NTNU – Trondheim
Norwegian University of
Science and Technology

Characterization of the Cytochrome p450 Family in the unique Diatom *Seminavis robusta*

Elise Strøm Midthun

Biology
Submission date: December 2012
Supervisor: Atle M. Bones, IBI
Co-supervisor: Per Winge, IBI

Norwegian University of Science and Technology
Department of Biology

ACKNOWLEDGEMENTS

I would like to thank my supervisor Atle M. Bones for giving me the opportunity to do my master's thesis at the Cell- and Molecular Biology Group. I would also like to thank my scientific supervisor, Per Winge, for sharing from of his enormous reservoir of knowledge. Though many a time overwhelmed, I somehow managed to find an understanding in the end.

Further, I would like to thank the staff and other members at the CMB-group for help and guidance, especially Tore Brembu and Leila Alipanah. Also, I would especially like to thank Torfinn Sparstad for his excellent technical guidance and peculiar sense of humour.

To my fellow students and friends; thank you for fruitful discussions and silly jokes.

Finally, I would like to thank my dear family for love and encouragement, especially my husband, Bjørnar.

Trondheim, 5th of December, 2012

Elise Strøm Midthun

"The more you learn, the more you realize how little you know"

Socrates

SAMMENDRAG

Denne oppgaven beskriver superfamilien cytokrom p450 i den pennate diatomen *Seminavis robusta,* som har en bentisk livsstil og kan adhere til overflater. *S. robusta* er en mulig ny modellorganisme for de pennate diatomene, og nylig har genomet blitt sekvensert og blir nå analysert. Dette har vist at genomet til *S. robusta* inneholder 68 CYP-kodende gen, noe som er uvanlig mange sammenliknet med andre nærliggende organismer. Videre ble det også funnet at CYPene i *S. robusta* består av en rekke unike familier og underfamilier, men samtidig ble det også funnet familier som er kjent fra både planter og dyr. Dette reflekterer diversiteten av faktorer som kan ha bidratt til evolusjonen av disse genene i diatomer, og muligens hvordan en bentisk livsstil kan ha styrt denne prosessen.

Både bioinformatiske verktøy og transkripsjonelle analyser har blitt brukt til å karakterisere og utforske denne superfamilien. Et eksperimentelt studie ble utført hvor *S. robusta* ble dyrket ved forskjellige forhold, inkludert normal temperatur og konstant lys, lavere temperatur og en dag/natt syklus. Dette bekreftet at noen av CYPene er involvert i lys-/mørke responser, og i andre responser som ser ut til å være koblet til temperatur.

Som del av studiet ble det utført en sammenlikning mellom to programvarer som blir brukt for å analysere genekspresjonsdata. Dette konkluderte med at å erstatte programvaren som nå brukes med en nyere programvare med flere muligheter vil sørge for en høyere grad av kvalitetssikring når det publiseres data fra RT-qPCR. Videre vil dette sørge for at data i større grad blir publisert i samsvar med MIQE-retningslinjene.

ABSTRACT

This thesis describes the cytochrome p450 superfamily in the pennate diatom *Seminavis robusta*. This diatom has a benthic way of life and can adhere to surfaces, and is a new candidate model organism representing the pennate diatoms. The genome of *S. robusta* is currently being sequenced and is characterized. This revealed that the genome of *S. robusta* contains 68 genes encoding CYPs, an unusually high number compared to other related organisms. Several unique families and subfamilies were discovered, but also known families from both plants and animals. This reflects the diversity of factors that may have contributed to the evolution of these genes in diatoms, and possibly how a benthic lifestyle may drive this process.

Both bioinformatical tools and transcriptional analysis have been used to characterize and explore this superfamily. An experimental study was performed by cultivating *S. robusta* at different conditions, including normal temperature and light, lower temperature and a day/night cycle. This confirmed some of the CYPs' involvement in light-/dark-responses, as well as some responses that seemed coupled to temperature.

As part of the study, a comparison was made between two software programs used for analyzing gene expression data. This concluded that replacing the software currently used with a new software program with more possibilities will be beneficial for ensuring a higher quality assurance when publishing RT-qPCR data, as well as one step closer publishing according to the MIQE-guidelines.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

Bp	Base pairs
BLAST	Basic Local Alignment Search Tool
cDNA	Complementary DNA
Cp	Crossing point
Cq	Quantification cycle
Ct	Threshold cycle
cTP	Chloroplast transit peptide
CYP	Cytochrome p450
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
dsDNA	Double stranded DNA
ER	Endoplasmic reticulum
E-value	Expect-value
FAD	Flavin adenine dinucleotide
FA-gel	Formaldehyde agarose gel
FMN	Flavin mononucleotide
gDNA	Genomic DNA
GFP	Green Fluorescent Protein
Mb	Mega basepairs / million basepairs
MIQE	Minimum Information for Publication of Quantitative Real-Time PCR Experiments
mRNA	Messenger RNA
mTP	Mitochondrial transit peptide
NAD	Nicotinamide adenine dinucleotide
NADH	Reduced form of NAD
NADP	Nicotinamide adenine dinucleotide phosphate
NADPH	Reduced form of NADP
NJ	Neighbor-joining
NN-score	Neural network-score
NTC	No-template control
PCR	Polymerase chain reaction
qPCR	Quantitative real-time PCR
q.s.	Quantity sufficient
RC	Reliability class
RNA	Ribonucleic acid
RNase	Ribonuclease
rRNA	Ribosomal RNA
RT	Reverse transcription
-RT	No-reverse transcription control
RT-qPCR	Reverse transcription qPCR
SP	Signal peptide
SRA	Sequence Read Archive
ssRNA	Single stranded RNA
TAE	Tris-acetate EDTA

T_A	Ambient temperature
T_m	Melting temperature

1. INTRODUCTION

1.1. Diatoms

The diatoms (Bacillariophyceae) form a group of photosynthetic unicellular organisms found in aquatic environments globally. Hundreds of genre and as much as 200 000 species are known. They are part of the eukaryotic division Heterokontophyta (Stramenopiles), are extremely diverse and of great ecological importance (Chepurnov *et al.*, 2008). At least 25 % of the inorganic carbon fixed in the oceans each year is due to diatoms, and they are thus a dominating part of the primary production of many marine and freshwater ecosystems (Granum *et al.*, 2005). The Stramenopiles is a eukaryotic linage well separated from fungi and animals, but also from other eukaryotes performing photosynthesis and can provide great insight in the evolution of cellular processes in eukaryotes.

A key feature of the diatoms is their ornate silica shells, the frustules, and the two distinct shapes seen in Figure 1.1.1 below are known; the centric and the pennate, both found in benthic as well as pelagic habitats. The centric diatoms have radially patterned frustules, and form a paraphyletic group (Gillard *et al.*, 2008). The pennate diatoms are a monophyletic group which can be recognized by their elongated, feather-like structure of the frustules. These special silica shells are widely used as environmental and stratigraphical indicators, and both biologists and chemists are intrigued by them, as they show great potential to be utilized in technological and industrial applications. As the diatoms contributes considerably to the petroleum reservoirs (Armbrust, 2009), these organisms are of great interest in several areas of science.

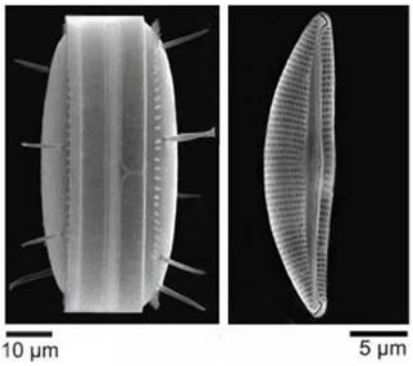


Figure 1.1.1 Representatives of the centric and the pennate diatoms. The centric diatom *Thalassiosira punctigera* to the left and the pennate diatom *Seminavis robusta* to the right (Chepurnov *et al.*, 2008).

1.1.1. Origin

Based on ribosomal genes, often used as molecular clocks, the diatoms are estimated to first have appeared as early as 250 million years ago, but this is not verified from fossils, as the silica shells of the diatoms are easily dissolved (Falkowski *et al.*, 2004, Armbrust, 2009). The phytoplankton community was dominated by cyanobacteria and small green algae before the appearance of the diatoms, together with the dinoflagellates and the coccolithophores. In the modern oceans, the cyanobacteria outnumber other phytoplankton by far, but the most diverse group is the diatoms.

The diatoms have long been thought to have arisen through a secondary endosymbiotic process between a red algae and hetertrophic eukaryote as described in Figure 1.1.2 below (Armbrust, 2009). The primary endosymbiosis is thought to have happened when a eukaryote heterotroph engulfed a cyanobacterium, which then formed the plastids of the Plantae. Plantae includes all the land plants, and the red and green algae. The later, secondary endosymbiosis assumedly took place when a different eukaryote heterotroph engulfed a red algae. This red alga was subsequently transformed to the plastids of the Stramenopiles, which include the brown macroalgae and plant parasites, in addition to diatoms.

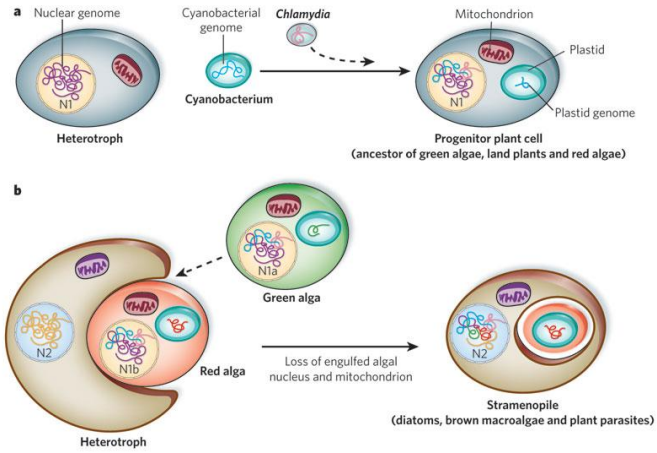


Figure 1.1.2 Endosymbiosis in diatoms. The diatom plastids originated through sequential primary (a) and secondary (b) endosymbiosis. A large amount of the engulfed cyanobacterial genome was transferred to the host nucleus (NI) during primary endosymbiosis **(a)**, and just a few of the original genes retained in the plastid. A potential chlamydial parasite invading the host is indicated, with the transfer of pink chlamydial genes to the nucleus. This progenitor plant cell further diverged into green and red algae, as well as land plants. During secondary endosymbiosis **(b)**, a different heterotroph engulfed a eukaryotic red algae. A possible engulfment of a green algae is also indicated. The algal mitochondrions are lost, as well at the nuclei, and algal nuclear and plastid genes are transferred to the host. Further gain and loss of genes from bacterial genomes has happened during diatom evolution, but is not indicated here (Armbrust, 2009)

This theory is subject to controversy now, though, as more organisms have their genome sequenced. Especially one study (Moustafa *et al.*, 2009) suggests that as much as 16 % of the diatom genes can have green algal origin. Further, this study used complete genome data and found that 85 % of the "green genes" in diatoms can be traced back to a common ancestor of both diatoms and other Stramenopiles. They concluded with a prasinophyte-like endosymbiont in the common ancestor of the putative supergroup Chromalveolata, comprising stramenopiles, cryptophytes, haptophytes and Alveolata (ciliates, apicomplexans and dinoflagellates). A more recent study by Deschamps and Moreira (2012) does not support the theory of a green algal endosymbiont, though, and claims that the results of Moustafa *et al.* are due to misinterpretations of data and a lack of sufficient genome data from red algae. The possibility of a green algal endosymbiont is not completely excluded, but the contribution is probably significantly lower than proposed by Moustafa *et al.* As more organisms' genomes are completely sequenced and analyzed, this will shed more light on the origin of the diatoms, and the theories of a red or green endosymbiont will probably be revised again several times.

About 100 million years ago, the diatoms started diverging, with the bipolar and multipolar centrics to first appear. Later, around 65 million years ago, the araphid pennates appeared, and finally, around 30 million years ago, the raphid pennates. The last group has a raphe in their silica walls, a slit that enables them to move along surfaces. This feature expanded the amount of ecological niches available for diatoms, and was as important for diatoms as the evolution of flight was for birds.

1.1.2. Life cycle

An interesting feature of the diatoms is their life cycle, switching between a sexual and a vegetative stage. This is triggered by size-specific threshold, as the vegetative stage leads to a gradual reduction of cell-size (Chepurnov *et al.*, 2008). When this threshold is reached, the reproduction is switched from vegetative to sexual, and after mating the population will again regain its original size.

The life cycle is dominated by the vegetative stage, which will last months, and even years. In contrast, the sexual stage will only last hours to days. While in vegetative stage, the diatoms

divide mitotically and experience a gradual reduction in cell-size, as seen in Figure 1.1.3 below. The frustule is what determines the cells shape and size, and is made up of two halves, thecae, which are overlapping, much like a Petri dish. In mitotic division, one daughter cell inherits the top of the theca, the epitheca, and the other daughter cell inherits the bottom, the hypotheca. Both daughter cells use the theca inherited from the parent as epitheca, and manufacture a new hypotheca. Consequently, the daughter cell which inherited the hypotheca will be smaller in size, whereas the other will be the same size as the parent. This will lead to a decreasing cellsize in the population as a whole. As cells reproduce further below the cardinal point, the extent of sexual reproduction increases. To restore size of the population, an auxospore is formed through sexual reproduction. The thecae are shedded during gametogenesis, and the zygote matures in to this specialized auxospore which now has the ability to expand. The auxospore reach maximum size and a new, initial cell is formed inside the auxospore's envelope. This initial cell is about three times bigger than the parental cells, and will now start a new round of vegetative multiplication.

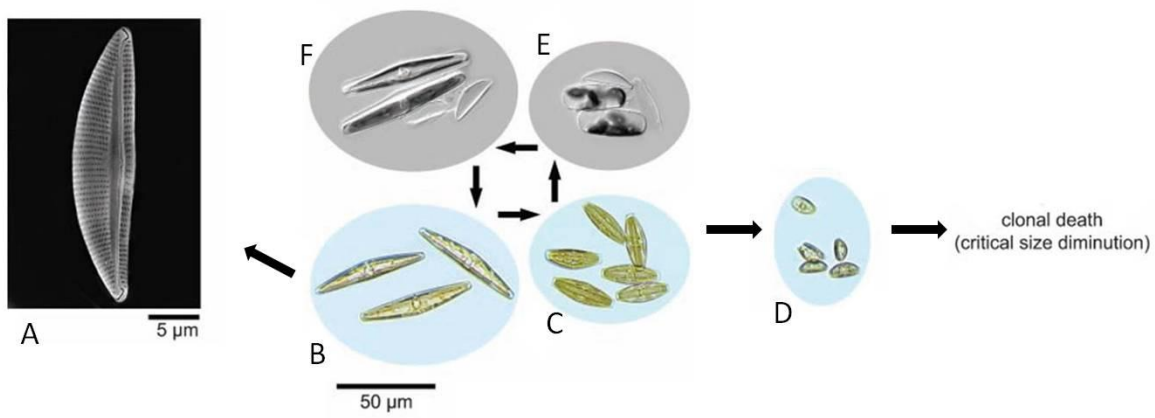


Figure 1.1.3 Principal stages in the lifecycle of the pennate diatom *Seminavis robusta*. A: Scanning electron micrograph of the frustules. **B – D:** Gradual diminution of the cell size over time in clonal cultures. **E:** Auxospores expanding bipolarly. **F:** The contents of the developed auxospores transform into enlarged, initial cells, still enclosed in the auxospore cell wall. These initial cells start a new round of vegetative multiplication (Modified after Chepurnov *et al.*, 2008).

In diatoms, these two processes are linked in a unique way. Sexual reproduction and size restitution are coupled in such a way that only cells within a particular size range have the ability to reproduce sexually. If this sexual reproduction for some reason is not switched on, the population will continue with vegetative reproduction, and thus reduction of cell-size, until the entire population has a critical minimum size, where the cells will die. Such specific

size thresholds are referred to as cardinal points, and include the critical minimum size described here and the critical maximum size for sexual reproduction as well as the maximum size of initial cells. These cardinal points are quite strict and species-specific, and control the entire lifecycle of the diatoms.

This particular trait of the diatoms is what makes it difficult to maintain them in cultures in the laboratory. Strains that are not able to reproduce sexually under the conditions given in the laboratory will die within a relatively short time. If sexual reproduction is able to occur intraclonally, though, this will maintain the population, but the genetic structure of the culture will be changed through recombination. A minority of the diatoms, though, are able to avoid this problem, by avoiding getting smaller. How this is achieved is still some unclear, but must involve some kind of elasticity in the region where the epitheca and hypotheca overlap. These strains are of course much easier to keep in laboratory cultures, and the preferred model organisms so far are amongst these, including *Thalassiosira pseudonana* and *Phaedactylum tricornutum*. Several attempts have been made to find or induce sex and auxosporulation in both, but so far, this has not been successful. *Thalassiosira weissflogii*, on the other hand, has a well-documented size-reduction restitution life cycle, and the production of sperm is readily controlled under experimental condition. Yet, production of eggs, fertilization and auxosporulation has still not been documented, and knowledge of this part of the cycle is sparse.

1.1.3. The diatom genome

T. pseudonana and *P. tricornutum* were the first diatoms to have their genomes completely sequenced (Chepurnov *et al.*, 2008). These are representatives of the centric and pennate diatoms, respectively, and their genomes have revealed the diatom's unique combination of animal- and plantlike characteristics. Further, it has been shown that these genomes contain several genes not known from any other organisms.

Armbrust *et al.* (2004) describes the genome of *T. pseudonana* to be diploid and about 34.5 million basepairs (Mb), with 24 chromosomepairs, containing novel genes for transport of silicic acid and formation of the silica-based cell walls. Also described are novel biosynthetic enzymes for several types of polyunsaturated fatty acids. Further, the authors states that

they were unable to identify the functions of about half of the genes, as it is difficult to deduce through comparison with other organisms, given that diatoms shows so many distinctive features.

P. tricornutum was the other diatom to be fully sequenced. As shown by Bowler *et al.* (2008), *P. tricornutum* has a genome of approximately 27.4 Mb, which is somewhat smaller than *T. pseudonana*. Bowler describes that the two diatoms share 57 % of their genes, but recent exploration of the genomes suggest that this number might be as high as 75 %. It was found that *P. tricornutum* has a higher number of species-specific genes, compared to *T. pseudonana*. This is reflected in the fact that the centric diatoms share features found in other eukaryotes, such as the flaggelar apparaturs (Merchant *et al.*, 2007), in contrast to the pennate. This might be due to the fact that the pennate diatoms often live in a benthic environment, and thus is exposed to a large number of potentially toxic or even lethal organisms to a much higher degree. Further, *P. tricornutum* has a high number of genes that have bacterial origin. Less than half of these are shared with *T. pseudonana*, and even fewer shared with more distant related organisms. The basic features of the diatoms came as a result of endosymbiotic events, but individual adaption amongst species is a result of subsequent gain (or loss) of genes, mainly from bacteria (Armbrust, 2009). The bacterial genes does not come from one specific source, but seem to have been acquired from a range of sources, including archaea and cyanobacteria. As more bacterial genomes are known, this will be further enlightened. This horizontal gene transfer from bacteria to diatoms is higher than what have been found in any other sequenced eukaryote. The diatoms live in an environment rich in bacteria, and the bacterial influence is therefore omnipresent. Several examples of symbiosis between bacteria and diatoms exist, where diatoms can be dependent on bacterial metabolites and vice versa (Croft *et al.*, 2005, Riemann *et al.*, 2000).

1.2. *Seminavis*

The diatom genus *Seminavis* is relatively recently established (Round *et al.*, 1990,p. 572 - 573), but species belonging to this genus was describes as early as 1857 by William Gregory (Gregory, 1857). Several species were described from sublittoral sediments from the Firth of

Clyde, off the west coast of Scotland, and were then assumed to be organisms belonging to the genus *Amphora*.

More recent reviews of these findings, including studies of Gregory's material, have revealed that some of these organisms were in fact discrete species under the *Seminavis* genus. Danielidis and Mann (2002) found that two of the species described as *Amphora* were known *Seminavis* species, whereas two species previously described as *A. macilenta* and *A. ventricosa* were new, and were named *S. macilenta* and *S. robusta*, respectively.

These misclassifications by Gregory were, naturally, the result of poor microscopic and analytical tools, but can also be related to the previously accepted broad species concept. Further misunderstandings have been added later, by other scientists following Gregory's misclassifications. As a result, these species have been misclassified as species belonging to the *Amphora* genus for about 150 years.

