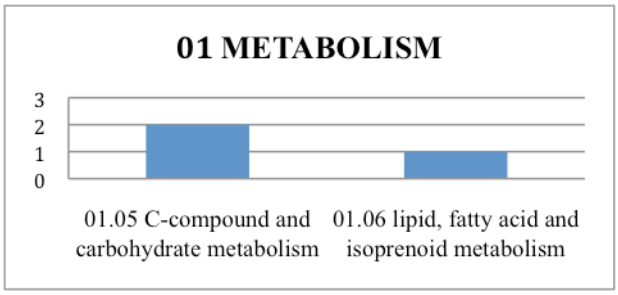
Gene ID	$DDNO_3$	$DDNH_4$	DDNV	$LLNH_4$	$LLNO_3$
NCU09040	3,993	3,042	3,117	3,513	3,235
NCU08087	3,234	2,707	2,867	2,020	2,391
NCU05627	2,052	2,455	2,047	3,340	2,971
NCU10941	-3,683	-2,105	-2,472	-2,674	-3,173
NCU04510	-2,461	-2,175	-2,495	-2,529	-2,958
NCU09508	-2,407	-2,510	-3,105	-2,409	Data is not found
NCU00732	-4,088	-2,552	-3,081	-2,783	-3,320
NCU02904	-2,813	-2,588	-2,317	-2,191	-2,278
NCU02930	-3,250	-3,363	-3,492	-3,590	-3,110
NCU06420	-3,456	-3,372	-2,646	-3,056	-2,649
NCU08129	-4,088	-5,206	-4,636	- 4,081	-4,513

Table 1. Genes expressed similarly at all conditions. List of genes that showed similar high or low expression values in all samples. The expression values are represented as log2 values for *Neurospora* genes. The columns represent different growth conditions. Values are rounded to three decimals, but full values were used in the analysis.

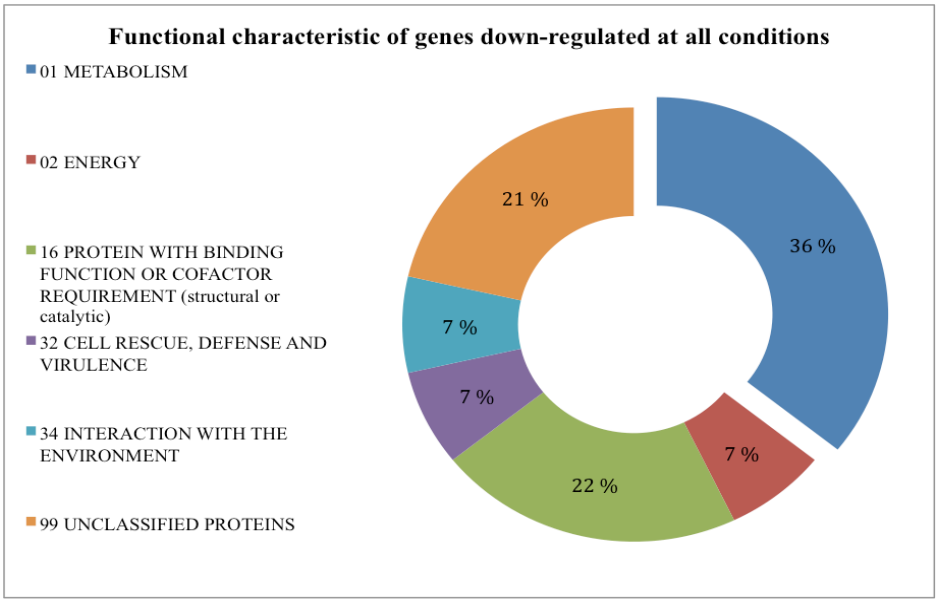
Genes that showed similar high/low expression levels at all conditions were subjected to the functional category (FunCat) analysis (Ruepp et al., 2004) and the results are represented in Figure 6.



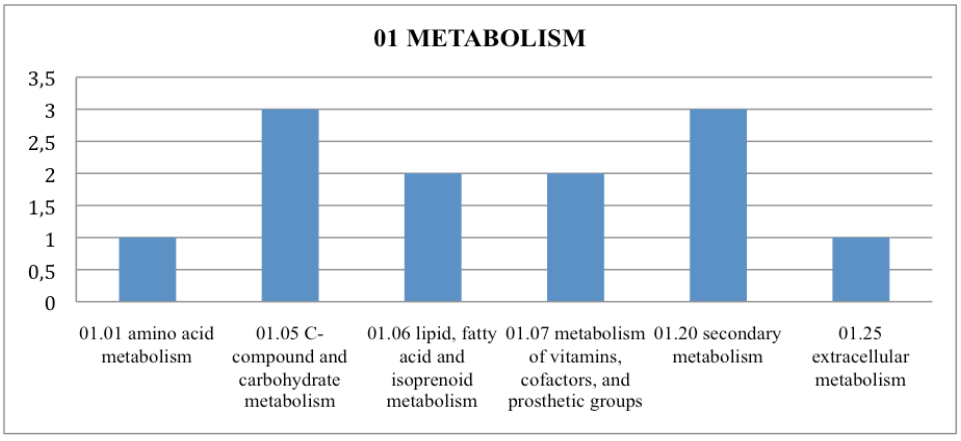
a



b



c



d

Figure 6. Functional analysis of genes up-regulated (a and b) and down-regulated (c and d) at all conditions. (a), (c) - Different sectors show the relative percentage of analyzed proteins which fell into each functional category. (b), (d) - Characteristics of the largest functional groups for proteins encoded by up-regulated (b) and down-regulated (d) genes. The scale to the left indicates the quantity of genes which fell into each subcategory.

It appears that the genes which up-regulated at all conditions are related to the C-compound and carbohydrate metabolism, particularly to polysaccharide metabolism and regulation of C-compound and carbohydrate metabolism. The cellular transport function is also related to the C-compound and carbohydrate transport, sugar transport. Another gene, up-regulated at all conditions (NCU09040), encodes the protein oxidoreductase, which plays a role in cell defense.

One unclassified protein (NCU08087) was described as a hypothetical protein with no assigned function in the Broad Institute database (http://www.broadinstitute.org). The Blast analysis (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?CMD=Web&PAGE_TYPE=BlastHome) showed that the protein contains the Glyco_hydro_26 superfamily, which is typical for proteins involved in carbohydrate transport and metabolism.

Genes which were down-regulated at all conditions seemed to fall into a variety of different functional categories. The largest of those is metabolism of C-compounds and carbohydrates. This data is in agreement with the fact that the medium contains a low concentration of sucrose, and the proteins involved in the metabolism of sugar are active/passive at all conditions.

The Broad Institute database and Blast analysis were used for predicting and identifying the functional assignment of the three unclassified genes which down-regulated at all conditions (NCU06420, NCU08129 and NCU10941). NCU06420 – described by Broad Institute database as a hypothetical protein, with the function related to response to light stimulus. No conserved domains are found for this gene by Blast. NCU08129 and NCU10941 were not found in the Broad Institute database. The Blast analysis of NCU08129 did not result in any conserved domains. The closest similarity to a protein with a known function – 36% (C6 transcription factor [*Colletotrichum gloeosporioides Nara gc5*]) is too low to make any predictions. The Blast analysis of the NCU10941 revealed three conserved Lysine Motifs, which characterize proteins involved in wide range of biological functions.

FunCat analysis was also performed for differently expressed genes. Generally, genes found by arbitrary cut-off more then two showed enrichment for proteins associated with C-compounds and carbohydrate metabolism, and genes encoding proteins which are responsible for cellular transport (Table 2).

It is clear that the quantity of up-regulated genes related to the stress response and detoxification, as well as the cellular transport, and particularly transported compounds (substrate), is higher in the nitrate medium. The functional subcategory “Cellular sensing and response to external stimulus” seems to be most represented among genes related to light and nitrate. Protein modification, cell cycle and translation are functions most represented in light on ammonium substrate.

The genes encoding proteins, which are related to nitrogen, sulfur and selenium metabolism are most represented in Normal Vogel medium in darkness, and in nitrate medium in light. That makes it reasonable to relate the genes involved in nitrogen metabolism to the presence of nitrate in the growth medium.

Genes found by arbitrary cut-off less then minus two showed enrichment for proteins associated with cellular transport, while the function of C-compounds and carbohydrate metabolism dominated in only one group of repressed genes - genes representing nitrate medium in darkness.

The highest amount of unclassified proteins was observed for genes in nitrate substrate, which makes it possible to suggest that there are many undiscovered mechanisms related to the nitrogen uptake and transport.

	$DDNH_4$ >2	$DDNH_4$ <-2	$DDNO_3$ >2	$DDNO_3$ <-2	DDNV >2	DDNV <-2	$LLNH_4$ >2	$LLNH_4$ <-2	$LLNO_3$ >2	$LLNO_3$ <-2
01 METABOLISM	5	2	8	9	5	3	1	7	9	3
01.01 amino acid metabolism			1	1			2	2		1
01.02 nitrogen, sulfur and selenium metabolism				1	2			2	2	1
01.03 nucleotide/nucleoside/nucleobase metabolism						1			1	
01.03.16 polynucleotide degradation						1				
01.04 phosphate metabolism	1		2	1	2		2			
01.05 C-compound and carbohydrate metabolism	4	1	6	6			7	3	5	3
01.06 lipid, fatty acid and isoprenoid metabolism			2	2		1		2		1
01.07 metabolism of vitamins, cofactors, and prosthetic groups	1	1	1	2	1	1		2	1	1
01.20 secondary metabolism		1		3		1				1
01.25 extracellular metabolism		1		1				1		1
02 ENERGY	1			1	1	1	1	1	1	1
02.01 glycolysis and gluconeogenesis	1						1			
02.10 tricarboxylic-acid pathway (citrate cycle, Krebs cycle, TCA cycle)								1		
02.11 electron transport and membrane-associated energy conservation				1						1
02.13 respiration				1	1				1	1
02.16 fermentation				1						1
02.45 energy conversion and regeneration				1		1				1
10 CELL CYCLE AND DNA PROCESSING							2			
10.03 cell cycle							2			
11 TRANSCRIPTION			1			1	1		1	
11.02 RNA synthesis			1			1	1		1	
12 PROTEIN SYNTHESIS						2	2			
12.04 translation						2	2			
12.10 aminoacyl-tRNA-synthetases						2				

	$DDNH_4$ >2	$DDNH_4$ <-2	$DDNO_3$ >2	$DDNO_3$ <-2	DDNV >2	DDNV <-2	$LLNH_4$ >2	$LLNH_4$ <-2	$LLNO_3$ >2	$LLNO_3$ <-2
14 PROTEIN FATE (folding, modification, destination)				1			5	1		1
14.01 protein folding and stabilization							1			
14.04 protein targeting, sorting and translocation							1			
14.07 protein modification				1			4	1		1
14.13 protein/peptide degradation							1			
16 PROTEIN WITH BINDING FUNCTION OR COFACTOR REQUIREMENT (structural or catalytic)		1	3	3		4	2	3	1	2
16.01 protein binding							1		1	
16.03 nucleic acid binding						1			1	
16.13 C-compound binding				1						1
16.17 metal binding		1		1		1	1	3		
16.19 nucleotide/nucleoside/nucleobase binding			1			2	1	1		
16.21 complex cofactor/cosubstrate/vitamine binding		1		3		2				2
18 REGULATION OF METABOLISM AND PROTEIN FUNCTION							2			
18.01. regulation by modification							1			
18.02 regulation of protein activity							2			
20 CELLULAR TRANSPORT, TRANSPORT FACILITIES AND TRANSPORT ROUTES	3		6	2	5	3	5	1	7	2
20.01 transported compounds (substrates)	3		4	2	5	2	4		7	1
20.03 transport facilities	3		6	1	3	2	4		6	1
20.09 transport routes	2		3	1	3	1	5		5	
30 CELLULAR COMMUNICATION/SIGNAL TRANSDUCTION MECHANISM			2				1			
30.01 cellular signalling			2				1			

	$DDNH_4$ >2	$DDNH_4$ <-2	$DDNO_3$ >2	$DDNO_3$ <-2	DDNV >2	DDNV <-2	$LLNH_4$ >2	$LLNH_4$ <-2	$LLNO_3$ >2	$LLNO_3$ <-2
32 CELL RESCUE, DEFENSE AND VIRULENCE	2	1	6	5	4	1	3		5	4
32.01 stress response	2		4	4	3		3		4	3
32.05 disease, virulence and defense			3	1	2				1	1
32.07 detoxification	1	1	2	4	2	1			2	3
34 INTERACTION WITH THE ENVIRONMENT	3	2	4	4	2	2	4		6	1
34.01 homeostasis		1	2	2		1	1		2	
34.07 cell adhesion				1		1				1
34.11 cellular sensing and response to external stimulus	3	1	2	2	2		3		4	
40 CELL FATE							2			
40.01 cell growth / morphogenesis							2			
41 DEVELOPMENT (Systemic)				1						
41.01 fungal/microorganismic development				1						
42 BIOGENESIS OF CELLULAR COMPONENTS			1	1			2			1
42.01 cell wall							1			
42.16 mitochondrion				1			1			1
42.25 vacuole or lysosome			1							
43 CELL TYPE DIFFERENTIATION	2		2		2		1		1	
43.01 fungal/microorganismic cell type differentiation	2		2		2		1		1	
99 UNCLASSIFIED PROTEINS	11	6	25	34	13	12	22	9	1	1

Table 2. Functional analysis of genes extracted by arbitrary cut-off function of MatLab. Differentially expressed genes identified by arbitrary cutoff were subjected to functional category representation analysis (FunCat). The saturation of red color indicates the quantity of genes representing each functional category, which can be found on the left. The columns represent the set of genes showing particular expression levels at the particular conditions, for example, $DDNH_4$ > 2 - the genes with expression levels more than 2 in the ammonium-rich medium in darkness. Unclassified proteins are shown in the bottom of the table.

Genes which showed some common patterns at different conditions were extracted with help of “or” and “and” functions of the MatLab script. That allowed the assignment of certain genes to specific conditions. Ten genes (Appendix 2) show similar high relative expression values in darkness. The FunCat analysis of these genes is shown in the Figure 7.

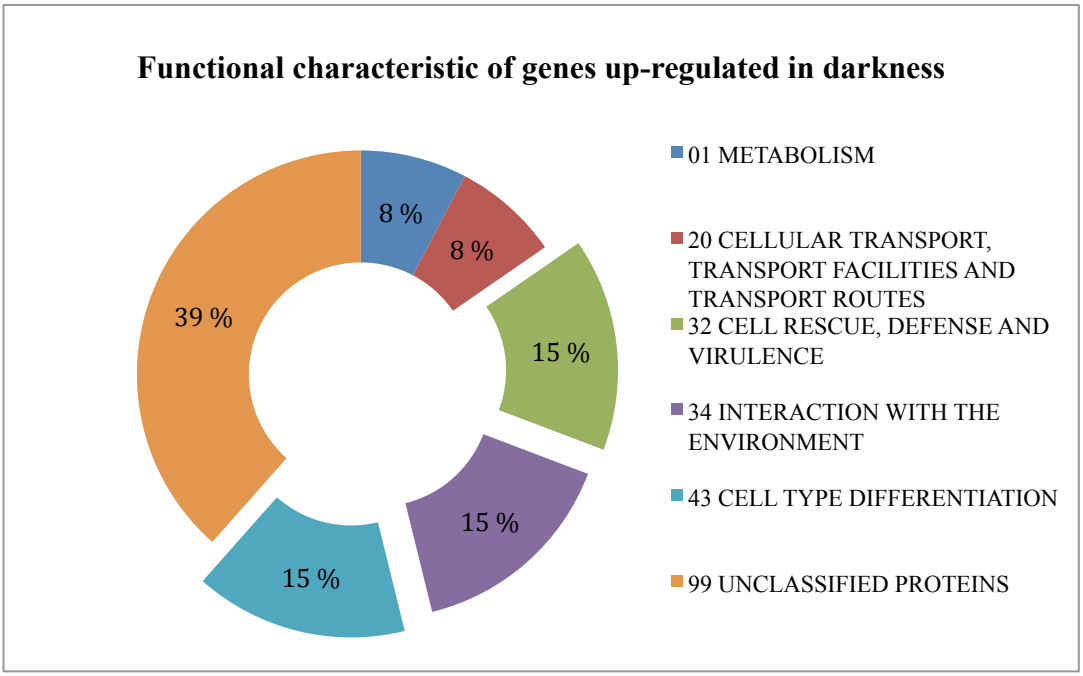


Figure 7. Functional characteristic of genes up-regulated in darkness. Genes which showed high expression values in darkness were subjected to functional category representation analysis (FunCat). Different sectors show the relative percentage of analyzed proteins which fell into each functional category, shown on the right.

Three of the functional categories are most represented: cell type differentiation, interaction with the environment and the cell rescue, defense and virulence. The most interesting category seems to be the interaction with the environment, since the analysis concern genes up-regulated in darkness. This functional category is represented by only one subcategory – cellular sensing and response to external stimulus.

There were also discovered six genes which down-regulated in dark conditions (arbitrary cut-off less then minus two for $DDNH_4$ and $DDNO_3$), also analyzed by FunCat. Analysis showed that the main functional category appeared to be metabolism, including C-compound and carbohydrate metabolism, metabolism of vitamin cofactors, and prosthetic groups, secondary metabolism and extracellular metabolism (data is shown in Appendix 3).

Eight genes (Appendix 2) showed similar high relative expression values while ammonium was present in the growth medium (arbitrary cut-off more then two for $DDNH_4$ and $LLNH_4$). The FunCat analysis of these genes is shown in the Figure 8.

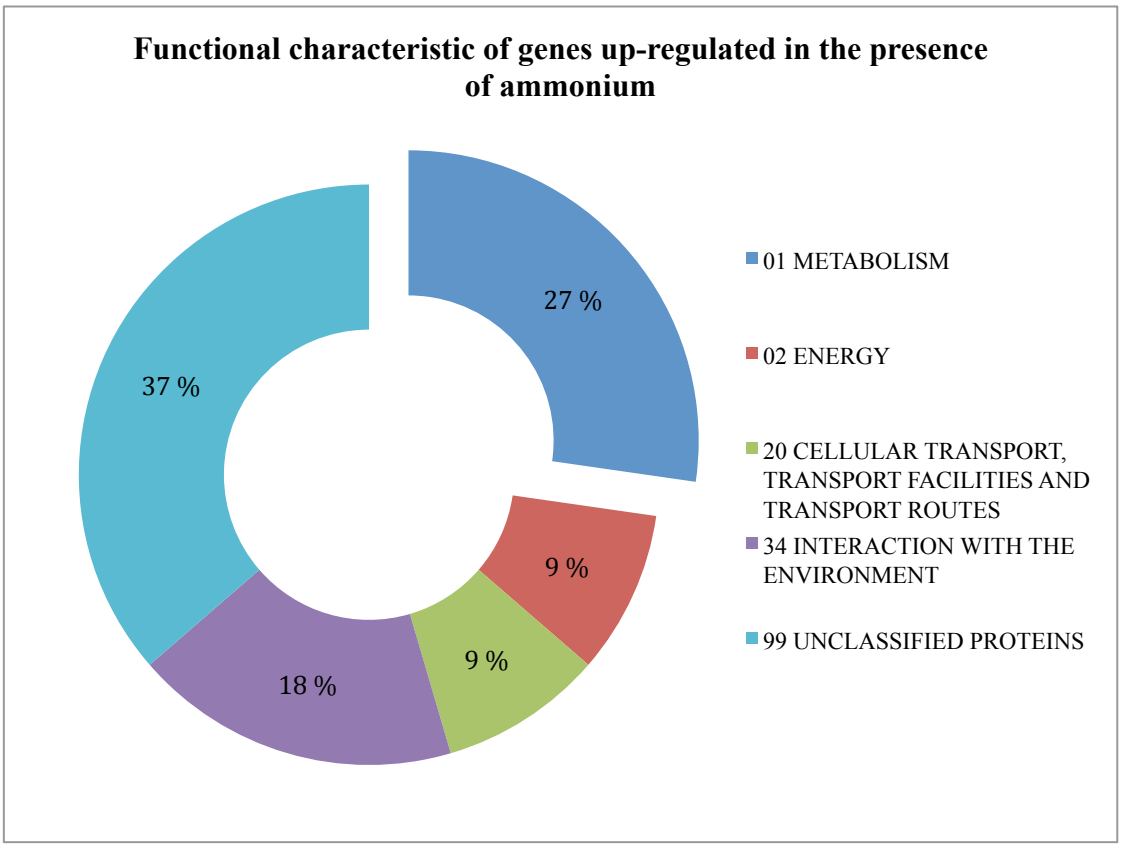


Figure 8. Functional analysis of genes up-regulated when only NH_4^+ is present in the medium. Genes, which showed high expression values in medium with only ammonium as nitrogen source were subjected to functional category representation analysis (FunCat). Different sectors show the relative percentage of analyzed proteins which fell into each functional category, shown on the right.

The main functional category appears to be metabolism, with the following subcategories: phosphate metabolism and C-compound and carbohydrate metabolism (Figure 9).

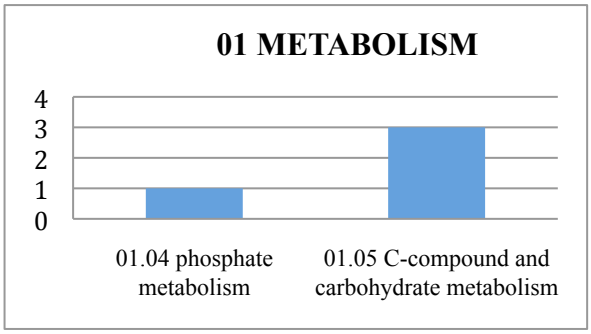


Figure 9. Characteristic of the largest functional group for genes up-regulated when only NH_4^+ is present in the medium. The largest functional category "Metabolism" is represented by two subcategories – phosphate metabolism and C-compound and carbohydrate metabolism. The scale to the left shows the quantity of genes which represent each particular function.

FunCat analysis of four genes which down-regulated in the presence of ammonium but showed significant difference in expression (arbitrary cut-off less then minus two for

$DDNH_4$ and $LLNH_4$), classified just one gene out of four. The function for the encoded protein is related to C-compound and carbohydrate metabolism, and extracellular metabolism (Appendix 3).

Ten genes (Appendix 2) show similar high relative expression values for genes expressed in the medium with the nitrate as the sole nitrogen source (arbitrary cut-off more then two for $DDNO_3$ and $LLNO_3$). The FunCat analysis of these genes is shown in Figure 10.

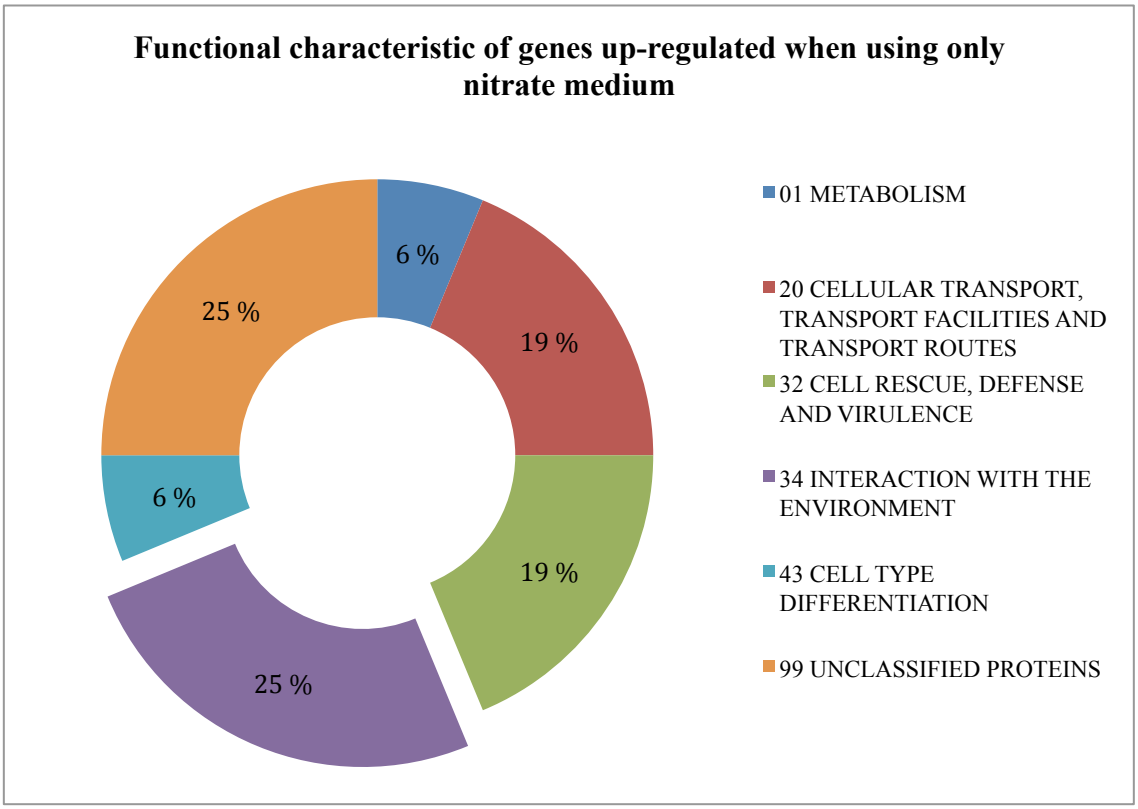


Figure 10. Functional analysis of genes up-regulated when nitrate was the only nitrogen source. Genes, which showed high expression values in medium with only nitrate as source of nitrogen were subjected to functional category representation analysis (FunCat). Different sectors show the relative percentages of analyzed proteins, which fell into each functional category, shown on the right.

The major functional category appears to be "interaction with the environment", with the following subcategories: homeostasis and cellular sensing and response to external stimulus (Figure 11).

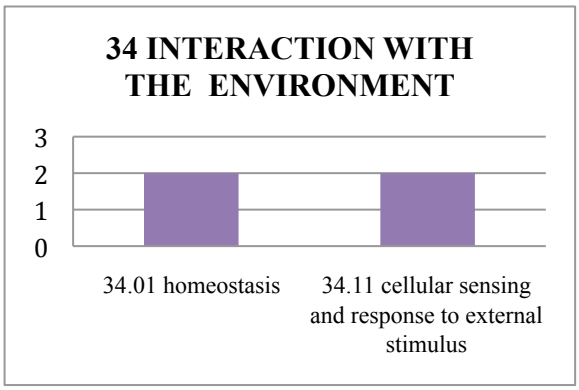


Figure 11. Characteristic of the largest functional group for genes up-regulated in the nitrate medium. The largest functional category "Interaction with the environment" is represented by two subcategories – homeostasis and cellular sensing and response to external stimulus. The scale to the left shows the quantity of genes which represent each particular function.

Arbitrary cut-off recognized thirteen genes as down-regulated in the nitrate conditions. Expression values of these genes are less then minus two while *Neurospora* was grown on the medium with nitrate as the only nitrogen source: $DDNO_3$ and $LLNO_3$. These genes were analyzed by FunCat. The two main functional categories appeared to be metabolism, including C-compound and carbohydrate metabolism, metabolism of vitamin cofactors, and prosthetic groups, secondary metabolism, extracellular metabolism, amino acid metabolism and cell rescue, defense and virulence, including stress response (Appendix 3).

Twelve genes (Appendix 2) showed similar high relative expression values in light (arbitrary cut-off more then two for $LLNH_4$ and $LLNO_3$). The FunCat analysis of these genes is shown in Figure 12.

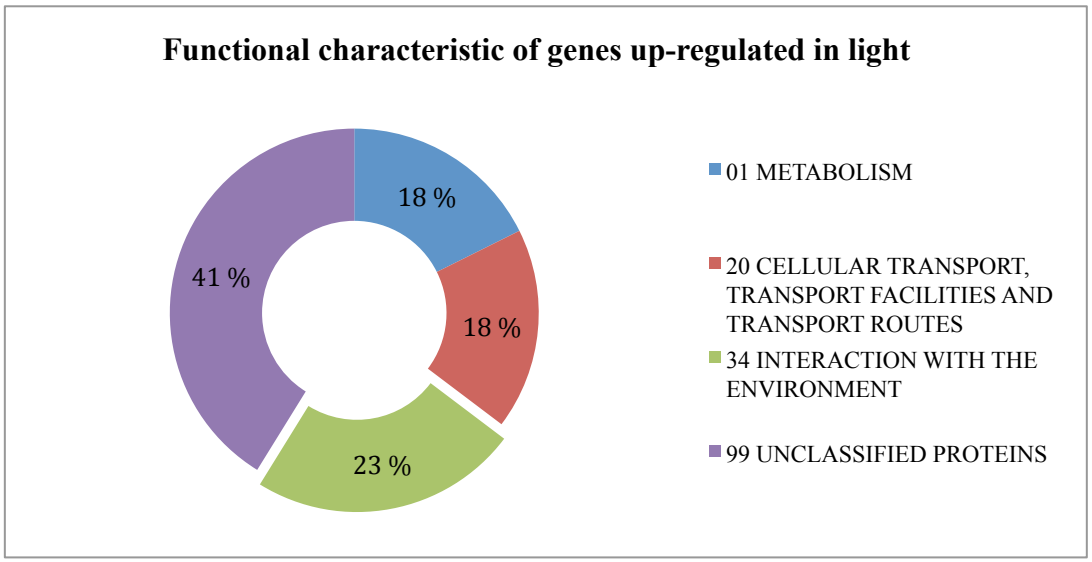


Figure 12. Functional analysis of genes up-regulated in light. Genes which showed high expression values at light were subjected to functional category representation analysis (FunCat). Different sectors show the relative percentage of analyzed proteins which fell into each functional category, shown on the right.

The functional category “Interaction with the environment” is represented most and includes homeostasis and cellular sensing and response to external stimulus (Figure 13).

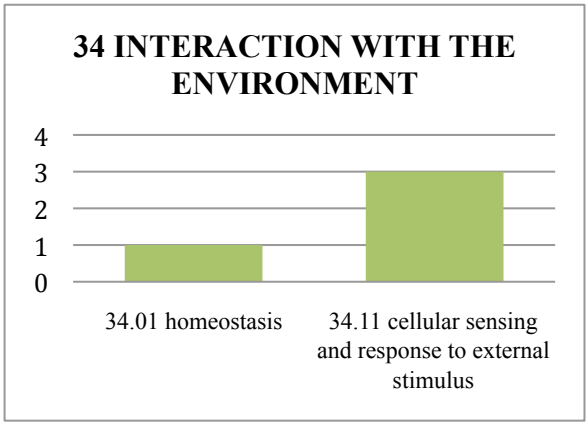


Figure 13. Characteristic of the largest functional group for genes up-regulated in light. The largest functional category “Interaction with the environment” is represented by two subcategories – homeostasis and cellular sensing and response to external stimulus. The scale to the left shows the quantity of genes which represent each particular function.

Six genes down-regulated in light which show significant statistical difference (arbitrary cut-off less then minus two for $LLNO_3$ and $LLNH_4$) were analyzed with help by FunCat. All six genes fell into the functional category ”Metabolism” and subcategory “C-compound and carbohydrate metabolism” (Appendix 3).

3.1.2 Group analysis using BRB-array tool

Since the data extracted by arbitrary cut-off did not include genes with expression values between 2 and -2, it can be assumed that it may not fully indicate the statistically significant differentially expressed genes. The arbitrary cut-off did not allow for calculation of the differences between expression values of single genes in different conditions., Therefore, genes which are highly up-regulated at some conditions, but did not cross the limit of 2 or -2, were not included in further analysis. On the other hand, genes which did not show a large variation among the five samples, but had expression values above 2/-2 at some particular conditions, were recognized by arbitrary cut-off and characterized. This is why one more technique of recognizing statistically significant differently expressed genes was used.

BRB-ArrayTools is an integrated software package for the analysis of DNA microarray data, developed by the Biometric Research Branch of the Division of Cancer Treatment & Diagnosis of the National Cancer Institute under the direction of Dr. Richard Simon. Among other functions, it allows the user to process expression data from multiple experiments (Dr. Richard Simon and Corporation, November, 2012).

Through filtering, normalization, and gene subsetting functions of the BRB-array tool it a total of 458 *N. crassa* genes were identified as differentially expressed (Apendix 4). For further analysis of the extracted genes, the mean values of three arrays were taken.

Analysis was performed by clustering - the most commonly used method for gene expression data analysis. Hierarchical clustering methods organize genes in a tree structure, based on their similarity. In *k*-means clustering, genes are organized into *k* clusters, where the number of clusters, *k,* needs to be chosen in advance. The software for hierarchical clustering and k-means clustering was uploaded from http://cluster2.software.informer.com/3.0/.

3.1.2.1 Unsupervised hierarchical clustering of the statistically different expressed genes

The basic idea of hierarchical clustering is to assemble a set of items (genes or arrays) into a tree. Short branches in the tree indicate the items, which are very similar to each other, and the long branches represent the items which are less similar. It starts with a calculation of the distance matrix between the gene expression data. Agglomerative hierarchical processing consists of repeated cycles where the two closest remaining items (those with the smallest distance) are joined by a node/branch of a tree, with the length of the branch set to the distance between the joined items. The two joined items are removed from the list of items being processed and replaced by an item that represents the new branch. The distances between this new item and all other remaining items are computed, and the process is repeated until only one item remains (Babu, 2004).

Similarity Metric for hierarchical clustering was set to correlation (centered) and the chosen clustering method was average linkage. With the help of hierarchical clustering, a cluster dendrogram was created, as well as a color image plot of all genes extracted by the BRB-array tool. For each cluster, the program provided a hyperlinked list of genes, which was saved and used in later analyses. Visually, hierarchical clustering showed that genes fell into six main clusters (Figure 14, Appendix 5).

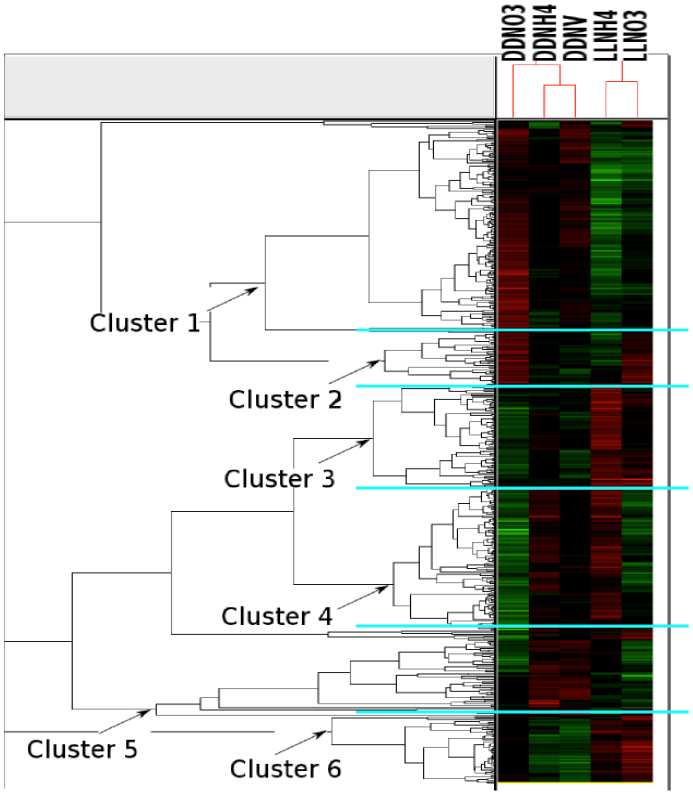


Figure 14. Hierarchical clustering analysis. Hierarchical clustering of 458 genes that show expression differences. Red indicates higher relative expression and green indicates lower relative expression. Lane 1: A 16 h *N. crassa* culture grown in darkness on medium with nitrate as the sole nitrogen source. Lane 2 and 3: A 16 h *N. crassa* culture grown in darkness on a medium with ammonium. Lanes 4 and 5: Expression profiles from cultures grown on ammonium-rich medium (Lane 4) and on a medium with nitrate as the sole nitrogen source (Line 5) in light.

In Figure 14 it can be seen that expression profile of genes at DDNV condition is most similar to the expression profile of $DDNH_4$ condition, which is outlined by the condition dendogram at the top of the figure. Conditions related to light, $LLNO_3$ and $LLNH_4$ are connected by one branch on the dendogram, while DDNV, $DDNH_4$ and $DDNO_3$ are connected by the other branch.

Genes from each cluster were associated to specific growth conditions by visual analysis based on Figure 14 and analyzed with the help of FunCat (Table 3). The visual analysis included comparison of the environmental conditions for clusters of genes with differentially high expression levels and grouping by a common pattern. For example, clusters of genes differentially high expressed at $LLNH_4$ and at $LLNO_3$ were associated with light; clusters of genes differentially high expressed at $LLNH_4$ and at $DDNH_4$ were associated with the presence of ammonium in substrate; et cetera.

	C1	C2	C3	C4	C5	C6
01 METABOLISM	30	4	12	23	12	12
01.01 amino acid metabolism	6	0	1	5	5	3
01.02 nitrogen, sulfur and selenium metabolism	3	0	2	2	0	1
01.03 nucleotide/nucleoside/nucleobase metabolism	4	0	2	0	1	2
01.04 phosphate metabolism	3	2	2	2	0	1
01.05 C-compound and carbohydrate metabolism	11	4	6	14	7	7
01.06 lipid, fatty acid and isoprenoid metabolism	4	2	1	6	3	4
01.07 metabolism of vitamins, cofactors, and prosthetic groups	2	0	0	2	2	2
01.20 secondary metabolism	6	0	2	6	3	2
01.25 extracellular metabolism	1	0	0	0	0	1
02 ENERGY	4	2	1	4	3	3
02.01 glycolysis and gluconeogenesis	0	0	0	0	1	2
02.07 pentose-phosphate pathway	0	0	0	0	1	1
02.09 anaplerotic reactions	0	0	0	1	0	0
02.10 tricarboxylic-acid pathway (citrate cycle, Krebs cycle, TCA cycle)	2	1	0	2	1	0
02.11 electron transport and membrane-associated energy conservation	0	0	0	2	0	0
02.13 respiration	1	1	1	1	0	0
02.16 fermentation	1	0	0	2	1	1
02.19 metabolism of energy reserves (e.g. glycogen, trehalose)	0	1	0	0	0	0
02.45 energy conversion and regeneration	1	0	0	1	0	1
10 CELL CYCLE AND DNA PROCESSING	4	0	2	5	0	3
10.01 DNA processing	2	0	0	3	0	1
10.03 cell cycle	3	0	2	2	0	2
11 TRANSCRIPTION	9	0	4	4	0	1
11.02 RNA synthesis	7	0	3	3	0	1
11.04 RNA processing	4	0	1	2	0	0
12 PROTEIN SYNTHESIS	2	0	3	3	0	1
12.01 ribosome biogenesis	0	0	2	3	0	1
12.04 translation	2	0	1	3	0	1
12.07 translational control	1	0	0	1	0	0
12.10 aminoacyl-tRNA-synthetases	0	0	1	0	0	0
14 PROTEIN FATE (folding, modification, destination)	14	4	1	6	2	3
14.01 protein folding and stabilization	3	1	0	0	0	1
14.04 protein targeting, sorting and translocation	4	2	0	2	0	0
14.07 protein modification	12	2	1	1	0	3
14.10 assembly of protein complexes	1	0	0	2	0	0
14.13 protein/peptide degradation	0	1	0	1	2	0

	C1	C2	C3	C4	C5	C6
16 PROTEIN WITH BINDING FUNCTION OR COFACTOR REQUIREMENT (structural or catalytic)	16	5	7	13	7	9
16.01 protein binding	6	1	1	4	1	4
16.03 nucleic acid binding	4	0	0	2	0	1
16.05 polysaccharide binding	2	0	1	0	0	0
16.09 lipid binding	0	1	2	0	0	0
16.13 C-compound binding	0	0	0	2	0	1
16.17 metal binding	5	1	0	4	2	4
16.19 nucleotide/nucleoside/nucleobase binding	4	3	0	2	1	1
16.21 complex cofactor/cosubstrate/vitamine binding	2	1	3	4	4	0
18 REGULATION OF METABOLISM AND PROTEIN FUNCTION	4	0	0	1	0	0
18.01 regulation by	1	0	0	0	0	0
18.02 regulation of protein activity	4	0	0	1	0	0
20 CELLULAR TRANSPORT, TRANSPORT FACILITIES AND TRANSPORT ROUTES	13	4	4	8	1	4
20.01 transported compounds (substrates)	8	2	3	8	0	2
20.03 transport facilities	2	3	1	3	0	1
20.09 transport routes	8	1	2	7	1	4
30 CELLULAR COMMUNICATION/SIGNAL TRANSDUCTION MECHANISM	3	3	1	3	2	2
30.01 cellular signalling	2	3	1	2	2	1
30.05 transmembrane signal transduction	0	0	0	1	0	1
30.07 regulation of signal transduction	1	0	0	0	0	0
32 CELL RESCUE, DEFENSE AND VIRULENCE	10	2	6	6	2	5
32.01 stress response	7	2	3	2	1	1
32.05 disease, virulence and defense	3	0	1	4	0	3
32.07 detoxification	2	0	3	4	1	1
32.10 degradation / modification of foreign (exogenous) compounds	1	0	0	0	1	1
34 INTERACTION WITH THE ENVIRONMENT	5	0	2	1	2	3
34.01 homeostasis	2	0	0	0	0	1
34.05 cell motility	1	0	1	0	0	0
34.11 cellular sensing and response to external stimulus	4	0	2	1	2	2
40 CELL FATE	2	1	1	0	0	1
40.01 cell growth / morphogenesis	1	1	1	0	0	1
40.10 cell death	2	1	0	0	0	0
41 DEVELOPMENT (Systemic)	0	0	0	0	1	0
41.01 fungal/microorganismic development	0	0	0	0	1	0

	C1	C2	C3	C4	C5	C6
42 BIOGENESIS OF CELLULAR COMPONENTS	6	3	2	5	1	1
42.01 cell wall	3	1	1	1	1	0
42.02 eukaryotic plasma membrane	1	0	0	0	0	0
42.04 cytoskeleton/structural proteins	1	1	1	0	0	0
42.07 endoplasmic reticulum	1	0	0	0	0	0
42.08 Golgi	1	0	0	0	0	0
42.10 nucleus	0	0	0	1	0	0
42.16 mitochondrion	0	1	1	2	0	0
42.25 vacuole or lysosome	0	0	0	1	0	0
42.27 extracellular / secretion proteins	1	0	0	0	0	1
43 CELL TYPE DIFFERENTIATION	2	2	1	1	0	1
43.01 fungal/microorganismic cell type differentiation	1	2	1	1	0	1
43.03 animal cell type differentiation	1	0	0	0	0	0
99 UNCLASSIFIED PROTEINS	70	21	31	53	28	20

Table 3. Functional analysis of gene clusters extracted by hierarchical clustering. Clusters of genes identified by hierarchical clustering were subjected to functional category representation analysis (FunCat).The saturation of red color indicates the quantity of genes representing each functional category, which can be found on the left. Genes from each cluster are associated to specific conditions by comparing their highest expression level. The first column represents Cluster 1, the second, Cluster 2, et cetera. Unclassified proteins are shown in the bottom of the table.

The first cluster of genes (Cluster 1; 141 genes) shows the highest expression levels in darkness. Functional category analysis of these genes showed an enrichment for proteins associated with primary metabolism - C-compound and carbohydrate metabolism and amino acid metabolism as well as proteins related to a secondary metabolism. Others represented functional categories – protein fate, including protein modification, protein and metal binding, biogenesis of cellular components, cellular transport and stress response.

The second cluster (C2) includes 35 genes that show the highest expression levels in medium with nitrate as the only nitrogen source. FunCat analysis of these genes shows that one functional category (C- compound and carbohydrate metabolism) was slightly enriched.

A third cluster (C3) of 72 genes shows the highest relative expression level with *N. crassa* grown on a Normal Vogel medium in light. FunCat analysis of these 72 genes shows an enrichment for proteins involved with carbon metabolism, as well as complex cofactor/cosubstance/vitamin binding and stress response.

The fourth cluster (C4; 99 genes) shows significant high expression level with *N.crassa* grown on ammonium-rich medium, except the Normal Vogel medium, which contains both ammonium and nitrate. This cluster includes variety of functional categories: C-compound and carbohydrate metabolism, amino acid metabolism, metabolism of vitamins, cell cycle and DNA processing, binding of nucleotides, cellular transport, detoxification and biogenesis of cellular components.

When *N.crassa* is grown on ammonium-rich and Normal Vogel medium, fifty-two genes, composing the fifth cluster (C5), are characterized by significant high expression levels in darkness. FunCat analysis of these 52 genes indicates that many of these proteins are involved in carbon metabolism, as well as in complex cofactor/vitamin binding.

The last cluster, C6, containing 47 genes, can be associated with the genes related to light, since it includes genes with the highest expression in light conditions. However, the FunCat analysis of these 47 genes shows many proteins involved in carbon metabolism, protein and metal binding, and cellular transport.

3.1.2.2 *k*-means clustering

According to Babu (Babu, 2004), one of the major criticisms of hierarchical clustering is that there is no compelling evidence that a hierarchical structure best suits grouping of the expression profiles. Therefore, an alternative non-hierarchical clustering was performed – *k*-means clustering, which requires pre-determination of the number of clusters.

In *k*-means clustering, the existing objects are grouped into predefined clusters instead of being organized into a hierarchical structure. Analysis starts with the random assignment of items (e.g. genes) into a chosen number of clusters (*k*). Further analysis proceeds by repeated application of a two-step process where:

1. the mean vector for all items in each cluster is computed;
2. items are reassigned to the cluster whose center is closest to the item.

In this case, the number of clusters was estimated first by performing hierarchical clustering of the data, as advised by Babu (Babu, 2004). The output of the software gave a graphical representation of genes in the form of a color image plot. All genes extracted by the BRB-array tool were arranged into the expected six clusters. Each cluster was assigned a

hyperlinked list of genes, which was saved and used in later analyses (Figure 15, Appendix 6).

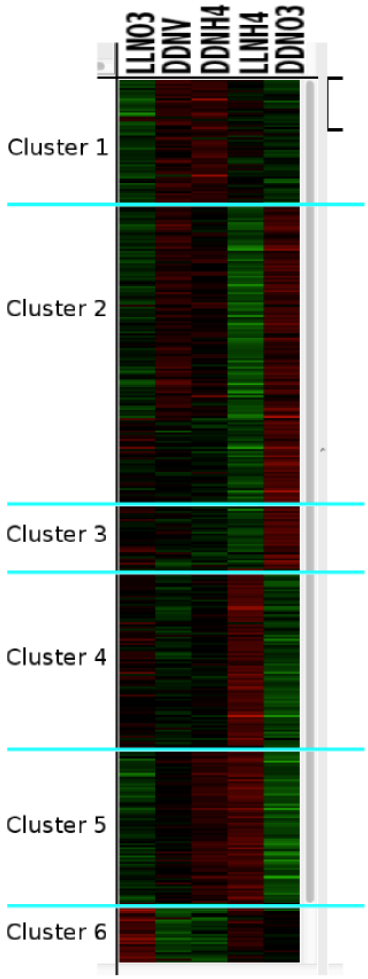


Figure 15. *k*-means clustering analysis of 458 genes that show expression differences. Red indicates higher relative expression and green indicates lower relative expression. Lane 1: A 16 h *N. crassa* culture grown in light on medium with nitrate as the sole nitrogen source. Lane 2 and 3: A 16 h *N. crassa* culture grown in darkness on ammonium and Normal Vogel medium respectively. Lanes 4: Expression profile from culture grown on ammonium-rich medium in light and Lane 5: A 16 h *N. crassa* culture grown in darkness on medium with nitrate as the sole nitrogen source.

Genes from each cluster revealed by *k*-means clustering were associated to specific conditions by visual analysis and subjected to functional category representation analysis (Table 4). The visual analysis included comparison of the environmental conditions for clusters of genes with differentially high expression levels and grouping them by a common pattern. For example, clusters of genes differentially high expressed at $LLNH_4$ and at $LLNO_3$ were associated with light; clusters of genes differentially high expressed at $LLNH_4$ and at $DDNH_4$ were associated with the presence of ammonium in substrate, et cetera.

	C1	C2	C3	C4	C5	C6
01 METABOLISM	13	18	15	17	21	14
01.01 amino acid metabolism	5	4	2	2	5	3
01.02 nitrogen, sulfur and selenium metabolism	0	1	2	2	2	1
01.03 nucleotide/nucleoside/nucleobase metabolism	2	2	2	2	0	2
01.04 phosphate metabolism	0	2	3	2	2	1
01.05 C-compound and carbohydrate metabolism	7	6	7	10	12	9
01.06 lipid, fatty acid and isoprenoid metabolism	3	2	5	2	5	4
01.07 metabolism of vitamins, cofactors, and prosthetic groups	2	2	0	0	2	2
01.20 secondary metabolism	3	6	0	3	6	2
01.25 extracellular metabolism	0	1	0	0	0	1
02 ENERGY	3	3	1	2	3	6
02.01 glycolysis and gluconeogenesis	1	0	0	0	0	2
02.07 pentose-phosphate pathway	1	0	0	0	0	1
02.09 anaplerotic reactions	0	0	0	0	1	0
02.10 tricarboxylic-acid pathway (citrate cycle, Krebs cycle, TCA cycle)	1	1	1	0	2	1
02.11 electron transport and membrane-associated energy conservation	0	0	0	0	2	1
02.13 respiration	0	1	0	1	1	2
02.16 fermentation	1	0	1	1	1	1
02.19 metabolism of energy reserves (e.g. glycogen, trehalose)	0	0	0	0	0	1
02.45 energy conversion and regeneration	0	1	2	0	1	1
10 CELL CYCLE AND DNA PROCESSING	2	2	2	2	3	4
10.01 DNA processing	1	1	1	0	2	2
10.03 cell cycle	1	1	2	2	1	3
11 TRANSCRIPTION	3	6	4	5	1	3
11.02 RNA synthesis	3	5	3	3	1	3
11.04 RNA processing	1	3	1	2	0	2
12 PROTEIN SYNTHESIS	1	1	1	4	1	1
12.01 ribosome biogenesis	1	0	0	3	1	1
12.04 translation	1	1	1	0	1	1
12.07 translational control	1	1	0	0	0	0
12.10 aminoacyl-tRNA-synthetases	0	0	0	1	0	0
14 PROTEIN FATE (folding, modification, destination)	4	7	10	2	4	4
14.01 protein folding and stabilization	0	1	3	0	0	1
14.04 protein targeting, sorting and translocation	0	1	4	1	2	1
14.07 protein modification	0	6	8	1	1	3
14.10 assembly of protein complexes	1	1	0	0	1	0
14.13 protein/peptide degradation	3	0	1	0	0	0
16 PROTEIN WITH BINDING FUNCTION OR COFACTOR REQUIREMENT (structural or catalytic)	9	9	12	11	9	12
16.01 protein binding	2	3	4	2	3	4
16.03 nucleic acid binding	2	2	2	0	1	2
16.05 polysaccharide binding	0	1	1	1	0	0
16.09 lipid binding	0	0	0	0	0	1
16.13 C-compound binding	0	0	0	1	1	1
16.17 metal binding	2	2	5	3	2	4
16.19 nucleotide/nucleoside/nucleobase binding	2	3	4	2	1	2
16.21 complex cofactor/cosubstrate/vitamine binding	4	1	2	4	3	1
18 REGULATION OF METABOLISM AND PROTEIN FUNCTION	0	1	3	0	1	0
18.01 regulation by	0	0	1	0	0	0
18.02 regulation of protein activity	0	1	3	0	1	0

	C1	C2	C3	C4	C5	C6
20 CELLULAR TRANSPORT, TRANSPORT FACILITIES AND TRANSPORT ROUTES	1	9	8	6	6	6
20.01 transported compounds (substrates)	1	5	3	4	6	4
20.03 transport facilities	0	1	3	1	3	2
20.09 transport routes	1	5	5	4	5	5
30 CELLULAR COMMUNICATION/SIGNAL TRANSDUCTION MECHANISM	2	2	4	2	2	2
30.01 cellular signalling	2	1	4	1	2	1
30.05 transmembrane signal transduction	0	0	0	1	0	1
30.07 regulation of signal transduction	0	1	0	0	0	0
32 CELL RESCUE, DEFENSE AND VIRULENCE	3	7	5	8	3	6
32.01 stress response	1	5	4	3	2	2
32.05 disease, virulence and defense	1	2	1	1	3	3
32.07 detoxification	1	2	1	5	2	1
32.10 degradation / modification of foreign (exogenous) compounds	1	1	0	0	0	1
34 INTERACTION WITH THE ENVIRONMENT	2	4	1	2	1	3
34.01 homeostasis	0	2	0	0	0	1
34.05 cell motility	0	1	0	1	0	0
34.11 cellular sensing and response to external stimulus	2	3	1	2	1	2
38 TRANSPOSABLE ELEMENTS, VIRAL AND PLASMID PROTEINS	0	0	0	0	0	2
38.07 proteins necessary for transposon movement	0	0	0	0	0	2
40 CELL FATE	0	1	2	1	0	1
40.01 cell growth / morphogenesis	0	0	2	1	0	1
40.10 cell death	0	1	2	0	0	0
41 DEVELOPMENT (Systemic)	1	0	0	0	0	0
41.01 fungal/microorganismic development	1	0	0	0	0	0
42 BIOGENESIS OF CELLULAR COMPONENTS	2	5	4	3	3	2
42.01 cell wall	1	3	1	2	0	1
42.02 eukaryotic plasma membrane	0	1	0	0	0	0
42.04 cytoskeleton/structural proteins	0	1	1	1	0	0
42.07 endoplasmic reticulum	0	0	1	0	0	0
42.08 Golgi	0	0	1	0	0	0
42.10 nucleus	1	0	0	0	0	0
42.16 mitochondrion	0	0	1	1	2	0
42.25 vacuole or lysosome	0	0	0	0	1	0
42.27 extracellular / secretion proteins	0	1	0	0	0	1
43 CELL TYPE DIFFERENTIATION	0	2	2	1	1	1
43.01 fungal/microorganismic cell type differentiation	0	1	2	1	1	1
43.03 animal cell type differentiation	0	1	0	0	0	0
99 UNCLASSIFIED PROTEINS	32	60	38	43	42	28

Table 4. Functional analysis of gene clusters extracted by *k*-means clustering. Clusters of genes identified by *k*-means clustering were subjected to functional category representation analysis (FunCat). The saturation of red color indicates the quantity of genes representing each functional category, which can be found on the left. Genes from each cluster are associated to specific conditions by comparing their highest expression level. The first column represents Cluster 1, the second, Cluster 2, et cetera. Unclassified proteins are shown in the bottom of the table.

The first cluster of genes (C1; 60 genes) shows the highest expression levels in darkness with *N.crassa* grown on ammonium-rich and Normal Vogel medium and can be

associated with Cluster 5 from the hierarchical clustering. Of these genes, 49 overlapped with the 52 genes in the C5 cluster from the hierarchical clustering. Functional category (FunCat) analysis of these genes showed an enrichment for proteins associated with primary metabolism - amino acid metabolism, C-compound and carbohydrate metabolism, as well as proteins responsible for RNA-synthesis, protein degradation and protein binding.

The second cluster (C2) includes 104 genes that show the highest expression levels in darkness, mainly on the medium with nitrate as the only nitrogen source. FunCat analysis of these genes showed variety of functional categories. The most represented are those for metabolism of C-compounds and carbohydrates, secondary metabolism, RNA synthesis and processing, protein modification, cellular transport, biogenesis of the cell wall and stress response. The cluster can be associated with Cluster 1 from the hierarchical clustering as 99 of its genes overlap with C1 from hierarchical clustering.

The third cluster of genes, had higher expression values in nitrate conditions, especially in darkness (C3; 73 genes). FunCat analysis shows the presence of genes encoding the following functional categories: metabolism of C-compounds and carbohydrates, metabolism of lipids, fatty acids and isoprenoids; protein modification, metal and nucleotide binding function. Functional categories which are represented less, but not least are cellular transport, communication and stress response. This cluster can be associated with Cluster 2 from the hierarchical clustering with an overlap of 29 genes.

Cluster 4 (C4) may be possibly related to light, since there are 87 genes, which showed higher levels of expression in light, and lower levels of expression in darkness. Functional category analysis of the up-regulated gene set revealed a significant overrepresentation of genes annotated to be involved in C-compound and carbohydrate metabolism and detoxification. Cluster 6 from the hierarchical clustering which seem to contain genes related to light does not overlap with Cluster 4 from *k*-means clustering..

Cluster 5 (C5; 75 genes): contains genes highly expressed at $LLNH_4$ and $DDNH_4$, but not at DDNV. It overlaps with Cluster 4 from hierarchical clustering (74 genes) and is also enriched with C-compound and carbohydrate metabolism and cellular transport.

Cluster 6 (C6; 60 genes) can also be associated with light, as well as Cluster 6 from hierarchical clustering with an overlap of 48 genes. Functional category analysis of the 60

genes from cluster 6 revealed an abundance of the genes related to the C-compound and carbohydrate metabolism. This gene set also seems to be enriched with energy-related genes. The high percentage of overlapping genes within clusters from two different clustering methods indicates the relative robustness of the clustering methods. The genes representing overlapping of clusters are listed in Appendix 7.