1.2.1. Seminavis robusta

Seminavis robusta is a pennate diatom of the family Naviculaceae, and is a benthic species which can adhere to substrates (Gillard *et al.*, 2008). It is quite a large diatom, and can reach 100 µm in size, which is 20 – 25 times larger than the current model diatoms. This makes it well suited for studies, especially as it is possible to control sexual mating in experimental conditions. Further, its size makes it well suited for microscopic studies. This diatom also shows the ability to move, and could thus show primitive responses, such as chemotaxis and flight-responses, which would require a complex signaling system.

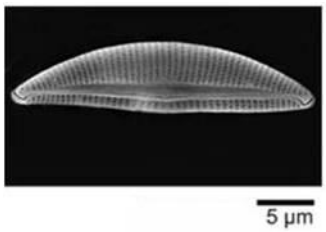


Figure 1.2.1 *Seminavis robusta*. *S. robusta* can reach 100 µm in size and is thus 20 – 25 times larger than the current model diatoms (Chepurnov *et al.*, 2008).

As many diatoms, *S. robusta* has a life cycle involving size reduction and both a vegetative and a sexual stage, as described in section 1.1.2 above. The vegetative stage is prolonged,

and involves mitotic multiplication, which gradually reduces the size of the cells as seen in Figure 1.1.3 above. To regain the original size of the cells, *S. robusta* undergo sexual auxosporulation, which is initiated after the cell size has reached a critical size of 50 µm (Chepurnov *et al.*, 2002). Two mating types are known, and it is thus possible to mate two individuals.

S. robusta is considered as a good candidate for a new model organism representing the pennate diatoms. This is both due to the size, and the proven sexual reproduction. The total genome of *S. robusta* is therefore sequenced, and is currently being analyzed.

1.3. Cytochrome P450

Cytochrome P450 (CYP) is a very large superfamily of enzyme proteins named for their absorption at 450 nm peak in the carbonmonoxide-bound form (Werck-Reichhart and Feyereisen, 2000). They catalyze activation of molecular oxygen by the use of electrons from the reduced forms of nicotinamide adenine dinucleotide (NAD) and nicotinamide adenine dinucleotide phosphate (NADP); NAD(P)H, which leads to regiospecific and stereospecific oxidative attack on a wide range of substrates. The CYPs are found in all types of tissue in almost all organisms. It was previously assumed that CYPs were present in every single living organism on Earth, but as the genome of more species have been sequenced, this assumption has been corrected, as some bacterial species do not appear to have any CYP-encoding genes (Meunier *et al.*, 2004, Werck-Reichhart and Feyereisen, 2000).

This thesis describes the CYP-family of the pennate diatom *S. robusta* which has expanded significantly compared to CYP-families found other eukaryotic algae. With 68 CYP-coding genes, *S. robusta* has almost ten times as many CYPs as the pennate diatom *P. tricornutum*, which was found to have 8 CYP-coding genes by exploration of the genome sequence.

1.3.1. Function

CYPs are generally monooxygenases, and catalyze the insertion of one atom from molecular oxygen into a substrate while the second oxygen gets reduced to water, as shown in Figure 1.3.1 (Anzenbacher and Anzenbacherová, 2001). The central heme iron atom, common for all cytochrome P450s, will bind two oxygen atoms from the substrate molecule peroxide.

CYPs have the ability to bind and activate two atoms of oxygen, usually the O_2-molecule, though exceptions occur. Even though some differences, all CYPs can utilize one oxygen atom from bound hydrogen peroxide or peroxides for monooxygenations. Figure 1.3.1 show a simple, schematic overview of most CYP-catalyzed reactions. Other results of CYP catalysis include cleavage of carbon-carbon bonds, isomerization, dimerization, dehydration, dealkylation, dehydrogenation and also reduction.

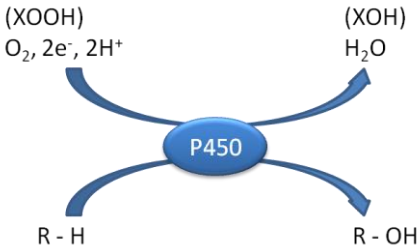


Figure 1.3.1 Schematic overview of P450 reactions. R-H, substrate; R-OH, hydroxylated product; XOOH, peroxide (X = H or organic residue); XOH, hydroxylated by-product (Anzenbacher and Anzenbacherová, 2001).

Another common feature amongst the CYPs is the heme-thiolate bond, where the central heme-molecule is bound to the protein via an anionic, thiolate sulfur of a cystein residue.

Human CYPs are involved in a wide range of processes, including carbon source assimilation, carcinogenesis and degradation of xenobiotics. They are also involved in biosynthesis of compounds such as steroids, prostaglandins, tromboxanes, fatty acid derivates and derivates of retinoic acid. As CYPs are involved in drug metabolism in humans, they are of great interest to the pharmaceutical industry where compounds used in drug development are screened as possible substrates of CYPs. A recent study (Schneider and Clark, 2013) describes the development of biosensors with immobilized CYPs that can detect CYP-activity in potential drug compounds. A future possible use of this might be a so-called patient-on-a-chip, where a CYP biosensor can accurately replicate the human response to a given drug.

In plants, CYPs are thought to have three main functions; synthesis of pigments, growth regulators and plant toxins. As plants are sedentary, they do not have the same ability to save their life by escaping as animals, and plants therefore synthesize toxins to help in their defense. This great diversity amongst the CYPs is thought to have arisen due to gene duplications during the evolution, and also probably lateral gene transfer, conversions, gene loss, gene amplification and genome duplications. The CYP superfamily has its origin in

prokaryotes, but, as mentioned, has now been found in a great majority of organisms so far characterized, from baker's yeast to humans.

1.3.2. Classification

Eukaryotic CYPs are divided in four classes, depending on localization and function, as well as how electrons are transferred from NAD(P)H to the catalytic site (Werck-Reichhart and Feyereisen, 2000). Enzymes classified in class I are associated with the mitochondrial inner membrane and are found to be involved as a catalyst in biosynthesis of steroid hormones and vitamin D_3 in mammals. Class I CYPs have been described in both insects and nematodes, as well as in mammals, but so far not in plants. Class I CYPs require both a flavin adenine dinucleotide (FAD)-containing reductase and an iron sulfur redoxin to deliver electrons.

Most common in the eukaryotes is the class II CYPs, which are found anchored to the outer face of the endoplasmic reticulum (ER). The CYPs belonging to class II have extremely diverse functions, and have been described in fungi and animals, as well as plants. In plants they have been described to be involved in catabolism or biosynthesis of all kinds of hormones, in oxygenation of fatty acids, and in all the pathways in the secondary metabolism. Further, they are involved in lignifications, synthesis of pigments and defense chemicals. In mammals, class II CYPs are involved biosynthesis and catabolism of several components, including steroid hormones, signaling molecules, retinoic acid and oxylipins. In fungi, these CYPs are responsible for synthesis of membrane sterols and mycotoxins, metabolism of lipid carbon sources and also detoxification of phytoalexins.

CYPs of both classes I and II in all organisms are part of the detoxification and, in some cases activation, of xenobiotics. Further, they have been found to be involved in carcinogenesis, and are also essential in metabolism and tolerance of drugs and pesticides. It has been shown that CYPs that metabolize xenobiotics actively are induced by endogenous chemicals. Class II CYPs require a FAD/flavin mononucleotide (FMN)-containing CYP reductase for electron transfer.

The CYPs in class III have been shown to be involved in synthesis of signaling molecules, such as jasmonate in plants and prostaglandins in mammals. These CYPs are so-called self-

sufficient; they do not require molecular oxygen, nor an external electron source. In plants they have diverse subcellular localization, including plastids (Werck-Reichhart and Feyereisen, 2000).

A last class of CYPs, class IV have been described to be found solely in fungi, and reduce NO to N_2O. These CYPs receive electrons directly from NAD(P)H. Both class III and IV are likely to be remains of the most ancestral forms of CYPs involved in detoxification of the damaging activated oxygen species.

1.3.3. Structure

Even though the functions of the CYPs are diverse, they share common features, like an overall similar structure and shape. This is due to the fact that most of the secondary-structure elements are conserved, as well as the overall folding pattern

Even though the amino acid sequence similarity between different CYPs can be extremely low, the structure is still highly conserved, especially in the core around the prosthetic heme group all cytochromes possess. This confirms the common mechanism of electron and proton transfer and oxygen activation shared by the CYPs. The core consists of a four-helix (D, E, I and L) bundle, helices J and K, two sets of β-sheets and a coil (Werck-Reichhart and Feyereisen, 2000). The most characteristic CYP consensus sequence (Phe-X-X-Gly-X-Arg-X-Cys-X-Gly), with cystein as highly conserved, is located in the heme-binding loop. Helix K also contains a highly conserved motif (Glu-X-X-Arg), and at last, helix I, which contains a consensus sequence (Ala/Gly-Gly-X-Asp/Glu-Thr-Thr/Ser), known as the CYP signature. More variable regions of the CYPs are associated with the N-terminal part, which is involved in amino-terminal anchoring, targeting of membrane-bound proteins, or parts involved in substrate binding and recognition. Further, variations related to differences in electron donors, reaction catalyzed and membrane localization have been observed.

1.3.4. Nomenclature

The CYPs are given a name that reflects family and subfamily, as well as the gene. The root *CYP* is followed by a number describing family, a letter for subfamily and finally a number for the gene, for example *CYP1A1*. Generally, CYPs of the same family should have an amino

acid similarity of at least 40 %, and a subfamily is formed by CYPs sharing at least 55 % amino acid similarity (Werck-Reichhart and Feyereisen, 2000). Further designations are also given to groups of *CYP*-families diverged from a single common ancestors. These are called clans, and are named from the lowest family number in the group. To ensure that this system is followed, The Human Cytochrome P450 Allele Nomenclature Committee has drawn up a series of guidelines, and all CYPs should be classified according to this. In plants, this categorization in to families and subfamilies does not, however, necessarily reflect enzymatic function. In some families, it has been shown that all members catalyze the same enzymatic reactions. Other families, though, may catalyze a variety of reactions, and the only thing in common is the amino acid similarity (de Montellano, 2005,p. 555). Animal CYPs are categorized by taking function in to account as well.

1.3.5. CYP diversity

The number of *CYP* genes present varies great among different organism, but multicellular organisms seem to have a somewhat similar amount, with about 80 genes in the nematode *Caenorhabditis elegans* and about 55 genes in humans (Werck-Reichhart and Feyereisen, 2000). Higher plants, though, have a much greater number of *CYP* genes, as *Arabidopsis thaliana* with over 250 genes (Nelson, 2009), and the CYPs make up the largest superfamily in plants (de Montellano, 2005,p. 554).

Although CYPs have been thoroughly described in several organisms, ranging from bacteria to humans and higher plants, very few CYPs have so far been characterized in eukaryotic algae. Yang *et al. (*2003) described a novel CYP from the marine diatom *Skeletonema costatum*, yet the function was not completely elucidated. The CYP from *S. costatum* was designated CYP97E1, due to a predicted relatedness to the CYP97B isoforms in plants. Later this CYP has also been found in *T. pseudonana*. CYP97B is most likely involved in metabolizing carotenoids, which is a function that probably evolved in algae before the land plants arose. The CYP97-family is found in the green algae Chlamydomonas, as well as in the land plants Arabidopsis, pine, rice and poplar (Nelson, 2006).

Further characterization of CYPs in eukaryotic algae is sparse, and this thesis is the very first detailed description of the CYP-family in a Stramenopile.

1.4. Gene expression analysis

1.4.1. RT-qPCR

Gene expression analysis was performed with real-time quantitative polymerase chain reaction (qPCR). This method measure the amount of synthesized DNA in real time during the run, in contrast to regular, end-point measuring of the target gene.

To measure levels of gene expression, complementary DNA (cDNA) must be synthesized from the RNA isolated. This is done by reverse transcription (RT), and is an essential step before qPCR can be performed. The enzyme used in the RT-reaction, reverse transcriptase, has the ability to synthesize cDNA from RNA. Reverse transcriptase is an enzyme originally associated with retrovirus, which uses this enzyme to convert its single stranded RNA (ssRNA) to a double stranded DNA (dsDNA) and then insert this to the host cell (Clark, 2010,p. 646 - 647). The Quantiscript Reverse Transcriptase used in this thesis does not have a retroviral source, but the manufacturer does not state which source it is derived from (Qiagen, 2009a). In vivo transcription of ssRNA to dsDNA is achieved by the three enzymatic activities of the multifunctional enzyme reverse transcriptase; an RNA-dependent DNA-polymerase, a hybrid-dependent exoribonuclease (RNase H) and a DNA-dependent DNA-polymerase. When using reverse transcriptase in vitro, the first two activities are used to produce ssDNA from ssRNA. The RNA-dependent DNA-polymerase is what actually transcribes the cDNA from the RNA-template, whereas the RNase H degrades the RNA in the RNA:DNA hybrids.

PCR is a method for rapid amplification of a known sequence of DNA (Clark, 2010,p. 635 - 639). It is based on the ability of DNA polymerase to manufacture new DNA using a pre-existing sequence of DNA as template. As new DNA is synthesized, each new DNA molecule is used as template, and the amount of DNA thus increase exponentially.

To initiate the polymerases activity, a small known sequence of the DNA, called primer, is needed. The primer is a short, single-stranded DNA-sequence chemically synthesized and designed to match the target sequence's 5'-end on both strands. Two primers are needed for each target sequence, one for forward synthesis and one for reverse synthesis. Further,

the primers are designed in such a way that they do not bind to themselves, creating primer-dimers or hairpin-loops. The primers used in RT-qPCR should ideally be designed to cross exon-exon boundaries. This will prevent them from binding to any genomic DNA (gDNA) present. Further, they have to be specific, and create only one product of amplification (amplicon). They should also have a GC-content of about 50 %, and more important; the GC-content should be the same in both the reverse and the forward primer, as this determines the annealing temperature.

A regular PCR reaction is performed in three main steps; denaturation, annealing and extension, as shown in Figure 1.4.1. Denaturation is commonly performed at approximately 90°C for one or two minutes. Heating the dsDNA melts the double strand and creates ssDNA. The temperature is then dropped to around 50°C to 60°C to allow the primers to anneal to the target sequence of the now ssDNA. The annealing temperature depends on the GC-content of the primers and is therefore decided when evaluating this. The final step, extension, is carried out around 70°C, the optimal temperature for the polymerase activity. The polymerase now elongates the DNA-strand, creating dsDNA. These steps are repeated 20 – 40 times in a thermocycler with the ability to rapidly change the temperature and create millions of copies of the target DNA sequence.

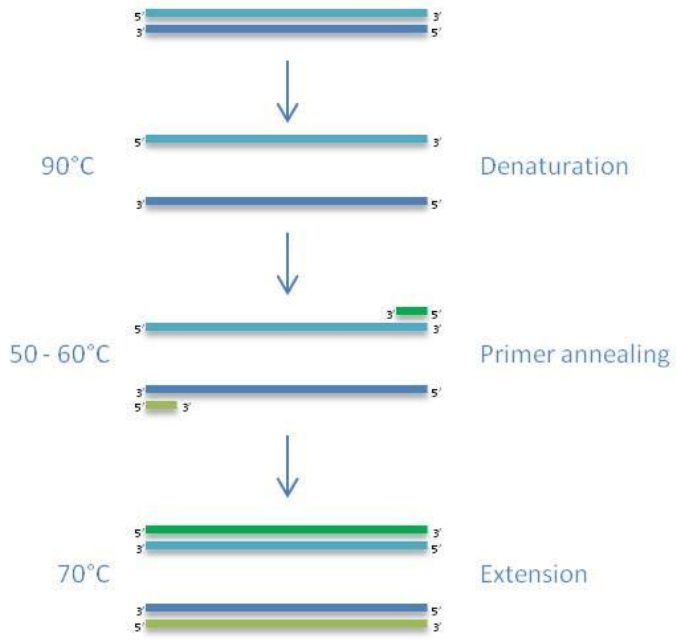


Figure 1.4.1 The steps of a regular PCR. These steps are repeated 20 – 40 times in a thermocycler creating millions of copies of the target DNA sequence.

Measuring the amount of synthesized DNA in real-time is done by the help of fluorescent probes that binds to target DNA. As the amount of newly synthesized DNA increases, so does the amount of fluorescence, as illustrated in Figure 1.4.2. This fluorescence is measured in the instrument, and can be directly related to the amount of DNA in the sample. Both sequence-specific and more general probes are available. SYBR® Green I, which is used in this thesis, is not sequence-specific, but binds only to dsDNA. As cDNA is ssDNA, the dye will not bind until the primers have annealed, Unbound dye emits very little fluorescence, but bound to dsDNA the fluorescence (530 nm) is greatly enhanced (Clark, 2010,p. 655 - 656, Roche, 2011).

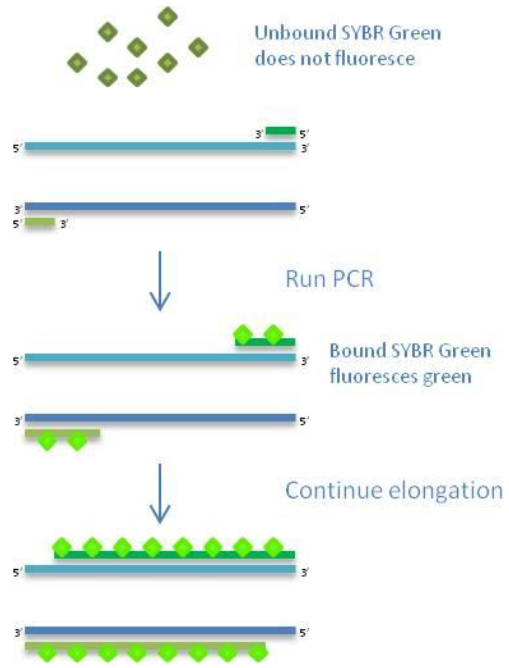


Figure 1.4.2 RT-qPCR with SYBR Green. The probe SYBR Green only fluoresce when bound to dsDNA, and the amount of fluorescence measured can thus be directly related to the amount DNA synthesized (Adapted from Clark, 2010).

To evaluate any contamination from gDNA, a no-reverse transcription control (-RT) is included. This includes template and all the reagents from the cDNA-synthesis, except reverse transcriptase. If gDNA has contaminated the samples during RNA isolation, this will be seen during the RT-qPCR. The –RT-control needs only be included once for each cDNA-synthesis, and should, if possible, be evaluated with a primer pair that does not cross exon-exon boundaries.

Further, to eliminate the possibility of contamination of the reagents used in the RT-qPCR, a no-template control (NTC) is included for each primer pair in each run. The NTC consists of the reagents used in the RT-qPCR and the primer pair, whereas the template is replaced with PCR-grade water. Ideally, the NTC should give no result in the RT-qPCR, but if the difference in Cq-value between the NTC and the samples is >5-6, it is usually considered as tolerable. The contribution of a NTC with such a high Cq-value will be minimal, as shown by the formula for calculating relative quantities (RQ): $RQ = 2^{\Delta Cq}$ (Biogazelle, 2012b). A difference in Cq-value of 6 will give an RQ of 64, and the NTC will thus contribute to a 64^{th} of the gene expression in the sample.

1.4.2. Quantification cycle, melting temperature and efficiency

The data which is generated from the instrument is called *threshold cycle* (Ct), *crossing point* (Cp) or *quantification cycle* (Cq). These terms all refer to the same; the numbers of cycles required to reach a defined intensity of fluorescence (Guenin *et al.*, 2009). This value will correspond to the number of amplicons created during the PCR reaction. According to the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines from Bustin *et al.* (2009), it is recommended to use the term Cq as a standard, and this will be followed in the rest of this thesis.

The Cq-value is calculated by the instrument's software, Light Cycler® 480 Software, version 1.5.1.62 (Roche, 2008). This is done by the Second Derivative Maximum method, which is based on the fact that during the exponential phase of the reaction, the fluorescence is increasing at an ever-increasing rate. This rate slows down as the reaction reaches a plateau, giving the reaction-curve its characteristic sigmoid shape. The Cq-value is identified as the point where the curve turns sharply upward, as shown in Figure 1.4.3, and corresponds to the maximum of the second derivative of the amplification curve, thus the name of the method. As described above, this value can be related to the number of amplicons created, and thus the amount of cDNA present in the first place. The Cq-value can therefore be used to compare the cDNA-level between different samples, and thus the expression of different genes.

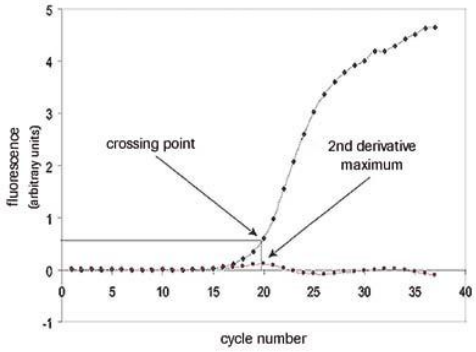


Figure 1.4.3 The Second Derivative Maximum Method. The Cq-value corresponds to the maximum of the second derivative of the amplification curve (Roche, 2008).