147 genes associated to light by visual analysis of clusters constructed by *k*-means clustering (cluster 4 and cluster 6) were compared to 314 genes, showed to have strong early or late light responces by Chen (Chen et al., 2009). Only five of the 147 genes appeared to be in common. The list of these five genes and their short descriptions are represented in Table 5.

Locus	Description
NCU05770.5	cat-2, peroxidase/catalase 2
NCU00766.5	hypothetical protein, no conserved domains
NCU08824.5	molybdopterin binding domain-containing protein
NCU03803.5	sorbitol utilization protein sou-2
NCU09306.5	hypothetical protein with domain of unknown function

Table 5. List of genes related to light (overlap). The list of genes which represents overlap between results from this thesis and experimental data by Chen (Chen et al., 2009). The first column indicates gene locus. The second column represents short description of encoded protein taken from Broad institute database.

3.2 Analysis of genes showing extreme expression values

Genes that showed extremely high or low values of expression at different conditions were investigated with the help of Broad Institute database and MIPS Functional catalogue. A short summary of results is shown in the table below (Table 6).

Locus	**Conditions**	**Expression value**	**Gene name**	**Functional characteristic (Broad Institute, MIPS)**
1	**2**	**3**	**4**	**5**
NCU05143	DDNH4	3,325	rds1	stress response
	DDNO3	3,830		
	DDNV	3,615		
	LLNO3	2,763		

1	2	3	4	5
NCU04963	DDNH4	2,809	high-affinity glucose transporter	regulation of C-compound and carbohydrate metabolism C-compound and carbohydrate transport cellular import perception of nutrients and nutritional adaptation
NCU07345	DDNH4	2,798	hypothetical protein	-
	DDNO3	3,669		
	DDNV	3,547		
NCU02500	DDNH4	2,717	clock-controlled gene-4	pheromone response, mating-type determination, sex-specific proteins rhythm (e.g. circadian, ultradian)
	DDNO3	3,304		
	LLNH4	3,787		
	LLNO3	3,781		
NCU01873	DDNH4	2,657	hypothetical protein	C-compound and carbohydrate metabolism polysaccharide metabolism
NCU05969	DDNH4	-2,142	glycosylhydrolase family 61-9	polysaccharide metabolism extracellular polysaccharide degradation
NCU05763	DDNH4	-2,203	hypothetical protein	-
NCU08037	DDNH4	-2,389	hypothetical protein	-
	DDNV	-2,315		
	LLNH4	-3,494		
NCU07449	DDNH4	-3,006	hypothetical protein	-
	DDNO3	-3,012		
	DDNV	-2,828		
NCU05768	DDNH4	-4,995	plenty of it-2	sporocarp development involved in sexual reproduction
NCU02329	DDNO3	3,768	hypothetical protein	-
	DDNV	2,886		
	LLNO3	2,778		
NCU07894	DDNO3	3,148	oligopeptide transporter 2	peptide transport cellular sensing and response to external stimulus chemoperception and response pheromone response, mating-type determination, sex-specific proteins development of asco- basidio- or zygospore
NCU04533	DDNO3	-2,893	abundant perithecial protein	cell-cell adhesion
	DDNV	-2,382		

1	2	3	4	5
NCU07053	DDNO3	-3,263	aldehyde dehydrogenase	degradation of polyamines C-2 compound and organic acid catabolism lipid, fatty acid and isoprenoid metabolism metabolism of vitamins, cofactors, and prosthetic groups metabolism of acetic acid derivatives metabolism of nonprotein amino acids electron transport and membrane-associated energy conservation respiration alcohol fermentation energy conversion and regeneration C-compound binding NAD/NADP binding stress response mitochondrion
NCU06301	DDNO3	-3,373	hypothetical protein	-
NCU09620	DDNO3	-3,483	hypothetical protein	-
NCU05706	DDNV	2,854	glutathione S-transferase	metabolism of vitamins stress response
NCU07928	DDNV	2,787	hypothetical protein	
NCU09495	DDNV	-2,300	set-domain histone methyltransferase-6	response to dsRNA
NCU00915	DDNV	-2,581	aspartyl-tRNA synthetase	translation aminoacyl-tRNA-synthetases nucleotide/nucleoside/nucleobase binding
NCU04998	LLNH4	3,469	hypothetical protein	-
NCU07869	LLNH4	3,355	INSIG domain-containing protein	function of proteins with this domain: egulation of cholesterol synthesis over a wide range of sterol concentrations
NCU04268	LLNH4	3,312	hypothetical protein	-
NCU08986	LLNH4	3,302	hypothetical protein	-
NCU07743	LLNH4	-2,501	hypothetical protein	-
NCU05788	LLNH4	-2,508	hypothetical protein	-
NCU05969	LLNH4	-2,576	hypothetical protein	-
NCU09620	LLNH4	-2,732	hypothetical protein	-
NCU07894	LLNO3	3,202	hypothetical protein	-
NCU05627	LLNO3	2,971	high affinity glucose transporter ght1	C-compound and carbohydrate transport non-vesicular cellular import eukaryotic plasma membrane / membrane attached
NCU06301	LLNO3	-2,457	hypothetical protein	-
NCU07569	LLNO3	-2,546	hypothetical protein	stress response

1	2	3	4	5
NCU07338	LLNO3	-2,584	alpha-1,6-mannosyltransferase Och1	C-compound and carbohydrate metabolism protein modification
NCU07053	LLNO3	-2,619	hypothetical protein	-
NCU00175	LLNO3	-3,313	hypothetical protein	-

Table 6. Analysis of genes which showed extremely high/low expression values. Genes showed extremely high or low expression values at different conditions were analyzed with the help of internet databases. The first column represents the list of genes which were extracted with help of cutoff function of MatLab. Second column indicates the environment conditions. Third column indicates expression value of each particular gene in particular condition (values are rounded to three decimals, but full values were used in the analysis) and the fourth and fifth columns represent the data found in internet databases Broad institute and MIPS related to each particular gene.

As can be seen from the table above, there are some genes, which are highly expressed/depressed at more than one condition. One genes, related to stress response (NCU05143), is highly up-regulated in darkness and in nitrate medium in light, while another gene, related to pheromone response, mating-type determination and rhythm (NCU02500), is highly up-regulated in all conditions, except Normal Vogel medium in darkness. A gene related to the cell-cell adhesion (NCU04533) is highly up-regulated in darkness, in both nitrate and Normal Vogel medium.

Genes, which were not found in the databases were explored for the presence of conserved domains or for similar genes which encode proteins with a known function (Table 7).

Lokus	Conditions	Expression value	Conserved domains	Identical proteins
1	2	3	4	5
NCU07345	DDNH4	2,798	No conserved domains identified	Endothelial cells scavenger receptor [*Crassostrea gigas*], Identity: 28%
	DDNO3	3,669		
	DDNV	3,547		
NCU05763	DDNH4	-2,203	No conserved domains identified	-
NCU08037	DDNH4	-2,389	No conserved domains identified	c6 zinc finger domain containing protein [*Ophiostoma piceae UAMH 11346*], Identity: 72%
	DDNV	-2,315		
	LLNH4	-3,494		
NCU07449	DDNH4	-3,006	No conserved domains identified	putative pathogenicity protein [*Fusarium oxysporum*], Identity: 42%
	DDNO3	-3,012		
	DDNV	-2,828		

1	2	3	4	5
NCU02329	DDNO3	3,768	DUF1993 - domain of unknown function	helix-turn-helix- domain containing protein type [*Ophiostoma piceae UAMH 11346*], Identity: 42%
	DDNV	2,886		
	LLNO3	2,778		
NCU06301	DDNO3	-3,373	Glutathione-dependent formaldehyde-activating enzyme	glutathione-dependent formaldehyde-activating enzyme [*Colletotrichum gloeosporioides Cg-14*], Identity: 49%
NCU09620	DDNO3	-3,483	No conserved domains identified	Aminotransferase class I and II [uncultured organism], Identity: 30%
NCU07928	DDNV	2,787	No conserved domains identified	Eukaryotic translation initiation factor 3 subunit-like protein [*Chaetomium thermophilum var. thermophilum DSM 1495*], Identity: 63%
NCU04998	LLNH4	3,469	K+-dependent Na+/Ca+ exchanger; Functions: Transport and binding proteins, Cations and iron carrying compounds	-
NCU04268	LLNH4	3,312	Peroxiredoxin (PRX)-like 2 family AhpC/TSA antioxidant enzyme Functions: protective antioxidant role in cells	fmHP [*Metarhizium acridum CQMa 102*], Identity: 53%
NCU08986	LLNH4	3,302	No conserved domains identified	Vacuolar fusion protein MON1 like protein A [*Chelonia mydas*], Identity: 34%
NCU07743	LLNH4	-2,501	Taurine catabolism dioxygenase TauD, TfdA family; Clavaminic acid synthetase (CAS) -like; gamma-butyrobetaine hydroxylase ;Members of this protein family are gamma-butyrobetaine hydroxylase, both bacterial and eukarytotic. This enzyme catalyzes the last step in the conversion of lysine to carnitine. Carnitine can serve as a compatible solvent in bacteria and also participates in fatty acid metabolism.	Taurine catabolism dioxygenase TauD [*Verticillium dahliae VdLs.17*], Identity: 44%
NCU05788	LLNH4	-2,508	No conserved domains identified	-
NCU05969	LLNH4	-2,576	Glycosyl hydrolase family 61	Glycoside hydrolase family 61 protein [*Thielavia terrestris NRRL 8126*], Identity: 75%

1	2	3	4	5
NCU09620	LLNH4	-2,732	No conserved domains identified	Aminotransferase class I and II [uncultured organism], Identity: 30%
NCU07894	LLNO3	3,202	OPT family; function: Transport and binding proteins, Amino acids, peptides and amines	oligopeptide transporter 2 [*Neurospora crassa OR74A*], Identity: 98%
NCU06301	LLNO3	-2,457	Glutathione-dependent formaldehyde-activating enzyme	Glutathione-dependent formaldehyde-activating [*Colletotrichum orbiculare MAFF 240422*], Identity: 48%
NCU07053	LLNO3	-2,619	Aldedh multi-domain Functions: catalyse the oxidation of a broad range of aldehydes into their corresponding carboxylic acids with the reduction of their cofactor	Aldehyde dehydrogenase [*Fusarium oxysporum f. sp. cubense race 4*], Identity: 66%
NCU00175	LLNO3	-3,313	No conserved domains identified	Repetitive proline-rich cell wall protein [*Ophiocordyceps sinensis CO18*],Identity: 38%

Table 7. Bioinformatical analysis of genes encoding hypothetical proteins. Genes, which showed extremely high or low values of expression and were not characterized by internet databases MIPS and Broad institute were analyzed by Blast for presence of conserved regions and the existence of identical proteins with known function. The first column represents the list of genes which were extracted with help of cutoff function of MatLab. Second column indicates the environment conditions. Third column indicates expression value of each particular gene in particular condition (values are rounded to three decimals, but full values were used in the analysis). Conserved domains and identical proteins with known function are noted in column 4 and 5 respectively.

Some hypothetical proteins (NCU07345 and NCU07449) expressed at darkness seem to be related to pathogenesis and cell defense, while another one (NCU08037), most expressed in the presence of ammonium, can be connected to a large scale of functions. The low identity percentage makes it difficult to predict functions of the hypothetical proteins.

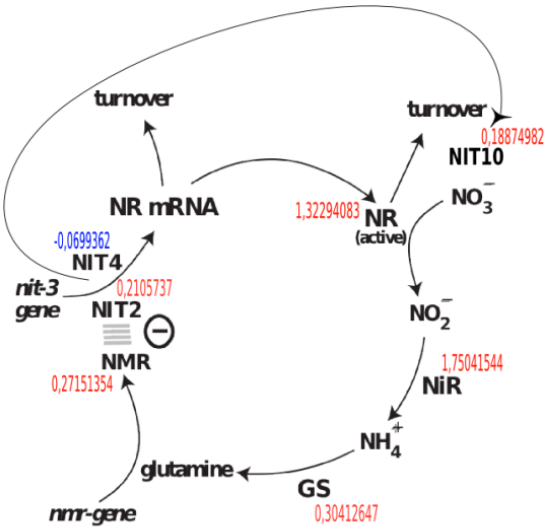
Summarizing the data above, we can conclude that the five most expressed genes in ammonium-rich medium represent proteins related to sugar transport, stress response, pheromone response and rhythm. The hypothetical protein highly expressed in the presence of ammonium, seems to be related to cell defense. Proteins depressed in that condition are involved in polysaccharide metabolism and sexual reproduction.

Generally, most proteins encoded by genes, showing extreme values of expression are involved in primary metabolism (C-compounds, lipids, aminoacids), stress response and biological rhythms.

3.3 Analysis of genes related to different functional activity

3.3.1 Genes related to nitrogen metabolism

Genes related to the uptake of nitrogen are predicted to be up-regulated under nitrate conditions, since in the absence of the preferable nitrogen source – ammonium, positively acting transcriptional factors NIT-2 and NIT-4 activate the transcription of the *nit-3* gene, the structural protein of nitrate reductase (NR) and *nit-10* – nitrate transporter (Gao-Rubinelli and Marzluf, 2004). The expression values of genes related to the nitrogen metabolism circuit were extracted from the microarray data with the help of Perl script created by Ruoff and presented in the figure below (Figure 16).



a

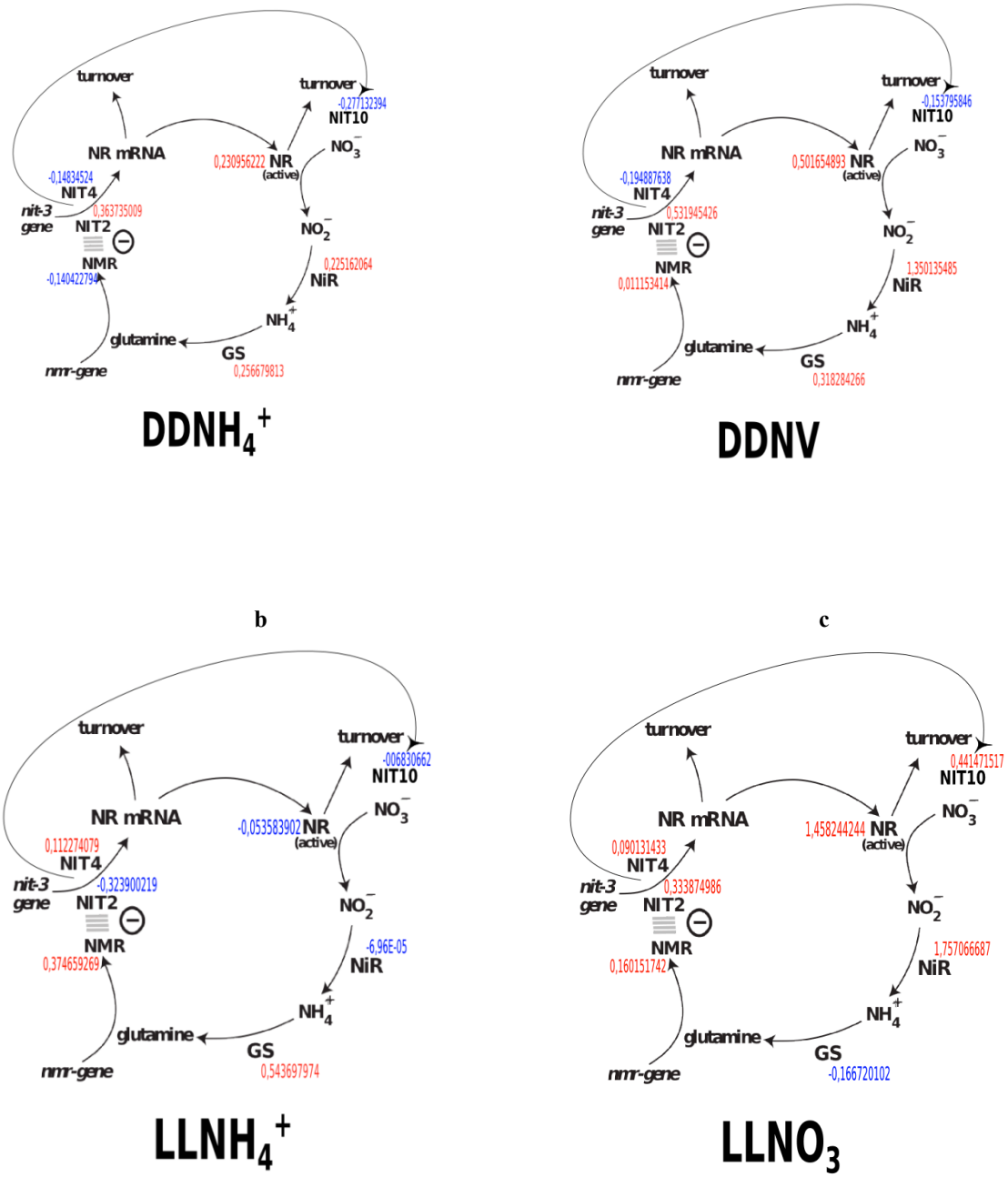
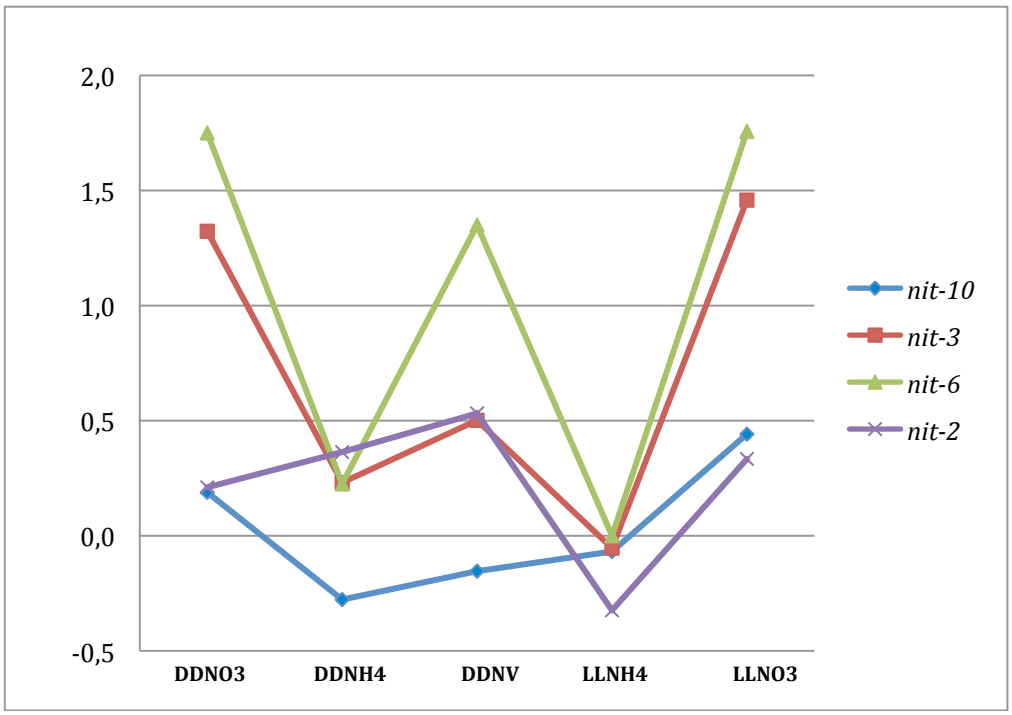


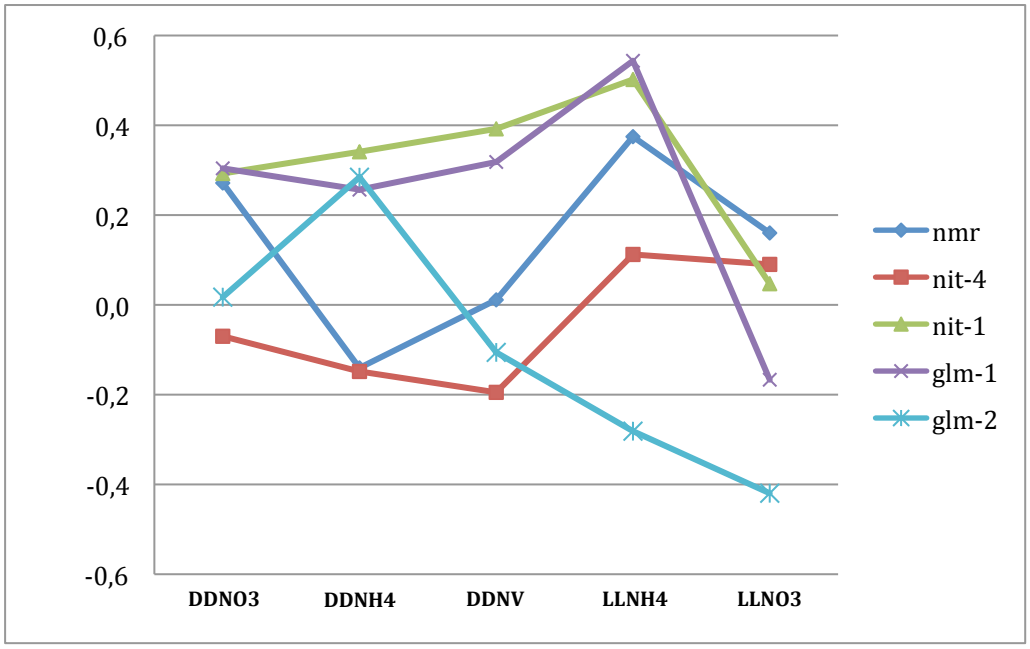
Figure 16. Expression values of genes from the nitrogen metabolism circuit, in different conditions. Expression values of genes related to nitrogen metabolism, extracted by Perl were applied on the model for the NR negative feedback loop, suggested by Christensen(Christensen et al., 2004). Figures a-e show the expression values of genes at different environmental conditions. The red color indicates expression values above zero. Blue color indicates negative expression values.

From Figure 16 it is clear that the expression of genes related to the nitrogen metabolism circuit do not show any statistically significant difference. Those genes were not

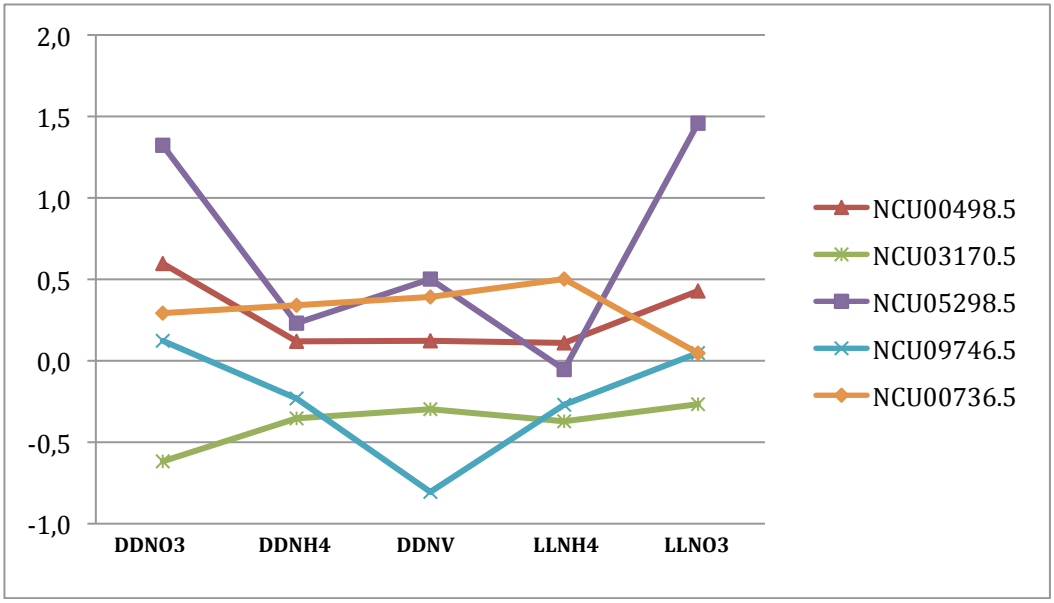
recognized by either arbitrary cut-off function of MatLab, or by use of filtering by the BRB array tool. However, some of the genes show similar behavior at different conditions (Figure 17) and it is possible to see patterns in the gene expression in different environmental conditions.



a



b



c

Figure 17. Expression values of genes related to the nitrogen metabolism circuit, which show relative similar behavior. Expression values of genes related to nitrogen metabolism were found by using a Perl script. The names of genes are shown on the right, while expression levels are indicated on the scale to the left. The growth conditions are represented on the x-axis. It is important to mention that there is no connection between expression values in different conditions, and the symbols on the graphs are joined by a line just to make it easier to see the pattern in each gene's expression. Genes are devided into three groups (a, b and c) in order to simplify the representation of data. Gene names and descriptions are listed in Table 8.

Gene name	Locus	Description
nit-10	NCU07205.5	nitrate transporter CRNA
nmr	NCU04158.5	nitrogen metabolite regulation
nit-3	NCU05298.5	nitrate nonutilizer-3, nitrate reductase
nit-6	NCU04720.5	nitrate nonutilizer-6, nitrite reductase
nit-2	NCU09068.5	nitrate nonutilizer-2, nitrogen catabolic enzyme regulatory protein
nit-4	NCU08294.5	nitrate nonutilizer-4, nitrogen assimilation transcription factor *nit-4*
nit-1	NCU00736.5	molybdenum cofactor biosynthetic protein
glm-1	NCU06724.5	glutamine synthetase, glutamine-1
glm-2	NCU04856.5	glutamine synthetase , glutamine-2
-	NCU00498.5	molybdenum cofactor biosynthesis protein 1 B
-	NCU03170.5	molybdopterin-converting factor subunit 2
-	NCU09746.5	gephyrin

Table 8. List of genes related to the nitrogen regulatory circuit. The list of genes related to the nitrogen metabolism (Borkovich et al., 2004). The list was updated with the help of the Broad Institute database. The first column indicates gene name. The second column represents gene locus and the third column contains its short description.

nit-6 and *nit-3* show similar behavior at different conditions, and both are up-regulated in the presence of nitrate, as was expected. The data also corresponds to the fact that the action of nitrite reductase is correlated with the action of nitrate reductase. However, the fact that the *nit-6* and *nit-3* are up-regulated in the Normal Vogel medium, which includes both nitrate and ammonium, does not allow for a direct relation between the presence of nitrate and the expression of nitrate/nitrite reductases. This observation will be discussed further.

nit-2 shows similar behavior to *nit-3* and *nit-6* in Normal Vogel medium and light conditions. It correlates partly with the fact that *nit-2* is involved in activation of transcription of many structural genes of the nitrate assimilatory pathway, including *nit-3*, *nit-6*, and *nit-10*.

From Figure 17 a, it is clear that the expression levels of *nit-2* do not change dramatically among the five different conditions. Therefore, the observation that the transcription of the *nit-2* increases 2 to 3 times under nitrogen repressed conditions, i.e., lack of a primary nitrogen source, described by Fu has not been confirmed in this instance (Fu and Marzluf, 1987).

According to Feng (Feng and Marzluf, 1996), *nit-4* and *nmr* are expressed at the cell constantly at low levels. From Figure 17 b, one can see some variations in the level of expressed NIT-4 and NMR in different conditions, and a common pattern in the variations for these two genes. This can indicate a possible regulation of the transcription of these proteins. Interestingly, the variations of expression levels of *nit-1* and *glm-1* also have a similar pattern.

It has been shown that the *nit-1*, *nit-7*, *nit*-8 and *nit-9* (Figure 17 c) genes are cofactors for the nitrate reductase. *nit-4* and *nit-5* genes are shown to be alleles from the same locus and regulate the induction of the pathway by nitrate and nitrite (Tomsett and Garrett, 1980). There are no signs of noticeable variations among the five samples for the NCU00736, NCU00498 and NCU03170 (proteins related to the molybdenum-cofactor – one of three domains of NIT-3). This suggests that the proteins may indeed be expressed in the cell constantly at low levels.

Ammonium transporters are proteins used for the transport of ammonium into the cell through the cell wall. In the absence of ammonium, this function is not in use, and one would expect low expression levels for these proteins. Expression values of the ammonium

transporters, predicted by Khademi (Khademi et al., 2004) were extracted with help of Perl script from the microarray data and presented in Figure 18.