To evaluate the melting curve of the target DNA, a melting program is included in the qPCR-program, as described in Table 2.4.4 in section 2.4.2 below. The melting temperature (T_m) is defined as the temperature where half of the DNA-strand has melted. The temperature which a double stranded DNA molecule denature (melts), depends on factors such as GC-content, the sequence itself and the length of the strand. Products of the same length can have varying T_m, and each target DNA should have a characteristic T_m. This is essential in designing primers, but can also be used to identify genotype products, as a sequence that perfectly match the primer will melt at a higher temperature than a mismatched sequence with, say, a single-base mutation. Melting Curve analysis is performed in the Light Cycler® 480 Software, and is called *T_m Calling*. A melting curve chart and a melting peak are given for each sample, as shown in Figure 1.4.4. The target DNA will give a strong, sharp peak, and any bi-products such as primer-dimers can be seen as weaker, broader peaks at a different temperature. All primer pairs were evaluated to amplify a unique amplicon when inspecting the T_m-curves and the temperatures, as well as the Cq-values (see Appendix A4 and A3, respectively).

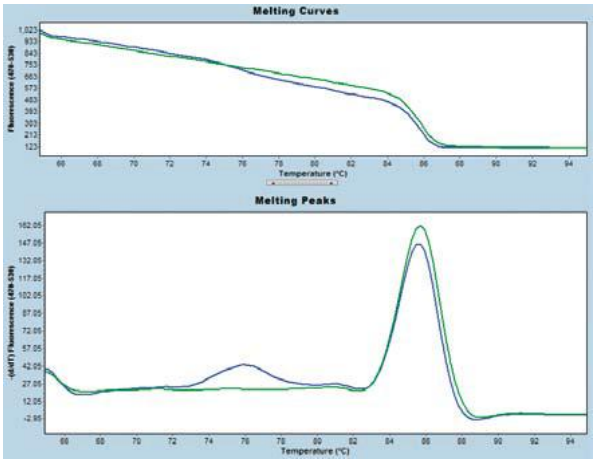


Figure 1.4.4 Tm-calling. A sharp peak is shown from a DNA-product with a T_m of approximately 86°C. One of the samples also show a weak, broad peak at 77°C, which is an unspecific bi-product such as primer-dimers (Roche, 2008).

The primers used in the RT-qPCR are evaluated by the efficiency, which is defined as the fold increase of amplicon per PCR-cycle (Ruijter *et al.*, 2009). Ideally, the PCR-reaction should double the number of amplicons in each cycle, and the efficiency of the reaction should therefore be 2, but this is seldom the case. The PCR efficiency is specific for each primer pair, and is calculated by LinRegPCR, version 2012.1 (available from http://LinRegPCR.nl). This is done by calculating the slope of the amplification curve in its exponential phase. If the PCR efficiency is calculated to be 2, this is considered as 100 % amplification; whereas an efficiency of 1 means that no amplification has taken place. LinReg calculates both the efficiency for each sample, but also gives a mean efficiency for each gene (see Appendix A3), which is used in further analysis.

1.5. Aim of study

This thesis aims to give a description of the CYP-family in *S. robusta*. *S. robusta* is a new candidate model organism representing the pennate diatoms, and the first draft genome sequence is currently being assembled. Analysis of cDNA and genome contigs identified a total of 68 CYPs. As this family is unusually expanded and contain several unique subfamilies it will provide a better understanding of how these genes have evolved in diatoms and how a benthic lifestyle may drive the evolution of these genes. Both bioinformatical tools and transcriptional analyses were used to characterize and explore this superfamily.

As part of the transcriptional analysis, a comparison has been made between two software programs for analyzing gene expression data. This is essential in publishing data from qPCR, and is one step closer publishing according to the MIQE guidelines.

2. MATERIALS AND METHODS

The materials and methods used in this thesis are described in this chapter. Background theory is given in detail in part 1.4. Detailed recipes of all media and buffers are provided in Appendix A1, and a list of primers used is given in Appendix A2.

2.1. Genome sequencing

The genome of *S. robusta* was sequenced by Professor Kjetill Sigurd Jakobsen at Centre for Ecological and Evolutionary Synthesis, Department of Biology at the University of Oslo. The 454 sequencing platform was used, a high-throughput pyrosequencing method, based on recording light burst as each nucleotide is incorporated (Rodriguez-Ezpeleta *et al.*, 2012, p. 16).

Additional data available through the NCBI Sequence Read Archive (SRA) was also used (available online at http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch&BLAST_SPEC=SRA&SHOW_DEFAULTS=on). These data were provided by Professor Wim Vyverman at the Department of Biology at Ghent University in Belgium, and were obtained by the Illumina sequencing platform. The Illumina platform is based on immobilization of linear sequencing library fragments by the help of solid support amplifications. By using fluorescent reversible terminator nucleotides, DNA-sequencing is enabled (Rodriguez-Ezpeleta *et al.*, 2012, p. 17).

Per Winge (PhD) at Department of Biology at the Norwegian University of Science and Technology set up a BLAST database for analyzing assembled contigs (cDNA and genomic) for use in the identification of complete genes and further aligned the amino acid sequences with CYPs of related organisms.

2.2. Experimental work on *Seminavis robusta*

An initial trial was conducted to get an overview of how the different *CYPs* were expressed (data not included). Cultures of axenic *S. robusta* were grown at regular conditions, at 18°C with a constant light of 100 µmol photons m^{-2} to serve as a control group. The methods used

for cultivation and axenization are described in sections 2.2.1 and 2.2.2 below, respectively. Three different experimental conditions were chosen in this initial trial; nitrogen starvation, dark conditions at regular temperature (18°) and dark conditions at a colder temperature (4°C). The cultures grown at 18°C were grown in a climate controlled growth room, whereas the cultures grown at 4°C were grown in a cold room. All cultures were cultivated for three days, before harvested as described in section 2.2.3 below. RNA was then purified, and gene-expression analysis was conducted as described in chapter 2.3 and 2.4 below.

This initial trial revealed that cultures grown in cold, dark conditions showed responses in the expression of some *CYPs.* As cultures grown in dark conditions at 18°C did not show this response, it was decided to include cultures grown in constant light of 100 µmol photons m^{-2} at 4°C in the main trial as this may introduce additional stress related responses in the diatoms. To see if this response was rapid, samples were harvested after a 2 hour exposure to cold conditions. It was also interesting to see if the diatoms were able to adapt to these cold conditions and a sampling after 24 hours exposure was therefore also included. This latter sampling would also reveal any slower responses to the colder temperature. Cultures grown at regular conditions, at 18°C with a constant light of 100 µmol photons m^{-2} are included as a control group, and are designated “const_0h_ctrl”, as they were exposed to zero hours of experimental conditions.

Another initial growth trial revealed a possible reaction to day/night conditions, as the cultures grown here showed a somewhat higher growth rate compared to cultures grown at constant light (data not included). Cultures grown at this condition were therefore included in this trial, to reveal any response in the expression of *CYPs*.

The sampling design is listed in table Table 2.2.1, and four biological replicas are included for each condition. A total of 16 cultivation flasks of 225 cm^3 were prepared with 22 ml of a dense culture of *S. robusta* and 128 ml of growth medium as described in section 2.2.1 below. The start cultures had been acclimated to the initial conditions described in the table below for approximately one week before initiating the experiment. The flasks were then placed in initial conditions for 72 hours to allow growth to start before placing them in the experimental conditions. Further description of the conditions is given in section 2.2.1 below.

Table 2.2.1 Sample design. Four biological replicas was included for each condition.

Sample	Initial conditions	Experimental conditions
Const_0h_ctrl_1	Constant light, 18°C, 72 hours	-
Const_0h_ctrl_2	Constant light, 18°C, 72 hours	-
Const_0h_ctrl_3	Constant light, 18°C, 72 hours	-
Const_0h_ctrl_4	Constant light, 18°C, 72 hours	-
Const_2h_1	Constant light, 18°C, 72 hours	Constant light, 4°C, 2 hours
Const_2h_2	Constant light, 18°C, 72 hours	Constant light, 4°C, 2 hours
Const_2h_3	Constant light, 18°C, 72 hours	Constant light, 4°C, 2 hours
Const_2h_4	Constant light, 18°C, 72 hours	Constant light, 4°C, 2 hours
Const_24h_1	Constant light, 18°C, 72 hours	Constant light, 4°C, 24 hours
Const_24h_2	Constant light, 18°C, 72 hours	Constant light, 4°C, 24 hours
Const_24h_3	Constant light, 18°C, 72 hours	Constant light, 4°C, 24 hours
Const_24h_4	Constant light, 18°C, 72 hours	Constant light, 4°C, 24 hours
D/n_1	-	Day/night cycle, 18°C, 72 hours
D/n_2	-	Day/night cycle, 18°C, 72 hours
D/n_3	-	Day/night cycle, 18°C, 72 hours
D/n_4	-	Day/night cycle, 18°C, 72 hours

After cultivation in the conditions given above, the diatoms were harvested as described in section 2.2.3 below, and further analysis were conducted. The samples cultivated in the day/ night cycle were harvested 3.5 hours into the new day cycle, thus after being exposed to light for 3.5 hours after the final night-period of 12 hours.

2.2.1. Method for cultivation

Cultures of *S. robusta* were prepared from cryopreserved cultures from the PAE Culture Collection of Diatoms from the Ghent University in Belgium as described below. Modified f/2 medium (see appendix A1.1) was used as a cultivation medium. This medium is a common medium used for marine algae, diatoms in particular (CCAP, 2007). The original formulation was named “f medium”, but the concentrations now used has been reduced to half, thus the name “f/2 medium”. Axenic cultures were prepared according to the protocol described in section 2.2.2 below.

The cryopreserved cultures were kept on liquid N_2 until they were thawed just before added f/2 medium. Gentle thawing was performed by a water bath kept at 35 - 37°C. 1 ml of the culture was transferred to 50 ml of f/2 medium within three minutes after thawing was initiated. The cultures were then cultivated at 18°C in 75 cm^2 sterile, BD Falcon™ Cell Culture

Flasks culture flasks with ventilated cap, lying down. As *S. robusta* is a benthic alga, the cultures were cultivated in culture flasks lying down, giving a large surface in which they could adhere to. To ensure a gentle recovery, the cultures were kept in the dark for 24 h, before exposed to a day-night cycle until proper growth was visible.

Further maintenance of the cultures was done by changing the medium approximately every 3 – 4 days. To make fresh cultures, 5 ml of a dense culture was transferred to a new culture flask with 45 ml f/2 medium. Larger and smaller culture flasks was also used, with a ratio of 1:10 of culture:f/2 medium.

A climate controlled growth room at approximately 18°C was used for maintaining cultures. Constant light of 100 µmol photons $m^{-2}s^{-1}$ and a 12 hour light / 12 hour dark photoperiod were both used to determine any difference in the growth rate. For experimental work, two growth cabinets with a temperature of 18°C and 4°C, respectively, were used, both with a constant light period of 100 µmol photons $m^{-2}s^{-1}$. The cultures exposed to the day/night cycle were grown in the growth room with the 12 hour light / 12 hour dark photoperiod and a light intensity of 100 µmol photons $m^{-2}s^{-1}$ in the light photoperiod.

2.2.2. Method for axenization

Axenization was done to ensure that the cultures were free of bacterial contamination. The method is modified after Andersen (2005).

The diatoms were grown to a dense, exponentially growing culture, before resuspended. 0.5 ml of the resuspended culture was transferred to a 15 ml falcon tube containing 10 ml f/2 added 1.5 or 2.0 ml antibiotic mixture (see appendix A1.1). To stimulate growth of any bacteria that might be present, 0.5 ml of f/2 with 1 mg/ml peptone was added. The cultures with antibiotics were incubated for approximately three days in horizontally oriented tubes.

The cultures were checked for diatom growth by microscoping, before resuspended and 0.5 ml was transferred to 10 ml f/2 in a small culturing flask (25 cm^2). After about a week of growth, the axenity was tested. This was done by adding 0.5 ml of culture to tubes containing 5 ml of f/2 added 1mg/ml peptone. The tubes were then kept in the dark, and checked visually for any bacterial growth. Any clouding of the solution would indicate

bacterial growth, and the axenization would not have been successful. The tubes were kept for a total of 14 days, and the cultures with no visual growth after this time period were considered axenic.

2.2.3. Method for harvesting

Harvesting of the algae was done relatively fast to ensure that the conditions in the harvesting area did not affect the RNA-pool. This was especially important while harvesting cultures that had been cultivated in different light- and/or temperature conditions than in the harvesting area.

The medium was poured off, leaving about 40 – 50 ml, before scraping the diatoms loose from the flask's surface with a cell scraper suitable for the flask size. The suspension was transferred to a 50 ml falcon tube and centrifuged at 3000 rpm for 5 minutes at 18°C, and the supernatant was poured off, leaving about 1.5 ml. The pellet was resuspended, and the suspension was transferred to a 2 ml tube, before centrifuged at 13 000 rpm for 1 minute at 4°C. The supernatant was removed, and the pellet was flash frozen in liquid N_2 before stored at -80°C.

2.2.4. Method for counting

Estimating the concentration of cells in the cultures was done by the help of a Nageotte counting chamber and a Nikon Eclipse E800 microscope. A total magnification of x100 was used, with phase-contrast adapted to this objective.

The cells of the culture were brought in to suspension, and approximately 1 – 2 ml culture was transferred to a small tube. Lugol's iodine solution (see Appendix A1.1) was added to immobilize the cells and distinguish between live and dead cells, as dead cells do not take up color. A small amount of cell culture added Lugol's solution was then transferred to a counting chamber and the cell concentration could be estimated.

The counting chamber consists of 40 stripes arranged as shown in Figure 2.2.1, and constitutes a total of 50 µl. A total of 8 stripes were counted, distributed evenly in the grid (shown as shaded in Figure 2.2.1). The average was then calculated, and total amount of cells in the culture could thus be estimated.

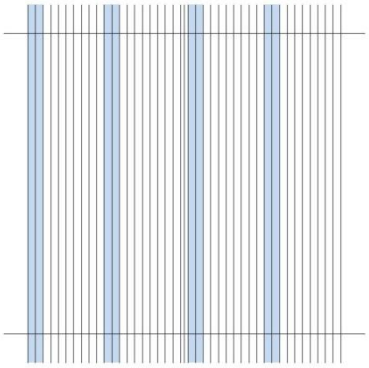


Figure 2.2.1 The grid of a Nageotte counting chamber. Those counted are shaded. (Not to scale).

The cultures were in exponential growth phase when the cell concentration reached 40-45 000 cells/ml.

2.3. RNA isolation and handling

Ribonucleic acid (RNA) is constantly made and degraded in the cells. In order to be able to say something about the genes expressed at a certain time, the total RNA-pool at that time point have to be isolated and used for further analysis. As RNA is very unstable compared to DNA, the samples have to be handled properly, and are therefore flash frozen in liquid N_2 as described in section 2.2.3. This is also essential to inactivate the ribonucleases (RNase) always present.

2.3.1. Method for RNA isolation

RNA was isolated with the Spectrum™ Plant Total RNA Kit (Sigma Aldrich, cat. no. STRN250) following the manufacturer's protocol. This method is a column-based method, and a series of binding- and washing steps are performed to ensure that the RNA isolated is free from proteins, salts, and other substances that can inhibit downstream analysis.

1. The dry algae-pellet (see section 2.2.3) was initially homogenized for 2x 1 minute at 25/s in a TissueLyser homogenizing instrument (Qiagen, cat. no. 85200). To keep the sample

cold, the homogenizing block was kept at -80°C for approximately 30 minutes before use, together with the 5 mm steel homogenizing beads used. After the first round, the homogenization block was turned over to ensure equal homogenization.

2. 500 µl Lysis Solution with 10 µl β-mercaptoethanol was then added to each sample, before another homogenizing step for 2x 1 minute at 25/s. The block was turned over after the first homogenization, as described in step 1. After this step the samples could be handled at ambient temperature (T_A).
3. The samples were then incubated at 56°C for 4 minutes, followed by centrifugation at 13 000 rpm for 3 minutes.
4. The lysate supernatant was then transferred to a filtration column placed in a 2 ml collection tube, followed by centrifugation at 13 000 rpm for 1 minute.
5. 750 µl Binding Solution was then added to the flow-through and mixed. 700 µl of the mix was then transferred to a binding column placed in a 2 ml collection tube and centrifuged at 13 000 rpm for 1 minute.
6. The flow-through was removed before the remaining mix from step 5 was transferred to the column and centrifuged at 13 000 rpm for 1 minute.
7. 300 µl Wash Solution I was added to the binding column, followed by centrifugation at 13 000 rpm for 1 minute. The flow-trough was discarded.
8. To remove any DNA present, the samples were treated with DNase on-column with RNase-Free DNase (Qiagen, cat. no. 79254). A mix of 11 µl RNase-Free DNase added 77 µl RDD-buffer was placed directly on the filter of the column to each sample, and incubated at ambient temperature for 15 – 30 minutes.
9. 500 µl Wash Solution I was added to the binding column, followed by centrifugation at 13 000 rpm for 1 minute. The flow-trough was discarded.
10. 500 µl Wash Solution II was added to the binding column, followed by centrifugation at 13 000 rpm for 30 seconds. The flow-trough was discarded, and this step was repeated.
11. The binding column was transferred to a new 2 ml collection tube and 50 µl Elution Buffer was added. The tube was left for 1 minute at ambient temperature, followed by centrifugation at 13 000 rpm for 1 minute.
12. Another 50 µl Elution Buffer was added and centrifuged at 13 000 rpm for 1 minute.

13. The samples were placed on ice and 2.5 µl RNasin (Progmega, cat. no. N2611) was added to each sample. The samples were mixed by vortexing. RNasin is a ribonuclease inhibitor designed to not interfere with downstream analysis.
14. After evaluation of the RNA-yield, the samples were kept at -80°C before further analysis.

Evaluation of the RNA-yield was done by the help of a NanoDrop 1000 from Thermo Fisher Scientific. The NanoDrop is a full spectrum spectrophotometer (220 – 750nm) able to measure the absorbance in a very small volume of sample (ThermoFisherScientific, 2011). The result is given in an absorbance curve, together with the nucleic acid concentration in the sample, given in ng/µl. The purity of the RNA isolated is evaluated by analyzing the ratio of the absorbance measured at 260 nm and 280 nm, and the ratio at 260 nm and 230 nm, designated 260/280 and 260/230 respectively. A 260/280 ratio of about 2.0 is generally considered as pure for RNA, but totally pure RNA will have a ratio of >2.1. A lower 260/280 ratio can indicate the presence of proteins, phenols or other contaminants. The 260/230 ratio is a secondary measure of nucleic acid purity. A pure sample will often give a higher number for the 260/230 than for the 260/280, and is usually in the range 1.8 – 2.2. A lower number can indicate co-purified contaminants.

As RNA is in danger of being degraded if not handled properly, the RNA-integrity must be evaluated before proceeding with downstream analysis. This was done by running the samples on a formaldehyde agarose gel (FA-gel), prepared as described in appendix A1.2, according to Qiagen's Guide To Analytical Gels (Qiagen, 2003). The ribosomal RNA (rRNA) bands should be visual as clear, sharp bands, and the 28S rRNA should show an intensity approximately twice as strong as the smaller 18S rRNA band. If the rRNA bands are not clearly separated, but instead appears as a smear of smaller RNA fragments the RNA is most likely degraded and should not be used in further analysis.

2.4. Gene expression analysis

Gene expression analysis was performed by RT-qPCR as described in section 2.4.1 and 2.4.2 below.

2.4.1. Method for cDNA synthesis

cDNA synthesis was performed with the QuantiTect® Reverse Transcription kit (Qiagen, cat. no. 205313).

Even though the RNA was treated with DNase, it might still be some DNA left in the sample. A reaction to eliminate gDNA was therefore performed according to Table 2.4.1, and incubated at 42°C for 2 minutes before placed on ice (Qiagen, 2009a).

Table 2.4.1 Reaction for elimination of genomic DNA

	Sample	-RT
RNA-template	1 µg	0.5 µg
gDNA Wipeout Buffer, 7x	2 µl	1 µl
RNase-free H_2O	q.s.*	q.s.
Total	14 µl	7 µl

*q.s.: quantity sufficient

After elimination of the genomic DNA, the RT-reaction was performed according to Table 2.4.2. The reaction was incubated at 42°C for 15 minutes, followed by 3 minutes at 95°C to inactivate the enzyme. The Quantiscript Reverse Transcriptase also contains RNase inhibitor, and the Quantiscript RT buffer includes both Mg^{2+} and dNTPs. RT can be performed with specific, oligo-dT or random primers. The primers used in this thesis was provided with the kit, and is a mix of random primers and oligo-dT primers (Qiagen, 2009a).