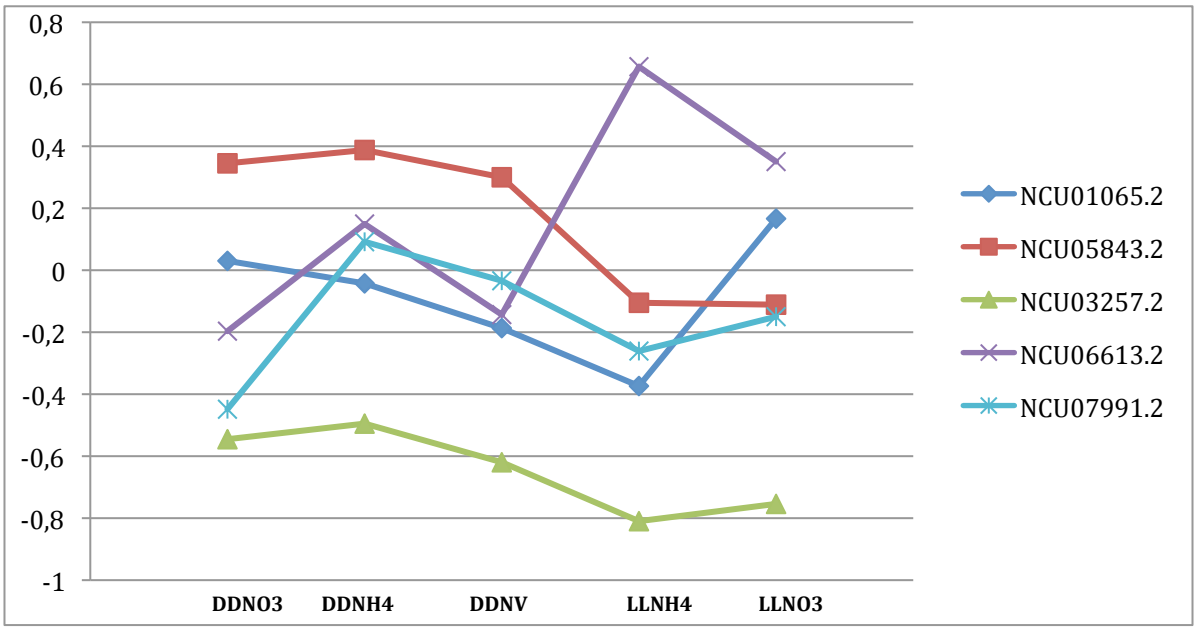


Figure 18. Expression values of ammonium transporters. Expression values of genes encoding ammonium transporters were extracted by Perl. The names of genes are shown on the right, while the expression levels are indicated on the scale to the left. The growth conditions are represented on the x-axis. It is important to mention that there is no connection between expression values in different conditions, and the symbols on the graphs are joined by a line just to make it easier to see the pattern in each gene's expression. Gene names and descriptions are listed in Table 9.

From Figure 18 it is clear that the nitrate conditions do not lower expression levels of the ammonium transporters. In fact, the ammonium transporter MEP2 (encoded by NCU01065) seems to be induced at the nitrate conditions.

Locus	Description
NCU01065.2	ammonium transporter MEP2
NCU05843.2	ammonium transporter 1
NCU03257.2	ammonium transporter MEP1
NCU06613.2	ammonium transporter
NCU07991.2	DUF292 domain-containing protein

Table 9. List of ammonium transporters. The list of ammonium transporters (Khademi et al., 2004). List is updated with help of the Broad Institute database. The first column represents gene locus and the second column contains short description of each gene.

3.3.2 Genes related to oxidative stress

When nitrate is the only nitrogen source, *Neurospora* appears to be under oxidative stress, leading to enchanced photoconidiation (Ruoff, 2011). That is why it was interesting to look at expression of genes related to oxidative stress defense mechanism. Their expression values were extracted from the microarray data with Perl script (Figure 19).

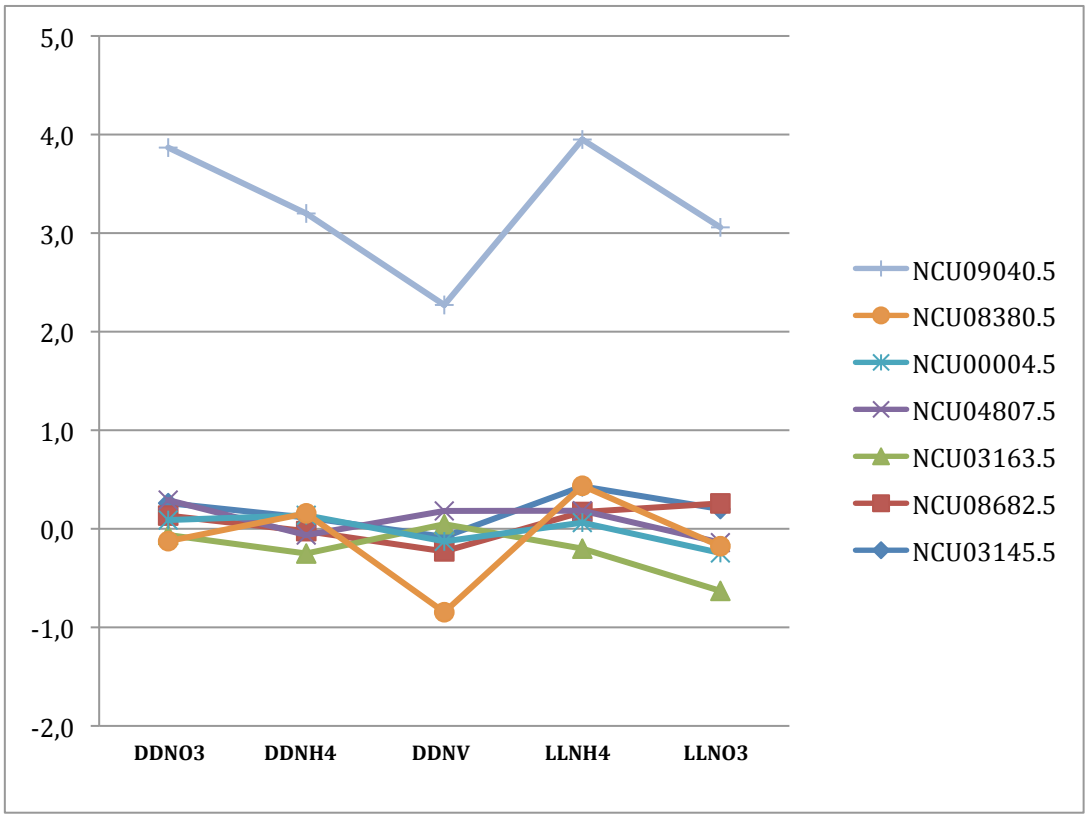


Figure 19. Expression values of genes related to the oxidative stress. Expression values of genes related to oxidative stress were extracted by Perl. The growth conditions are represented on the x-axis. From the left to right: DDNO3, nitrate only, darkness; DDNH4, ammonium only, darkness; DDNV, Normal Vogel medium (ammonium nitrate), darkness; LLNH4, ammonium only, light; LLNO3, nitrate only, light. It is important to mention that there is no connection between expression values in different conditions, and the symbols on the graphs are joined by a line just to make it easier to see the pattern in each gene's expression.

Locus	Description
NCU03145.5	oxidative stress resistance
NCU08682.5	stress response protein nst-1
NCU03163.5	stress responsive A/B barrel domain-containing protein
NCU04807.5	universal stress protein family domain-containing protein

Locus	Description
NCU00004.5	universal stress protein family domain-containing protein
NCU08380.5	plasma membrane phosphatase required for sodium stress response
NCU09040.5	oxidoreductase

Table 10. List of genes related to the oxidative stress defense mechanism. The list of genes related to the oxidative stress (Borkovich et al., 2004). List is updated with help of the Broad Institute database. The first column represents gene locus and the second column contains short description of each gene.

It appears that the most of the genes related to the oxidative stress are both up- and down-regulated while nitrate is the only nitrogen source in the medium, and do not show any significant difference with the expression values in ammonium-rich medium. These results confirm results from the study of Ruoff, where genes related to the oxidative stress were both up- and down-regulated (Ruoff, 2011). However, the oxidoreductase NCU09040 showed the high expression values at all conditions.

3.3.3 Genes related to conidiation

Oxidative stress and nitrogen starvation induce conidiation at *Neurospora*, and genes related to conidiation might be expressed in the medium containing nitrate as the only source of nitrogen (Davis, 2000). Expression values of genes related to conidiation were extracted from the microarray data (Figure 20).

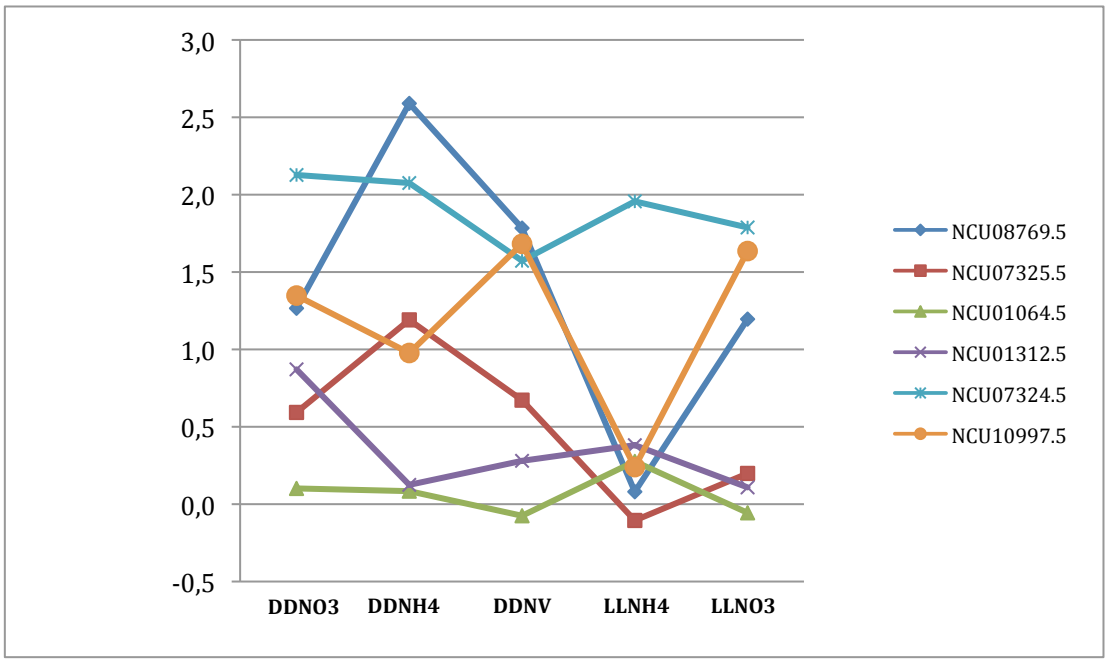


Figure 20. Expression values of genes related to conidiation. Expression values of genes related to conidiation were extracted by Perl. The growth conditions are represented on the x-axis. From the left to right: $DDNO_3$, nitrate only, darkness; $DDNH_4$, ammonium only, darkness; DDNV, Normal Vogel medium (ammonium nitrate), darkness; $LLNH_4$, ammonium only, light; $LLNO_3$, nitrate only, light. It is important to mention that there is no connection between expression values in different conditions, and the symbols on the graphs are joined by a line just to make it easier to see the pattern in each gene's expression.

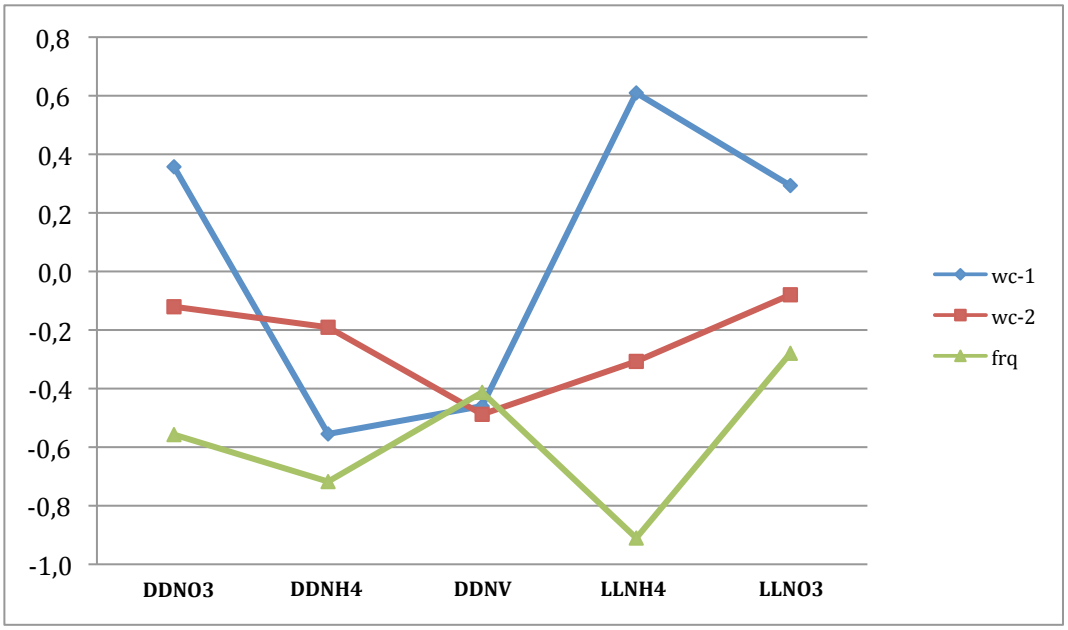
Gene name	Locus	Description
con 6	NCU08769.5	conidiation-6
con 10	NCU07325.5	conidiation-10
con 8	NCU10997.5	conidiation-specific protein 8
-	NCU01064.5	related to conidiation protein *con-6*
rca-1	NCU01312.5	RCA-1 regulator of conidiation
con 13	NCU07324.5	conidiation-specific protein conidiation-13

Table 11. List of genes related to conidiation. The list of genes related to conidiation (Borkovich et al., 2004). List is updated with help of the Broad Institute database. The first column represents gene locus and the second column contains short description of each gene.

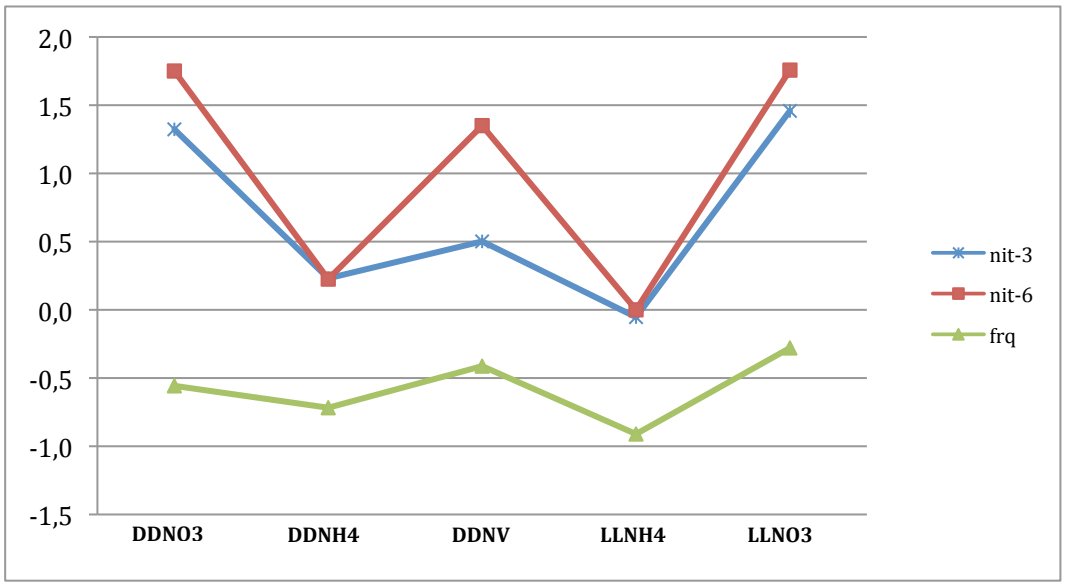
The genes *con-6* and *con-10* show similar behavior at different conditions, with the highest expression values in darkness. In addition, the *con-6* gene and *con-13* also show a similar expression pattern. However, there is no clear relation between the expression values of the genes, conidiation and the substitute of the medium. The data does not confirm the results from a study by Ruoff (Ruoff, 2011), where the upregulation of the conidiation-related genes when nitrate is the sole nitrogen source for *Neurospora* was observed.

3.3.4 Genes related to circadian rhythm

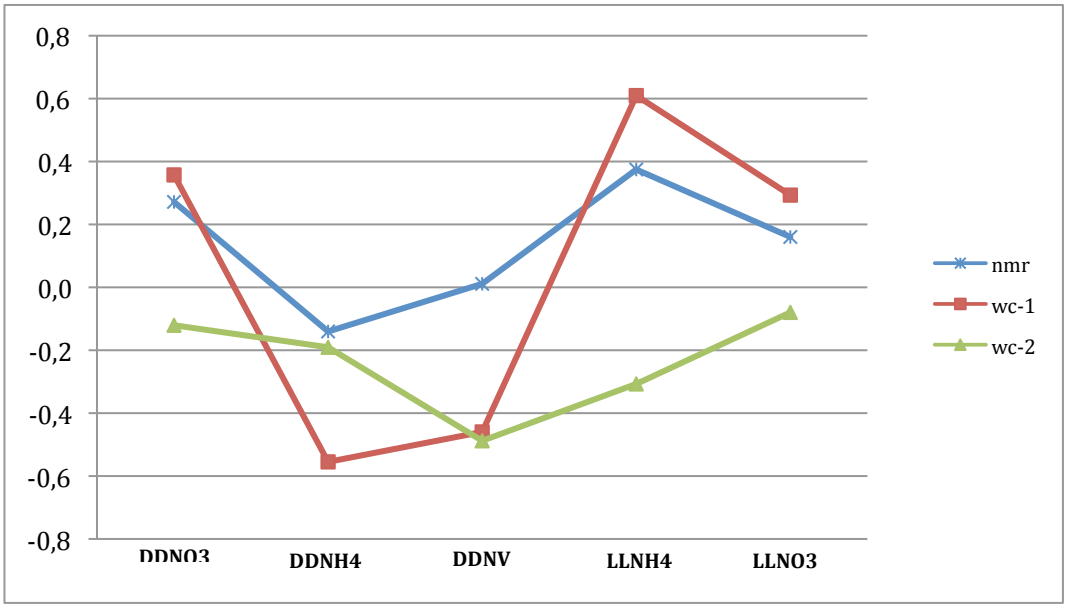
Expression values of genes related to circadian rhythm were extracted from the microarray data and analyzed (Figure 14). White collar-1 (WC-1) and white collar-2 (WC-2) are transcription factors essential for light-mediated responses in *Neurospora crassa* (Chen et al., 2009) and are expected to be highly expressed in light. From Figure 21 a below, it is clear that *wc-1* is up-regulated in light conditions on both substrates, and in darkness on a nitrate medium. *wc-2* is most expressed in light on a nitrate medium. On the ammonium-rich medium the expression value does not confirm the relation between light and expression of the *wc-2* gene. However, this observation can be explained by abundance of WC-1, which negatively regulates the transcription of *wc-2* (Cheng et al., 2003).



a



b



c

Figure 21. ***Neurospora´*****s gene expression pattern for genes related to the circadian rhythm, when different nitrogen sources and light/dark conditions are used.** Expression values of genes related to circadian rhythm were extracted by Perl. The growth conditions are represented on the x-axis. From the left to right: $DDNO_3$, nitrate only, darkness; $DDNH_4$, ammonium only, darkness; DDNV, Normal Vogel medium (ammonium nitrate), darkness; $LLNH_4$, ammonium only, light; $LLNO_3$, nitrate only, light. It is important to mention that there is no connection between expression values in different conditions, and the symbols on the graphs are joined by a line just to make it easier to see the pattern in each gene's expression. Figures b and c indicate comparison of expression levels between genes related to circadiam rhythm and genes from nitrogen metabolism.

Gene name	Locus	Description
wc-1	NCU02356.5	white collar-1
wc-2	NCU00902.5	white collar-2, zinc finger white collar protein WC-2
frq	NCU02265.5	frequency, period clock protein FRQ

Table 12. List of genes related to circadian rhythm. The list of genes related to circadian rhythm (Borkovich et al., 2004). List is updated with help of the Broad Institute database. The first column indicates gene name. The second column represents gene locus and the third column contains it´s short description.

frq seems to be up-regulated in the presence of nitrate and there is clearly a similar pattern in expression values of *frq* and *nit-3* and *nit-6* (Figure 21 b).

3.4 Comparison of expression of nitrate reductase at DDNV and $DDNH_4$

From Figure 22 it is clear that the gene encoding nitrate reductase protein, *nit-3,* is slightly up-regulated in the conditions where nitrate is the only source of nitrogen. This confirms the fact that the genes related to the nitrogen metabolism circuit are being activated in the absence of the primary source of nitrogen for *Neurospora* – ammonium (Davis, 2000). However, the level of expression of *nit-3* in the Normal Vogel medium in darkness is also seen to be high (Figure 22). This is in agreement with the data by Ruoff (Ruoff, 2011), where the concentration of the *nit-3* is clearly high in the Normal Vogel medium after 16 hours (Figure 23 a).

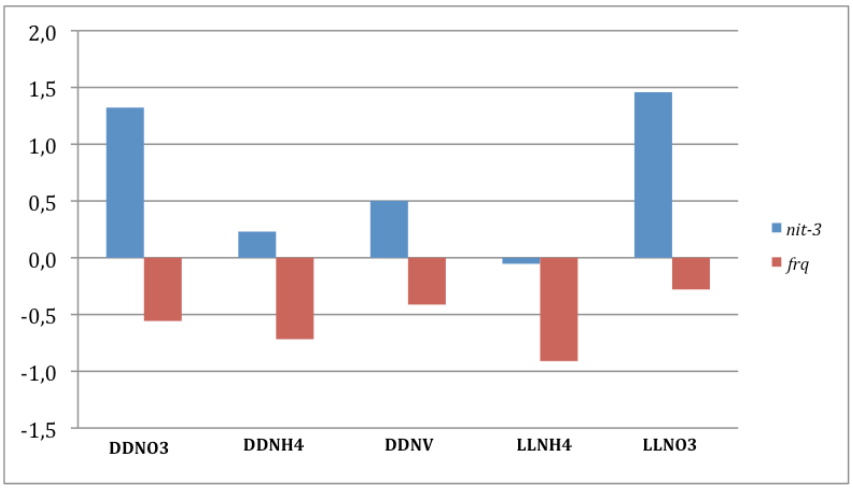
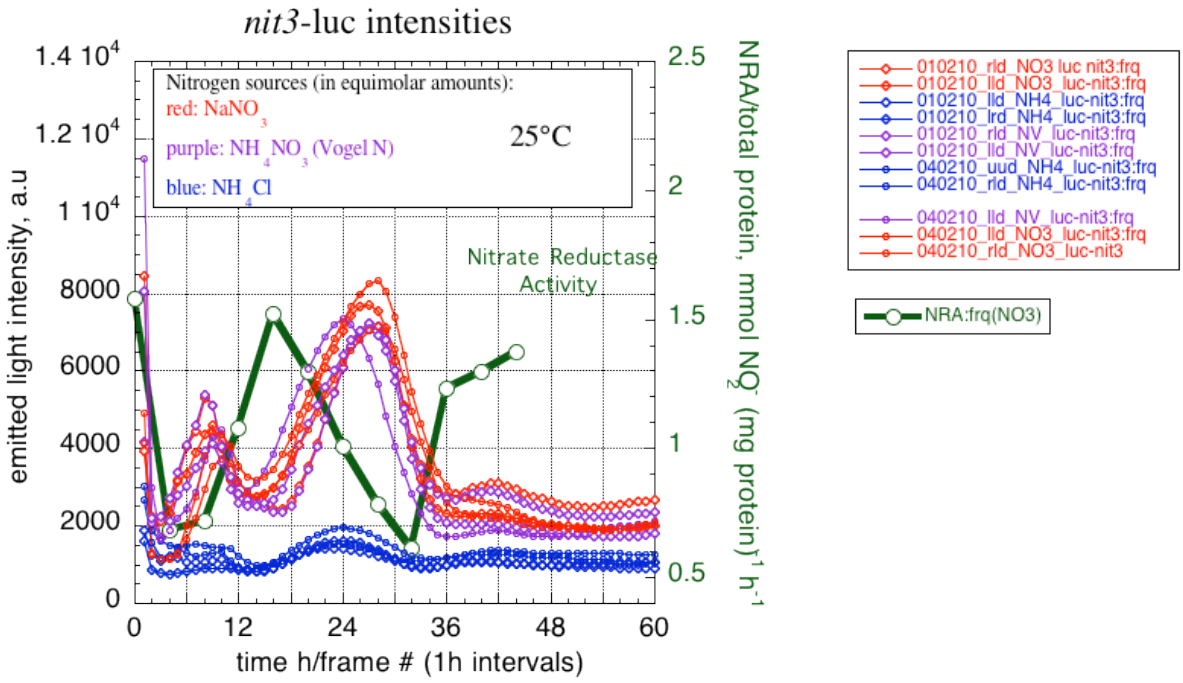
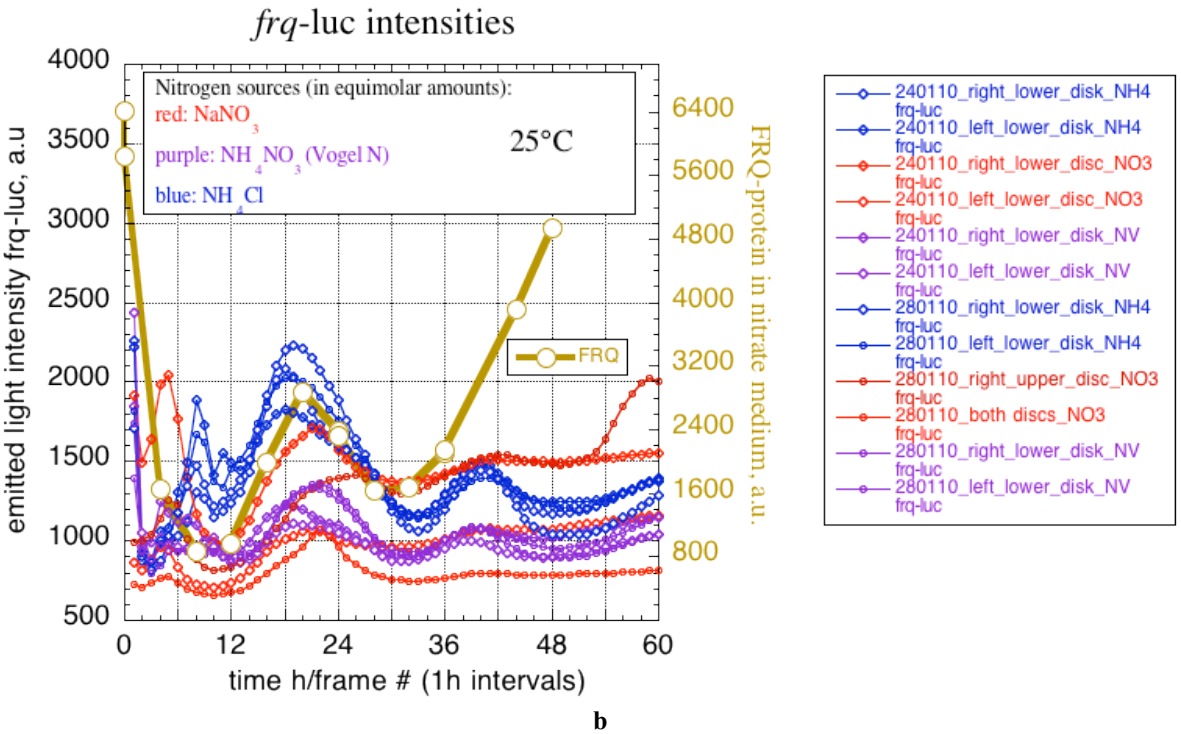


Figure 22. Expression levels of the *nit-3* and *frq*-gene in different conditions. Expression values of genes of interest were extracted by Perl. Expression levels of the *nit-3* are clearly higher in the nitrate substrate and on

the Normal Vogel medium. *frq* is down-regulated at all conditions, with the least expression values on ammonium substrate.



a



b

Figure 23. Experimental data from *nit-3* and *frq* luciferase reporters by Ruoff (unpublished data). The point that corresponds with the experimental conditions from this thesis is 16 hours. a) Emission intensity of *nit-3* luciferase reporter from different mycelial discs as a function of time from different mycelial disks in medium containing nitrate (red lines) or ammonium (blue lines) as the only nitrogen sources. The violet lines indicate

emission intensity of *nit-3* luciferase reporter from different mycelial discs as a function of time from mycelial disks in Normal Vogel medium. The green solid line shows the average of nitrate reductase activity from earlier study by Ruoff.
b) Emission intensity of *frq* luciferase reporter from different mycelial discs as a function of time from different mycelial disks in medium containing nitrate (red lines) or ammonium (blue lines) as the only nitrogen sources. The violet lines indicate emission intensity of frq luciferase reporter from different mycelial discs as a function of time from mycelial disks in Normal Vogel medium. The brown solid line shows the average of nitrate reductase activity from earlier study by Ruoff.