Table 2.4.2 RT-reaction

	Sample	-RT
gDNA elimination reaction (Table 2.4.1)	14 µl	7 µl
Quantiscript Reverse Transcriptase	1 µl	-
Quantiscript RT Buffer, 5x	4 µl	2 µl
RT Primer Mix	1 µl	0.5 µl
RNase free H_2O	-	0.5 µl
Total	20 µl	10 µl

After cDNA synthesis, the samples were stored at -20°C for further analysis.

2.4.2. Method for RT-qPCR

Real time qPCR was carried out with the LightCycler® 480 SYBR Green Master kit (Roche Applied Science, cat. no. 04 707 516 001) and the LightCycler® 480 instrument from Roche Applied Science. The reaction was prepared according to Table 2.4.3, and performed in a white 96-well plate. LightCycler® 480 SYBR Green I Master is a hot-start PCR-mix, consisting of FastStart *Taq* DNA polymerase, reaction buffer, dNTPs (with dUTP instead of dTTP), SYBR Green I dye and $MgCl_2$ (Roche, 2011). The gene-specific primers are listed in appendix A2.1, and were designed to have an optimal T_m of approximately 55°C.

Table 2.4.3 qRT-PCR mix

	Volume
Template (1:10 cDNA)	5 µl
LightCycler® 480 SYBR Green I Master	3 µl
PCR primer mix, 5 µM	2 µl
H_2O, PCR-grade	10 µl
Total	20 µl

Before loading the multiwell plate on the instrument, the plate was centrifuged at 1500g for 2 minutes. The qPCR-program performed on the LightCycler is described in Table 2.4.4.

Table 2.4.4 RT-qPCR-program

Step	Temperature	Time	Cycles
Pre-Incubation	95°C	5 minutes	1
Amplification	95°C	10 seconds	45
	55°C	10 seconds	
	72°C	10 seconds	
Melting Curve	95°C	5 seconds	1
	65°C	1 minute	
	97°C	∞	
Cooling	40°C	10 seconds	1

2.5. Confirmation of putative introns in *Cyp19A1*

The *Cyp19A1* gene was predicted to contain two introns of 94 and 87 bp by analyzing the genomic sequence. No cDNA-sequence existed, though, and a PCR-reaction was thus performed to possibly confirm these introns. cDNA synthesized as described in Section 2.4.1 and diluted 1:10 was used as template. According to an initial trial, *Cyp19A1* showed low expression according to the RT-qPCR, and the two samples showing highest expression were used as templates.

Primers flanking the two putative introns were designed and are listed in Appendix A2.2. The primers were expected to give products of 133 and 102 bp, respectively, if the predicted introns existed. If these introns were coding regions, though, the primers would give products of 227 and 189 bp.

To obtain a complete cDNA sequence for the *Cyp19A1* gene, the PCR product could then be sequenced.

2.5.1. Method for PCR

A standard PCR-reaction was prepared according to Table 2.5.1 with the *Taq* DNA Polymerase with ThermoPol Buffer (BioLabs, cat. no. M0267).

Table 2.5.1 PCR-reaction

	Volume
Template	1 µl
10x ThermoPol Reaction Buffer	5 µl
10 mM dNTPs	1 µl
10 µM forward primer	1 µl
10 µM reverse primer	1 µl
Taq DNA polymerase	0.25 µl
H_2O, PCR-grade	40.75 µl
Total	50 µl

A one-step PCR was performed according to Table 2.5.2.

Table 2.5.2 PCR-program

Step	Temperature	Time	Cycles
Denaturation	95°C	30 seconds	
Annealing	55°C	30 seconds	45
Extension	70°C	30 seconds	
Cooling	4°C	Hold	

The PCR-products were visualized on a 1 % agarose gel (see appendix A1), with a 100 bp GeneRuler ladder (Thermo Scientific, cat. no. R0491). Further analysis was performed on the samples giving visible products on the gel.

2.5.2. Method for cloning

Cloning was performed with the TOPO® TA Cloning® Kit (Invitrogen, cat. no. 45-0640) and One Shot® Mach1^{TM}-T1^{R} Chemically Competent *E. coli* (Invitrogen, cat. no. C8620-03) according to the protocol from the manufacturer (Invitrogen, 2011). This was done in an attempt to increase the amount of template for sequencing. Isolation of plasmid DNA from the bacterial pellet was performed with the QIAprep® Spin Miniprep (Qiagen, cat. no. 27106).

A thorough description of this method is not given, as it did not yield any results, as described in part 3.2 below.

2.5.3. Method for sequencing

Plasmid DNA isolated as described above was used as template in sequencing reaction. Sequencing was performed with BigDye v3.1 protocol described by the sequencing facility at University Hospital of North Norway (http://www.unn.no/bigdye-v3-1-sequencing-protocol/category25527.html).

A thorough description of this method is not given, as it did not yield any results, as described in part 3.2 below.

2.6. Analysis of RT-qPCR data

To analyze the gene expression data retrieved from the LightCycler® 480, two different software programs were used and compared; REST2009, version 2.0.13 from Qiagen and qBase^{Plus} from Biogazelle. Both software programs were used to determine whether a significant difference existed between the controls and the samples exposed to experimental conditions, and the results were then compared to reveal any differences. Previously, REST2009 was the most common program used in the research group, but due to bad compatibility with newer computers, it was determined to compare it with qBase^{Plus} to see if this software was a good replacement.

2.6.1. REST2009

REST2009 (available from https://www.qiagen.com/forms/rest/default.aspx?) use bootstrapping techniques and randomization to eliminate the uncertainty related to analyzing gene expression data (Qiagen, 2009b). Multiple reference and target genes can be analyzed, but it is only possible to analyze one subgroup at a time.

The result is presented in both a table and a whisker-dot-plot. A hypothesis test, P(H1), is given in the results table, and represents the probability (p-value) of the alternate hypothesis. This alternate hypothesis states that the difference indicated between the samples and the control is due only to chance. To evaluate this alternate hypothesis, a randomization test is performed, where the scenario is that if the difference indicated between two groups is due to chance, random values could be swapped in the two groups, and the difference should be the same as between the initial groups. A total of 10 000 random reallocations of samples and controls are performed, and the resulting p-value is used to evaluate whether the difference is significant or due to chance. Samples generating a p-value < 0.05 are considered as significantly up- or down-regulated compared to the control.

2.6.2. qBase^{Plus}

qBase^{Plus} (available from http://www.biogazelle.com/products) , as REST2009, has the possibility of using multiple reference genes, but has a larger range of statistic analysis. The

software includes, to mention some, unpaired t-test, paired t-test, Mann-Whitney test, Pearson correlation and One-way ANOVA. Another feature of qBase^{Plus} is the possibility of analyzing multiple targets and subgroups at the same time, with the same test.

As mentioned, the statistics module of qBase^{Plus} has a wide range of possibilities (Biogazelle, 2012a). A statistics wizard guides the user through the different choices, and depending on the features of the dataset, a suitable statistic analysis is chosen. Further, the software corrects the p-values when the same test is performed for several targets. This ensures that the amount of false positives is as low as possible (≤ 5 %). When several subgroups are compared with ANOVA, a multiple comparison post-test is also performed to correct the p-value further.

Also included in the qBase^{Plus} software is geNorm^{Plus} (Biogazelle, 2011). geNorm^{Plus} evaluate the expression stability of candidate reference genes, and gives an interpretation of the results, and thus a recommendation for optimal reference gene(s). geNorm was earlier based on Microsoft Excel, and the scientist had to interpret the result, but as geNorm^{Plus} now does this, this will ensure a more uniform evaluation of reference genes. Three results are given when performing a geNorm^{Plus} analysis on a set of candidate reference genes. The geNorm M values rank the candidate reference genes according to stability as seen in Figure 2.6.1. The genes are arranged after declining M-value from left to right, with the most stable reference gene(s) giving the lowest M-value.

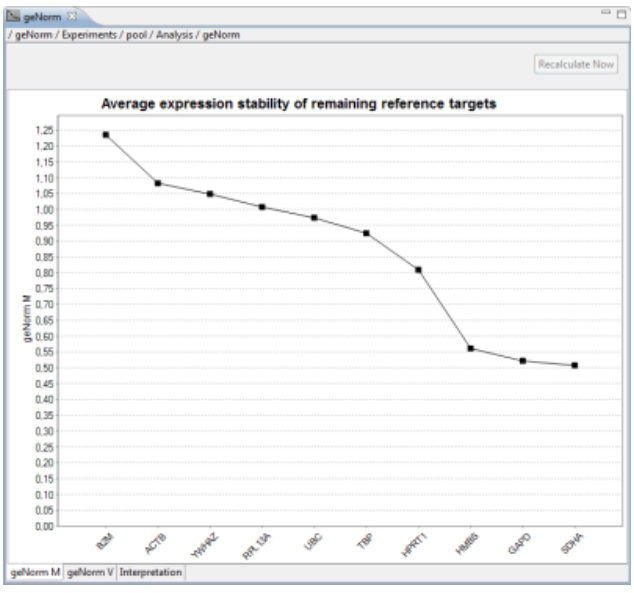


Figure 2.6.1 Example of geNorm^{Plus} results; the M-value. The geNorm M-value rank the candidate reference genes according to stability. The genes are arranged after declining M-value from left to right, with the most stable reference gene(s) giving the lowest M-value (Biogazelle, 2011)

The second result from geNorm^{Plus}, geNorm V, is a bar chart which shows the optimal number of reference genes that is recommended, as shown in Figure 2.6.2 below. This bar chart is based on the geNorm V-value, $V_{n/n+1}$, which is shown for each comparison between two consecutive numbers (n and n+1) of reference genes. The geNorm V-value gives an indication of the possible benefit of using an extra (n+1) reference gene. When the V-value drops below the threshold of 0.15, the benefit is limited.

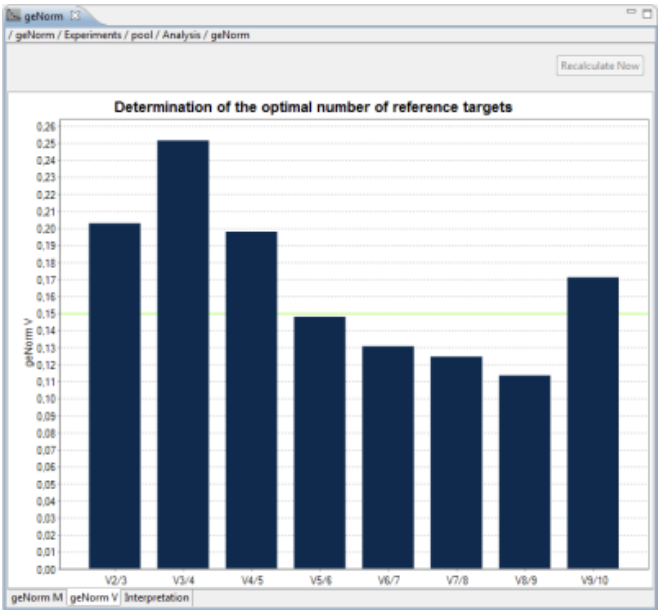


Figure 2.6.2 Example of geNormPlus results; the V-value. The geNorm V-value, $V_{n/n+1}$, gives an indication of the possible benefit of using an extra (n+1) reference gene. When the V-value drops below the threshold of 0.15, the benefit is limited (Biogazelle, 2011).

The third result window is an interpretation of the evaluation of the candidate reference genes. This section contains a summary of the results, as well as a recommendation of which and how many reference gene(s) that should be used.

2.7. Sequence analysis

Several tools available online were used when analyzing the sequences of the *S. robusta* genome as described below.

2.7.1. BLAST

The Basic Local Alignment Search Tool (BLAST) is an online tool (available from http://blast.ncbi.nlm.nih.gov/) that finds similar regions between one or more query sequences and sequences in the database. The statistical significance of the matches is calculated and an Expect (E)-value is given for each result (Madden, 2003). The E-value describes the number of hits that can be expected to be by chance, and decreases exponentially as the score of the sequences increases. The closer the E-value is to zero, the better the match between the query sequence and the result from the database. It is important to mention, thought, that short sequences can get a relatively high E-value even when the score is good. This is because the length of the sequence is taken to account when calculating the E-value, and as shorter sequences will have a higher probability of occurring by chance, the E-value will be higher. BLAST is a useful tool when determining which superfamilies a number of sequences belongs to.

2.7.2. MACAW

Multiple Alignment Construction & Analysis Workbench (MACAW) is a software program for analyzing multiple alignments by searching for and linking matching blocks of sequences. It is an excellent tool for identifying introns in a given sequence if both a genomic and a cDNA sequence are known. Both protein and nucleotide sequences can be analyzed and visualized in a schematic view and in detail in the alignment. A typical MACAW-file is seen in Figure 2.7.1 below, with the genomic sequence and the cDNA-sequence, as well as three primer pairs designed, and the position of these.

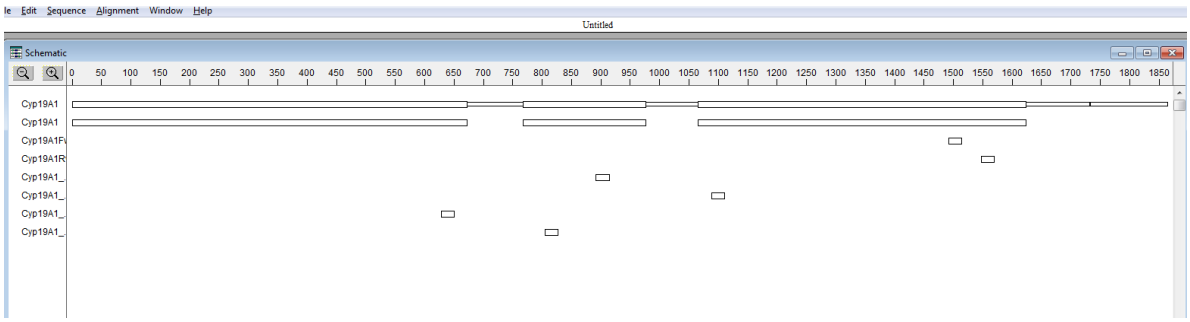


Figure 2.7.1 MACAW-file for *Cyp19A1*. The genomic sequence is seen at the top, followed by the cDNA-sequence, and the two introns are easily recognized. The next, smaller sequences are a total of three primer pairs designed for *Cyp19A1*, and the positions of the forward and reverse primers are visualized.

2.7.3. Phylogenetic analysis

A phylogenetic tree was made by Per Winge (PhD), and was used to show lines of descent and evolutionary relationship between the CYPs of *S. robusta* and other organisms.

A neighbor joining (NJ) tree, that included 68 predicted *S. robusta* CYP proteins and 99 CYP proteins from other organisms, was produced by the ClustalX program (Larkin *et al.*, 2007). The basic principle of the neighbor-joining (NJ) method (Saitou and Nei, 1987) is to find pairs of operational taxonomic units (=neighbors) that minimize the total length of the branch. This method is particular useful when evaluating a group of sequences with varying rates of evolution. The pair-wise alignment was made using the Gonnet 250 score-matrix, and the neighbor joining tree was corrected for multiple substitutions using Kimura's correction method. 1000 bootstrap replicates were run, testing the reliablilty in the multiple sequence alignment. This is done by analyzing whether the whole dataset supports the tree created, or if many equal alternatives exists and the current tree is just a marginal winner (Baldauf, 2003). Random sampling with replacement is used to create subsample trees. If one group appears in every subsample of the tree, the bootstrap support is 100 %, and will give a bootstrap value of 1000 in our case where 1000 bootstrap tests were run. Generally, values of 70 %, or higher is considered as reliable groupings.

For visualization the NJ-tree was imported into Dendroscope version 3.2.2 (Huson *et al.*, 2007). The NJ-tree was drawn as a circular cladogram and nodes that have bootstrap values above 500 are shown in the tree, as seen in Figure 3.1.1 in part 3.1. Phylogenetic analysis was performed by Per Winge.

2.7.4. GeneDoc

The assignment of CYPs to specific subfamilies was initially done based on results from a phylogenetic analysis of 167 CYP-proteins form various organisms, as described in part 2.7.3. GeneDoc was then used to align the amino acid sequences of the CYP-family in *S. robusta* with homologous sequences from other organisms.

GeneDoc is a multiple sequence alignment tool available online (http://www.nrbsc.org/gfx/genedoc/) used to investigate characteristics and properties in a group of sequences. GeneDoc is particular useful when you have to do manual editing of protein alignments which is later used in phylogenetic studies (Nicholas *et al.*, 2007). GeneDoc was also used to find the amino acid similarity between CYPs predicted to belong to the same family, as the report module gives a statistical report where the amino acid similarity is listed. The results are given in Appendix A8, and as mentioned in section 1.3.4, CYPs of the same family generally should have an amino acid similarity of at least 40 %, and CYPs belonging to the same subfamily a similarity of at least 55 %.

The complete GeneDoc alignment is given in Appendix A7, but only contains the CYPs of *S. robusta* as the original alignment was too large to include.

2.8. Prediction of proteinlocalization

Different tools available online were used in a putative prediction of the localization of the different CYPs as described below.

2.8.1. WoLF PSORT

WoLF PSORT is an online tool (available from http://wolfpsort.org/) that predicts the subcellular localization of the proteins based on the amino acid sequence (Horton, 2007, Horton *et al.*, 2007). The amino acid sequence of the protein is converted to numerical localization features, based on the composition of amino acids and functional motifs, like DNA-binding motifs. The results are given in both a list of proteins with known localization that share some features, and tables with more detailed information about specific localization features of the submitted protein sequence.

2.8.2. SecretomeP

SecretomeP 2.0 is another tool available online (available from http://www.cbs.dtu.dk/services/SecretomeP-2.0/), and gives a prediction of whether the protein in query is secreted, and also if this is classical; i.e. based on signal peptide, or non-classical (Bendtsen *et al.*, 2004). Non-classical protein secretion is not triggered by an N-

terminal signal peptide, as it has been shown that also protein without this signal peptide can be exported. Such non-classical secretion pathways can include amongst others direct translocation across the plasma membrane and lysosomal secretion. SecretomeP is based on protein sequence mapping. The protein is scanned for features such as predicted structure, size, composition, charge, post-translational modifications and more, as extracellular proteins share some features and properties related not to the secretory process itself, but to the function outside the cell. The result is given in a SecP-score, or neural network-score (NN-score). A value above 0.500 indicates possible secretion, and if a signal peptide at the same time is not predicted, the protein is possibly a non-classical secreted protein.

2.8.3. TargetP

The third program used, TargetP 1.1 (available from http://www.cbs.dtu.dk/services/TargetP/), predicts subcellular localization of eukaryotic proteins based on predicted presence of N-terminal pre-sequences; mitochondrial targeting peptide (mTP), secretory pathway signal peptide (SP) or chloroplast transit peptide (cTP) (Emanuelsson *et al.*, 2000). The score given for these pre-sequences are not probabilities, and do not necessarily add to one, but the highest score is the most likely localization. The program also gives a value for the reliability class (RC), in a range of 1 – 5; the lower the RC-value, the safer the prediction.

3. RESULTS

3.1. Families and subfamilies

A total of 68 CYPs were discovered in *S. robusta* and classified in families and subfamilies according to phylogenetic relationship and amino acid similarity. The complete lists of amino acid similarity in the groups are given in Appendix A8. The phylogenetic tree was made as described in section 2.7.3 and is seen in Figure 3.1.1 below. A total of 25 families of CYPs were shown to exist in *S. robusta*.