The experiment by Ruoff was performed with help of luminescent reporter *luciferase (luc)* gene. Encoded by the *luc* gene protein LUC catalyses the oxidation of the bioluminescent substrate luciferin in the presence of oxygen, ATP and Mg^{2+}. This reaction undergoes with the photon emission. *luc* can be coupled to an endogenous promoter and in such way its the expression allows to dynamically reflect transcriptional activity of promoter (Jensen, 2012).

It is clear from Figure 23 b that the level of the *frq*-expression should be high after 16 hours. Although our data seems to show lower values of *frq* expression (Figure 22), it is not clear if it contradicts the data extracted by Ruoff. The reason for this uncertainty is the impossibility of comparing our experimental data with data from other time-points, which may possibly have been lower.

4. Discussion

The responses of the fungal cell to nitrogen starvation is an important aspect in understanding the biology of *Neurospora*. Gene expression profiling can help to better understand assimilatory pathways and their relations. Gene expression profiles of *Neurospora*, grown on nitrate and ammonium medium were compared in order to identify genes and pathways that contribute to the transition from ammonium to nitrate uptake.

Two methods were used for identifying genes of interest: arbitrary cut-off (MatLab) and filtering, normalization and gene subsetting software from BRB-array tool.

The data extracted by arbitrary cut-off function of MatLab was mostly useful in recognizing genes that showed extremely high or low values of expression and did not allow for identification of all genes which are significantly differently expressed. That point is supported by variety of studies which show the importance of the accuracy in recogniztion of statistically differently expressed genes and ability to identify small, but biologically important, changes (Sweet-Cordero et al., 2005; Yao et al., 2004). The use of software from the BRB-array tool allowed for the extraction of genes that showed statistically significant differences in their expression. Filtering, normalization, and gene subsetting functions identified a total of 458 *N. crassa* genes.

Data clustering, a common technique for statistical data analysis was used for further investigation of distinct genes. Clustering is the grouping of objects based on similarity, and in this manner, clustering of gene expression data helps in identifying genes of similar function. Genes with a known function which co-express with either poorly characterized, predicted or entirely novel genes may provide a simple means of gaining insight to the functions of many genes for which information is not currently available (Eisen et al., 1998). It has also been demonstrated that co-regulated families of genes cluster together (Alon et al., 1999).

Sample clustering techniques have been widely used to group samples with similar expression patterns and have contributed immensely to our understanding of the different assimilatory pathways in *Neurospora crassa*. In particular, such analysis has suggested mechanisms of early and late light-responses (Chen et al., 2009), helped to identify and

describe genes related to development and conidiation (Greenwald, 2010), and contributed in identifying and investigating genes related to stress response (Sokolovsky, 2000).

Clustering analysis of identified genes was performed in two ways: hierarchical clustering and *k*-means clustering. Hierarchical clustering was used to see how well samples clustered together, to visually identify clusters in the samples and identify correlated genes. The dendogram made by hierarchical clustering showed close relation between samples performed in light, which supports the assumption made by Chen that the genes that are truly responsive to light should behave similarly to each other (Chen et al., 2009).

Since hierarchical clustering has been criticized for offering no compelling evidence that a hierarchical structure best suits grouping of the expression profiles (Babu, 2004), an alternative method, non-hierarchical *k*-means clustering, was also performed. The necessity of two means of clustering can also be explained by the wish to obtain more robust results. One of the most commonly used clustering methods, *k*-means clustering, has a wide application in microarray studies. *k*-means clustering groups existing objects into predefined clusters rather than organizing them into a hierarchical structure (Do and Choi, 2008). In this thesis, *k*-means clustering allowed for identification of correlation between clustered genes and particular parameters, giving insight into the relationship between each particular gene and growth condition. Nevertheless, the six clusters recognized by *k*-means showed a significant overlap with the six clusters recognized by hierarchical clustering.

Both clustering methods recognized groups of genes related to light and dark, and nitrate or ammonium environments. FunCat analysis was performed for clusters of genes associated with these parameters, in order to find their functional distribution. Functional description of genes from the different clusters was compared in order to find clusters enriched with genes belonging to a particular gene ontology or pathway.

As such, the functional analysis of groups of genes with high expression levels on ammonium substrate showed enrichment of genes involved in secondary metabolism; in particular, the metabolism of polyketides, alkaloids and secondary products derived from L-glutamic acid, L-lysine et cetera. On the other hand, genes involved in secondary metabolism were also up-regulated on nitrate-medium in darkness. These genes are related to metabolism of such secondary products as compounds derived from L-phenylalanine and L-tyrosine, melanins and peptide derived compounds.

Secondary metabolism is commonly associated with the stage when fungus has completed its initial growth phase and is begins the stage of development represented by the formation of spores (Calvo et al., 2002). Calvo describes the availability and type of nitrogen source as the primary factor affecting the production of the secondary metabolites. For example, in *A. nidulans,* where the production of secondary metabolites sterigmatocystin and aflatoxin increases in ammonium-based medium and decreases in nitrate-based medium, or in *A. Flavus* grown on agar media where the development of sclerotia occurs with nitrate as the sole nitrogen source but not ammonium (Calvo et al., 2002). The results from this thesis suggest the existence of different regulation mechanisms which, based on the type of nitrogen source, exert influence on *Neurospora´s* secondary metabolism pathways. That suggesting is also supported by observed by Park increased levels of mRNA and protein for tyrosinase in the mutants under nitrogen starvation, a condition favoring sexual differentiation. Tyrosinase is an enzyme that catalyzes production of the secondary metabolite l-DOPA melanin (Park et al., 2008). These results implicate the nitrogen oscillatory pathway in regulation of development and secondary metabolism in *Neurospora*.

Functional category analysis of genes, which showed to be related to light by visual analysis of clusters comprised by *k*-means clustering revealed a variety of data, which considers the relation between light induction and such functions as C-compound and carbohydrate metabolism, RNA and protein synthesis, detoxification (for approximately half of the genes identified as light responsive the function is not classified). According to Chen (Chen et al., 2009), there is clear correspondense between the timing of light induction and underlying nature of the biological process. Such, majority of classified early light-response genes, identified by Chen are related to lipid, fatty acid and isoprenoid metabolism. Genes highly enriched among the late light-response group involved in carbohydrate metabolism, oxidation and detoxification. These data suggests light related genes, revealed by this thesis belong to late light-responsive genes.

However, the comparison of 147 light-related genes, provided by this thesis (cluster 4 and cluster 6 from *k*-means clustering) with the 314 genes, described by Chen as strong early or late light responsive indicated just 5 common genes (NCU05770, NCU00766, NCU08824, NCU03803 and NCU09306). Brod institute database describes one of these proteins, NCU08824, as molybdopterin binding domain-containing protein. Considering the fact that

one of three domains of nitrate reductase is molybdopterin-containing domain (Davis, 2000), it is reasonable to suggest the possible regulation of NIT-3 by light.

Despite the high percentage of unidentified proteins or hypothetical proteins, it is clear that the genes that showed extreme values of expression are mostly involved in primary metabolism (C-compounds, lipids, amino acids), stress response and biological rhythms. The highest amount of unclassified proteins was observed for genes on nitrate substrate, which suggests that there are many undiscovered mechanisms related to nitrogen uptake and transport.

Functional category analysis of genes relatively high or low expressed at all conditions, identified C-compound and carbohydrate related pathways to be dominant among both up-regulated and down-regulated genes in all five samples. This is not surprising considering the fact that the growth medium was a low-sucrose medium, and genes affected by carbon starvation may act as both negative and positive elements in the carbon-related pathways (Christensen, 2007).

The analysis of genes up-regulated on the medium with nitrate as the only source of nitrogen, showed that the transition from ammonium to nitrate uptake resulted in a large quantity of up-regulated genes related to stress response and detoxification. Nitrate, when chosen as the alternative nitrogen source for *Neurospora* is an inorganic compound, acting as an oxidative element in many chemical reactions (Moore, 2009). Therefore, it is reasonable to suggest that the environment with nitrate as the only nitrogen source affects not just genes related to nitrogen metabolism, but also genes involved in the oxidative-stress response. However, the relatively constant expression level for genes related to oxidative stress in our study did not confirm that idea. Only oxidoreductase (encoded by NCU09040) showed high expression levels, observed in all five samples.

Oxidative stress and nitrogen starvation are shown to be strong inducers of conidiation (Davis, 2000). Confirming that fact, Sokolovsky (Sokolovsky, 2000) observed remarkable increase in the level of mRNA for genes from *con* group, in the dark and nitrogen limited environment. In fact, the level of mRNA of *con-8* increased 2.3 times, while the level of mRNA of *con-10* increased 5.8 times. In that experiment, the nitrogen starvation was induced by low concentrations of ammonium, while in our study, nitrate was used as an alternative source of nitrogen. When analyzing the expression levels of *con*-genes where nitrate was the

solo source of nitrogen in this thesis, there was not observed any significant increase in expression levels.

A possible explanation to this discrepancy can be found in an article by Greenwald (Greenwald, 2010), who describes media with low glucose as a potential for promoting gluconeogenesis and conidiation. Also in the study by Sokolovsky (Sokolovsky, 2000), the expression levels of *con*-genes in glucose-starvation conditions are shown to be increased more than 3 times (the experiment was performed in darkness). The low sucrose concentration and over-representation of proteins related to C-compound and carbohydrate metabolism in the experiment from this study suggests carbon starvation in all investigated samples that may result in upregulation of *con*-genes in all studied conditions. Consequently it is impossible to recognize the increase of expression values for *con*-genes caused by nitrate. Therefore, the suggestion that nitrogen starvation condition induces *con*-genes cannot be confirmed or refuted by the obtained results, since a reference sample with a high concentration of glucose is not available.

A closer look at the genes related to the nitrogen metabolism pathway did not fully confirm the expectations about de-repression of nitrate-related genes (Gao-Rubinelli and Marzluf, 2004). The nitrate-transporter *nit-10* is slightly up-regulated under nitrate conditions, as was expected. On the other hand, the expression levels of ammonium transporters do not seem to be lower under nitrate conditions. In addition, the ammonium transporter MEP2 (encoded by NCU01065) seems to be induced under nitrate conditions. This may suggest an additional role of this transporter in the nitrate assimilation pathway. This suggestion can be supported through a study by Lorenz (Lorenz and Heitman, 1998), who describes the ammonium transporter MEP2 of the yeast *Saccharomyces cerevisiae*. In the study Lorenz shows that MEP2, a high affinity ammonium permease, is required for pseudo-hyphal differentiation in response to ammonium limitation. In contrast, MEP1 and MEP3, which are lower affinity ammonium permeases, are not required for filamentous growth. Therefore, Lorenz proposes that MEP2 is an ammonium sensor, generating a signal to regulate filamentous growth in response to ammonium starvation. In a similar way, the ammonium transporter MEP2 of *Neurospora crassa* could function as an ammonium sensor, acting as a signal protein in the absence of ammonium. Full sequence alignment of ammonium transporter MEP2 of *Neurospora crassa* and ammonium transporter MEP2 of the yeast *Saccharomyces cerevisiae* shows 53 % identity.

Although the observed increase in the expression levels of *nit-6* and *nit-3* in the medium with nitrate as the only nitrogen source indicates their activation, there were also increased levels of mRNA to *nit-6* and *nit-3* in Normal Vogel medium, which contains both nitrate and ammonium. This observation is in agreement with the experimental data from Ruoff (unpublished), where similar expression levels of *nit-3* were found. That may suggest the functionality of nitrogen oscillatory pathway in the presence of both nitrate and ammonium in medium. Explanation to that can be found in low concentrations of NMR at the experimental time-point and abundance of NIT-2, which is essential for the activation of nitrate reductase.

nit-2 shows similar expression levels to *nit-3* and *nit-6* in Normal Vogel medium and light conditions. This correlates partly with the fact that *nit-2* is involved in transcription activation of many structural genes of the nitrate assimilatory pathway, including *nit-3*, *nit-6*, and *nit-10*. On the other hand, *nit-2* expression level does not vary significantly among the five different conditions. This contradicts the experimental data from Fu, which showed that the transcription of *nit-2* increases by 2 to 3 times under nitrogen repressed conditions, i.e., lack of a primary nitrogen source (Fu and Marzluf, 1987). The explanation of that can be in the early or late responses to nitrate or abundance of NIT-2 in the cell due to long half-life of NIT-2 protein and *nit-2* mRNA in vivo, that is shown to be not affected by different nitrogen sources, including those that lead to N-repression or N-depression (Tao and Marzluf, 1999).

According to Feng (Feng and Marzluf, 1996), *nit-4* and *nmr* are expressed at the cell constitutively at low levels. In this study, there was a common pattern of expression levels in different conditions for these two genes. This could indicate a regulation of the transcription of these proteins. Interestingly, the variations of expression levels of *nit-1* and *glm-1* also show a similar pattern. Greenwald in his study (Greenwald, 2010) describes the expression levels of *glm-1* under conditions of nitrogen deficiency as low, whereas the levels for *glm-2* were high, which is not consistent with this study.

Although cofactors for the nitrate reductase, *nit-1*, *nit-7*, *nit*-8 and *nit-9* are expected to be up-regulated under nitrate conditions (Tomsett and Garrett, 1980), no significant variation among expression levels in all five conditions was observed. This also suggests that these genes could be expressed constitutively at low levels.

White collar-1 (WC-1), the blue-light receptor in *Neurospora,* is up-regulated in light, as was expected. *wc-2* is most expressed in light on a nitrate medium, but on the ammonium-rich medium the expression value of *wc-2* is low. This does not confirm a relationship between light and expression of the *wc-2* gene. This observation can be explained by the abundance of WC-1, which negatively regulates the transcription of *wc-2* (Cheng et al., 2003).

WC-1 and WC-2 have been shown to be under control of clock (Chen et al., 2009). On the other hand, the action of WC-1 and WC-2 is also shown to drive expression of the *frq* gene. WC-2 enters the promoter region of *frq* coincident with increases in *frq* expression and then exits when the cycle of transcription is over, whereas WC-1 can always be found there. FRQ promotes the phosphorylation of the WCs, thereby decreasing their activity, and phosphorylation of FRQ then leads to its turnover, allowing the cycle to reinitiate (Dunlap et al., 2007). Expression patterns of these genes, which is partly in agreement with the proposed model, was observed in this study; relative level of *wc-1* transcripts is higher when the expression level of *frq* is low, and is lower when the expression level of *frq* is high at light. This data can be explained by a low level of WC-1 in the cell at the experimental timepoint, therefore the level of *frq* mRNA is low while the transcription of WC-1 is activated.

Interestingly, the *nit-3*, *nit-6* and *frq* seem to have the same expression pattern at different conditions. This observation, together with the connection between WC-1 and *frq* described above, and the experimental data of Christensen (Christensen, 2007) showing a significant decrease in NRA-levels in a *wc-1* mutant, confirms Christensen's suggestion, about a possible "overlap" between the nitrate reductase oscillatory system and the FRQ/WCC transcriptional/translational oscillator.

The most represented functions for genes that showed high expression values on medium with nitrate as the solo source of nitrogen are seem to be RNA synthhesis, protein modification, binding of metals and nucleotides, together with transport routes and stress response. The function of protein modification, which includes modification by phosphorylation and protein/peptide degradation can be possibly related to promoting of phosphorylation of WCC by FRQ-FRH complex (de Paula et al., 2007; Dunlap et al., 2007). That is another point in this job, that may confirm the existence of the connection between nitrate reductase oscillatory system and the FRQ/WCC transcriptional/translational oscillator.

Surprisingly, the expression pattern of *wc-1* at different conditions is similar to the *nmr* and *nit-4*. The NMR protein is involved in the repression of nitrate reductase by interacting with the positively acting transcription factor NIT-2, which activates the transcription of *nit-3*. *Neurospora nmr* mutant display elevated levels of NR and NiR due to greater enzyme concentrations (Pan et al., 1997). That observation confirms the previously discussed relationship between the nitrate reductase oscillatory system and the FRQ/WCC transcriptional/translational oscillator, and suggests a possible role of *nmr* and *nit-4* in this relationship.

Disagreements with the experimental data from other studies discussed above can be possibly explained by inability to compare the expression values with another time-points. Experiment for this thesis was performed at 16 hour time-point and may not indicate the oscillations in the expression of genes of interest. It is also not known the initial concentrations of the corresponding proteins. Concidering the possible inhibition or inducing of transcription wherether the concentration of protein is high or low and different half-life of proteins it is reasonable to suggest the existence of deviations in relation between expression levels of genes and concentrations of corresponding proteins in the cell.

Conclusion

Change of the nitrogen source and light conditions leads to the difference in expression of big amount of genes. That indicates differences in the development of fungus, formation of the reproductive organs on the vegetative mycelium, and apparently is a defensive reaction of the organism to the action of stressors. Under the influence of stressors the endogenous circadian rhythms are also changed.

In this study the functional categories of the vast majority of genes identified as statistically differently expressed have not been found in reviewed literature. This indicated the existence of alternate pathways, which are not yet investigated.

In the case of the nitrate assimilatory pathway in *Neurospora crassa*, the ability to utilize a wide range of nitrogen sources will increase survivability. To date, the nitrogen metabolism pathway is suggested to be related to circadian system (Christensen, 2007). In this study the existence of the connection between nitrate reductase oscillatory system and the FRQ/WCC transcriptional/translational oscillator is confirmed and a possible role of *nmr* and *nit-4* in this relationship is suggested.

However, some of the obtained results are not in agreement with experimental data from a number of reviewed studies. Therefore it is possible to hypothesize that gene expression is influenced by such factors as the abundance of gene transcript, the amount of translation product and the biological activity of the protein.

List of Figures

Figure 22. Expression levels of the *nit-3* and *frq*-gene in different conditions

Figure 23. Experimental data from *nit-3* and *frq* luciferase reporters by Ruoff (unpublished data)

List of Tables

References

Alon, U., N. Barkai, D.A. Notterman, K. Gish, S. Ybarra, D. Mack, and A.J. Levine. 1999. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proceedings of the National Academy of Sciences of the United States of America*. 96:6745-6750.

Babu, M.M. 2004. An Introduction to Microarray Data Analysis. Horizon Press, U.K.

Bell-Pedersen, D. 1998. Keeping pace with Neurospora circadian rhythms. *Microbiology*. 144 (Pt 7):1699-1711.

Borkovich, K.A., L.A. Alex, O. Yarden, M. Freitag, G.E. Turner, N.D. Read, S. Seiler, D. Bell-Pedersen, J. Paietta, N. Plesofsky, M. Plamann, M. Goodrich-Tanrikulu, U. Schulte, G. Mannhaupt, F.E. Nargang, A. Radford, C. Selitrennikoff, J.E. Galagan, J.C. Dunlap, J.J. Loros, D. Catcheside, H. Inoue, R. Aramayo, M. Polymenis, E.U. Selker, M.S. Sachs, G.A. Marzluf, I. Paulsen, R. Davis, D.J. Ebbole, A. Zelter, E.R. Kalkman, R. O'Rourke, F. Bowring, J. Yeadon, C. Ishii, K. Suzuki, W. Sakai, and R. Pratt. 2004. Lessons from the genome sequence of Neurospora crassa: tracing the path from genomic blueprint to multicellular organism. *Microbiology and molecular biology reviews : MMBR*. 68:1-108.

Calvo, A.M., R.A. Wilson, J.W. Bok, and N.P. Keller. 2002. Relationship between secondary metabolism and fungal development. *Microbiology and molecular biology reviews : MMBR*. 66:447-459, table of contents.

Campbell, W.H., and K.R. Kinghorn. 1990. Functional domains of assimilatory nitrate reductases and nitrite reductases. *Trends in biochemical sciences*. 15:315-319.

Chen, C.H., C.S. Ringelberg, R.H. Gross, J.C. Dunlap, and J.J. Loros. 2009. Genome-wide analysis of light-inducible responses reveals hierarchical light signalling in Neurospora. *The EMBO journal*. 28:1029-1042.

Cheng, P., Q. He, Q. He, L. Wang, and Y. Liu. 2005. Regulation of the Neurospora circadian clock by an RNA helicase. *Genes & development*. 19:234-241.

Cheng, P., Y. Yang, K.H. Gardner, and Y. Liu. 2002. PAS domain-mediated WC-1/WC-2 interaction is essential for maintaining the steady-state level of WC-1 and the function of both proteins in circadian clock and light responses of Neurospora. *Molecular and cellular biology*. 22:517-524.

Cheng, P., Y. Yang, L. Wang, Q. He, and Y. Liu. 2003. WHITE COLLAR-1, a multifunctional neurospora protein involved in the circadian feedback loops, light sensing, and transcription repression of wc-2. *The Journal of biological chemistry*. 278:3801-3808.

Christensen, M.K. 2007. A frequency-independent nitrate reductase rhythm in Neurospora crassa. Vol. PhD. Universitet i Stavanger.

Christensen, M.K., G. Falkeid, J.J. Loros, J.C. Dunlap, C. Lillo, and P. Ruoff. 2004. A nitrate-induced frq-less oscillator in Neurospora crassa. *Journal of biological rhythms*. 19:280-286.

Clarke, P.A., R. te Poele, R. Wooster, and P. Workman. 2001. Gene expression microarray analysis in cancer biology, pharmacology, and drug development: progress and potential. *Biochemical pharmacology*. 62:1311-1336.

Davis, R.H. 2000. Neurospora. Contributions of a model organism. Oxford University Press, Oxford.

de Paula, R.M., M.W. Vitalini, R.H. Gomer, and D. Bell-Pedersen. 2007. Complexity of the Neurospora crassa circadian clock system: multiple loops and oscillators. *Cold Spring Harbor symposia on quantitative biology*. 72:345-351.

Do, J.H., and D.K. Choi. 2008. Clustering approaches to identifying gene expression patterns from DNA microarray data. *Molecules and cells*. 25:279-288.

Dr. Richard Simon, B.-A.D.T., and T.E. Corporation. November, 2012. BRB-ArrayTools Version 4.3: User's Manual

.

Draghici, S. 2011 Statistics and Data Analysis for Microarrays Using R and Bioconductor. CRC Press, Boca Raton London New York.

Dunlap, J.C., J.J. Loros, H.V. Colot, A. Mehra, W.J. Belden, M. Shi, C.I. Hong, L.F. Larrondo, C.L. Baker, C.H. Chen, C. Schwerdtfeger, P.D. Collopy, J.J. Gamsby, and R. Lambreghts. 2007. A circadian clock in Neurospora: how genes and proteins cooperate to produce a sustained, entrainable, and compensated biological oscillator with a period of about a day. *Cold Spring Harbor symposia on quantitative biology*. 72:57-68.

Dunlap, J.C., Loros,Jennifer J. , DeCoursey,Patricia J. . 2004. Chronobiology: Biological Timekeeping. Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts, U.S.A.

Eisen, M.B., P.T. Spellman, P.O. Brown, and D. Botstein. 1998. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America*. 95:14863-14868.

Feng, B., and G.A. Marzluf. 1996. The regulatory protein NIT4 that mediates nitrate induction in Neurospora crassa contains a complex tripartite activation domain with a novel leucine-rich, acidic motif. *Current genetics*. 29:537-548.

Fu, Y.H., and G.A. Marzluf. 1987. Molecular cloning and analysis of the regulation of nit-3, the structural gene for nitrate reductase in Neurospora crassa. *Proceedings of the National Academy of Sciences of the United States of America*. 84:8243-8247.

Galagan, J.E., S.E. Calvo, K.A. Borkovich, E.U. Selker, N.D. Read, D. Jaffe, W. FitzHugh, L.J. Ma, S. Smirnov, S. Purcell, B. Rehman, T. Elkins, R. Engels, S. Wang, C.B. Nielsen, J. Butler, M. Endrizzi, D. Qui, P. Ianakiev, D. Bell-Pedersen, M.A. Nelson, M. Werner-Washburne, C.P. Selitrennikoff, J.A. Kinsey, E.L. Braun, A. Zelter, U. Schulte, G.O. Kothe, G. Jedd, W. Mewes, C. Staben, E. Marcotte, D. Greenberg, A. Roy, K. Foley, J. Naylor, N. Stange-Thomann, R. Barrett, S. Gnerre, M. Kamal, M. Kamvysselis, E. Mauceli, C. Bielke, S. Rudd, D. Frishman, S. Krystofova, C. Rasmussen, R.L. Metzenberg, D.D. Perkins, S. Kroken, C. Cogoni, G. Macino, D. Catcheside, W. Li, R.J. Pratt, S.A. Osmani, C.P. DeSouza, L. Glass, M.J. Orbach, J.A. Berglund, R. Voelker, O. Yarden, M. Plamann, S. Seiler, J. Dunlap, A. Radford, R. Aramayo, D.O. Natvig, L.A. Alex, G. Mannhaupt, D.J. Ebbole, M. Freitag, I. Paulsen, M.S. Sachs, E.S. Lander, C. Nusbaum, and B. Birren. 2003. The genome sequence of the filamentous fungus Neurospora crassa. *Nature*. 422:859-868.

Gao-Rubinelli, F., and G.A. Marzluf. 2004. Identification and characterization of a nitrate transporter gene in Neurospora crassa. *Biochemical genetics*. 42:21-34.

Greenwald, C.J. 2010. Temporal and Spatial Regulation of Gene Expression During Asexual Development of Neurospora crassa. *Genetics*.

Jensen, K.J. 2012. On the circadian organization of nitrogen and sulphur metabolism in Neurospora crassa. *In* Faculty of Science and Technology Department of Mathematics and Natural Science. Vol. PhD. Universitet i Stavanger.

Jinhu, G., and L. Yi. 2010. Molecular mechanism of the Neurospora circadian oscillator. *Protein & cell*. 1:331-341.
Khademi, S., J. O'Connell, 3rd, J. Remis, Y. Robles-Colmenares, L.J. Miercke, and R.M. Stroud. 2004. Mechanism of ammonia transport by Amt/MEP/Rh: structure of AmtB at 1.35 A. *Science*. 305:1587-1594.
Lakin-Thomas, P.L., and S. Brody. 2004. Circadian rhythms in microorganisms: new complexities. *Annual review of microbiology*. 58:489-519.
Lillo, C., C. Meyer, and P. Ruoff. 2001. The nitrate reductase circadian system. The central clock dogma contra multiple oscillatory feedback loops. *Plant physiology*. 125:1554-1557.
Liu, Y., and D. Bell-Pedersen. 2006. Circadian rhythms in Neurospora crassa and other filamentous fungi. *Eukaryotic cell*. 5:1184-1193.
Lorenz, M.C., and J. Heitman. 1998. The MEP2 ammonium permease regulates pseudohyphal differentiation in Saccharomyces cerevisiae. *The EMBO journal*. 17:1236-1247.
Maheshwari, R. 1999. Microconidia of Neurospora crassa. *Fungal genetics and biology : FG & B*. 26:1-18.
Marzluf, G.A. 1997. Genetic regulation of nitrogen metabolism in the fungi. *Microbiology and molecular biology reviews : MMBR*. 61:17-32.
Moler, C. 2004. The Origins of MATLAB. *Mathworks Newsletter*.
Moore, J., Stanitski, Conrad, Jurs, Peter. 2009. Principles of Chemistry: The Molecular Science. Cengage Learning, Belmont, USA.
Pan, H., B. Feng, and G.A. Marzluf. 1997. Two distinct protein-protein interactions between the NIT2 and NMR regulatory proteins are required to establish nitrogen metabolite repression in Neurospora crassa. *Molecular microbiology*. 26:721-729.
Park, G., S. Pan, and K.A. Borkovich. 2008. Mitogen-activated protein kinase cascade required for regulation of development and secondary metabolism in Neurospora crassa. *Eukaryotic cell*. 7:2113-2122.
Perkins, D.D., Radford A., Sachs MS. 2001. The Neurospora Compendium. Chromosomal Loci. Academic Press, San Diego.
Ruepp, A., A. Zollner, D. Maier, K. Albermann, J. Hani, M. Mokrejs, I. Tetko, U. Guldener, G. Mannhaupt, M. Munsterkotter, and H.W. Mewes. 2004. The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. *Nucleic acids research*. 32:5539-5545.
Ruoff, P. 2011. Homeostatic Controll Mechanisms and their Oscillatory behaviors: nitrate homeostasis and circadian nitrogen assimilation.
Seale, T. 1973. Life cycle of Neurospora crassa viewed by scanning electron microscopy. *Journal of bacteriology*. 113:1015-1025.
Selker, E.U., N.A. Tountas, S.H. Cross, B.S. Margolin, J.G. Murphy, A.P. Bird, and M. Freitag. 2003. The methylated component of the Neurospora crassa genome. *Nature*. 422:893-897.
Sokolovsky, V.J. 2000. Dejstvie stressorov na differencial´nuju ekspressiju genov v hode razvitija Neuropora crassa. *Uspehi biologicheskoj himii* :85-152.
Springer, M.L., and C. Yanofsky. 1989. A morphological and genetic analysis of conidiophore development in Neurospora crassa. *Genes & development*. 3:559-571.

Sweet-Cordero, A., S. Mukherjee, A. Subramanian, H. You, J.J. Roix, C. Ladd-Acosta, J. Mesirov, T.R. Golub, and T. Jacks. 2005. An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis. *Nature genetics*. 37:48-55.

Tao, Y., and G.A. Marzluf. 1999. The NIT2 nitrogen regulatory protein of Neurospora: expression and stability of nit-2 mRNA and protein. *Current genetics*. 36:153-158.

Tisdall, J. 2001. Beginning Perl for Bioinformatics. O'Reilly. 384 pp.

Tomsett, A.B., and R.H. Garrett. 1980. The isolation and characterization of mutants defective in nitrate assimilation in Neurospora crassa. *Genetics*. 95:649-660.

Wijnen, H., and M.W. Young. 2006. Interplay of circadian clocks and metabolic rhythms. *Annual review of genetics*. 40:409-448.

Wiltgen, M., and G.P. Tilz. 2007. DNA microarray analysis: principles and clinical impact. *Hematology*. 12:271-287.

Yao, B., S.N. Rakhade, Q. Li, S. Ahmed, R. Krauss, S. Draghici, and J.A. Loeb. 2004. Accuracy of cDNA microarray methods to detect small gene expression changes induced by neuregulin on breast epithelial cells. *BMC bioinformatics*. 5:99.

Young, J.L., G. Jarai, Y.H. Fu, and G.A. Marzluf. 1990. Nucleotide sequence and analysis of NMR, a negative-acting regulatory gene in the nitrogen circuit of Neurospora crassa. *Molecular & general genetics : MGG*. 222:120-128.

Appendix 1. Statistically significantly expressed genes, extracted by arbirtary cut-off

DDNH4		DDNO3		DDNV		LLNH4		DDNO3	
GeneNames	Mean ratio	GeneNames	Mean ratio	GeneNames	Mean ratio	GeneNames	Mean ratio	GeneNames	Mean ratio
NCU05897	3,693	NCU09040	3,993	NCU05143	3,616	NCU02500	3,787	NCU02500	3,782
NCU05143	3,325	NCU05143	3,830	NCU07345	3,547	NCU05897	3,716	NCU05897	3,396
NCU09040	3,042	NCU02329	3,768	NCU05897	3,149	NCU09040	3,513	NCU09040	3,235
NCU04963	2,809	NCU07345	3,669	NCU09040	3,117	NCU04998	3,469	NCU07894	3,202
NCU07345	2,798	NCU02500	3,304	NCU02329	2,886	NCU07869	3,355	NCU05627	2,971
NCU02500	2,717	NCU08087	3,234	NCU08087	2,867	NCU05627	3,340	NCU02329	2,779
NCU08087	2,707	NCU07894	3,148	NCU05706	2,854	NCU04268	3,313	NCU05143	2,763
NCU01873	2,657	NCU08727	2,989	NCU07928	2,788	NCU08986	3,302	NCU03509	2,735
NCU08769	2,590	NCU00659	2,878	NCU05732	2,742	NCU01517	3,046	NCU10051	2,715
NCU05627	2,455	NCU04538	2,865	NCU00659	2,727	NCU09873	3,005	NCU08727	2,658
NCU09873	2,449	NCU07928	2,856	NCU03684	2,604	NCU03732	2,992	NCU04998	2,653
NCU03684	2,383	NCU07966	2,840	NCU04963	2,470	NCU04963	2,926	NCU01517	2,611
NCU04722	2,352	NCU05706	2,746	NCU09049	2,411	NCU04722	2,904	NCU04963	2,520
NCU02329	2,264	NCU04431	2,725	NCU07894	2,365	NCU07232	2,762	NCU08986	2,442
NCU05706	2,253	NCU09049	2,655	NCU08986	2,301	NCU01620	2,534	NCU08087	2,391
NCU07869	2,192	NCU08670	2,639	NCU01873	2,299	NCU04265	2,499	NCU04562	2,364
NCU00659	2,171	NCU02142	2,564	NCU04722	2,294	NCU07337	2,430	NCU07966	2,362
NCU09525	2,166	NCU05619	2,558	NCU09686	2,286	NCU08561	2,424	NCU01050	2,361
NCU09049	2,143	NCU03509	2,553	NCU10051	2,282	NCU02798	2,422	NCU10021	2,325
NCU01517	2,104	NCU05897	2,546	NCU00875	2,222	NCU09525	2,376	NCU09751	2,322
NCU07928	2,103	NCU07995	2,459	NCU10016	2,205	NCU00176	2,375	NCU07337	2,316
NCU07894	2,100	NCU07337	2,446	NCU02080	2,181	NCU10021	2,318	NCU10016	2,254
NCU09698	2,083	NCU05732	2,433	NCU11136	2,094	NCU03197	2,277	NCU09698	2,236
NCU07324	2,075	NCU06005	2,429	NCU08770	2,064	NCU03509	2,271	NCU04234	2,221
NCU08986	2,074	NCU06128	2,367	NCU05627	2,047	NCU06815	2,258	NCU06295	2,209
NCU04482	2,004	NCU04727	2,343	NCU09698	2,019	NCU06651	2,235	NCU05706	2,182
NCU01833	2,003	NCU08770	2,312	NCU04667	2,008	NCU01559	2,233	NCU04722	2,130
NCU08291	-2,021	NCU04167	2,294	NCU04656	-2,001	NCU09686	2,187	NCU07869	2,120
NCU07718	-2,022	NCU04883	2,227	NCU05137	-2,003	NCU09364	2,169	NCU02046	2,110
NCU07097	-2,099	NCU00586	2,215	NCU00585	-2,023	NCU02897	2,161	NCU07345	2,022
NCU10941	-2,105	NCU05510	2,201	NCU00790	-2,032	NCU01560	2,159	NCU02533	2,003
NCU05770	-2,123	NCU03254	2,179	4nc453_010	-2,082	NCU01391	2,157	NCU08013	-2,014
NCU05969	-2,142	NCU08075	2,169	NCU05788	-2,102	NCU04562	2,148	NCU00355	-2,029
NCU04510	-2,175	NCU00638	2,158	NCU08291	-2,127	NCU08434	2,147	NCU09620	-2,038
NCU05763	-2,203	NCU00852	2,157	NCU06555	-2,178	4nc450_150	2,079	NCU08037	-2,051
NCU08037	-2,389	NCU05040	2,150	NCU03107	-2,193	NCU09364	2,070	NCU04533	-2,063
NCU09508	-2,407	NCU04667	2,150	NCU02380	-2,194	NCU01640	2,067	4nc453_010	-2,070
NCU00732	-2,552	NCU07324	2,127	NCU04929	-2,198	NCU05977	2,047	NCU05788	-2,086
NCU02904	-2,588	NCU04471	2,112	NCU09620	-2,208	NCU00180	2,047	NCU03408	-2,089
NCU07449	-3,006	NCU11136	2,107	NCU07718	-2,211	NCU07237	2,043	NCU07257	-2,151
NCU02930	-3,363	NCU01095	2,106	NCU04452	-2,249	NCU04805	2,037	NCU04677	-2,177
NCU06420	-3,372	NCU05627	2,052	NCU04895	-2,264	NCU05291	2,022	NCU03107	-2,207
NCU05768	-4,995	NCU03098	2,021	NCU09495	-2,301	NCU08087	2,020	NCU02904	-2,278
NCU08129	-5,206	NCU03830	2,004	NCU08037	-2,316	NCU06295	2,015	NCU05969	-2,346
		3nc440_690	-2,004	NCU02904	-2,317	NCU04955	2,009	NCU04895	-2,366
		NCU01972	-2,005	NCU04533	-2,383	NCU07253	2,001	NCU06555	-2,384

		NCU04603	-2,020	NCU10941	-2,472	NCU09575	-2,020	NCU09508	-2,409
		NCU04929	-2,022	NCU04510	-2,495	NCU04931	-2,060	NCU06301	-2,457
		NCU01720	-2,025	NCU09508	-2,510	NCU07718	-2,099	NCU07569	-2,547
		NCU09761	-2,034	NCU00915	-2,581	NCU07338	-2,134	NCU07338	-2,585
		NCU07718	-2,045	NCU06420	-2,646	NCU02333	-2,134	NCU07053	-2,619
		NCU05768	-2,049	NCU07449	-2,829	NCU05768	-2,156	NCU06420	-2,649
		NCU03152	-2,067	NCU00732	-3,081	NCU06785	-2,165	NCU04510	-2,958
		NCU05598	-2,069	NCU02930	-3,492	NCU02904	-2,191	NCU02930	-3,110
		NCU07126	-2,086	NCU08129	-4,636	NCU01830	-2,346	NCU10941	-3,173
		NCU04342	-2,093			NCU09495	-2,470	NCU00175	-3,314
		NCU04931	-2,096			NCU06895	-2,485	NCU00732	-3,320
		NCU00790	-2,096			NCU07743	-2,501	NCU08129	-4,513
		NCU00763	-2,105			NCU05788	-2,509		
		NCU01898	-2,109			NCU04510	-2,529		
		NCU03725	-2,122			NCU05969	-2,577		
		NCU06170	-2,129			NCU10941	-2,674		
		NCU06912	-2,141			NCU09620	-2,733		
		NCU02328	-2,154			NCU00732	-2,783		
		NCU05841	-2,157			NCU06420	-3,056		
		NCU09495	-2,163			NCU09508	-3,105		
		NCU04895	-2,175			NCU08037	-3,495		
		NCU07338	-2,184			NCU02930	-3,590		
		NCU07454	-2,185			NCU08129	-4,081		
		NCU06940	-2,207						
		NCU06911	-2,208						
		NCU07198	-2,255						
		NCU03408	-2,257						
		NCU09775	-2,317						
		NCU00585	-2,326						
		NCU08291	-2,328						
		NCU07569	-2,394						
		NCU06555	-2,431						
		NCU05788	-2,438						
		NCU05018	-2,448						
		NCU04510	-2,461						
		NCU05137	-2,573						
		NCU08390	-2,576						
		NCU03151	-2,592						
		NCU00355	-2,630						
		NCU07097	-2,630						
		NCU10656	-2,651						
		NCU08183	-2,660						
		NCU07257	-2,664						
		NCU09210	-2,767						
		NCU05969	-2,787						
		NCU02904	-2,813						
		4nc453_010	-2,841						
		NCU09508	-2,876						
		NCU04533	-2,893						
		NCU07449	-3,012						
		NCU02930	-3,250						
		NCU07053	-3,263						
		NCU06301	-3,373						
		NCU06420	-3,456						
		NCU09620	-3,483						

		NCU10941	-3,683						
		NCU08129	-4,088						
		NCU00732	-4,088						

Appendix 2. Genes which show similar high/low expression values at particular conditions

GeneNames	Conditions	Mean ratio	Mean ratio
Conditions		**DDNH4**	**DDNO3**
NCU05143	**Genes, up-regulated in dark**	3,325	3,830
NCU07345		2,798	3,669
NCU02500		2,717	3,304
NCU02329		2,264	3,768
NCU05706		2,253	2,746
NCU00659		2,171	2,878
NCU09049		2,143	2,655
NCU07928		2,103	2,856
NCU07894		2,100	3,148
NCU07324		2,075	2,127
NCU08291	**Genes, down-regulated in dark**	-2,021	-2,328
NCU07718		-2,022	-2,045
NCU07097		-2,099	-2,630
NCU05969		-2,142	-2,787
NCU07449		-3,006	-3,012
NCU05768		-4,995	-2,049
Conditions		**LNH4**	**DDNH4**
NCU04963	**Genes, up-regulated in the presence of ammonium**	2,926	2,809
NCU02500		3,787	2,717
NCU09873		3,005	2,449
NCU04722		2,904	2,352
NCU07869		3,355	2,192
NCU09525		2,376	2,166
NCU01517		3,046	2,104
NCU08986		3,302	2,074
NCU07718	**down-regulated in the presence of ammonium**	-2,099	-2,022
NCU05969		-2,577	-2,142
NCU08037		-3,495	-2,389
NCU05768		-2,156	-4,995
Conditions		**LLNO3**	**DDNO3**
NCU05143	**Genes, up-regulated in the nitrate medium**	2,763	3,830
NCU02329		2,779	3,768
NCU07345		2,022	3,669
NCU02500		3,782	3,304
NCU07894		3,202	3,148
NCU08727		2,658	2,989

NCU07966		2,362	2,840
NCU05706		2,182	2,746
NCU03509		2,735	2,553
NCU07337		2,316	2,446
NCU04895	**Genes, down-regulated in the nitrate medium**	-2,366	-2,175
NCU07338		-2,585	-2,184
NCU03408		-2,089	-2,257
NCU07569		-2,547	-2,394
NCU06555		-2,384	-2,431
NCU05788		-2,086	-2,438
NCU00355		-2,029	-2,630
NCU07257		-2,151	-2,664
NCU05969		-2,346	-2,787
NCU04533		-2,063	-2,893
NCU07053		-2,619	-3,263
NCU06301		-2,457	-3,373
NCU09620		-2,038	-3,483
Conditions	**Genes, up-regulated in light**	**LLNH4**	**LLNO3**
NCU02500		3,787	3,782
NCU04998		3,469	2,653
NCU07869		3,355	2,120
NCU08986		3,302	2,442
NCU01517		3,046	2,611
NCU04963		2,926	2,520
NCU04722		2,904	2,130
NCU07337		2,430	2,316
NCU10021		2,318	2,325
NCU03509		2,271	2,735
NCU04562		2,148	2,364
NCU06295		2,015	2,209
NCU07338	**Genes, down-regulated in light**	-2,134	-2,585
NCU02904		-2,191	-2,278
NCU05788		-2,509	-2,086
NCU05969		-2,577	-2,346
NCU09620		-2,733	-2,038
NCU08037		-3,495	-2,051

Appendix 3. Functional Analysis of genes showing low expression values at particular conditions

DDNO3 and DDNH4, expresison values <-2			
FUNCTIONAL CATEGORY	abs SET	rel SET	genes SET
01 METABOLISM	2	33.3	NCU08291 NCU05969
01.05 C-compound and carbohydrate metabolism	1	16.6	NCU05969
01.05.03 polysaccharide metabolism	1	16.6	NCU05969
01.07 metabolism of vitamins, cofactors, and prosthetic groups	1	16.6	NCU08291
01.07.01 biosynthesis of vitamins, cofactors, and prosthetic groups	1	16.6	NCU08291
01.20 secondary metabolism	1	16.6	NCU08291
01.20.19 metabolism of secondary products derived from glycine, L-serine and L-alanine	1	16.6	NCU08291
01.20.19.01 metabolism of porphyrins	1	16.6	NCU08291
01.25 extracellular metabolism	1	16.6	NCU05969
01.25.01 extracellular polysaccharide degradation	1	16.6	NCU05969
16 PROTEIN WITH BINDING FUNCTION OR COFACTOR REQUIREMENT (structural or catalytic)	1	16.6	NCU08291
16.17 metal binding	1	16.6	NCU08291
16.21 complex cofactor/cosubstrate/vitamine binding	1	16.6	NCU08291
16.21.08 Fe/S binding	1	16.6	NCU08291
34 INTERACTION WITH THE ENVIRONMENT	1	16.6	NCU08291
34.01 homeostasis	1	16.6	NCU08291
34.01.01 homeostasis of cations	1	16.6	NCU08291
34.01.01.01 homeostasis of metal ions (Na, K, Ca etc.)	1	16.6	NCU08291
99 UNCLASSIFIED PROTEINS	4	66.6	NCU21132 NCU05768 NCU07449 NCU21133
LLNH4 and DDNH4, expresison values <-2			
FUNCTIONAL CATEGORY	abs SET	rel SET	genes SET
01 METABOLISM	1	20	NCU05969
01.05 C-compound and carbohydrate metabolism	1	20	NCU05969
01.05.03 polysaccharide metabolism	1	20	NCU05969
01.25 extracellular metabolism	1	20	NCU05969

01.25.01 extracellular polysaccharide degradation	1	20	NCU05969
99 UNCLASSIFIED PROTEINS	4	80	NCU21132 NCU05768 NCU08037 NCU21133
LLNO3 and DDNO3, expresison values <-2			
FUNCTIONAL CATEGORY	abs SET	rel SET	genes SET
01 METABOLISM	3	23.0	NCU05969 NCU07338 NCU07053
01.01 amino acid metabolism	1	7.69	NCU07053
01.01.03 assimilation of ammonia, metabolism of the glutamate group	1	7.69	NCU07053
01.01.03.02 metabolism of glutamate	1	7.69	NCU07053
01.01.03.02.02 degradation of glutamate	1	7.69	NCU07053
01.01.03.03 metabolism of proline	1	7.69	NCU07053
01.01.03.03.01 biosynthesis of proline	1	7.69	NCU07053
01.01.05 metabolism of urea cycle, creatine and polyamines	1	7.69	NCU07053
01.01.05.01 metabolism of polyamines	1	7.69	NCU07053
01.01.05.01.02 degradation of polyamines	1	7.69	NCU07053
01.01.09 metabolism of the cysteine - aromatic group	1	7.69	NCU07053
01.01.09.05 metabolism of tyrosine	1	7.69	NCU07053
01.02 nitrogen, sulfur and selenium metabolism	1	7.69	NCU07053
01.05 C-compound and carbohydrate metabolism	3	23.0	NCU07338 NCU07053 NCU05969
01.05.03 polysaccharide metabolism	1	7.69	NCU05969
01.05.06 C-2 compound and organic acid metabolism	1	7.69	NCU07053
01.05.06.07 C-2 compound and organic acid catabolism	1	7.69	NCU07053
01.06 lipid, fatty acid and isoprenoid metabolism	1	7.69	NCU07053
01.06.06 isoprenoid metabolism	1	7.69	NCU07053
01.07 metabolism of vitamins, cofactors, and prosthetic groups	1	7.69	NCU07053
01.07.01 biosynthesis of vitamins, cofactors, and prosthetic groups	1	7.69	NCU07053
01.20 secondary metabolism	1	7.69	NCU07053
01.20.05 metabolism of acetic acid derivatives	1	7.69	NCU07053
01.20.17 metabolism of secondary products derived from primary amino acids	1	7.69	NCU07053
01.20.17.01 metabolism of nonprotein amino acids	1	7.69	NCU07053

01.20.29 metabolism of secondary products derived from L-glutamic acid, L-proline and L-ornithine	1	7.69	NCU07053
01.20.31 metabolism of secondary products derived from L-lysine, L-arginine and L-histidine	1	7.69	NCU07053
01.25 extracellular metabolism	1	7.69	NCU05969
01.25.01 extracellular polysaccharide degradation	1	7.69	NCU05969
02 ENERGY	1	7.69	NCU07053
02.11 electron transport and membrane-associated energy conservation	1	7.69	NCU07053
02.13 respiration	1	7.69	NCU07053
02.16 fermentation	1	7.69	NCU07053
02.16.01 alcohol fermentation	1	7.69	NCU07053
02.45 energy conversion and regeneration	1	7.69	NCU07053
14 PROTEIN FATE (folding, modification, destination)	1	7.69	NCU07338
14.07 protein modification	1	7.69	NCU07338
14.07.02 modification with sugar residues (e.g. glycosylation, deglycosylation)	1	7.69	NCU07338
16 PROTEIN WITH BINDING FUNCTION OR COFACTOR REQUIREMENT (structural or catalytic)	2	15.3	NCU00355 NCU07053
16.13 C-compound binding	1	7.69	NCU07053
16.21 complex cofactor/cosubstrate/vitamine binding	2	15.3	NCU00355 NCU07053
16.21.01 heme binding	1	7.69	NCU00355
16.21.07 NAD/NADP binding	1	7.69	NCU07053
20 CELLULAR TRANSPORT, TRANSPORT FACILITIES AND TRANSPORT ROUTES	1	7.69	NCU07053
20.01 transported compounds (substrates)	1	7.69	NCU07053
20.01.15 electron transport	1	7.69	NCU07053
32 CELL RESCUE, DEFENSE AND VIRULENCE	3	23.0	NCU00355 NCU07053 NCU07569
32.01 stress response	3	23.0	NCU00355 NCU07569 NCU07053
32.01.01 oxidative stress response	2	15.3	NCU07053 NCU00355
32.01.03 osmotic and salt stress response	1	7.69	NCU07053
32.05 disease, virulence and defense	1	7.69	NCU07053
32.05.01 resistance proteins	1	7.69	NCU07053
32.05.03 defense related proteins	1	7.69	NCU07053
32.07 detoxification	2	15.3	NCU07053 NCU00355
32.07.03 detoxification by modification	1	7.69	NCU07053
32.07.07 oxygen and radical detoxification	1	7.69	NCU00355

32.07.07.01 catalase reaction	1	7.69	NCU00355
32.07.09 detoxification by degradation	1	7.69	NCU07053
34 INTERACTION WITH THE ENVIRONMENT	1	7.69	NCU04533
34.07 cell adhesion	1	7.69	NCU04533
34.07.01 cell-cell adhesion	1	7.69	NCU04533
42 BIOGENESIS OF CELLULAR COMPONENTS	1	7.69	NCU07053
42.16 mitochondrion	1	7.69	NCU07053
99 UNCLASSIFIED PROTEINS	7	53.8	NCU05788 NCU20925 NCU04895 NCU21503 NCU06555 NCU06301 NCU21561
LLNO3 and LLNH4, expresison values <-2			
FUNCTIONAL CATEGORY	abs SET	rel SET	genes SET
01 METABOLISM	3	50	NCU07338 NCU02904 NCU05969
01.05 C-compound and carbohydrate metabolism	3	50	NCU02904 NCU07338 NCU05969
01.05.03 polysaccharide metabolism	1	16.6	NCU05969
01.05.11 aromate metabolism	1	16.6	NCU02904
01.05.11.07 aromate catabolism	1	16.6	NCU02904

Appendix 4. Statistically significantly expressed genes, extracted by BRB-array tool

	Gene ID_Gene Name	DDNH4	DDNO3	DDNV	LLNH4	LLNO3
1.	2023__NCU02343	-1,391	-3,263	-1,940	-1,946	-2,619
2.	10890__NCU00381	0,572	0,728	0,666	-0,717	-0,232
3.	2990__NCU06991	-0,992	-2,141	-0,786	-1,108	-1,928
4.	3660__NCU08791	0,272	1,295	1,448	0,165	1,844
5.	4133__NCU01235	-0,420	-0,170	-0,233	-1,322	-0,911
6.	2317__NCU06167	0,174	0,003	0,038	1,591	-0,227
7.	1486__NCU02773	-0,855	-0,144	-0,809	0,401	0,386
8.	1637__NCU01088	1,135	1,610	1,612	0,309	1,294
9.	6656__NCU04149	1,131	1,995	1,617	0,644	1,150
10.	10870__NCU06447	-0,265	1,015	0,283	-0,044	-0,034
11.	11260__NCU06588	-0,652	0,548	-1,061	-1,323	-0,471
12.	1168__NCU00330	0,965	0,328	1,424	2,534	0,609
13.	148__NCU03183	1,247	-0,061	0,777	1,244	0,379
14.	486__NCU05291	0,360	0,585	0,805	-0,398	-0,179
15.	6531__NCU02877	-0,547	-1,711	-0,312	-1,302	-2,014
16.	6736__NCU03967	0,828	-0,677	0,571	0,932	-0,237
17.	6989__NCU10188	-0,004	1,365	0,357	0,333	0,949
18.	7156__NCU09004	0,745	0,314	1,325	1,955	1,094
19.	7515__NCU08997	0,276	0,752	0,823	-0,170	-0,104
20.	8679__NCU00847	0,813	1,583	1,140	0,340	0,168
21.	8952__NCU08199	0,480	-0,836	0,249	1,025	0,241
22.	9192__NCU01972	0,383	1,258	0,543	-0,564	0,249
23.	9275__NCU03945	0,655	1,577	1,584	0,437	0,602
24.	11199__NCU00990	-0,095	1,438	0,678	0,019	0,642
25.	6893__NCU02939	1,158	1,730	1,495	-0,136	0,484
26.	880__NCU02378	0,343	0,955	1,376	-0,765	-0,705
27.	8688__NCU02817	0,241	1,479	1,060	-0,610	-0,351
28.	1035__NCU08239	1,095	2,312	2,064	0,390	0,594
29.	5453__NCU02938	-0,022	1,029	0,754	-0,161	-0,129
30.	4607__NCU07838	-0,684	-2,394	-1,235	-0,271	-2,547
31.	7844__NCU06642	0,908	-0,104	0,585	0,350	-0,186
32.	8439__NCU11161	0,388	-0,703	0,426	-0,218	-1,021
33.	10734__NCU09592	1,639	-1,176	1,329	1,817	0,248
34.	10006__NCU07025	0,381	-0,661	0,145	0,516	-0,500
35.	10443__NCU01176	0,357	-0,564	0,151	0,689	0,229
36.	11228__NCU07819	-0,100	-1,141	-0,817	0,525	0,051
37.	2275__NCU07334	0,527	0,467	0,128	-0,648	-0,141
38.	2595__NCU07598	-0,182	-1,444	-0,758	-0,226	-1,216
39.	2791__NCU02446	-2,123	-0,752	-1,740	-0,694	0,001
40.	3611__NCU08095	0,974	0,640	0,871	-0,055	0,176
41.	5214__NCU00771	0,448	0,614	1,053	-0,071	-0,139
42.	9523__NCU05764	-0,276	-0,423	-0,230	0,721	0,271
43.	9533__NCU00776	0,633	1,249	1,398	0,111	0,459

44.	11064__NCU00323	0,814	1,561	1,705	0,033	0,323
45.	3521__NCU05474	-0,636	-2,122	-0,934	-1,592	-1,567
46.	7828__NCU01887	0,365	0,816	0,592	-0,575	-0,040
47.	8199__NCU04839	-0,003	-0,700	0,456	0,218	-0,957
48.	11312__NCU06702	0,544	1,700	1,032	-0,522	0,326
49.	1925__NCU05282	1,341	2,150	1,962	-0,439	0,370
50.	3906__NCU06001	0,655	-0,268	0,708	-0,135	-0,337
51.	2802__NCU05373	1,418	2,865	1,371	-0,466	1,201
52.	5931__NCU04603	0,216	-0,327	-0,140	-0,800	-1,051
53.	816__NCU07910	0,664	0,555	0,622	-0,167	-0,416
54.	5087__NCU00534	1,308	1,434	1,313	0,315	0,359
55.	2044__NCU01391	1,092	1,358	1,531	0,060	0,109
56.	2607__NCU00251	0,737	0,237	1,055	0,182	-0,301
57.	2896__NCU01944	-0,354	-1,243	-0,897	0,094	-1,167
58.	7108__NCU01618	0,666	0,725	0,622	-0,397	0,127
59.	7242__NCU00732	0,662	-1,130	-1,104	-1,261	-2,177
60.	8347__NCU05732	0,867	-0,926	1,238	0,663	-0,741
61.	9605__NCU03543	0,361	1,068	0,470	-0,352	0,387
62.	9854__NCU09686	1,552	0,704	1,167	2,009	0,552
63.	11450__NCU08263	-0,151	-0,956	-0,139	-0,298	-1,337
64.	11444__NCU06093	0,214	-0,254	-0,224	0,811	0,369
65.	1188__NCU04547	1,603	2,840	0,751	1,356	2,362
66.	1358__NCU04536	1,936	1,964	1,742	-0,059	0,520
67.	4602__NCU05747	1,297	1,559	0,549	0,181	-0,121
68.	511__NCU09577	0,653	-0,324	0,074	0,961	-0,190
69.	10630__NCU01989	-0,248	0,280	-0,332	0,456	1,016
70.	10966__NCU05861	1,335	-0,153	0,282	1,268	-1,195
71.	1268__NCU08728	0,879	-0,470	0,016	1,276	0,683
72.	2025__NCU07400	0,931	1,545	1,314	0,221	0,520
73.	3354__NCU07279	-3,006	-3,012	-2,829	-0,688	0,043
74.	3608__NCU06650	0,627	-0,398	0,589	1,548	-0,622
75.	3610__NCU08087	1,585	2,564	1,688	0,597	0,983
76.	3761__NCU02328	-0,297	0,461	-0,060	-0,914	-0,630
77.	4217__NCU01057	-1,185	-0,299	-1,369	0,109	0,932
78.	4630__NCU07572	1,775	0,833	1,645	1,753	0,878
79.	4852__NCU00763	0,698	-0,420	0,599	1,208	0,785
80.	5884__NCU05303	1,634	2,639	1,864	0,864	1,434
81.	6640__NCU02323	1,659	1,457	1,574	-0,081	0,491
82.	6970__NCU02283	-0,254	0,835	0,090	-0,498	-0,286
83.	7449__8l21_010	0,421	-0,203	0,687	1,806	0,918
84.	7837__NCU04895	-1,022	-0,066	-1,128	-0,957	-0,250
85.	8161__NCU04168	1,949	1,806	1,806	0,529	0,757
86.	8766__NCU05319	-1,491	-1,397	-2,249	-1,102	-1,165
87.	9561__NCU06785	0,739	1,660	1,304	-0,562	0,269
88.	11217__NCU08953	0,230	-1,001	-0,162	0,204	-0,525
89.	1132__NCU02926	0,191	-0,190	-0,698	0,822	0,394
90.	1627__NCU01350	0,965	0,195	0,752	1,968	1,129
91.	1967__NCU05887	0,766	-0,865	0,141	1,507	-0,202
92.	1996__NCU01095	0,923	-0,391	0,690	0,938	-0,073
93.	2249__NCU00367	0,295	0,176	0,312	1,435	0,495
94.	2588__NCU05742	1,357	1,807	2,222	0,527	0,724