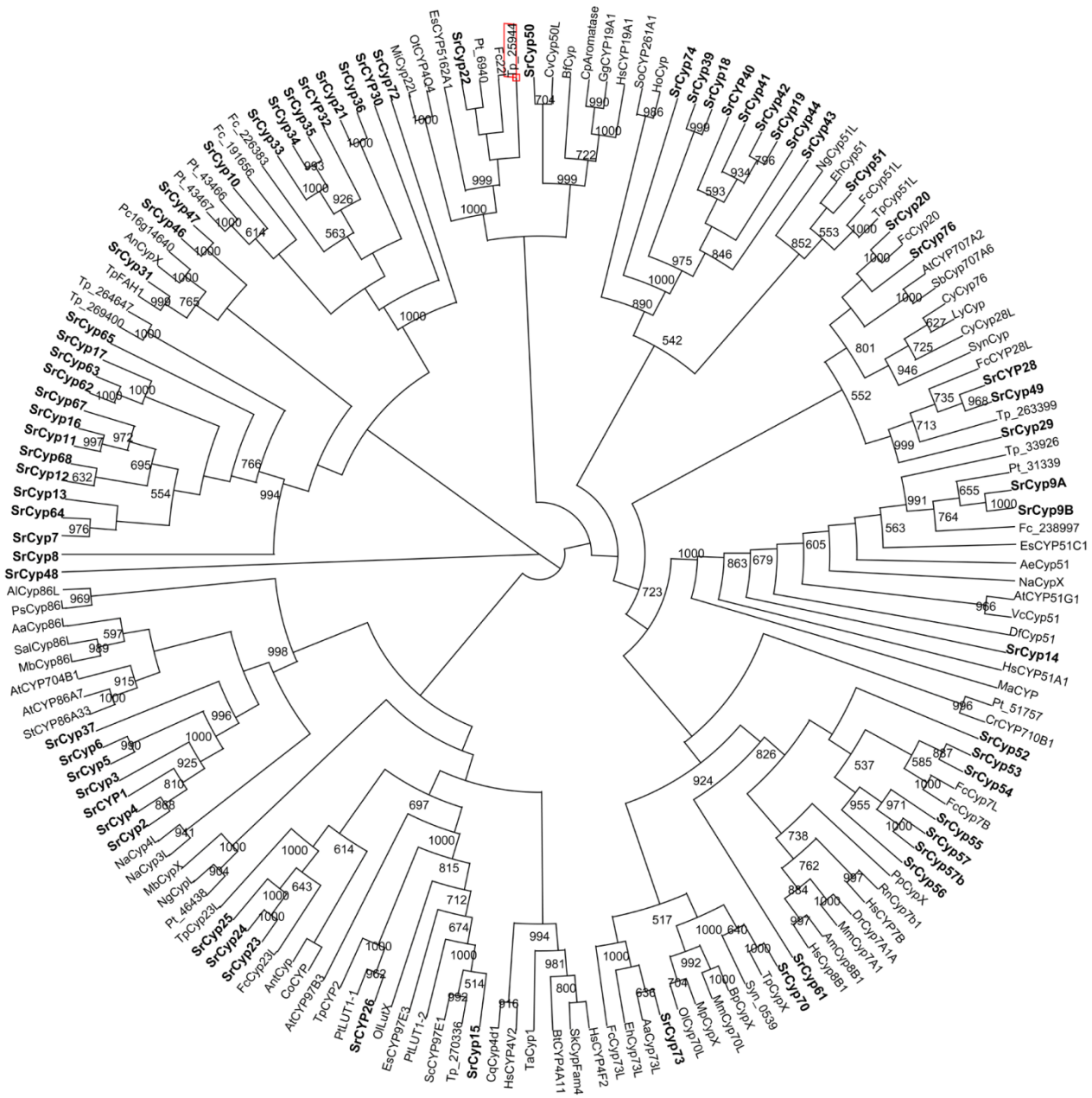


Figure 3.1.1 Phylogenetic tree of the 68 CYPs in S. robusta and related genes in other organisms. Original nomenclature has been used when creating this tree.

Initially, the CYPs were named by Per Winge (PhD) in order of discovery when exploring the *S. robusta* genome; Cyp1 was discovered first, then Cyp2 and so on. After grouping in families according to phylogenetic relationship, the amino acid similarity between the CYPs in the different sub-groups was explored and families and subfamilies were created.

Four of the families found are already known families. As a result, thirteen CYPs have been given a new nomenclature and are listed in Table 3.1.1 below.

Table 3.1.1 CYPs in *S. robusta* placed in known CYP-families.

Initial nomenclature	CYP-family	Subfamily	New nomenclature	Function
SrCyp50	Family 19	Subfamily A	Cyp19A1	Aromatase
SrCyp14	Family 51	Subfamily A	Cyp51A1	Sterol demetylase
SrCyp9a		Subfamily B	Cyp51B1	
SrCyp9b			Cyp51B2	
SrCyp2	Family 86	Subfamily A	Cyp86A1	Fatty acid hydroxylase
SrCyp4			Cyp86A2	
SrCyp1			Cyp86A3	
SrCyp3			Cyp86A4	
SrCyp5			Cyp86A5	
SrCyp6			Cyp86A6	
SrCyp37		Subfamily B	Cyp86B1	
SrCyp15	Family 97	Subfamily A	Cyp97A1	Carotenoid
SrCyp26		Subfamily B	Cyp97B1	

The remaining fifty five CYPs in *S. robusta* have been placed in families and subfamilies, but have not been given a final nomenclature. As the complete sequence is still not known for some of the CYPs, their suggested subfamily may change when these data are available. These CYPs are marked with a * in the tables below.

Initial grouping was based on phylogenetic relationship, and as seen in Table 3.1.2 and Table 3.1.3 below, amino acid similarity revealed that some of the subgroups were in fact several families. *SrCyp43* and *SrCyp44* possibly constitute a subfamily D in family XV when the sequences are complete.

Table 3.1.2 Subgroups consisting of two families. *SrCyp43* and *SrCyp44* possibly constitute a subfamily D in family XV when the sequences are complete.

Initial subgroup	Family	Subfamily	Initial nomenclature
Subgroup 1	Family I	Subfamily A	SrCyp32*
			SrCyp33
			SrCyp34
			SrCyp35
		Subfamily B	SrCyp21
			SrCyp36
		Subfamily C	SrCyp10
		Subfamily D	SrCyp30
	Family II	Subfamily A	SrCyp72
Subgroup 2	Family V	Subfamily A	SrCyp31
	Family VI	Subfamily A	SrCyp46
			SrCyp47
Subgroup 3	Family VII	Subfamily A	SrCyp20
	Family VIII	Subfamily A	SrCyp76
Subgroup 5	Family XIII	Subfamily A	SrCyp7
			SrCyp64
		Subfamily B	SrCyp13
		Subfamily C	SrCyp11
			SrCyp16
			SrCyp67
		Subfamily D	SrCyp12
			SrCyp68
		Subfamily E	SrCyp17
			SrCyp62
			SrCyp63
		Subfamily F	SrCyp65
	Family XIV	Subfamily A	SrCyp8
Subgroup 6	Family XV	Subfamily A	SrCyp19
			SrCyp42
			SrCyp40
			SrCyp41
		Subfamily B	SrCyp18
			SrCyp39
		Subfamily C	SrCyp74
	Family XVI	Subfamily A	SrCyp43*
			SrCyp44*

One subgroup was in fact four families, as seen in Table 3.1.3 below.

Table 3.1.3 Subgroup 4 was shown to consist of four families.

Initial subgroup	Family	Subfamily	Initial nomenclature
Subgroup 4	Family IX	Subfamily A	SrCyp52
	Family X	Subfamily A	SrCyp53
		Subfamily B	SrCyp54
	Family XI	Subfamily A	SrCyp55
			SrCyp56
			SrCyp57
			SrCyp57b
	Family XII	Subfamily A	SrCyp61

The subgroups seen in Table 3.1.4 below were shown to consist of only one subfamily in each.

Table 3.1.4 Subgroups consisting of only one CYP-family.

Subfamily	Initial nomenclature	Initial subgroup	Family
Subgroup 7	Family III	Subfamily A	SrCyp23
			SrCyp24
			SrCyp25
Subgroup 8	Family IV	Subfamily A	SrCyp28
			SrCyp49*
		Subfamily B	SrCyp29
Subgroup 9	Family XVII	Subfamily A	SrCyp70
Subgroup 10	Family XVIII	Subfamily A	SrCyp73
Subgroup 11	Family XIX	Subfamily A	SrCyp48
Subgroup 12	Family XX	Subfamily A	SrCyp51
Subgroup 13	Family XXI	Subfamily A	SrCyp22

3.2. Gene structure

The structure of the genes was elucidated as described in part 2.7.

3.2.1. Introns

The structure of the *CYP*s in *S. robusta* was elucidated by different tools as described in section 2.7 and revealed that *CYP*s both with and without introns existed. As seen in Figure 3.2.1., introns in position A are roughly in the first third of the gene, in position B in the middle of the gene and introns in position C are in the last third part of the gene.

Figure 3.2.1 Position of introns. Introns in position A are roughly in the first third of the gene, in position B in the middle of the gene and introns in position C are in the last third part of the gene.

Of the 68 *CYP*s described in *S. robusta*, 40 are without introns, 14 have one intron, 6 have two and 2 genes have three introns. Most of the genes are approximately between 1400 and 2300 bp, and the introns are all in the range of 15 to 130 bp. For some of the genes it was not possible to determine whether they contain introns or not, as complete genomic sequences was not obtained from the sequencing. Further, not all genes are yet complete, and neither the length of the coding region, nor any possible introns could therefore be determined. These genes are *SrCyp32, SrCyp34, SrCyp43, SrCyp44*, *SrCyp49* and *SrCyp57b*.

The genes without introns are listed in Table 3.2.1 below, along with the length of the coding region.

Table 3.2.1 *CYP*s without introns in *S. robusta*

Gene	Length of coding region	Gene	Length of coding region
Cyp51B1	1473 bp	*SrCyp46*	1821 bp
Cyp51B2	1473 bp	*SrCyp47*	1845 bp
Cyp86A1	1539 bp	*SrCyp48*	1578 bp
Cyp86A2	1557 bp	*SrCyp52*	1494 bp
Cyp86A3	1557 bp	*SrCyp53*	1521 bp
Cyp86A4	1620 bp	*SrCyp54*	1566 bp
Cyp86A5	1530 bp	*SrCyp55*	1545 bp
Cyp86A6	1590 bp	*SrCyp57*	1539 bp
Cyp97B1	1608 bp	*SrCyp61*	1602 bp
SrCyp8	1578 bp	*SrCyp62*	1653 bp
SrCyp11	1581 bp	*SrCyp63*	1647 bp
SrCyp16	1581 bp	*SrCyp64*	1656 bp
SrCyp17	1818 bp	*SrCyp65*	1602 bp
SrCyp18	1632 bp	*SrCyp67*	1587 bp
SrCyp20	1599 bp	*SrCyp68*	1647 bp
SrCyp21	1428 bp	*SrCyp70*	1650 bp
SrCyp22	1833 bp	*SrCyp72*	1401 bp
SrCyp28	1842 bp	*SrCyp73*	1548 bp
SrCyp30	1518 bp	*SrCyp74*	1626 bp
SrCyp36	1443 bp	*SrCyp76*	1518 bp

Listed in Table 3.2.2 below are the 14 *CYP*s with one intron, as well as the size of the introns, and the approximate position according to Figure 3.2.1 above. The introns are positioned in all three thirds of the genes, and are between 15 and 119 bp long.

Table 3.2.2 *CYP*s with one intron in *S. robusta*

Gene	Introns	Size of intron	Position	Length of coding region
Cyp86B1	1	90 bp	A	1773 bp
SrCyp7	1	104 bp	C	1680 bp
SrCyp12	1	89 bp	B	1569 bp
SrCyp13	1	87 bp	B	1614 bp
SrCyp19	*1*	83 bp	B	1602 bp
SrCyp23	1	114 bp	A	1641 bp
SrCyp24	1	114 bp	A	1647 bp
SrCyp25	1	114 bp	A	1629 bp
SrCyp29	1	84 bp	C	1620 bp
SrCyp31	1	78 bp	A	1581 bp
SrCyp39	1	15 bp	B	1776 bp
SrCyp41	1	84 bp	C	1635 bp
SrCyp42	1	77 bp	B	1602 bp
SrCyp56	1	119 bp	B	1527 bp

The six *CYP*s with two introns are listed in Table 3.2.3 below. The introns are positioned in all three thirds of the genes, and are between 18 and 113 bp long. The sizes of the introns are given after position, respectively.

Table 3.2.3 *CYP*s with two introns in *S. robusta*

Gene	Introns	Size of introns	Position	Length of coding region
Cyp19A1	2	94 bp & 87 bp	B & B	1437 bp
SrCyp51A1	2	93 bp & 94 bp	A & C	1542 bp
SrCyp10	2	85 bp & 71 bp	B & B	1359 bp
SrCyp33	2	86 bp & 92 bp	A & B	1524 bp
SrCyp40	2	84 bp &18 bp	B & C	1596 bp
SrCyp51	2	59 bp & 113 bp	B & B	1494 bp

A few *CYP*s contained as much as three introns, and are listed in Table 3.2.4 below. All introns are positioned in positions B or C and are between 85 and 126 bp. The sizes of the introns are given after position, respectively.

Table 3.2.4 *CYP*s with three introns in *S. robusta*

Gene	Introns	Size of introns	Position	Length of coding region
Cyp97A1	3	126 bp, 88 bp & 85 bp	B, C & C	2262 bp
SrCyp35	3	96 bp, 113 bp & 113 bp	B, B & C	1701 bp

3.2.2. Putative introns in Cyp19A1

The predicted introns in *Cyp19A1* (initially named *Cyp50*) were confirmed by the PCR-analysis and subsequent agarose gel. If the predicted introns existed, the products should be approximately 130 bp for the primer pair *Cyp19A1_1* and approximately 100 bp for the primer pair *Cyp19A1_2*. If the predicted introns instead were coding regions, the products should be approximately 230 and 190 bp, respectively. As seen in Figure 3.2.2, the PCR-products of primer pair *Cyp19A1_2* both show bands at approximately 100 bp, whereas one of the PCR-products of primer pair *Cyp19A1_1* show a weak band of a somewhat higher bp. One PCR-product with *Cyp19A1_1* did not show any bands. The slightly uneven bands are probably caused by the voltage which might have been applied somewhat uneven.

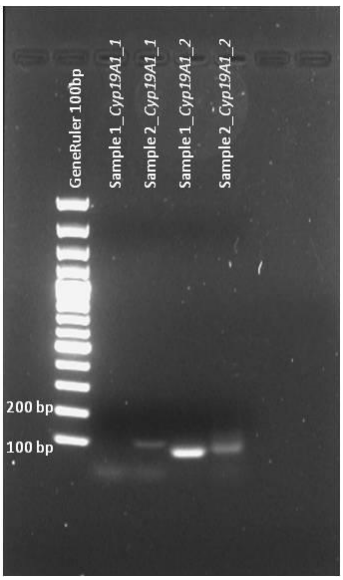


Figure 3.2.2 1 % agarose gel stained with GelRed. Four samples analyzed with two different primer pairs (*Cyp19A1_1*) and (*Cyp19A1_2*) are shown. To the far right is the GeneRuler 100 bp ladder with one band for each 100 bp.

Further analysis was performed on the three PCR-products with visible bonds. Cloning and subsequent sequencing was attempted to further confirm the predicted introns, and to obtain a cDNA sequence of this particular area. But as confirmed by RT-qPCR-analysis, this gene is particular low expressed in the conditions *S. robusta* was grown at in this trial, and attempt to clone the PCR-product and sequence it was not successful. New cDNA sequences of *S. robusta* from the NCBI Sequence Read Archive were later obtained and the two intron-exon junctions for *Cyp19A1* were verified.

3.3. Transcriptional analysis

As described in part 2.2, experimental studies were done to reveal any differential transcriptional regulation of the *CYP*s exposed to different environmental conditions. RT-qPCR was used to determine the gene expression level as described in part 2.4, and the data analyzed as described in part 2.6. The results are presented as a whole in Appendix A3 through A5, but a summary of the results is given in this part. For the genes *Cyp51B1* and *Cyp51B2*, one primer pair recognizing both genes were designed, thus the expression data does not distinguish between these two, but give a joint result. Seven of the 27 genes analyzed did not show a significant response to any of the experimental conditions; *Cyp19A1, Cyp51A1, Cyp86A6, SrCyp13_1, SrCyp16, SrCyp52* and *SrCyp73*.

After 2 hours in 4°C, ten genes showed a significant transcriptional response, nine in which were down-regulated and one up-regulated compared to the samples cultivated in control conditions. The results are presented in Table 3.3.1 below and are calculated by One-Way ANOVA. The ratio represents the fold change between the sample exposed to experimental conditions and the control sample, where the control is given a value of 1.000. A ratio of >1.000 will thus represent a sample being up-regulated, whereas a ratio of <1.000 means that the sample is down-regulated. Also presented is the 95% confidence interval (C.I.) for the ratio, as well as an indication of the significance.

Table 3.3.1 Genes showing transcriptional response after 2 hours in 4°C.

Target	Ratio	95% C.I. of the ratio	Significant (p <0,05)	Result
SrCyp22	0.044	0.017 - 0.115	Yes	DOWN
SrCyp10	0.101	0.053 - 0.194	Yes	DOWN
SrCyp44	0.142	0.040 - 0.508	Yes	DOWN
SrCyp46	0.210	0.055 - 0.799	Yes	DOWN
Cyp86A3	0.242	0.092 - 0.633	Yes	DOWN
Cyp86A2	0.268	0.104 - 0.693	Yes	DOWN
SrCyp17	0.269	0.120 - 0.602	Yes	DOWN
SrCyp18	0.359	0.149 - 0.865	Yes	DOWN
Cyp51B1+2	0.363	0.175 - 0.751	Yes	DOWN
SrCyp21	7.095	2.552 - 19.723	Yes	UP

After 24 hours at 4°C, thirteen genes showed a significant transcriptional response compared to those cultivated in control conditions, and are presented in Table 3.3.2 below. Six of the genes were down-regulated, whereas seven were up-regulated.

Table 3.3.2 Genes showing transcriptional response after 24 hours in 4°C.

Target	Ratio	95% C.I. of the ratio	Significant (p <0,05)	Result
SrCyp10	0.201	0.105 - 0.386	Yes	DOWN
SrCyp18	0.212	0.088 - 0.512	Yes	DOWN
SrCyp54	0.219	0.069 - 0.692	Yes	DOWN
Cyp97A1	0.250	0.119 - 0.526	Yes	DOWN
Cyp51B1+2	0.312	0.151 - 0.646	Yes	DOWN
SrCyp13_2	0.448	0.219 - 0.916	Yes	DOWN
SrCyp28	2.849	1.533 - 5.295	Yes	UP
SrCyp20	3.675	1.360 - 9.930	Yes	UP
SrCyp70	3.775	1.031 - 13.831	Yes	UP
SrCyp12_1	4.332	1.886 - 9.954	Yes	UP
SrCyp12_2	4.450	1.993 - 9.937	Yes	UP
SrCyp23	5.717	1.821 - 17.952	Yes	UP
SrCyp21	38.169	13.731 - 106.104	Yes	UP

Eight genes showed response to the day/night cycle when compared to the samples cultivated in control conditions, and are listed in Table 3.3.3 below. Three genes were down-regulated, and five genes were up-regulated.

Table 3.3.3 Genes showing transcriptional response to day/night cycle.

Target	Ratio	95% C.I. of the ratio	Significant (p <0,05)	Result
Cyp97B1	0.048	0.019 - 0.124	Yes	DOWN
SrCyp22	0.110	0.042 - 0.286	Yes	DOWN
Cyp97A1	0.143	0.068 - 0.301	Yes	DOWN
SrCyp21	2.929	1.054 - 8.141	Yes	UP
SrCyp20	3.369	1.247 - 9.104	Yes	UP
SrCyp54	3.373	1.067 - 10.664	Yes	UP
SrCyp18	3.518	1.459 - 8.483	Yes	UP
SrCyp23	8.187	2.607 - 25.705	Yes	UP

Several of the genes showing a rapid response to the cold conditions after two hours, shows an adaptive response after 24 hours in 4°C, and are presented in Table 3.3.4 below.

Table 3.3.4 Genes showing an adaptive response to cold conditions after 24 hours.

Target	Condition	Ratio	95% C.I. of the ratio	Significant (p <0,05)	Result
SrCyp22	Const_0h_ctrl				
	Const_2h	0.044	0.017 - 0.115	Yes	DOWN
	Const_24h	0.540	0.208 - 1.401	No	
SrCyp17	Const_0h_ctrl				
	Const_2h	0.269	0.120 - 0.602	Yes	DOWN
	Const_24h	2.062	0.923 - 4.609	No	
Cyp86A3	Const_0h_ctrl				
	Const_2h	0.242	0.092 - 0.633	Yes	DOWN
	Const_24h	0.460	0.176 - 1.205	No	
Cyp86A2	Const_0h_ctrl				
	Const_2h	0.268	0.104 - 0.693	Yes	DOWN
	Const_24h	0.498	0.192 - 1.288	No	
SrCyp44	Const_0h_ctrl				
	Const_2h	0.142	0.040 - 0.508	Yes	DOWN
	Const_24h	0.745	0.209 - 2.659	No	
SrCyp46	Const_0h_ctrl				
	Const_2h	0.210	0.055 - 0.799	Yes	DOWN
	Const_24h	0.430	0.113 - 1.634	No	

This can also be visualized in the bar charts for the given genes, as seen for *SrCyp44* in below. Bar charts for all genes are presented in Appendix A5.