95.	468__NCU04476	-0,174	0,096	-0,912	-0,184	0,273
96.	7054__NCU02486	0,827	1,768	1,024	-0,014	0,534
97.	264__NCU00816	1,368	0,274	1,166	1,085	0,216
98.	3842__NCU04163	-0,225	0,720	-0,154	-0,588	0,281
99.	4949__NCU01925	1,933	0,888	1,874	2,424	0,810
100.	5322__NCU05443	-0,163	-0,799	-0,222	0,661	-0,241
101.	5341__NCU04961	0,584	0,823	0,806	-0,380	0,478
102.	5676__NCU03671	3,325	3,830	3,616	1,374	2,763
103.	5694__NCU09019	0,331	0,214	0,508	-0,035	-0,895
104.	5715__xnc090_300	-0,619	-1,118	-0,719	0,163	-0,101
105.	7111__NCU02462	-0,254	-0,641	-0,959	0,426	-0,739
106.	7838__NCU04978	0,966	1,640	1,256	-0,380	-0,244
107.	8304__1nc380_040	1,011	1,508	1,235	0,008	0,348
108.	10937__NCU07132	-0,436	0,064	-0,321	-1,020	0,532
109.	11079__NCU04846	0,128	0,206	0,561	1,665	0,862
110.	11433__NCU02760	0,495	1,680	0,622	-0,255	-0,177
111.	2692__NCU04608	0,849	0,397	0,869	0,009	-0,153
112.	3680__NCU00071	0,331	1,115	0,025	0,078	0,686
113.	5332__NCU09110	0,762	0,106	0,459	1,601	0,666
114.	7057__NCU09926	-0,395	-1,694	-0,944	-1,210	-3,314
115.	7619__2nc610_050	0,027	0,665	0,057	-0,792	-0,088
116.	8297__NCU01852	-0,307	-0,577	-0,465	0,731	-0,511
117.	9314__NCU08260	0,682	1,259	1,438	-0,546	0,226
118.	9427__NCU07834	1,155	1,293	1,085	0,074	0,525
119.	9659__NCU04876	-2,389	-1,485	-2,316	-3,495	-2,051
120.	10034__xnc090_250	1,358	0,673	1,103	3,313	1,976
121.	10641__14h13_010	1,349	0,256	0,406	2,992	1,339
122.	10883__3nc435_040	0,089	-0,491	-0,380	0,858	0,309
123.	11192__NCU03163	-0,449	-1,685	-0,892	0,122	-1,446
124.	11443__NCU05913	0,569	-1,060	-0,066	1,243	-0,882
125.	11457__NCU09905	-0,258	-1,805	-0,800	-0,053	-1,553
126.	1214__NCU09691	-0,177	0,030	-0,010	-1,089	-0,165
127.	1942__NCU05557	0,260	0,138	-0,337	1,219	0,270
128.	3045__7nc605_490	1,378	0,328	0,528	2,037	0,960
129.	3322__NCU00185	0,052	-1,421	-0,524	0,418	-0,937
130.	3542__NCU05924	-0,316	-1,001	-0,957	0,121	-0,549
131.	372__NCU07431	1,025	2,989	1,110	-0,212	2,658
132.	4383__NCU08394	2,657	1,325	2,299	1,824	0,801
133.	7045__NCU05388	0,067	1,040	0,071	-0,290	-0,053
134.	7075__NCU09420	0,304	0,152	-0,264	1,146	-0,011
135.	746__NCU10490	1,256	-0,882	0,597	1,338	-0,094
136.	8553__NCU02667	0,775	0,930	1,024	-0,245	0,193
137.	9796__NCU06221	-0,758	-1,313	-0,969	0,091	-1,086
138.	9825__4nc400_050	-0,383	0,544	0,381	-0,797	-0,911
139.	10007__NCU07053	0,593	-0,433	0,302	1,395	-0,854
140.	10013__NCU08957	-1,647	-2,448	-1,573	-1,967	-0,917
141.	10025__NCU05974	0,503	1,725	2,282	1,368	2,715
142.	10041__NCU09168	0,498	0,464	0,351	-1,500	-0,947
143.	10044__NCU09460	-0,199	-0,919	-0,387	0,504	-0,344
144.	10102__NCU02548	-0,305	-1,022	-1,074	-0,139	-0,102
145.	10125__NCU00695	0,225	1,750	1,350	0,000	1,757

146.	10151__NCU06505	2,449	1,309	1,952	3,005	1,465
147.	10188__4nc453_010	0,076	0,967	0,224	-0,364	0,478
148.	10263__NCU09536	-0,992	-1,633	-1,505	-2,501	-1,082
149.	10311__NCU00465	-0,605	1,056	-0,203	-1,586	-0,438
150.	10317__NCU01620	-0,061	1,009	-0,463	1,295	2,322
151.	10354__NCU04181	-0,846	-1,959	-1,223	-0,437	-1,819
152.	10358__NCU05258	0,244	0,293	-0,054	-1,316	-1,145
153.	10361__NCU05739	0,477	0,441	0,224	1,325	0,098
154.	10371__NCU08824	0,414	0,777	-0,136	1,733	1,464
155.	10395__NCU08999	-0,472	0,254	-0,558	0,243	0,793
156.	10456__NCU01921	2,103	2,856	2,788	0,859	1,812
157.	10457__1nc315_070	0,920	0,269	-0,027	1,328	1,166
158.	10463__NCU02533	-0,665	-1,740	-1,422	-1,477	-0,806
159.	10464__NCU02540	0,225	0,836	0,153	-0,486	-0,300
160.	10478__NCU00881	-0,583	-0,768	0,165	-0,822	0,231
161.	10503__NCU08403	1,750	0,572	1,233	1,744	1,087
162.	10550__NCU06970	-0,084	0,457	-0,398	0,342	0,944
163.	10575__NCU07067	0,175	0,108	0,171	1,638	0,788
164.	10580__NCU09864	1,087	0,014	1,282	0,587	-0,154
165.	10586__NCU03872	0,691	1,458	0,417	0,335	1,164
166.	10655__NCU00972	0,098	1,126	0,313	-0,463	-0,116
167.	10742__NCU05928	2,166	0,140	1,815	2,376	1,537
168.	10801__NCU07159	-0,223	-0,785	0,273	-0,783	-1,602
169.	10810__NCU08092	2,707	3,234	2,867	2,020	2,391
170.	10889__NCU00373	-0,130	-1,011	-0,257	0,006	-0,856
171.	10914__NCU07259	0,632	2,158	0,882	0,459	0,450
172.	10917__NCU06295	0,151	1,838	0,465	-0,604	0,330
173.	10959__NCU02430	1,998	-0,696	1,395	2,375	0,138
174.	10964__NCU05770	-1,533	-2,651	-1,366	-1,409	-1,497
175.	10981__NCU03423	-0,302	0,222	-0,258	0,733	1,163
176.	10988__NCU02703	-0,552	-1,213	-1,212	0,225	-0,365
177.	11010__NCU05132	-0,903	-2,086	-1,279	-0,473	-1,933
178.	11026__24g5_030	-0,253	0,674	-0,278	-0,913	-0,513
179.	11027__NCU01747	-0,850	-1,753	-1,164	-0,753	-1,739
180.	11053__9j10_200	-1,648	-1,409	-2,194	-1,094	-1,086
181.	11099__NCU04877	-0,692	-2,154	-1,048	-0,612	-1,434
182.	11113__NCU08230	-1,576	-1,611	-2,581	-0,910	-1,100
183.	11163__NCU07151	-1,282	-2,767	-1,200	-0,729	-0,258
184.	11166__NCU01334	0,991	1,768	0,902	0,547	0,740
185.	11172__NCU08084	0,557	0,054	0,159	1,451	0,600
186.	11173__NCU07471	0,374	-0,143	0,173	1,392	1,123
187.	11178__NCU01886	-0,731	-0,362	-0,706	0,373	1,084
188.	1118__NCU08380	2,143	2,655	2,411	0,286	1,565
189.	11186__NCU02627	-0,482	-0,259	-0,621	0,250	0,930
190.	11194__NCU03278	-1,055	-2,576	-1,517	-0,112	-1,498
191.	11196__1nc495_190	0,824	2,157	1,234	0,625	0,716
192.	11216__NCU08944	-0,492	0,529	-0,783	1,133	2,221
193.	11250__NCU00356	0,011	0,020	-1,064	-1,022	-0,739
194.	11279__NCU06361	0,299	1,477	0,149	0,331	0,660
195.	11315__NCU06057	-0,845	-0,493	-0,882	0,045	0,523
196.	11335__NCU03803	-1,006	-0,110	-0,821	0,602	1,224

197.	11352__NCU00422	1,125	1,648	1,286	0,616	0,096
198.	11375__NCU02474	-0,211	0,458	-0,151	-0,764	0,118
199.	11412__NCU04270	-0,010	-0,266	-0,408	0,815	-0,062
200.	11461__NCU04940	-0,410	-0,098	-0,587	0,598	0,302
201.	11475__NCU08223	1,460	2,179	1,456	0,715	1,058
202.	11483__NCU09452	0,586	1,435	0,494	-0,118	0,500
203.	1182__NCU06597	-0,060	0,604	-0,191	-0,966	-0,315
204.	1192__NCU07005	-1,274	-1,732	-1,588	-0,487	-1,693
205.	1313__NCU00455	-0,368	0,211	-0,380	0,584	1,159
206.	1398__NCU09020	-0,266	-1,427	-0,782	-0,481	-1,458
207.	141__NCU05499	-0,588	0,024	-0,671	0,305	0,814
208.	1456__NCU01942	0,607	-0,031	0,694	1,481	0,510
209.	1475__NCU08357	0,192	-0,841	-0,221	0,220	-0,812
210.	1494__NCU08641	0,312	1,421	0,274	-0,665	0,861
211.	1507__3nc195_080	-0,146	-0,826	-0,745	0,259	0,216
212.	1598__NCU02390	1,114	1,969	1,700	0,369	0,916
213.	1604__NCU05814	0,241	1,618	0,930	0,074	1,052
214.	1711__NCU03259	0,186	1,138	0,231	-1,525	-0,012
215.	1720__NCU05652	-0,075	0,677	-0,073	-0,686	-0,161
216.	180__NCU04287	-0,170	0,506	-0,897	-0,060	0,653
217.	1803__NCU07787	0,771	2,112	0,935	0,701	0,438
218.	1824__NCU02535	-0,734	-1,981	-1,326	-0,841	-1,400
219.	1847__NCU00506	0,073	0,675	0,488	-1,008	-0,285
220.	1948__NCU04090	-0,704	-0,501	-0,646	-0,784	0,465
221.	1956__NCU06110	-0,480	-0,064	-0,502	0,349	1,028
222.	2030__NCU03011	-1,011	-0,483	-1,058	0,389	1,134
223.	2038__NCU09620	-0,593	1,638	-0,337	-1,655	-0,288
224.	2050__NCU03360	-0,078	0,718	1,412	0,748	1,594
225.	2087__NCU07650	0,032	-0,142	-0,599	1,151	0,668
226.	2100__NCU04904	-0,082	0,696	-0,264	-0,852	0,195
227.	2120__NCU09281	0,915	-0,174	-0,152	-0,518	-0,779
228.	2244__NCU00175	0,382	0,985	0,465	-0,919	-0,166
229.	2335__NCU03821	-1,032	-1,546	-1,556	-0,548	-0,719
230.	2383__NCU03928	-0,775	-0,121	-0,787	0,129	0,711
231.	2410__NCU04268	-0,480	-0,353	-1,033	-0,394	1,131
232.	2420__NCU05197	0,599	1,286	0,163	0,064	0,298
233.	2527__NCU03732	2,264	3,768	2,886	1,394	2,779
234.	2528__NCU03721	-0,919	-0,329	-0,975	-0,434	0,189
235.	2538__NCU01856	-0,345	-0,506	-0,625	0,974	0,625
236.	2551__NCU03118	-0,475	-1,966	-1,150	0,047	-1,126
237.	2558__1nc542_010	-0,433	-1,467	-0,520	-0,784	-1,362
238.	2616__NCU09663	-0,368	0,836	-0,032	-0,847	0,007
239.	2647__5.8S_forw rd	0,340	1,889	0,530	0,200	0,871
240.	2660__NCU04805	0,300	0,830	0,280	-0,575	-0,018
241.	2744__NCU03905	-0,255	0,536	0,332	-0,159	1,679
242.	2754__NCU04062	0,741	1,104	0,650	-0,627	-0,398
243.	2918__NCU00940	-0,400	-1,228	-0,524	0,133	-0,581
244.	2955__NCU06207	0,543	-0,632	0,448	0,967	0,191
245.	2967__NCU00322	0,335	-0,323	0,364	0,473	-0,530
246.	2999__NCU08671	-0,382	-0,807	-0,697	0,068	0,318
247.	3039__NCU02400	0,286	0,876	-0,204	-0,283	0,566

248.	3085__NCU03783	-0,751	-2,067	-1,176	-0,783	-0,962
249.	3249__18S_forw rd	-0,555	0,357	-0,460	0,610	0,293
250.	3289__NCU02883	-0,734	0,043	-0,976	-1,004	-0,064
251.	3299__NCU04998	0,464	-0,132	0,164	-0,825	0,054
252.	3369__NCU08854	0,259	-0,198	0,380	1,429	0,514
253.	3390__NCU09353	-0,330	-0,330	-0,494	-0,082	-1,477
254.	3393__NCU10007	-0,262	0,022	-0,461	0,890	2,110
255.	3396__NCU06111	-0,645	0,562	-0,270	-0,362	0,061
256.	3469__NCU03003	2,104	1,551	1,695	3,046	2,611
257.	3482__NCU05418	0,865	1,370	0,884	2,067	1,821
258.	3485__NCU05448	-1,147	-2,255	-1,152	-1,050	-1,871
259.	3558__xnc105_090	-0,017	-0,762	-0,224	0,999	0,263
260.	3696__NCU07912	-1,049	-0,214	-0,878	-0,267	0,171
261.	3746__NCU05650	-0,457	-1,773	-0,457	0,393	1,065
262.	3760__NCU02380	0,440	0,072	0,277	1,849	0,278
263.	3812__NCU09210	-0,333	-1,224	-0,524	-1,512	-1,014
264.	3819__NCU00589	0,339	0,289	1,702	0,991	0,511
265.	3961__NCU06660	-0,856	-1,067	-0,395	0,664	0,232
266.	3965__NCU01321	-0,741	-1,444	-0,283	0,216	0,167
267.	3999__NCU08390	0,631	0,028	0,195	1,882	0,788
268.	4014__NCU02945	-0,527	-0,653	-0,516	-1,598	-1,289
269.	4018__NCU04234	-0,042	-0,275	-0,115	1,027	0,288
270.	4025__NCU09772	0,013	-0,205	-0,056	1,040	0,552
271.	4049__NCU00340	-1,399	-2,020	-1,590	0,039	-1,737
272.	4050__NCU00350	-0,102	-1,274	-0,622	0,819	-0,519
273.	4072__NCU07008	-0,128	-0,990	-0,253	0,454	-0,472
274.	4114__NCU06042	-0,690	0,846	-0,689	-0,264	1,091
275.	4149__NCU03443	0,922	0,173	1,063	1,381	0,789
276.	4154__NCU02257	0,384	1,714	0,924	-0,685	0,454
277.	4160__NCU01063	-1,142	0,012	-1,040	-2,485	-0,685
278.	4172__NCU01689	0,615	-0,445	0,286	0,949	-0,058
279.	4173__NCU09195	-1,076	-1,896	-1,014	-0,288	-1,576
280.	4238__NCU04558	-0,664	0,104	-0,799	-0,937	0,134
281.	4240__NCU05330	0,049	-0,734	-0,170	-0,958	-0,930
282.	4264__NCU05962	-0,404	0,494	-1,046	-0,272	0,465
283.	4435__NCU06323	1,425	2,553	1,651	2,271	2,735
284.	4458__NCU04791	-1,647	-2,573	-2,003	-0,971	-1,899
285.	4474__NCU06121	0,227	1,726	0,885	-0,127	0,516
286.	4500__NCU04284	1,226	0,006	0,481	0,294	-0,060
287.	4513__NCU01271	0,309	-0,996	-0,006	-0,575	-0,847
288.	4787__NCU04471	0,821	-0,022	0,178	0,565	-0,312
289.	4797__NCU06315	0,819	0,087	0,551	1,205	-0,025
290.	4939__NCU05179	-0,684	-0,056	-1,012	-0,313	0,111
291.	4947__NCU01428	0,183	0,936	0,284	-0,488	0,071
292.	4977__NCU10045	1,458	2,215	1,989	0,250	1,158
293.	4987__NCU06946	-0,458	0,261	-0,484	0,217	0,519
294.	499__NCU04850	-0,730	-0,107	-0,828	-0,052	0,411
295.	4990__NCU07547	-0,113	-0,853	-0,372	0,741	0,722
296.	4991__NCU07557	-1,113	-2,025	-1,577	-0,273	-1,818
297.	5048__NCU06651	0,603	0,539	0,311	-0,871	0,054
298.	5052__NCU08104	0,217	1,204	0,333	0,500	1,386

299.	5057__NCU01901	-1,038	-0,928	-0,628	0,794	0,914
300.	5078__NCU00947	-0,816	-2,207	-0,970	-0,065	-1,206
301.	5097__NCU04405	0,548	0,839	-0,378	2,169	1,698
302.	5140__NCU06608	0,877	0,366	1,006	0,384	-0,418
303.	5158__NCU06373	1,096	-0,042	0,889	1,715	1,059
304.	5168__NCU03134	0,117	1,720	0,805	-0,037	0,620
305.	5170__NCU02307	0,084	-0,756	-0,279	0,940	-0,033
306.	5208__NCU05874	-0,932	-0,234	-1,336	-1,195	-0,221
307.	5298__NCU03949	0,130	1,269	0,047	-0,108	0,113
308.	5305__NCU01137	1,240	-0,028	1,104	1,920	1,126
309.	532__NCU00766	-0,055	-1,124	-0,350	0,242	-0,345
310.	5357__NCU09040	-1,575	-2,592	-1,744	-1,089	-1,810
311.	5414__NCU07475	-1,139	-1,009	-1,113	-0,403	-0,124
312.	5434__NCU03273	0,295	1,686	0,729	0,025	0,388
313.	5454__NCU02946	0,113	1,206	0,788	0,331	0,291
314.	5463__NCU05052	-0,149	0,089	-0,172	-1,169	-0,181
315.	5474__NCU07593	0,530	0,118	0,463	1,502	0,838
316.	5507__NCU04542	-0,380	-0,236	-0,370	0,695	0,835
317.	5512__NCU07010	0,446	-0,290	0,487	0,938	0,462
318.	5576__NCU03807	0,346	-0,840	-0,059	0,001	-0,639
319.	5581__NCU08501	-0,029	-1,252	-0,749	-0,489	-1,102
320.	5637__NCU02046	0,985	2,021	0,815	0,335	1,061
321.	5638__NCU01612	-0,231	-1,796	-1,022	-0,811	-1,385
322.	5648__NCU07198	-0,222	0,387	0,262	-0,992	-0,268
323.	5680__NCU05333	1,154	-0,241	0,286	-0,014	-0,231
324.	5700__NCU04854	0,241	-0,820	-0,330	1,265	0,432
325.	5775__NCU09810	-0,248	-2,109	-1,122	0,048	-0,874
326.	5816__NCU04367	1,060	1,219	1,862	0,185	0,468
327.	5829__NCU06413	1,018	0,117	0,558	2,147	1,055
328.	5855__NCU07953	0,214	0,936	0,777	-0,368	0,067
329.	5874__NCU07307	0,781	1,338	1,261	0,045	0,620
330.	5941__11e5_200	0,536	0,010	0,569	-0,400	-0,284
331.	6047__NCU07840	0,556	-0,062	1,050	-0,616	-0,466
332.	6066__NCU06895	-0,734	-0,815	-0,937	-1,937	-1,271
333.	6124__NCU01114	0,342	0,151	0,485	-0,911	-0,335
334.	6188__NCU07869	0,242	-0,193	0,495	-0,654	-0,304
335.	6192__NCU06485	0,385	1,626	0,856	-0,143	0,475
336.	6227__NCU04472	0,422	0,761	0,929	-0,089	-0,106
337.	6259__NCU09791	0,704	1,361	1,147	-0,731	0,429
338.	6299__NCU09467	-1,055	-3,373	-1,693	-0,542	-2,457
339.	6328__NCU05137	0,251	-0,940	0,770	2,762	0,752
340.	6344__NCU07395	0,009	-0,844	-0,493	-0,644	0,638
341.	6420__NCU04907	0,623	-0,602	-0,210	0,201	-1,018
342.	6591__NCU07994	0,849	-0,984	0,030	0,597	0,153
343.	6635__NCU06079	0,252	-0,972	-0,019	-0,627	-0,864
344.	6692__NCU09212	-0,092	-0,723	-0,361	0,679	-0,054
345.	6699__NCU00586	0,839	-0,505	0,214	1,711	-0,510
346.	6700__NCU05119	-0,284	-1,369	-0,599	-0,455	-1,286
347.	6709__NCU08750	1,024	1,393	1,685	0,435	0,799
348.	6847__NCU03728	0,595	1,614	1,619	0,321	0,777
349.	6849__NCU03709	2,798	3,669	3,547	1,473	2,022

350.	6917__NCU06226	-0,746	-0,531	-0,896	0,161	0,625
351.	6923__NCU00147	0,353	-0,411	0,436	-0,549	-0,494
352.	6926__NCU00246	0,980	1,682	1,153	0,376	0,469
353.	6942__NCU06601	0,659	0,296	0,768	2,047	0,955
354.	6958__NCU06348	0,099	-0,877	-0,397	0,194	-0,749
355.	6998__NCU06155	0,716	2,459	1,795	0,538	1,375
356.	7028__NCU03437	-1,462	-2,184	-1,169	-2,134	-2,585
357.	7036__NCU01868	0,214	0,996	0,515	-0,474	0,292
358.	7132__NCU08847	0,043	-2,004	-0,095	0,077	-1,228
359.	7160__NCU09158	1,333	1,753	1,807	0,057	0,050
360.	7228__NCU03098	-0,744	-1,382	-0,674	0,217	-0,756
361.	7237__NCU00935	-0,484	-2,660	-0,924	0,206	-1,109
362.	7272__NCU06512	0,371	1,174	0,655	-1,112	-0,331
363.	7315__NCU11009	-1,244	-1,318	-1,309	-0,435	-0,437
364.	737__NCU01898	0,353	1,574	0,681	0,365	0,684
365.	7388__NCU08434	1,149	2,343	1,733	0,873	1,598
366.	7401__NCU05338	-0,197	-0,903	-0,662	0,403	-0,003
367.	7482__NCU05763	-0,155	0,690	0,511	-0,914	-0,132
368.	7497__NCU11090	0,172	-1,552	-0,236	-0,020	-1,601
369.	7520__NCU09306	-0,352	0,042	-0,749	0,300	1,225
370.	7591__NCU03222	-0,096	-1,368	-0,941	-0,073	-1,058
371.	7654__NCU07923	0,700	1,434	0,896	-0,166	0,588
372.	7673__NCU07273	2,624	0,522	0,194	0,474	0,046
373.	7723__NCU05784	-0,018	-0,238	0,264	-1,140	-1,095
374.	775__NCU08954	0,053	-2,317	-0,194	0,060	-1,432
375.	7770__NCU05156	1,672	2,433	2,742	0,954	0,641
376.	7780__NCU01845	0,654	0,139	0,758	0,144	-0,794
377.	78__NCU00028	0,240	-0,326	0,790	1,193	0,499
378.	7805__NCU05435	-0,285	0,240	-0,802	-0,058	0,561
379.	7807__NCU09230	0,125	-0,287	-0,542	0,701	0,149
380.	7831__NCU02995	-0,196	-0,052	-0,780	0,633	0,650
381.	8009__NCU00371	-0,907	0,070	-0,888	-0,710	0,256
382.	8024__NCU07680	-0,808	-2,157	-1,173	-0,116	-0,791
383.	8146__NCU09762	0,727	-0,892	0,108	1,232	0,033
384.	8168__NCU04727	-0,581	-2,129	-0,849	-1,169	-1,397
385.	8196__NCU03865	1,689	2,150	2,008	-1,034	0,050
386.	8227__NCU06926	0,554	-0,829	0,315	0,859	-0,596
387.	8236__NCU09034	-0,488	-2,208	-0,896	-1,191	-1,181
388.	826__NCU04510	1,217	0,461	1,594	0,939	-0,203
389.	828__NCU04526	1,009	0,089	0,894	1,372	0,346
390.	8288__NCU03716	-0,599	-2,185	-1,029	-0,676	-0,914
391.	8338__NCU04237	0,157	-0,460	-0,070	1,248	-0,005
392.	8345__NCU09775	1,185	0,135	0,728	1,067	0,354
393.	8386__4nc446_110	0,377	-0,619	0,049	1,032	0,829
394.	8389__NCU04334	0,334	1,582	0,442	-0,090	0,844
395.	8419__NCU04801	-0,408	-0,272	-0,149	-1,542	-1,373
396.	8441__NCU02366	0,028	0,850	-0,142	-0,605	-0,092
397.	8460__10h18_100	0,060	0,456	0,497	-0,613	-0,226
398.	8477__NCU01876	0,606	1,411	0,973	-0,623	-0,009
399.	8479__NCU01068	-0,491	-0,147	-0,520	1,191	-0,375
400.	8543__mito_170	-1,063	-0,453	-1,740	0,301	0,886

401.	8565__NCU06170	0,324	-0,363	0,233	1,003	-0,289
402.	8567__NCU07052	0,932	0,692	1,562	0,289	0,253
403.	8578__NCU04667	-0,503	-1,435	-0,455	-0,234	-1,004
404.	8590__xnc073_190	-0,667	-0,386	-1,316	-0,599	-0,096
405.	860__NCU04830	2,171	2,878	2,727	0,189	1,711
406.	8680__NCU00719	0,371	-0,393	0,385	-0,022	-0,620
407.	8694__NCU08927	1,258	2,558	1,422	0,479	1,474
408.	8706__NCU05721	-0,155	0,350	-0,300	-0,664	0,371
409.	8714__NCU06203	2,004	-1,328	1,251	1,839	0,789
410.	8748__NCU04502	0,231	1,323	0,502	-0,054	1,458
411.	8825__NCU01011	0,122	-1,092	-0,579	0,396	-0,432
412.	8844__NCU02962	1,125	1,346	0,738	0,274	0,331
413.	889__NCU08168	-0,285	0,151	-0,567	0,594	1,083
414.	9013__NCU09075	-0,034	-0,723	-0,245	0,698	-0,351
415.	9018__NCU01939	2,590	1,266	1,784	0,080	1,196
416.	9019__NCU02111	0,276	0,412	0,727	-0,194	-0,449
417.	9040__NCU00754	0,439	0,481	0,703	-0,512	-0,075
418.	9044__NCU00659	0,724	0,227	0,286	1,809	1,322
419.	9085__NCU00289	-0,980	-1,927	-1,329	-0,761	-1,623
420.	9099__NCU06555	-0,070	-1,108	-0,001	-0,183	-0,635
421.	916__NCU04180	0,319	-0,517	0,845	0,821	-0,251
422.	920__NCU06803	0,686	1,565	1,365	0,201	0,695
423.	9220__NCU01844	-0,766	-1,954	-1,293	-1,123	-1,780
424.	9231__NCU03535	-0,936	-0,836	-1,418	-0,456	-0,110
425.	9232__NCU03651	-0,262	-1,595	-0,764	0,278	-0,996
426.	9253__NCU10468	0,124	1,204	1,297	-0,096	0,573
427.	9304__NCU06013	0,304	-1,334	-1,090	0,696	-1,162
428.	9366__NCU01121	-1,355	-0,564	-1,620	-2,020	-0,306
429.	9397__NCU00943	-1,105	-0,168	-1,792	-1,196	-0,208
430.	9401__NCU00801	-0,089	-1,252	-0,280	-0,046	-0,409
431.	946__NCU01546	-0,690	-0,279	-1,248	-0,210	0,469
432.	947__NCU09760	-1,023	-0,069	0,229	-0,312	-0,090
433.	9474__NCU07258	0,839	-0,676	-0,017	1,136	0,020
434.	9526__NCU05860	-0,307	-2,034	-0,764	-0,095	-1,064
435.	9531__NCU08184	1,237	1,746	1,770	0,167	0,357
436.	9535__NCU04155	0,400	2,294	1,037	0,508	1,071
437.	9574__NCU10144	0,982	2,429	1,367	0,639	0,810
438.	9586__NCU09761	-0,178	0,700	-0,230	-0,070	0,640
439.	9636__NCU03871	-1,327	-2,257	-1,384	-0,906	-2,089
440.	9638__NCU04793	0,296	-0,871	-0,023	1,320	-1,095
441.	9643__NCU06219	0,277	-0,646	0,052	0,870	0,046
442.	9666__NCU06005	0,654	2,169	0,141	-0,256	0,940
443.	9681__NCU09270	0,503	1,541	0,381	-0,630	-0,566
444.	9682__NCU09506	1,731	1,013	2,181	1,066	0,667
445.	9690__NCU03408	0,207	-0,511	0,488	0,922	-0,575
446.	9732__NCU08083	0,734	1,716	0,812	-0,302	0,393
447.	9760__NCU00837	0,012	-1,031	-0,150	-0,537	-1,026
448.	9763__NCU00741	-0,529	1,327	-0,737	-0,240	1,131
449.	9774__NCU03038	0,102	-1,159	-0,593	-0,179	-1,005
450.	9797__NCU06228	0,990	-0,072	1,000	0,892	-0,306
451.	9806__NCU00248	0,962	1,111	-0,234	0,416	0,356