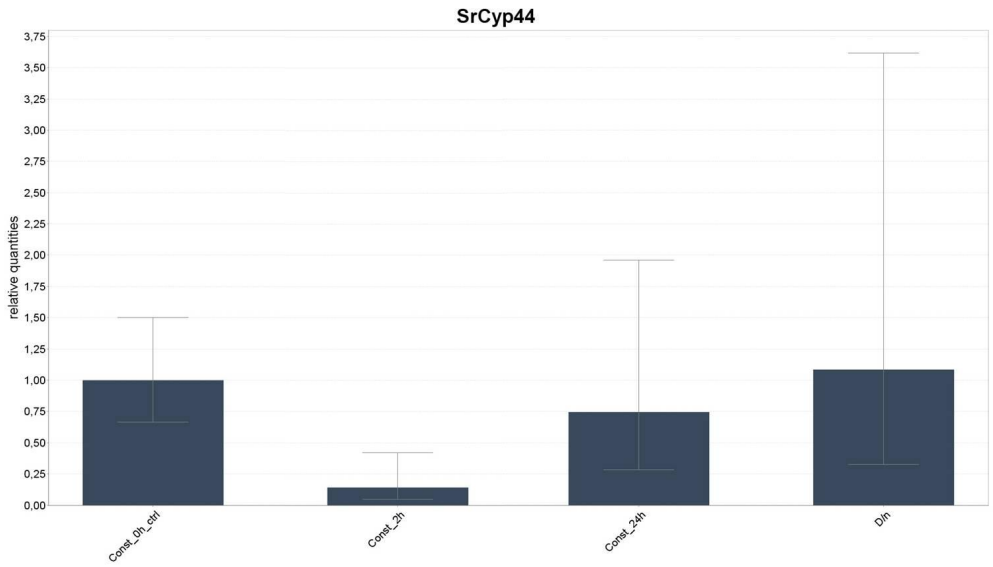


Figure 3.3.1 Linear bar chart of *Cyp44*, one of the genes showing an adaptive response to cold conditions. On the Y-axis is given the relative quantities of the genes when the quantity of the control sample is set to 1.000.

Five different reference genes were chosen according to previously experience form other diatoms; *qSrAtpE*, *qSrAtpB*, *qSrPbsO, qSrVps35* and *qSrMFS*. *qSrPbsO* was not suitable as the results from the RT-qPCR analysis gave a Cq-value >40 (results not shown). The remaining four candidate reference genes were evaluated by the help of the geNorm^{Plus} module of qBase^{Plus}. As seen in Figure 3.3.1 below, all four genes showed a high M-value, and geNorm^{Plus} did in fact recommend that other reference genes would be evaluated. Due to time limititations, though, this was not done, and the two reference genes with lowest M-value were chosen; *qSrMFS* and *qSrVps35*.

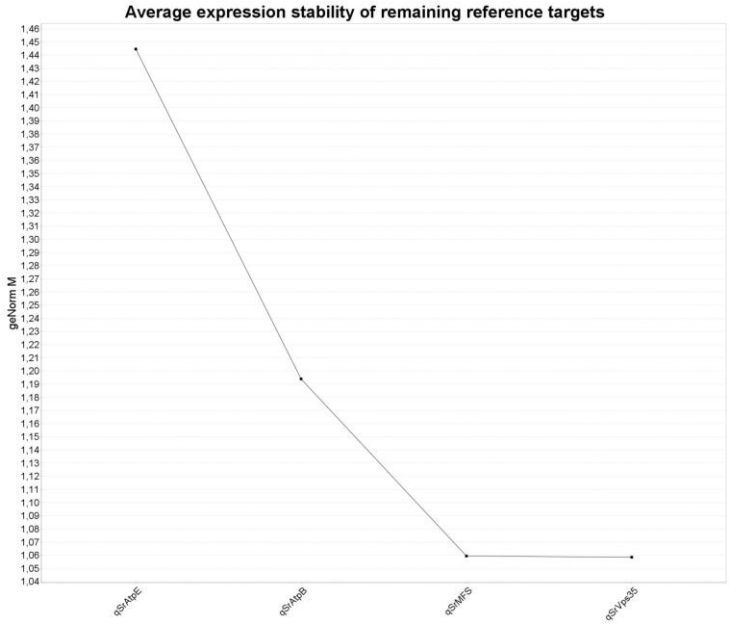


Figure 3.3.1 geNormPlus evaluation of candidate reference genes. The two reference genes with lowest M-value were chosen; *qSrMFS* and *qSrVps35*.

3.4. Comparison between qBase^{Plus} and REST2009

A comparison was made between the two software programs qBase^{Plus} and REST2009 for analyzing data from RT-qPCR. A total of 26 genes were analyzed, representing 15 of the initial 17 subgroups created, and 17 of the 25 CYP-families found in *S. robusta*. For the genes *Cyp51B1* and *Cyp51B2*, one primer pair recognizing both genes was designed, and for the genes *SrCyp12* and *SrCyp13*, two primer pairs for each gene were designed.

The results calculated by qBase^{Plus} shows the ratio, which represents the fold change between the sample exposed to experimental conditions and the control sample, where the control is given a value of 1.000. A ratio of >1.000 will thus represent a sample being up-regulated, whereas a ratio of <1.000 means that the sample is down-regulated. Also presented is the 95% confidence interval (C.I.) for the ratio, as well as an indication of the significance.

The results calculated by REST2009 shows the expression, which is basically the same as the ratio given from qBase^{Plus}. The P(H1)-value given in the results table represents the probability (p-value) of the alternate hypothesis. Samples generating a p-value < 0.05 are considered as significantly up- or down-regulated compared to the control. The 95 % confidence interval (C.I.) is also presented, as well as the standard error.

The results are presented in Table 3.4.1 below. Red marking indicates different results between qBase^{Plus} and REST2009. The results from qBase^{Plus} are calculated by One-Way ANOVA.

Table 3.4.1 Statistic results from qBasePlus and REST2009.

		qBase^{Plus}					REST2009				
Target	Condition	p	r2	Ratio	95% C.I. of the ratio	Significant (p <0,05)	Expression	Std. Error	95% C.I.	P(H1)	Result
Cyp19A1	Const_0h_ctrl	1.90E-01	0.338								
	Const_2h			0.376	0.102 - 1.382	No	0.383	0.190 - 0.720	0.149 - 1.318	0.066	
	Const_24h			0.917	0.249 - 3.371	No	1.006	0.461 - 2.627	0.250 - 3.092	0.994	
	D/n			0.656	0.179 - 2.413	No	0.679	0.371 - 1.082	0.286 - 1.928	0.257	
Cyp51A1	Const_0h_ctrl	4.30E-01	0.205								
	Const_2h			1.137	0.536 - 2.413	No	1.221	0.792 - 1.797	0.769 - 2.048	0.394	
	Const_24h			1.148	0.541 - 2.435	No	1.280	0.886 - 2.507	0.824 - 3.010	0.493	
	D/n			1.538	0.725 - 3.263	No	1.571	1.110 - 2.008	0.947 - 2.289	0.080	
Cyp51B1+2	Const_0h_ctrl	3.40E-04	0.813								
	Const_2h			0.363	0.175 - 0.751	Yes	0.342	0.258 - 0.447	0.211 - 0.533	0.002	DOWN
	Const_24h			0.312	0.151 - 0.646	Yes	0.346	0.213 - 0.801	0.185 - 1.073	0.032	DOWN
	D/n			1.334	0.644 - 2.764	No	1.299	0.999 - 1.780	0.859 - 2.296	0.170	
Cyp86A2	Const_0h_ctrl	4.06E-04	0.795								
	Const_2h			0.268	0.104 - 0.693	Yes	0.263	0.158 - 0.563	0.124 - 0.811	0.023	DOWN
	Const_24h			0.498	0.192 - 1.288	No	0.556	0.346 - 1.055	0.239 - 1.521	0.123	
	D/n			2.126	0.822 - 5.499	No	2.056	1.400 - 2.992	1.040 - 3.757	0.048	UP
Cyp86A3	Const_0h_ctrl	3.68E-04	0.801								
	Const_2h			0.242	0.092 - 0.633	Yes	0.262	0.170 - 0.488	0.123 - 0.657	0.000	DOWN
	Const_24h			0.460	0.176 - 1.205	No	0.518	0.334 - 0.806	0.241 - 0.997	0.041	DOWN
	D/n			2.012	0.769 - 5.266	No	2.078	1.203 - 3.146	0.868 - 4.040	0.087	
Cyp86A6	Const_0h_ctrl	3.27E-01	0.253								
	Const_2h			1.218	0.416 - 3.567	No	1.317	0.600 - 2.474	0.428 - 5.368	0.563	
	Const_24h			0.862	0.294 - 2.523	No	0.970	0.547 - 1.692	0.357 - 2.338	0.931	
	D/n			1.717	0.586 - 5.027	No	1.773	0.892 - 2.589	0.737 - 5.334	0.175	

Cyp97A1	Const_0h_ctrl	5.35E-05	0.873								
	Const_2h			0.796	0.378 - 1.677	No	0.774	0.662 - 0.969	0.516 - 1.040	0.070	
	Const_24h			0.250	0.119 - 0.526	Yes	0.243	0.116 - 0.419	0.090 - 0.490	0.011	DOWN
	D/n			0.143	0.068 - 0.301	Yes	0.146	0.103 - 0.248	0.098 - 0.325	0.003	DOWN
Cyp97B1	Const_0h_ctrl	1.12E-05	0.912								
	Const_2h			0.441	0.171 - 1.137	No	0.451	0.386 - 0.559	0.298 - 0.611	0.000	DOWN
	Const_24h			1.086	0.421 - 2.800	No	1.258	0.713 - 2.838	0.534 - 3.594	0.624	
	D/n			0.048	0.019 - 0.124	Yes	0.047	0.029 - 0.084	0.022 - 0.094	0.001	DOWN
SrCyp10	Const_0h_ctrl	1.12E-05	0.920								
	Const_2h			0.101	0.053 - 0.194	Yes	0.109	0.070 - 0.162	0.053 - 0.233	0.012	DOWN
	Const_24h			0.201	0.105 - 0.386	Yes	0.227	0.118 - 0.388	0.109 - 0.601	0.018	DOWN
	D/n			0.646	0.337 - 1.240	No	0.668	0.455 - 0.994	0.434 - 1.232	0.089	
SrCyp12_1	Const_0h_ctrl	5.74E-04	0.780								
	Const_2h			0.806	0.351 - 1.853	No	0.886	0.670 - 1.165	0.576 - 1.447	0.452	
	Const_24h			4.332	1.886 - 9.954	Yes	5.288	2.933 - 10.860	2.404 - 14.228	0.011	UP
	D/n			1.495	0.651 - 3.435	No	1.554	0.905 - 2.468	0.682 - 3.207	0.141	
SrCyp12_2	Const_0h_ctrl	3.68E-04	0.803								
	Const_2h			0.779	0.349 - 1.740	No	0.724	0.515 - 1.022	0.459 - 1.508	0.169	
	Const_24h			4.450	1.993 - 9.937	Yes	4.744	2.681 - 8.358	2.359 - 13.151	0.024	UP
	D/n			1.314	0.589 - 2.935	No	1.347	0.787 - 2.289	0.632 - 3.457	0.374	
SrCyp13_1	Const_0h_ctrl	2.59E-01	0.292								
	Const_2h			0.806	0.333 - 1.953	No	0.836	0.547 - 1.293	0.384 - 1.875	0.487	
	Const_24h			0.521	0.215 - 1.263	No	0.56	0.322 - 0.984	0.253 - 1.640	0.114	
	D/n			0.749	0.309 - 1.814	No	0.727	0.433 - 1.165	0.361 - 1.975	0.281	
SrCyp13_2	Const_0h_ctrl	3.49E-02	0.517								
	Const_2h			0.778	0.381 - 1.591	No	0.862	0.577 - 1.112	0.539 - 1.761	0.476	
	Const_24h			0.448	0.219 - 0.916	Yes	0.535	0.326 - 0.832	0.244 - 1.251	0.046	DOWN
	D/n			0.566	0.277 - 1.156	No	0.609	0.404 - 0.911	0.342 - 1.437	0.068	

SrCyp16	Const_0h_ctrl	5.37E-01	0.160								
	Const_2h			1.151	0.405 - 3.270	No	1.211	0.862 - 1.847	0.666 - 2.374	0.433	
	Const_24h			0.695	0.245 - 1.973	No	0.784	0.388 - 1.591	0.327 - 2.550	0.549	
	D/n			0.851	0.300 - 2.418	No	0.929	0.487 - 1.684	0.403 - 2.597	0.833	
SrCyp17	Const_0h_ctrl	7.83E-05	0.860								
	Const_2h			0.269	0.120 - 0.602	Yes	0.283	0.226 - 0.342	0.194 - 0.365	0.003	DOWN
	Const_24h			2.062	0.923 - 4.609	No	2.127	1.273 - 4.011	1.091 - 4.791	0.017	UP
	D/n			1.993	0.892 - 4.455	No	1.986	1.229 - 2.818	1.053 - 3.366	0.029	UP
SrCyp18	Const_0h_ctrl	1.80E-05	0.897								
	Const_2h			0.359	0.149 - 0.865	Yes	0.391	0.287 - 0.553	0.211 - 0.682	0.008	DOWN
	Const_24h			0.212	0.088 - 0.512	Yes	0.247	0.158 - 0.403	0.109 - 0.510	0.002	DOWN
	D/n			3.518	1.459 - 8.483	Yes	3.66	2.063 - 5.678	1.306 - 6.983	0.027	UP
SrCyp20	Const_0h_ctrl	9.33E-04	0.756								
	Const_2h			0.718	0.266 - 1.941	No	0.773	0.547 - 1.025	0.400 - 1.302	0.421	
	Const_24h			3.675	1.360 - 9.930	Yes	4.116	2.283 - 9.172	1.658 - 12.780	0.012	UP
	D/n			3.369	1.247 - 9.104	Yes	3.301	1.876 - 7.435	1.458 - 10.360	0.029	UP
SrCyp21	Const_0h_ctrl	1.12E-05	0.909								
	Const_2h			7.095	2.552 - 19.723	Yes	7.497	4.202 - 11.852	2.833 - 18.407	0.019	UP
	Const_24h			38.169	13.731 - 106.104	Yes	42.382	20.695 - 97.147	14.386 - 172.696	0.019	UP
	D/n			2.929	1.054 - 8.141	Yes	2.962	1.951 - 4.377	1.311 - 6.591	0.008	UP
SrCyp22	Const_0h_ctrl	1.12E-05	0.908								
	Const_2h			0.044	0.017 - 0.115	Yes	0.043	0.032 - 0.057	0.029 - 0.064	0.000	DOWN
	Const_24h			0.540	0.208 - 1.401	No	0.577	0.285 - 1.201	0.215 - 1.611	0.261	
	D/n			0.110	0.042 - 0.286	Yes	0.109	0.067 - 0.176	0.053 - 0.232	0.000	DOWN
SrCyp23	Const_0h_ctrl	7.68E-04	0.766								
	Const_2h			1.742	0.555 - 5.468	No	1.796	0.980 - 2.918	0.757 - 3.770	0.100	
	Const_24h			5.717	1.821 - 17.952	Yes	5.574	2.927 - 12.768	2.427 - 17.164	0.012	UP
	D/n			8.187	2.607 - 25.705	Yes	7.281	3.257 - 11.284	2.517 - 13.436	0.010	UP

SrCyp28	Const_0h_ctrl	2.39E-04	0.828								
	Const_2h			0.643	0.346 - 1.195	No	0.615	0.472 - 0.796	0.419 - 0.907	0.006	DOWN
	Const_24h			2.849	1.533 - 5.295	Yes	3.159	1.892 - 6.135	1.585 - 8.829	0.020	UP
	D/n			0.867	0.467 - 1.612	No	0.865	0.672 - 1.139	0.570 - 1.275	0.343	
SrCyp44	Const_0h_ctrl	2.32E-03	0.712								
	Const_2h			0.142	0.040 - 0.508	Yes	0.151	0.097 - 0.310	0.071 - 0.490	0.005	DOWN
	Const_24h			0.745	0.209 - 2.659	No	0.866	0.475 - 1.675	0.344 - 2.105	0.713	
	D/n			1.085	0.304 - 3.872	No	1.068	0.436 - 2.037	0.288 - 2.558	0.930	
SrCyp46	Const_0h_ctrl	1.49E-02	0.595								
	Const_2h			0.210	0.055 - 0.799	Yes	0.224	0.107 - 0.578	0.073 - 0.824	0.000	DOWN
	Const_24h			0.430	0.113 - 1.634	No	0.499	0.286 - 1.186	0.210 - 1.692	0.125	
	D/n			1.068	0.281 - 4.057	No	1.087	0.435 - 1.855	0.297 - 2.058	0.854	
SrCyp52	Const_0h_ctrl	9.16E-03	0.631								
	Const_2h			0.523	0.265 - 1.032	No	0.573	0.418 - 0.726	0.366 - 0.887	0.002	DOWN
	Const_24h			0.936	0.475 - 1.846	No	0.989	0.657 - 1.317	0.550 - 1.536	0.890	
	D/n			1.458	0.739 - 2.875	No	1.461	1.008 - 2.570	0.937 - 3.163	0.154	
SrCyp54	Const_0h_ctrl	3.68E-04	0.806								
	Const_2h			0.994	0.314 - 3.141	No	1.004	0.490 - 1.783	0.377 - 3.071	0.969	
	Const_24h			0.219	0.069 - 0.692	Yes	0.238	0.115 - 0.652	0.050 - 0.744	0.014	DOWN
	D/n			3.373	1.067 - 10.664	Yes	3.484	1.844 - 5.181	1.542 - 9.058	0.019	UP
SrCyp70	Const_0h_ctrl	2.37E-02	0.553								
	Const_2h			1.023	0.279 - 3.747	No	1.062	0.666 - 1.658	0.425 - 1.900	0.792	
	Const_24h			3.775	1.031 - 13.831	Yes	4.124	2.226 - 7.556	1.802 - 11.537	0.015	UP
	D/n			0.869	0.237 - 3.184	No	0.867	0.330 - 2.093	0.210 - 3.147	0.810	
SrCyp73	Const_0h_ctrl	1.70E-02	0.581								
	Const_2h			0.925	0.507 - 1.688	No	1.046	0.763 - 1.332	0.655 - 1.727	0.763	
	Const_24h			1.800	0.986 - 3.285	No	2.218	1.367 - 4.075	1.209 - 5.507	0.011	UP
	D/n			1.635	0.896 - 2.984	No	1.777	1.239 - 2.169	1.087 - 3.052	0.027	UP

3.5. Protein characterization

A putative prediction of the localization of the different CYPs was done by online tools, as described in section 2.8. The results are given in Appendix A6, and were in most cases inconclusive, especially in predicting a specific localization. Both plant and non-plant networks were used in the prediction. Most of the CYPs were predicted to be localized in the ER, and only three were predicted to be localized in the mitochondria; Cyp86A4, SrCyp51 and SrCyp52.

Of the 68 CYPs, SecretomeP predicted a signal peptide in 34 of them, suggesting that these might be classical secreted proteins. For a total of 21 CYPs, SecretomeP did not predict a signal peptide, but an NN-score of >0.500 suggest that these might be secreted proteins in a non-classical pathway. Only nine CYPs were not predicted to be secreted, in either a classical or non-classical pathway.

These results are not reliable, though, as the results were inconclusive in all cases, and needs verification. The severe uncertainty and inconclusiveness related to these analyses are discussed in more detail below.

4. DISCUSSION

4.1. Families and subfamilies of CYPs in *S. robusta*

The CYPs of *S. robusta* were classified in families and subfamilies according to phylogenetic analysis and protein comparisons. A total of 25 CYP-families were identified, which could be grouped into 17 sub-groups. Some of these families showed a clear relationship with known CYP-families in other organisms, and could thus be placed in these families and a suitable nomenclature was given. Other CYP-families did not show any significant relationship with known CYP-families, and could not be given any nomenclature related to known CYPs. One of the CYPs, SrCyp48, was in fact so special that is was impossible to find even a distant relationship to any other CYPs, both in *S. robusta* or in any other organisms, and this CYP is therefore classified as a novel CYP, so far only found in *S. robusta*.

The phylogenetic analysis revealed that the CYPs of *S. robusta* shared similarities with CYPs of plants and animals, as well as genes from bacteria, other diatoms, green algae and cyanobacteria. This broad diversity of CYPs in *S. robusta* is reflecting the complicated evolutionary origin of the diatoms, is and also a great source of confusion when trying to classify the *CYP*s.

The CYPs were all analyzed by the protein characterization tools WoLF PSORT, TargetP and SecretomeP available online, but this gave no conclusive results. Several of the CYPs where characterized as secreted in a classical or non-classical pathway, but given that no CYPs so far are shown to be secreted, this is very unlikely. The NN-score of >0.500 which contributes to this prediction is an indication of a good leader sequence present in these proteins. They are most likely localized in ER and transported via the Golgi apparatus to other organelles. The majority of CYPs were predicted to be ER-localized, and as most eukaryote CYPs are Class II CYPs, this might be accurate predictions. A few CYPs were predicted to be located to the mitochondria, suggesting that these might be Class I CYPs. Further work needs to be done in order to elucidate the localizations of the CYPs in *S. robusta*, as discussed in chapter 6.