452.	989__NCU01711	-0,240	0,444	-0,505	0,673	1,652
453.	9894__NCU01219	0,380	0,431	0,654	-0,638	-0,328
454.	9920__NCU01060	0,535	1,126	0,674	-1,264	-0,218
455.	9930__NCU08505	0,488	-0,149	0,775	-0,228	-0,362
456.	9960__NCU01584	0,066	1,285	-0,105	-0,275	0,456
457.	9973__NCU07804	0,516	-0,263	0,573	0,526	0,923
458.	9978__NCU01045	1,067	-0,567	0,937	0,594	-0,251
459.	9979__NCU04431	0,779	2,725	1,087	-0,514	-0,036

Appendix 5. Gene clusters, identified by Hierarchical clustering

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
9275 NCU03945	9659 NCU04876	7111 NCU02462	2896 NCU01944	6531 NCU02877	1948 NCU04090
5453 NCU02938	4160 NCU01063	7075 NCU09420	511 NCU09577	10580 NCU09864	2744 NCU03905
8688 NCU02817	11375 NCU02474	1942 NCU05557	11192 NCU03163	8439 NCU11161	2410 NCU04268
1035 NCU08239	1494 NCU08641	11412 NCU04270	10151 NCU06505	8680 NCU00719	1486 NCU02773
9825 4nc400_050	2100 NCU04904	2317 NCU06167	11443 NCU05913	826_NCU04510	11461 NCU04940
6847 NCU03728	10188 4nc453_010	10361 NCU05739	10354 NCU04181	4383 NCU08394	3482 NCU05418
9253 NCU10468	3842 NCU04163	8479 NCU01068	11010 NCU05132	7844 NCU06642	3249 18S_forw rd
8679 NCU00847	8389 NCU04334	11079 NCU04846	6699 NCU00586	9760 NCU00837	8766 NCU05319
11352 NCU00422	11279 NCU06361	3961 NCU06660	9854 NCU09686	8199 NCU04839	11113 NCU08230
7515 NCU08997	3680 NCU00071	2249 NCU00367	4797 NCU06315	2967 NCU00322	11053 9j10_200
11064 NCU00323	10586 NCU03872	9523 NCU05764	10007 NCU07053	11450 NCU08263	10371 NCU08824
486 NCU05291	3039 NCU02400	4025 NCU09772	9638 NCU04793	9797 NCU06228	5097 NCU04405
2588 NCU05742	11260 NCU06588	10575 NCU07067	4607 NCU07838	2990 NCU06991	7831 NCU02995
880 NCU02378	9666 NCU06005	10034 xnc090_250	11457 NCU09905	8347 NCU05732	2791 NCU02446
6227 NCU04472	9960 NCU01584	5474 NCU07593	11027 NCU01747	2558 1nc542_010	3696 NCU07912
7838 NCU04978	10937 NCU07132	4018 NCU04234	1475 NCU08357	9978 NCU01045	4987 NCU06946
2044 NCU01391	3289 NCU02883	7449 8l21_010	2595 NCU07598	9099 NCU06555	4435 NCU06323
7160 NCU09158	4238 NCU04558	6328 NCU05137	8345 NCU09775	3521 NCU05474	10981 NCU03423
8419 NCU04801	5208 NCU05874	3369 NCU08854	6958 NCU06348	4513 NCU01271	11335 NCU03803
9531 NCU08184	372 NCU07431	6942 NCU06601	7591 NCU03222	6635 NCU06079	1956 NCU06110
5214 NCU00771	8706 NCU05721	1268 NCU08728	9304 NCU06013	8168 NCU04727	11186 NCU02627
9019 NCU02111	9366 NCU01121	3045 7nc605_490	6736 NCU03967	8236 NCU09034	3393 NCU10007
7770 NCU05156	8748 NCU04502	3542 NCU05924	1996 NCU01095	5931 NCU04603	11178 NCU01886
5816 NCU04367	6989 NCU10188	10641 14h13_010	10959 NCU02430	816_NCU07910	11315 NCU06057

5087 NCU00534	3396 NCU06111	11172 NCU08084	10889 NCU00373	7057 NCU09926	2030 NCU03011
10358 NCU05258	1188 NCU04547	3999 NCU08390	10006 NCU07025	5694 NCU09019	6917 NCU06226
2754 NCU04062	180 NCU04287	1627 NCU01350	8227 NCU06926	5140 NCU06608	10630 NCU01989
8161 NCU04168	4264 NCU05962	3558 xnc105_090	828 NCU04526	7780 NCU01845	2383 NCU03928
1358 NCU04536	9397 NCU00943	5332 NCU09110	9636 NCU03871	3906 NCU06001	4217 NCU01057
10041 NCU09168	5052 NCU08104	5170 NCU02307	10734 NCU09592	6923 NCU00147	141 NCU05499
6640 NCU02323	4114 NCU06042	5829 NCU06413	3485 NCU05448	2692 NCU04608	11216 NCU08944
4014 NCU02945	8009 NCU00371	6692 NCU09212	7132 NCU08847	5941_11e5_200	1313 NCU00455
7108 NCU01618	7837 NCU04895	5322 NCU05443	775 NCU08954	4240 NCU05330	10317 NCU01620
9427 NCU07834	9586 NCU09761	10044 NCU09460	8578 NCU04667	2607 NCU00251	10395 NCU08999
10890 NCU00381	9763 NCU00741	1456 NCU01942	148 NCU03183	10801 NCU07159	499 NCU04850
8196 NCU03865		7228 NCU03098	746 NCU10490	7028 NCU03437	7520 NCU09306
6124 NCU01114		4050 NCU00350	11217 NCU08953	9682 NCU09506	989 NCU01711
8553 NCU02667		4458 NCU04791	11099 NCU04877	9930 NCU08505	2528 NCU03721
9040 NCU00754		9013 NCU09075	9526 NCU05860	6047 NCU07840	8543_mito_170
9894 NCU01219		8297 NCU01852	10503 NCU08403	6188 NCU07869	889 NCU08168
3611 NCU08095		3760 NCU02380	1824 NCU02535	8567 NCU07052	9231 NCU03535
7723 NCU05784		9796 NCU06221	1967 NCU05887	7242 NCU00732	468 NCU04476
2275 NCU07334		4049 NCU00340	9232 NCU03651	2120 NCU09281	8590 xnc073_190
5048 NCU06651		8338 NCU04237	3322 NCU00185	7673 NCU07273	10550 NCU06970
6066 NCU06895		1192 NCU07005	9085 NCU00289	4500 NCU04284	946 NCU01546
1637 NCU01088		4991 NCU07557	4172 NCU01689	5680 NCU05333	4939 NCU05179
5341 NCU04961		10102 NCU02548	6299 NCU09467	3390 NCU09353	7805 NCU05435
1214 NCU09691		10457 1nc315_070	5775 NCU09810	10263 NCU09536	
5676 NCU03671		1507 3nc195_080	2551 NCU03118	3812 NCU09210	
1118 NCU08380		2335 NCU03821	8146 NCU09762	3299 NCU04998	
6259 NCU09791		2999 NCU08671	8825 NCU01011	9018 NCU01939	
860 NCU04830		11444 NCU06093	9474 NCU07258	3819 NCU00589	

5463 NCU05052		10988 NCU02703	3608 NCU06650		
9533 NCU00776		10883 3nc435_040	4949 NCU01925		
6709 NCU08750		3469 NCU03003	9690 NCU03408		
9314 NCU08260		9044 NCU00659	1168 NCU00330		
8460 10h18_100		11228 NCU07819	4173 NCU09195		
4133 NCU01235		7401 NCU05338	8565 NCU06170		
6893 NCU02939		5700 NCU04854	916 NCU04180		
1925 NCU05282		5715 xnc090_300	10966 NCU05861		
8304 1nc380_040		11173 NCU07471	4787 NCU04471		
2025 NCU07400		4990 NCU07547	6420 NCU04907		
6849 NCU03709		8386 4nc446_110	1398 NCU09020		
10810 NCU08092		2087 NCU07650	264 NCU00816		
7828 NCU01887		1132 NCU02926	4630 NCU07572		
7272 NCU06512		7807 NCU09230	7497 NCU11090		
8477 NCU01876		3354 NCU07279	6700 NCU05119		
2244 NCU00175		5414 NCU07475	2023 NCU02343		
9920 NCU01060		5507 NCU04542	5638 NCU01612		
9561 NCU06785		5057 NCU01901	5581 NCU08501		
1847 NCU00506		2538 NCU01856	9774 NCU03038		
4977 NCU10045		7315 NCU11009	5576 NCU03807		
10456 NCU01921c			9220 NCU01844		
5874 NCU07307			7156 NCU09004		
5648 NCU07198			78 NCU00028		
7482 NCU05763			4852 NCU00763		
1598 NCU02390			10443 NCU01176		
5855 NCU07953			5158 NCU06373		
920 NCU06803			5305 NCU01137		
6656 NCU04149			8024 NCU07680		

11312 NCU06702			5512 NCU07010		
4154 NCU02257			2955 NCU06207		
7036 NCU01868			8952 NCU08199		
9605 NCU03543			5357 NCU09040		
2802 NCU05373			532 NCU00766		
7619 2nc610_050			7237 NCU00935		
9192 NCU01972			4072 NCU07008		
7654 NCU07923			2918 NCU00940		
1711 NCU03259			5078 NCU00947		
1182 NCU06597			11194 NCU03278		
2660 NCU04805			9643 NCU06219		
10655 NCU00972			4149 NCU03443		
11026 24g5_030			10964 NCU05770		
5884 NCU05303			3085 NCU03783		
7054 NCU02486			8288 NCU03716		
4947 NCU01428			6591 NCU07994		
9732 NCU08083			10742 NCU05928		
2527 NCU03732			9401 NCU00801		
10311 NCU00465			8714 NCU06203		
10917 NCU06295					
2038 NCU09620					
8694 NCU08927					
2616 NCU09663					
11483 NCU09452					
1720 NCU05652					
8441 NCU02366					
5637 NCU02046					
3610 NCU08087					

11475 NCU08223					
11433 NCU02760					
9979 NCU04431					
3761 NCU02328					
6926 NCU00246					
10464 NCU02540					
9681 NCU09270					
7045 NCU05388					
11166 NCU01334					
5298 NCU03949					
2420 NCU05197					
10914 NCU07259					
1803 NCU07787					
6970 NCU02283					
5434 NCU03273					
11196 1nc495_190					
9574 NCU10144					
10870 NCU06447					
5454 NCU02946					
11199 NCU00990					
1604 NCU05814					
4474 NCU06121					
6192 NCU06485					
5168 NCU03134					
6998 NCU06155					
7388 NCU08434					
2647 5.8S_forw rd					
737 NCU01898					

9535 NCU04155					
4602 NCU05747					
8844 NCU02962					
11250 NCU00356					
9806 NCU00248					

Appendix 6. Gene clusters, identified by *k*-means clustering

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
2990 NCU06991	10890 NCU00381	3660 NCU08791	2317 NCU06167	2023 NCU02343	1486 NCU02773
6531 NCU02877	4133 NCU01235	10870 NCU06447	7156 NCU09004	1168 NCU00330	2791 NCU02446
7844 NCU06642	1637 NCU01088	11260 NCU06588	10443 NCU01176	148 NCU03183	10630 NCU01989
8439 NCU11161	6656 NCU04149	6989 NCU10188	11228 NCU07819	6736 NCU03967	3354 NCU07279
3521 NCU05474	486 NCU05291	11199 NCU00990	9523 NCU05764	8952 NCU08199	4217 NCU01057
8199 NCU04839	7515 NCU08997	9605 NCU03543	11444 NCU06093	4607 NCU07838	8766 NCU05319
3906 NCU06001	8679 NCU00847	1188 NCU04547	1268 NCU08728	10734 NCU09592	468 NCU04476
5931 NCU04603	9192 NCU01972	6970 NCU02283	4852 NCU00763	10006 NCU07025	10025 NCU05974
2607 NCU00251	9275 NCU03945	7837 NCU04895	7449 8l21 010	2595 NCU07598	10317 NCU01620
7242 NCU00732	6893 NCU02939	3842 NCU04163	1132 NCU02926	2896 NCU01944	10371 NCU08824
8347 NCU05732	880 NCU02378	10937 NCU07132	1627 NCU01350	9854 NCU09686	10395 NCU08999
11450 NCU08263	8688 NCU02817	3680 NCU00071	2249 NCU00367	511 NCU09577	10550 NCU06970
4630 NCU07572	1035 NCU08239	7619 2nc610 050	5322 NCU05443	10966 NCU05861	10981 NCU03423
264 NCU00816	5453 NCU02938	9659 NCU04876	5715 xnc090 300	3608 NCU06650	11053 9j10 200
5694 NCU09019	2275 NCU07334	372 NCU07431	7111 NCU02462	11217 NCU08953	11113 NCU08230
2692 NCU04608	3611 NCU08095	7045 NCU05388	11079 NCU04846	1967 NCU05887	11178 NCU01886
7057 NCU09926	5214 NCU00771	10125 NCU00695	5332 NCU09110	1996 NCU01095	11186 NCU02627
4383 NCU08394	9533 NCU00776	10188 4nc453 010	8297 NCU01852	4949 NCU01925	11216 NCU08944
10463 NCU02533	11064 NCU00323	10311 NCU00465	10034 xnc090 250	11192 NCU03163	11315 NCU06057
10478 NCU00881	7828 NCU01887	10586 NCU03872	10641 14h13 010	11443 NCU05913	11335 NCU03803
10580 NCU09864	11312 NCU06702	10655 NCU00972	10883 3nc435 040	11457 NCU09905	11461 NCU04940
10801 NCU07159	1925 NCU05282	10914 NCU07259	1942 NCU05557	3322 NCU00185	1313 NCU00455
1398 NCU09020	2802 NCU05373	10917 NCU06295	3045 7nc605 490	746 NCU10490	141 NCU05499
2120 NCU09281	816 NCU07910	11026 24g5 030	3542 NCU05924	10007 NCU07053	180 NCU04287

2558 1nc542 010	5087 NCU00534	11166 NCU01334	7075 NCU09420	10151 NCU06505	1948 NCU04090
2967 NCU00322	2044 NCU01391	11196 1nc495 190	9796 NCU06221	10354 NCU04181	1956 NCU06110
3390 NCU09353	7108 NCU01618	11250 NCU00356	10013 NCU08957	10503 NCU08403	2030 NCU03011
3812 NCU09210	1358 NCU04536	11279 NCU06361	10044 NCU09460	10742 NCU05928	2050 NCU03360
3819 NCU00589	4602 NCU05747	11375 NCU02474	10102 NCU02548	10889 NCU00373	2383 NCU03928
4240 NCU05330	2025 NCU07400	11483 NCU09452	10361 NCU05739	10959 NCU02430	2410 NCU04268
4500 NCU04284	3610 NCU08087	1182 NCU06597	10457 1nc315 070	11010 NCU05132	2528 NCU03721
4513 NCU01271	3761 NCU02328	1494 NCU08641	10575 NCU07067	11027 NCU01747	2744 NCU03905
4787 NCU04471	5884 NCU05303	1604 NCU05814	10964 NCU05770	11099 NCU04877	3249 18S forw rd
5140 NCU06608	6640 NCU02323	1720 NCU05652	10988 NCU02703	11194 NCU03278	3393 NCU10007
5576 NCU03807	8161 NCU04168	1803 NCU07787	11163 NCU07151	1475 NCU08357	3482 NCU05418
5581 NCU08501	9561 NCU06785	2038 NCU09620	11172 NCU08084	1824 NCU02535	3696 NCU07912
5638 NCU01612	2588 NCU05742	2100 NCU04904	11173 NCU07471	2551 NCU03118	4114 NCU06042
5680 NCU05333	7054 NCU02486	2420 NCU05197	11412 NCU04270	2955 NCU06207	4264 NCU05962
5941 11e5 200	5341 NCU04961	2527 NCU03732	1192 NCU07005	3485 NCU05448	4435 NCU06323
6047 NCU07840	5676 NCU03671	2616 NCU09663	1456 NCU01942	4072 NCU07008	4939 NCU05179
6188 NCU07869	7838 NCU04978	2647 5.8S forw rd	1507 3nc195 080	4172 NCU01689	4987 NCU06946
6420 NCU04907	8304 1nc380 040	3039 NCU02400	2087 NCU07650	4173 NCU09195	499 NCU04850
6635 NCU06079	11433 NCU02760	3289 NCU02883	2335 NCU03821	4797 NCU06315	5052 NCU08104
6700 NCU05119	9314 NCU08260	3396 NCU06111	2538 NCU01856	4991 NCU07557	5057 NCU01901
6923 NCU00147	9427 NCU07834	4160 NCU01063	2918 NCU00940	5078 NCU00947	5097 NCU04405
7028 NCU03437	1214 NCU09691	4238 NCU04558	2999 NCU08671	5775 NCU09810	5414 NCU07475
7673 NCU07273	8553 NCU02667	4474 NCU06121	3085 NCU03783	6299 NCU09467	5507 NCU04542
7780 NCU01845	9825 4nc400 050	4947 NCU01428	3369 NCU08854	6591 NCU07994	6344 NCU07395
8168 NCU04727	10041 NCU09168	5168 NCU03134	3469 NCU03003	6699 NCU00586	6917 NCU06226
8236 NCU09034	10263 NCU09536	5208 NCU05874	3558 xnc105 090	6958 NCU06348	7520 NCU09306
826 NCU04510	10358 NCU05258	5298 NCU03949	3746 NCU05650	7132 NCU08847	7805 NCU05435
8567 NCU07052	10456 NCU01921c	5434 NCU03273	3760 NCU02380	7237 NCU00935	7831 NCU02995

8680 NCU00719	10464 NCU02540	5454 NCU02946	3961 NCU06660	7497 NCU11090	8009 NCU00371
9099 NCU06555	10810 NCU08092	5637 NCU02046	3965 NCU01321	7591 NCU03222	8543 mito 170
9220 NCU01844	1118 NCU08380	6192 NCU06485	3999 NCU08390	775 NCU08954	8590 xnc073 190
9682 NCU09506	11352 NCU00422	6998 NCU06155	4018 NCU04234	8146 NCU09762	889 NCU08168
9760 NCU00837	11475 NCU08223	7036 NCU01868	4025 NCU09772	8227 NCU06926	9231 NCU03535
9797 NCU06228	1598 NCU02390	737 NCU01898	4049 NCU00340	828 NCU04526	9397 NCU00943
9930 NCU08505	1711 NCU03259	7388 NCU08434	4050 NCU00350	8288 NCU03716	946 NCU01546
9978 NCU01045	1847 NCU00506	8389 NCU04334	4149 NCU03443	8345 NCU09775	989 NCU01711
	2244 NCU00175	8441 NCU02366	4458 NCU04791	8565 NCU06170	
	2660 NCU04805	8694 NCU08927	4990 NCU07547	8578 NCU04667	
	2754 NCU04062	8706 NCU05721	5158 NCU06373	8714 NCU06203	
	3299 NCU04998	8748 NCU04502	5170 NCU02307	8825 NCU01011	
	4014 NCU02945	9366 NCU01121	5305 NCU01137	9085 NCU00289	
	4154 NCU02257	947 NCU09760	532 NCU00766	916 NCU04180	
	4977 NCU10045	9535 NCU04155	5357 NCU09040	9232 NCU03651	
	5048 NCU06651	9574 NCU10144	5474 NCU07593	9304 NCU06013	
	5463 NCU05052	9586 NCU09761	5512 NCU07010	9401 NCU00801	
	5648 NCU07198	9666 NCU06005	5700 NCU04854	9474 NCU07258	
	5816 NCU04367	9763 NCU00741	5829 NCU06413	9526 NCU05860	
	5855 NCU07953	9806 NCU00248	6328 NCU05137	9636 NCU03871	
	5874 NCU07307	9960 NCU01584	6692 NCU09212	9638 NCU04793	
	6066 NCU06895		6942 NCU06601	9690 NCU03408	
	6124 NCU01114		7228 NCU03098	9774 NCU03038	
	6227 NCU04472		7315 NCU11009		
	6259 NCU09791		7401 NCU05338		
	6709 NCU08750		78 NCU00028		
	6847 NCU03728		7807 NCU09230		
	6849 NCU03709		8024 NCU07680		

	6926 NCU00246		8338 NCU04237		
	7160 NCU09158		8386 4nc446 110		
	7272 NCU06512		8479 NCU01068		
	7482 NCU05763		9013 NCU09075		
	7654 NCU07923		9044 NCU00659		
	7723 NCU05784		9643 NCU06219		
	7770 NCU05156		9973 NCU07804		
	8196 NCU03865				
	8419 NCU04801				
	8460 10h18 100				
	8477 NCU01876				
	860 NCU04830				
	8844 NCU02962				
	9018 NCU01939				
	9019 NCU02111				
	9040 NCU00754				
	920 NCU06803				
	9253 NCU10468				
	9531 NCU08184				
	9681 NCU09270				
	9732 NCU08083				
	9894 NCU01219				
	9920 NCU01060				
	9979 NCU04431				

Appendix 7. Gene which represent overlapping of clusters, identified by hierarchical clustering and *k*-means clustering

	Hierarchical Clustering/ *k*-means clustering					
	C1/C2	**C2/C3**	**C6/C4**	**C4/C5**	**C6/C6**	**C5/C1**
1.	5855 NCU07953	9960 NCU01584	-	9854 NCU09686	10317 NCU01620	10263 NCU09536
2.	5816 NCU04367	9763 NCU00741		9774 NCU03038	10371 NCU08824	10580 NCU09864
3.	5676 NCU03671	9666 NCU06005		9690 NCU03408	10395 NCU08999	10801 NCU07159
4.	5648 NCU07198	9659 NCU04876		9638 NCU04793	10550 NCU06970	11450 NCU08263
5.	5463 NCU05052	9586 NCU09761		9636 NCU03871	10630 NCU01989	2120 NCU09281
6.	5453 NCU02938	9366 NCU01121		9526 NCU05860	10981 NCU03423	2558 1nc542_010
7.	5341 NCU04961	8748 NCU04502		9474 NCU07258	11053 b9j10_200	2607 NCU00251
8.	5214 NCU00771	8706 NCU05721		9401 NCU00801	11113 NCU08230	2692 NCU04608
9.	5087 NCU00534	8389 NCU04334		9304 NCU06013	11178 NCU01886	2967 NCU00322
10.	5048 NCU06651	7837 NCU04895		9232 NCU03651	11186 NCU02627	2990 NCU06991
11.	4977 NCU10045	6989 NCU10188		916 NCU04180	11216 NCU08944	3390 NCU09353
12.	486 NCU05291	5208 NCU05874		9085 NCU00289	11315 NCU06057	3521 NCU05474
13.	4602 NCU05747	4238 NCU04558		8952 NCU08199	11335 NCU03803	3812 NCU09210
14.	4154 NCU02257	4160 NCU01063		8825 NCU01011	11461 NCU04940	3819 NCU00589
15.	4133 NCU01235	3842 NCU04163		8714 NCU06203	1313 NCU00455	3906 NCU06001
16.	4014 NCU02945	372 NCU07431		8578 NCU04667	141 NCU05499	4240 NCU05330
17.	3761 NCU02328	3680 NCU00071		8565 NCU06170	1486 NCU02773	4383 NCU08394
18.	3611 NCU08095	3396 NCU06111		8345 NCU09775	1948 NCU04090	4500 NCU04284
19.	3610 NCU08087	3289 NCU02883		8288 NCU03716	1956 NCU06110	4513 NCU01271
20.	2802 NCU05373	3039 NCU02400		828 NCU04526	2030 NCU03011	5140 NCU06608
21.	2754 NCU04062	2100 NCU04904		8227 NCU06926	2383 NCU03928	5680 NCU05333
22.	2660 NCU04805	1494 NCU08641		8146 NCU09762	2410 NCU04268	5694 NCU09019
23.	2588 NCU05742	1188 NCU04547		775 NCU08954	2528 NCU03721	5931 NCU04603
24.	2275	11375		7591	2744	5941

	NCU07334	NCU02474		NCU03222	NCU03905	b11e5_200
25.	2244 NCU00175	11279 NCU06361		7497 NCU11090	2791 NCU02446	6047 NCU07840
26.	2044 NCU01391	11260 NCU06588		746 NCU10490	3249 18S_forward	6188 NCU07869
27.	2025 NCU07400	10937 NCU07132		7237 NCU00935	3393 NCU10007	6531 NCU02877
28.	1925 NCU05282	10586 NCU03872		7132 NCU08847	3482 NCU05418	6635 NCU06079
29.	1847 NCU00506	10188 4nc453_010		6958 NCU06348	3696 NCU07912	6923 NCU00147
30.	1711 NCU03259			6736 NCU03967	4217 NCU01057	7028 NCU03437
31.	1637 NCU01088			6699 NCU00586	4435 NCU06323	7057 NCU09926
32.	1598 NCU02390			6591 NCU07994	468 NCU04476	7242 NCU00732
33.	1358 NCU04536			6299 NCU09467	4939 NCU05179	7673 NCU07273
34.	1214 NCU09691			5775 NCU09810	4987 NCU06946	7780 NCU01845
35.	11475 NCU08223			511 NCU09577	499 NCU04850	7844 NCU06642
36.	11433 NCU02760			5078 NCU00947	5097 NCU04405	8168 NCU04727
37.	11352 NCU00422			4949 NCU01925	6917 NCU06226	8199 NCU04839
38.	11312 NCU06702			4797 NCU06315	7520 NCU09306	8236 NCU09034
39.	1118 NCU08380			4607 NCU07838	7805 NCU05435	826 NCU04510
40.	11064 NCU00323			4173 NCU09195	7831 NCU02995	8347 NCU05732
41.	10890 NCU00381			4172 NCU01689	8543 mito_170	8439 NCU11161
42.	10810 NCU08092			4072 NCU07008	8590 xnc073_190	8567 NCU07052
43.	10464 NCU02540			3608 NCU06650	8766 NCU05319	8680 NCU00719
44.	10456 NCU01921			3485 NCU05448	889 NCU08168	9099 NCU06555
45.	10358 NCU05258			3322 NCU00185	9231 NCU03535	9682 NCU09506
46.	1035 NCU08239			2955 NCU06207	946 NCU01546	9760 NCU00837
47.	10041 NCU09168			2896 NCU01944	946 NCU01546	9797 NCU06228
48.				2595 NCU07598	989 NCU01711	9930 NCU08505
49.				2551 NCU03118		9978 NCU01045
50.				2023 NCU02343		
51.				1996 NCU01095		
52.				1967 NCU05887		
53.				1824 NCU02535		

54.				148 NCU03183		
55.				1475 NCU08357		
56.				1168 NCU00330		
57.				11457 NCU09905		
58.				11443 NCU05913		
59.				11217 NCU08953		
60.				11194 NCU03278		
61.				11192 NCU03163		
62.				11099 NCU04877		
63.				11027 NCU01747		
64.				11010 NCU05132		
65.				10966 NCU05861		
66.				10959 NCU02430		
67.				10889 NCU00373		
68.				10742 NCU05928		
69.				10734 NCU09592		
70.				10503 NCU08403		
71.				10354 NCU04181		
72.				10151 NCU06505		
73.				10007 NCU07053		
74.				10006 NCU07025		