The *CYP*s of *S. robusta* are all relatively low expressed, which suggest that they have quite a tight regulation. Given that this group of proteins is significantly larger in *S. robusta* than in other, related diatoms so far sequenced, it is reasonable to think that the *CYP*s have quite specific functions. Compared to *P. tricornutum*, which was found to have only 8 CYP-coding genes, *S. robusta* show a much more complex life cycle, which can explain both the expansion of this superfamiliy in *S. robusta*, as well as the low expression level. As *S. robusta* have a life cycle involving both vegetative multiplication and sexual reproduction, some of the *CYP*s might be related to specific stages in the life cycle, and thus show a higher expression level on specific time points. As *S. robusta* has the ability to move, some CYPs might also be coupled to primitive responses such as chemotaxis and flight-responses.

After two hours in 4 °C, ten genes showed significant responses; nine genes were down-regulated and one was up-regulated. After 24 hours in 4 °C, thirteen genes showed significant responses; six genes were down-regulated and seven genes were up-regulated. Some of the genes that were down-regulated after two hours showed a possible adaptive response to the cold conditions after 24 hours. These responses indicate that some of the CYPs in *S. robusta* are coupled to processes involved in temperature-responses.

The day/night cycle gave significant responses in eight genes; three were down-regulated, and five was up-regulated. This indicates that some of the CYPs in *S. robusta* are coupled to responses involved in light-/dark-responses and display circadian regulation. This is as expected, as diatoms can experience a varity in light conditions according to depth and time of day and season. When experiencing less amounts of light, they have to adapt by enhancing the photosyntethic apparatus in such a way that they harvest photons more efficient. This was shown by Nymark *et al.* (2009) when exposing *P. tricornutum* to low light and high light conditions, and similar responses can be expected to be seen in *S. robusta*. The diatoms were harvested 3.5 hours into the light cycle after the final night-period. If the harvesting was done at the end of the dark period, the responses might be somewhat different.

4.1.1. CYP19 Aromatase

The *Cyp19A1*-gene, originally named *Cyp50,* has been of particular interest while analyzing the *CYP*s of *S. robusta*. A surprising relationship appeared between this gene and the aromatase-encoding genes in other organisms, from *Crysemys picta* (turtle) to *Homo sapiens*. In humans, this CYP is known as the classic "aromatase", an enzyme that oxidizes the androgens testosterone and androstandione to the estrogens 17β-estradial and estrone, respectively (de Montellano, 2005,p. 450 - 452). Inhibitors of CYP19A1 are of interest in treatment of estrogen-dependent cancers in humans, as the *CYP19A1* gene has been shown to be expressed in different tumors. The expression of aromatase in mammals plays an important role in regulation of the reproductive cycle in females (Stocco, 2012). In the fish species *Odontesthes bonariensis* it plays an important role in the temperature-dependent sex-determination and gonadal sex differentiation period (Karube *et al.*, 2007). The existence of CYP19A1 in *S. robusta* could in theory be linked to the switching between the vegetative and sexual reproduction, which could also be an explanation of why this gene is so low expressed.

Initial analysis showed low expression of this particular gene in *S. robusta*, corresponding to a low number of sequence reads from the cDNA sequencing. Sequence analysis revealed two putative introns, and cloning and subsequent restriction enzyme analysis and sequencing was attemped to confirm this. This gene's low expression level, combined with problems to clone the PCR amplified exon-intron junctions made it difficult to analyze it. However, analysis of high throughput Illumina cDNA sequencing data confirmed the splice sites. The coding sequence is 1437 bp long, with two introns of 94 and 87 bp, respectively, both located in the middle third of the gene. Transcriptional analysis performed showed that *Cyp19A1* is very low expressed in *S. robusta*, both in normal conditions and when exposed to colder and darker conditions. A slight down-regulation was possibly seen after two hours in 4°C, but as the uncertainty of such high Cq-values (32 – 35) is high, this regulation was not significant.

Analyzing Cyp19A1 to possibly predict a localization gave no conclusive results. SecretomeP predicted this protein not to be secretory, but WoLF PSORT showed some similarity to secreted proteins, and the results do not give any conclusion.

The findings in this experimental trial do not reveal anything about the possible functions *Cyp19A1* have in *S. robusta.* Further analysis is needed to explore what type of environmental conditions that regulate this particular gene and what biochemical / molecular function it has in *S. robusta*.

4.1.2. CYP51 Sterol demetylase

Three genes in *S. robusta* were shown to be members of the *CYP51*-family. Initially, these genes were named *SrCyp9a*, *SrCyp9b* and *SrCyp14*, after discovery, with the new nomenclature *Cyp51B1*, *Cyp51B2* and *Cyp51A1* respectively. This family was originally discovered in yeast, but has later been found throughout eukaryotes, with some exceptions (de Montellano, 2005,p. 590). *CYP51* encodes the sterol 14-α-demetylase, and substrates differ some between kingdoms, with C4-dimethyl in fungi and animals and C4-methyl sterol obtusifoliol in plants.

Two of the genes, *Cyp51B1* and *Cyp51B2*, in *S. robusta* are both 1473 bp long and have no introns, whereas the last, *Cyp51A1*, is 1542 bp long, with two introns of 93 and 94 bp, located in the first and last third of the gene, respectively. *Cyp51B1* and *Cyp51B2* were analyzed with one primer pair recognizing both genes, and the expression data thus represents both genes. All three genes are medium high expressed in all conditions. *Cyp51B1* and *Cyp51B2* are significantly down-regulated after both 2 hours and 24 hours in 4°C and are expressed about three times lower than in the control conditions. *Cyp51A1*, on the other hand, did not show any significant response to any of the experimental conditions.

When analyzing the protein sequences, Cyp51B1 and Cyp51B2 were predicted to be classical secreted proteins, whereas Cyp51A1 was predicted to be non-classical. Further, they were all predicted to be membrane-bound on ER, which suggest that they might be Class II CYPs.

4.1.3. CYP86 Fatty acid hydroxylase

One CYP-family in *S. robusta* showed quite a close relationship with the CYP86-family in *A. thaliana*, more specifically *CYP86A1*. *CYP86A1* show high homology with the animal CYP families 4 and 52, and was a new CYP-family in plants when discovered, with only a distant relationship to the animal CYPs. *Cyp86A1* encodes a ω-hydroxylase with a broad range of

substrates (Kandel *et al.*, 2006). The ω-hydroxylases catalyze the hydroxylation of the terminal methyl of aliphatics acids, the ω-position. It has been shown that members of the CYP86-family in *A. thaliana* can be involved in many aspects of plant development, and probably mediate quite diverse biological processes.

Seven genes of *S. robusta* are predicted to be part of this family, and three of them were analyzed in transcriptional analysis. Six of the genes showed a closer relationship than the last. These six, *Cyp1 – 6*, are all without introns and approximately 1550 bp long. The last gene in this family*, Cyp86B1*, is 1773 bp long, and have one intron of 90 bp positioned in the first third of the gene. The three genes analyzed experimentally are *Cyp86A3, Cyp86A2* and *Cyp86A6. Cyp86A6* is medium high expressed, and did not show any significant response to any of the experimental conditions. A slight up-regulation was seen in day/night cycle, but this was not significant. Both *Cyp86A3* and *Cyp86A2* showed a significant down-regulation after two hours in 4°C, but after 24 hours, though, the expression level was increasing, which suggest a possible adaptation to the colder conditions.

4.1.4. CYP97 Carotenoid

Two genes in *S. robusta* show a close relationship with the CYP97-family, and are thus placed in this family. CYP97 is found in, amongst others, plants and cyanobacteria, and Yang *et al.* (2003) also found a CYP97 in the marine diatom *S. costatum*, CYP97E1. Phylogenetic analysis performed by Kim *et al.* (2009) suggest that the CYP97-family is a result of two consecutive gene duplications of ancestral *CYP97* genes that took place before the higher plant/green algae split. Further linage-specific gene duplication seems to have occurred in some organisms, like *Chlamydomonas reinhardtii* and *Osterococcus tauri*. These genes in *S. robusta* also show relationship to the LUT-genes in *P. tricornutum*

The *Cyp97* genes are encoding carotenoid hydroxylases, responsible for hydroxylation of α-carotenes (Kim *et al.*, 2009). Carotenoids are, together with chlorophylls, pigments in the photosynthetic complexes, responsible for absorbing light in oxygenic photosynthesis. Studies have suggested that the enzymes CYP97A3 and CYP97C1 in *A. thaliana* are mainly responsible for catalyzing hydroxylation of the β- and ε-rings of α-carotene. It is thus likely to

assume that the two members of this family in *S. robusta* are involved in light-harvesting and/ or light protection.

Cyp97A1 is 2262 bp long, with three introns of 126 bp, 88 bp and 85 bp, two of which are located in the middle third of the gene, while the last intron is located in the latter third of the gene, respectively. *Cyp97B1* is 1608 bp long and without introns.

Both genes are medium high expressed in the control conditions. *Cyp97A1* showed a significant down-regulation after 24 hours in 4°C, whereas *Cyp97B1* did not show any significant response to the colder conditions. The day/night cycle on the other hand, gave significant responses in both genes. A dramatic down-regulation was shown in both *Cyp97A1* and *Cyp97B1*, which confirms that these genes are most likely coupled to light-harvesting processes and display circadian regulation.

4.1.5. Families I and II

A quite large subgroup of the *CYP*s in *S. robusta* does not shown any relationship with known CYPs, though a relationship is been shown with genes from the diatoms *T. pseudonana, P. tricornutum* and *Fragilariopsis cylindrus*. A total of eight *CYP*s of *S. robusta* are placed in this subgroup, both with and without introns. When analyzing the amino acid similarity, it was shown that this subgroup consists of two families, with *Cyp72* as the most divergent member in a family of its own. The largest family, given the preliminary name Family I, consists of four subfamilies. These two families may be diatom specific CYPs.

Two of the genes, *SrCyp10* and *SrCyp21*, were analyzed experimentally.

SrCyp10 showed a medium low expression level in the control conditions, and were further down-regulated significantly in 4°C after both two hours and 24 hours. A slight increase in expression level was shown after 24 hours compared to after two hours, though, which may suggest that this gene also adapts to the colder conditions after some time.

SrCyp21 was very low expressed in the control conditions, possibly turned off almost completely. It showed a significant up-regulation in all the experimental conditions, though,

especially after 24 hours in 4°C, with a ratio of 38. This dramatic up-regulation in the cold conditions suggests that this gene may be coupled to cold-responses.

4.1.6. Family III

One subgroup created based on phylogenetic relationship was shown to consist of three CYPs, all in the same family and subfamily. This family, named Family III, does not show any close relationship with known CYPs of other organisms. Putative homologues have been found in *T. pseudonana* and *P. tricornutum*, though, and it is reasonable to assume that this family may be a diatom-specific CYP-family.

SrCyp23, SrCyp24 and *SrCyp25* are approximately 1600 bp and all have one intron of 114 bp each, located in the first third of the gene. *SrCyp23* was analyzed experimentally, and was very low, almost turned off, in the control conditions. A significant up-regulation was shown in both day/night conditions and after 24 hours in 4°C. A slight up-regulation was also seen after two hours in colder conditions, suggesting that the response is slow, but starts relatively rapid after low temperature exposure.

Responses to both temperature and light conditions suggest that *SrCyp23* might be involved in light coupled responses and the phylogenetic analysis suggest that these genes probably are distant related to the *Cyp97* carotenoid hydroxylases.

4.1.7. Family IV

Distant related proteins to this family are found in Cyanobacteria and plants. In *S. robusta*, this family, IV, consists of three CYPs all in the same family. Based on amino acid similarity, *SrCyp28* and *SrCyp49* are placed in subfamily A, and *SrCyp29* in subfamily B.

SrCyp28 have no introns and are 1842 bp long, whereas *SrCyp29* has one intron of 84 bp located in the latter third of the gene. The third gene of this subgroup*, SrCyp49*, is still not complete, and the length is thus not known, nor any possible introns.

SrCyp28 was analyzed experimentally, and was quite low expressed in the control conditions. A significant up-regulation was shown after 24 hours in 4°C. No response was

seen after two hours though, so this seems to be a response that is triggered after some time in colder conditions.

4.1.8. Families V and VI

Initially, one subgroup containing genes *SrCyp46, SrCyp47* and *SrCyp31* were created according to phylogenetic relationship. When analyzing the amino acid similarity amongst these three genes, though, it became clear that this subgroup consisted of two families. Family V contains only *SrCyp31*, whereas Family VI consists of *SrCyp46* and *SrCyp47*, both in the same subfamily. These families show a possible distant relationship to yeast CYPs, but might also be novel to diatoms or *S. robusta*.

SrCyp46 and *SrCyp47* are both approximately 1800 bp long and without introns, whereas *SrCyp31* have one intron of 78 bp located in the first third of the gene and is 1581 bp long.

SrCyp46 was analyzed experimentally, and was quite low expressed in the control conditions. After two hours in 4°C a further, significant, down-regulation was shown, and the gene was basically turned off.

4.1.9. Families VII and VIII

The phylogenetic relationship between *SrCyp20* and *SrCyp76* showed that they might be grouped together, but the amino acid similarity revealed that they in fact belong to two different families. *SrCyp20* has a possible homologue in *F. cylindrus*, and both *SrCyp20* and *SrCyp76* both show a possible relationship with the *CYP707*-family in *A. thaliana*, which includes ABA-β-hydroxylases.

SrCyp20 and *SrCyp76* are both without introns, and 1599 bp and 1518 bp in length, respectively. *SrCyp20* was analyzed experimentally, and showed quite a low expression level in the control conditions. A significant up-regulation was seen when diatoms were exposed to day/night cycles and after 24 hours in 4°C, though. The response to the colder conditions does not seem to be rapid, as no response was seen after two hours.

This response seen in both colder and darker conditions suggest that this gene might be coupled to both light- and cold-responses.

4.1.10. Families IX, X, XI and XII

A subgroup of eight *CYP*s was created according to phylogenetic relationship, but was divided in to four families based on amino acid similarity. Some of these might be involved in steroid cholesterol metabolism as it shows relationship with the *CYP7*-family in other organisms, mainly animals. CYP7A1 in *H. sapiens* catalyzes cholesterol 7α-hydroxylation, and is the rate-limiting step in bile acid synthesis (de Montellano, 2005,p. 439 - 441). Studies in rodents have shown that this enzyme is critical in proper uptake of dietary lipids and fat-soluble vitamins, but does not seem to be essential for maintaining cholesterol and lipid levels.

Seven of these genes in *S. robusta* are approximately 1500 – 1600 bp long, and six of them are without introns. Only *SrCyp56* has one intron of 119 bp, located in the middle third of the gene. *SrCyp57b* is not completely characterized yet due to lack of a complete genomic sequence. Two of the genes were analyzed experimentally, *SrCyp52* and *SrCyp54*, both in which showed a very low expression level in the control conditions.

SrCyp52 did not show any significant response to any of the experimental conditions, though a slight down-regulation might be seen after two hours in 4°C. As the gene was already very low expressed, any further down-regulation was hard to show, as the uncertainty with such high Cq-values is very high.

SrCyp54 showed a significant down-regulation after 24 hours in 4°C, and seemed practically turned off. In the day/night cycle, though, the gene was significantly up-regulated, suggesting a possible relation to light-responses.

4.1.11. Families XIII and XIV

A large subgroup consisting of 14 CYPs were created according to phylogenetic relationship. Further evaluation of the amino acid similarity amongst these 14 CYPs showed that 13 of them were in the same family, distributed in six different subfamilies. Only SrCyp8 was predicted to be in another family. These CYPs seems to be Stramenopile-specific, as the only close relationship so far discovered is with the centric diatom *T. pseudonana*.

All of the genes are approximately 1600 – 1800 bp in length, and ten of them are without introns. Three genes, *SrCyp12, SrCyp13* and *SrCyp7* all have one intron each, of 89 bp, 87 bp and 104 bp, respectively. The introns in *SrCyp12* and *SrCyp13* are located in the middle third of the gene, whereas *SrCyp7* has its intron located in the latter third. The last gene, *SrCyp35*, has three introns, one of 96 bp, located in the middle third of the gene, and two of 113 bp, one located in the middle third and one located in the last third of the gene.

Three of four genes analyzed with transcriptional analysis are all medium low expressed in the control conditions, particular *SrCyp17* is low expressed. *SrCyp13* on the other hand, is analyzed with two primer pairs, where one of them (*SrCyp13_1*) show a medium high expression in all conditions, where the other (*SrCyp13_2*) show a somewhat lower expression level. This can be due to uncertainty coupled to a possibly sub-optimal PCR-reaction.

SrCyp16 did not show any significant responses to any of the experimental conditions given, though a slight down-regulation could be seen after 24 hours in 4°C.

SrCyp12 was analyzed with two primer pairs, which both showed a significant up-regulation after 24 hours in 4°C.

Also *SrCyp13* was analyzed with two primer pairs. One of them (*SrCyp13_2*) showed a significant down-regulation after 24 hours in 4°C. The other primer pair (*SrCyp13_1*) also showed a slight down-regulation in this condition, but this response was not significant.

As mentioned, *SrCyp17* showed low expression in the control conditions. After two hours in 4°C it showed a significant down-regulation, and the expression level was so low that the gene seemed turned off. After 24 hours in 4°C, though, the expression level has increased, and an up-regulation is seen, but not significant. This can seem to be a possible adaptation to the colder condition after the initial, rapid “cold-shock” response.

4.1.12. Families XV and XVI

Another quite large group of CYPs in *S. robusta* is relatively poor described as no close relationship has been shown to other organisms. Nine CYPs of *S. robusta* are grouped

together in this subgroup, all around 1600 – 1800 bp long, both with and without introns. Amino acid similarity revealed that this subgroup consists of two families, with SrCyp43 and SrCyp44 in another family than the rest. These CYPs, though, are not complete yet, and this may therefore change as the sequences are complete.

Three genes, *SrCyp19, SrCyp39* and *SrCyp41* have one intron each, located in the middle third of the genes in *SrCyp19* and *SrCyp39*, and in the last third of the gene in *SrCyp41*. *SrCyp40* appears to have two introns, located in the middle and last third of the gene. Two of the genes, *SrCyp43* and *SrCyp44*, are still not complete, and both length and possible introns are thus not known. Two of the genes were analyzed experimentally, both of which showing a relatively low expression level in the control conditions, especially *SrCyp44*.

SrCyp18 showed responses to all of the experimental conditions. After both two hours and 24 hours in 4°C, the gene was significantly down-regulated. The day/night cycle on the other hand, up-regulated the gene significantly, suggesting that this is a gene displaying circadian regulation.

SrCyp44 was generally low expressed in all experimental conditions. After two hours in 4°C, though, the expression level was even lower, and the gene seemed turned off. As seen with other *CYP*s in this experiment, this response seemed to adjust after 24 hours in the cold condtions, and the gene seemed to be switched back on, although the expression level is still low.

4.1.13. Family XVII

SrCyp70 does not show any clear relationship with any known CYP-families, but is evolutionary related to CYPs previously found in cyanobacteria and green algae.

SrCyp70 has no introns and are 1650 bp in length. This gene is quite low expressed in the control conditions, but shows a significant up-regulation after 24 hours in 4°C. No response was seen after two hours in 4°C, though, suggesting that *SrCyp70* is coupled to a response taking place after some time in the colder conditions.

4.1.14. Family XVIII

Family XVIII contains only one *CYP* in *S. robusta*; *SrCyp73*. Genes related to *SrCyp73* is so far only found in Stramenopiles, suggesting that this is a novel Stramenopile-specific *CYP*.

SrCyp73 is 1548 bp long and has no introns. The expression level is quite low in the control conditions, and no significant responses were shown, although a slight up-regulation could be seen in day/night conditions and after 24 hours in 4°C.

4.1.15. Family XIX

One gene showed absolutely no relationship with other organisms; expect that it was confirmed to be a *CYP*. This gene, *SrCyp48*, is without introns and 1578 bp long. It seems to be quite low expressed when considering the low number of sequencing reads form this gene found in transcriptional data sets. No experimental analysis was conducted with this gene. Based on sequencing data available so far, this is a CYP unique for *S. robusta.*

4.1.16. Family XX

Family XX in *S. robusta* consists of one *CYP*; *SrCyp51*. Related genes exist in organisms within the Heterolobosea (*Naegleria gruberi*), Haptophyceae (*Emiliania huxleyi*) and other diatoms such as *F. cylindrus*.

SrCyp51 is relatively undescribed and has not been analyzed experimentally. The gene is 1494 bp long, with two introns of 59 bp and 113 bp located in the middle third of the gene.

4.1.17. Family XXI

Family XXI in *S. robusta* consists of one *CYP*; *SrCyp22*. Related genes have been found in other diatoms, such as *F. cylindrus* and *P. tricornutum*, and a related CYP has been found in the brown algae *Ectocarpus siliculosus*.

SrCyp22 is 1833 bp long with no introns and was medium high expressed under control conditions. After two hours in 4°C, a significant down-regulation was shown, but after 24 hours this response seemed to be reversing, suggesting a possible adaptation to colder

conditions after some time. A significant down-regulation was also seen in the day/night conditions.

This response to both colder and darker conditions suggests that this gene might be involved in both temperature- and light-responses.

4.2. qBase^{Plus} and REST2009

I most cases, REST2009 and qBase^{Plus} came to the same conclusion, but a few cases were shown where REST2009 indicated that a gene was regulated, whereas qBase^{Plus} concluded that the gene did not show a significant response, as seen in Table 3.4.1. When examining these cases more closely, it was clear that qBase^{Plus} was “stricter” in its evaluation. It appears as though qBase^{Plus} also take into consideration the uncertainty coupled to high Cq-values, as well as the deviation between the Cq-values. These differences is also probably due to the fact that qBase^{Plus} corrects the p-value when multiple targets are tested at the same time, as well as performing another multiple test correction when several subgroups are being compared, whereas REST2009 do not take this into consideration. In those cases where REST2009 showed a significant response whereas qBase^{Plus} did not, it is reasonable to assume that REST2009 gave false positives, as qBase^{Plus} does a more thorough analysis when calculating the significance of the response. A negative feature of qBase^{Plus} is that the p-value for the significance is not given, only a “yes/no”. In some cases, it is desirable to know the degree of the significance, rather than a simple yes or no.

The user interface of qBase^{Plus} is easier to work with, and gives the possibility of handling several runs easy. Further, the software has the possibility of handling several projects and experiments in the same window, unlike REST2009 where only one experiment can be open at a time.

When using REST2009, the candidate reference genes’ stability has to be analyzed and evaluated by each scientist, and different methods and opinions may affect choice of, and optimal number of reference genes. Using qBase^{Plus} with the built-in geNorm^{Plus} module to analyze candidate reference genes will ensure a more uniform evaluation of reference

genes, which again will contribute to a higher level of quality assurance when analyzing data from RT-qPCR.

5. CONCLUSION

The cytochrome p450 superfamily in *S. robusta* is unusually expanded compared to other related organisms. The CYPs in *S. robusta* consists of several unique families and subfamilies, but also families known from animals and plants, as well as bacteria, other diatoms, green algae and cyanobacteria. This reflects the evolutionary origin of the diatoms, which complicates the characterization of CYPs in *S. robusta*. Further works needs to be done in order to completely elucidate the characteristics and functions of these CYPs. As the genomes from a broader range of organisms become available, this will contribute to further understanding of the diatoms' origin.

Transcriptional analyses revealed a relatively low expression level, suggesting a tight regulation and specific functions of the different CYPs. This is possibly related to the complex lifecycle of *S. robusta*, involving both vegetative multiplication as well as sexual reproduction. A total of 20 *CYP*s analyzed showed a significant response to one or more of the experimental conditions, suggesting involvement in responses coupled to changes in temperature and/or light conditions.

As a conclusion from the comparison between the two software programs used for analyzing RT-qPCR data, it is highly recommended that the software qBase^{Plus} is implemented in the science group. Especially when analyzing multiple targets at the same time, with several subgroups, this software show great advantage compared to REST2009. Further, this software will ensure a more uniform interpretation of RT-qPCR data amongst the scientists. The built-in analysis of candidate reference genes will also contribute to a higher quality assurance in publishing RT-qPCR data. Replacing REST2009 with qBase^{Plus} will not only improve the quality of the data published, but will also be one step closer to publishing according to the MIQE-guidelines.

6. RECOMMENDATIONS FOR FURTHER WORK

Still, a lot of work needs to be done to completely characterize the cytochrome p450 superfamily in *S. robusta*. Especially the CYP preliminary named SrCyp48 needs to be further analyzed, as this appear to be a novel CYP without any possible homologs in organisms so far total sequenced.

To further elucidate the functions of the CYPs, it would be recommended to test a range of substrates to see if they can be metabolized by *S. robusta*. This could include exposing the diatom to different pollutants or possible toxic substrates in order to identify CYPs coupled to detoxification.

To enable transcriptional profiling in a higher level than by simply performing RT-qPCR, it is recommended to design a microarray chip for *S. robusta*. This would be especially useful to determine global gene expression when exposing the diatom to possible pollutants. A repetition of the low light / high light experiment performed by Nymark *et al.* (2009) can also be performed in order to evaluate the same pathways in *S. robusta* as was investigated in *P. tricornutum*. Evaluating the expression levels of CYPs in axenic vs. non-axenic cultures could be a useful approach in characterizing responses coupled to defense against microorganisms.

The protein localization tools used in this thesis did not give any conclusive results, and more work needs to be done in order to elucidate the intracellular localizations of the CYPs in *S. robusta*. One approach could be cloning and production of Green Fluorescent Protein (GFP)-CYP fusionprotein and thus localize the CYPs. In order to do this, proper vectors for *S. robusta* needs to be designed. One possibility is to transform *P. tricornutum*, but this is a sub-optimal solution as this diatom is very small and thus not suitable for localization studies. I addition to this, *P. tricornutum* does not normally express the majority of CYPs found in *S. robusta*.

REFERENCES

ANDERSEN, R. A. 2005. *Algal Culturing Techniques*, Academic Press.

ANZENBACHER, P. & ANZENBACHEROVÁ, E. 2001. Cytochromes P450 and metabolism of xenobiotics. *Cellular and Molecular Life Sciences,* 58**,** 737-747.

ARMBRUST, E. V. 2009. The life of diatoms in the world's oceans. *Nature.,* 459**,** 185-192.

ARMBRUST, E. V., BERGES, J. A., BOWLER, C., GREEN, B. R., MARTINEZ, D., ROKHSAR, D. S. *et al.* 2004. The genome of the diatom Thalassiosira pseudonana: Ecology, evolution, and metabolism. *Science,* 306**,** 79-86.

BALDAUF, S. L. 2003. Phylogeny for the faint of heart: a tutorial. *Trends in Genetics,* 19**,** 345-351.

BENDTSEN, J. D., JENSEN, L. J., BLOM, N., VON HEIJNE, G. & BRUNAK, S. 2004. Feature-based prediction of non-classical and leaderless protein secretion. *Protein Engineering Design & Selection,* 17**,** 349-356.

BIOGAZELLE. 2011. *qBasePlus geNormPlus* [Online]. Biogazelle.com. Available: http://www.biogazelle.com/files/manual/genormPLUS.pdf [Accessed 25.11 2012].

BIOGAZELLE. 2012a. *qBasePlus Statistics* [Online]. Biogazelle.com. Available: http://www.biogazelle.com/files/manual/statistics.pdf [Accessed 25.11 2012].

BIOGAZELLE. 2012b. *qPCR-analysis* [Online]. Biogazelle.com. Available: http://www.biogazelle.com/course/4_qPCR_analysis.pdf [Accessed 01.12 2012].

BOWLER, C., ALLEN, A. E., BADGER, J. H., GRIMWOOD, J., JABBARI, K., KUO, A., GRIGORIEV, I. V. 2008. The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature,* 456**,** 239-244.

BUSTIN, S. A., BENES, V., GARSON, J. A., HELLEMANS, J., HUGGETT, J., KUBISTA, M., MUELLER, R., NOLAN, T., PFAFFL, M. W., SHIPLEY, G. L., VANDESOMPELE, J. & WITTWER, C. T. 2009. The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. *Clinical Chemistry,* 55**,** 611-622.

CCAP. 2007. *f/2 + Si (Guillard's medium for diatoms)* [Online]. Ccap.ac.uk. Available: http://www.ccap.ac.uk/media/documents/f2_Si_001.pdf [Accessed 28.11 2012].

CHEPURNOV, V. A., MANN, D. G., VON DASSOW, P., VANORMELINGEN, P., GILLARD, J., INZE, D., SABBE, K. & VYVERMAN, W. 2008. In search of new tractable diatoms for experimental biology. *Bioessays,* 30**,** 692-702.

CHEPURNOV, V. A., MANN, D. G., VYVERMAN, W., SABBE, K. & DANIELIDIS, D. B. 2002. Sexual reproduction, mating system, and protoplast dynamics of Seminavis (Bacillariophyceae). *Journal of Phycology,* 38**,** 1004-1019.

CLARK, D. P. 2010. *Molecular Biology*, APCell.

CROFT, M. T., LAWRENCE, A. D., RAUX-DEERY, E., WARREN, M. J. & SMITH, A. G. 2005. Algae acquire vitamin B-12 through a symbiotic relationship with bacteria. *Nature,* 438**,** 90-93.

DANIELIDIS, D. B. & MANN, D. G. 2002. The systematics of Seminavis (Bacillariophyta): the lost identities of Amphora angusta, A-ventricosa and A-macilenta. *European Journal of Phycology,* 37**,** 429-448.

DE MONTELLANO, P. R. O. 2005. Cytochrome P450 Structure, Mechanism, and Biochemistry. *In:* MONTELLANO, P. R. O. D. (ed.) Third ed. Springerlink.com: Kluwer Academic/ Plenum Publishers.

DESCHAMPS, P. & MOREIRA, D. 2012. Reevaluating the green contribution to diatom genomes. *Genome biology and evolution,* 4**,** 683-688.

EMANUELSSON, O., NIELSEN, H., BRUNAK, S. & VON HEIJNE, G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology,* 300**,** 1005-1016.

FALKOWSKI, P. G., KATZ, M. E., KNOLL, A. H., QUIGG, A., RAVEN, J. A., SCHOFIELD, O. & TAYLOR, F. J. R. 2004. The Evolution of Modern Eukaryotic Phytoplankton. *Science,* 305**,** 354-360.

GILLARD, J., DEVOS, V., HUYSMAN, M. J. J., DE VEYLDER, L., D'HONDT, S., MARTENS, C., VANORMELINGEN, P., VANNERUM, K., SABBE, K., CHEPURNOV, V. A., INZE, D., VUYLSTEKE, M. & VYVERMAN, W. 2008. Physiological and Transcriptomic Evidence for a Close Coupling between Chloroplast Ontogeny and Cell Cycle Progression in the Pennate Diatom Seminavis robusta. *Plant Physiology,* 148**,** 1394-1411.

GRANUM, E., RAVEN, J. A. & LEEGOOD, R. C. 2005. How do marine diatoms fix 10 billion tonnes of inorganic carbon per year? *Canadian Journal of Botany-Revue Canadienne De Botanique,* 83**,** 898-908.

GREGORY, W. 1857. On New Forms of Marine Diatomaceae found in the Firth of Clyde and in Loch Fine. *Trans. R. Soc. Edinb.,* 21**,** 473 - 542.

GUENIN, S., MAURIAT, M., PELLOUX, J., VAN WUYTSWINKEL, O., BELLINI, C. & GUTIERREZ, L. 2009. Normalization of qRT-PCR data: the necessity of adopting a systematic, experimental conditions-specific, validation of references. *Journal of Experimental Botany,* 60**,** 487-493.

HORTON, P. 2007. *WoLF PSORT* [Online]. National Institute of Advanced Science and Technology. Available: http://wolfpsort.org/ [Accessed 10.09 2012].

HORTON, P., PARK, K. J., OBAYASHI, T., FUJITA, N., HARADA, H., ADAMS-COLLIER, C. J. & NAKAI, K. 2007. WoLF PSORT: protein localization predictor. *Nucleic Acids Research,* 35**,** W585-W587.

HUSON, D. H., RICHTER, D. C., RAUSCH, C., DEZULIAN, T., FRANZ, M. & RUPP, R. 2007. Dendroscope: An interactive viewer for large phylogenetic trees. *Bmc Bioinformatics,* 8.

INVITROGEN. 2011. *TOPO® TA Cloning® Kit* [Online]. Invitrogen.com. [Accessed 15.03 2012].

KANDEL, S., SAUVEPLANE, V., OLRY, A., DISS, L., BENVENISTE, I. & PINOT, F. 2006. Cytochrome P450-dependent fatty acid hydroxylases in plants. *Phytochemistry Reviews,* 5**,** 359-372.

KARUBE, M., FERNANDINO, J. I., STROBL-MAZZULLA, P., STRUSSMANN, C. A., YOSHIZAKI, G., SOMOZA, G. M. & PATINO, R. 2007. Characterization and expression profile of the ovarian cytochrome p-450 aromatase (cyp19A1) gene during thermolabile sex determination in pejerrey, Odontesthes bonariensis. *Journal of Experimental Zoology Part a-Ecological Genetics and Physiology,* 307A**,** 625-636.

KIM, J., SMITH, J. J., TIAN, L. & DELLAPENNA, D. 2009. The Evolution and Function of Carotenoid Hydroxylases in Arabidopsis. *Plant and Cell Physiology,* 50**,** 463-479.

MADDEN, T. 2003. *The BLAST Sequence Analysis Tool* [Online]. Ncbi.nlm.nih.gov. Available: http://www.ncbi.nlm.nih.gov/books/NBK21097/#A614 [Accessed 24.10 2012].

MERCHANT, S. S., PROCHNIK, S. E., VALLON, O., HARRIS, E. H., KARPOWICZ, S. J., BROKSTEIN, P., *et al.* 2007. The Chlamydomonas genome reveals the evolution of key animal and plant functions. *Science,* 318**,** 245-251.

MEUNIER, B., DE VISSER, S. P. & SHAIK, S. 2004. Mechanism of oxidation reactions catalyzed by cytochrome P450 enzymes. *Chemical Reviews,* 104**,** 3947-3980.

MOUSTAFA, A., BESZTERI, B., MAIER, U. G., BOWLER, C., VALENTIN, K. & BHATTACHARYA, D. 2009. Genomic Footprints of a Cryptic Plastid Endosymbiosis in Diatoms. *Science,* 324**,** 1724-1726.

NELSON, D. R. 2006. Plant cytochrome P450s from moss to poplar. *Phytochemistry Reviews,* 5**,** 193-204.

NELSON, D. R. 2009. *The Cytochrome P450 Homepage. Human Genomics 4 59-65* [Online]. Available: http://drnelson.uthsc.edu/CytochromeP450.html [Accessed 22.08 2012].

NICHOLAS, K. B., JR., H. B. N. & II, D. W. D. 2007. *GeneDoc: Analysis and Visualization of Genetic Variation* [Online]. Nrbsc.org. Available: http://www.nrbsc.org/gfx/genedoc/ebinet.htm [Accessed 05.12 2012].

NYMARK, M., VALLE, K. C., BREMBU, T., HANCKE, K., WINGE, P., ANDRESEN, K., JOHNSEN, G. & BONES, A. M. 2009. An Integrated Analysis of Molecular Acclimation to High Light in the Marine Diatom Phaeodactylum tricornutum. *Plos One,* 4.

QIAGEN. 2003. *A Guide to Analytical Gels* [Online]. Qiagen.com. Available: http://www.qiagen.com/literature/render.aspx?id=23577 [Accessed 01.12 2012].

QIAGEN. 2009a. *QuantiTect Reverse Transcription Handbook* [Online]. Qiagen.com. Available: http://www.qiagen.com/literature/render.aspx?id=252 [Accessed 20.01 2012].

QIAGEN. 2009b. *REST 2009 Software User Guide* [Online]. Qiagen.com. Available: http://www.qiagen.com/products/rest2009software.aspx#Tabs=t2 [Accessed 26.09 2012].

RIEMANN, L., STEWARD, G. F. & AZAM, F. 2000. Dynamics of bacterial community composition and activity during a mesocosm diatom bloom. *Applied and Environmental Microbiology,* 66**,** 578-587.

ROCHE 2008. *LightCycler® 480 Instrument Operator's Manual Software Version 1.5*.

ROCHE. 2011. *LightCycler 480 SYBR Green I Master* [Online]. Roche-Applied-Science.com. Available: https://e-labdoc.roche.com/LFR_PublicDocs/ras/04707516001_en_12.pdf [Accessed 20.01 2012].

RODRIGUEZ-EZPELETA, N., HACKENBERG, M. & ARANSAY, A. M. 2012. *Bioinformatics for High Throughput Sequencing*, Springer.

ROUND, F. E., CRAWFORD, R. M. & MANN, D. G. 1990. *The Diatoms. Biology & Morphology of the Genra*, Cambridge University Press.

RUIJTER, J. M., RAMAKERS, C., HOOGAARS, W. M. H., KARLEN, Y., BAKKER, O., VAN DEN HOFF, M. J. B. & MOORMAN, A. F. M. 2009. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Research,* 37.

SAITOU, N. & NEI, M. 1987. THE NEIGHBOR-JOINING METHOD - A NEW METHOD FOR RECONSTRUCTING PHYLOGENETIC TREES. *Molecular Biology and Evolution,* 4**,** 406-425.

SCHNEIDER, E. & CLARK, D. S. 2013. Cytochrome P450 (CYP) enzymes and the development of CYP biosensors. *Biosensors & Bioelectronics,* 39**,** 1-13.

STOCCO, C. 2012. Tissue physiology and pathology of aromatase. *Steroids,* 77**,** 27-35.

THERMOFISHERSCIENTIFIC. 2011. *NanoDrop 1000 Spectrophotometer V3.8 User's Manual* [Online]. Nanodrop.com. Available: http://www.nanodrop.com/Library/nd-1000-v3.8-users-manual-8%205x11.pdf [Accessed 24.10 2012].

WERCK-REICHHART, D. & FEYEREISEN, R. 2000. Cytochromes P450: a success story. *Genome Biology,* 1**,** reviews3003.3001 - reviews3003.3009.

YANG, S., WU, R. S. S., MOK, H. O. L., ZHANG, Z. P. & KONG, R. Y. C. 2003. Identification of a novel cytochrome P450 cDNA, CYP97E1, from the marine diatom Skeletonema costatum Bacillariophyceae. *Journal of Phycology,* 39**,** 555-560.

APPENDIX

A1. MEDIA AND SOLUTIONS

The media and solutions used in this thesis are listed after area of application below.

A1.1 CULTIVATION OF DIATOMS

The different media and solutions used for cultivation of the diatoms are listed alphabetically below.

Antibiotic mixture for axenization

Gentamycin (20 mg/ml)	1	ml
Penicillin G (100 000 U/ml)	0.5	ml
Polymixin B (250 000 U/ml)	50	µl
Cefortaxim (Claforan) (250 mg/ml)	0.5	ml
Milli Q-water	5	ml

Cefotaxim is to be added as the final ingredient.

f/2 medium

$NaNO_3$ (75 g/L dH_2O)	1	ml
NaH_2PO_4 x H_2O (5.65 g/L dH_2O)	1	ml
Na_2SiO_3 x $9H_2O$ (30 g/L dH_2O)	1	ml
Mineral mix (see below)	1	ml
Vitamin mix (see below)	0.5	ml
Autoclaved, sterile filtered seawater	1000	ml

The solution was stored at ambient temperature.

f/2 trace elements (chelated)

NA_2EDTA	4.36	g
$FeCl_2 \times 6H_2O$	3.15	g
$CuSO_4 \times 5H_2O$	0.01	g
$ZnSO_4 \times 7H_2O$	0.022	g
$CoCl_2 \times 6H_2O$	0.01	g
$MnCl_2 \times 4H_2O$	0.18	g
$Na_2Mo_4 \times 2H_2O$	0.006	g
Autoclaved, sterile filtered seawater	1000	ml

Adjust to pH 8 with 1M HCl or NaOH. Sterile filter through a 0.20 µm filter before use. Store at 4 °C.

f/2 vitamin mix

Cyanocolbalamin (Vitamin B_{12})	0.0005	g
Thiamine HCl (Vitamin B_1)	0.1	g
Biotin	0.0005	g
Autoclaved, sterile filtered seawater	1000	ml

Adjust to pH 8 with 1M HCl or NaOH. Sterile filter through a 0.20 µm filter before use. Store at 4°C.

Lugol's iodine solution

KI	100	g
Crystalline iodine	50	g
Glacial acetic acid	100	ml
Milli-Q water	1000	ml

Dissolve the KI in Milli-Q water before dissolving the crystalline iodine in this solution. Add glacial acetic acid and mix. Any possible precipitation should be removed by decanting the solution before use.

A1.2 GEL ELECTROPHORESIS

The different solutions and gels used in gel electrophoresis for both DNA- and RNA-gels are listed alphabetically below.

1% (w/v) agarose gel (50 ml)

Agarose (Sigma, cat. no. A9539-500)	0.50	g
1xTAE Buffer	50	ml
GelRed (Biotium, cat. no. 41003-0.5ml)	5	µl

Agarose was dissolved in 1 x TAE buffer by heating in a microwave oven. The solution was then allowed to cool for a couple of minutes before adding the GelRed.

1.2% (w/v) FA-gel (50 ml)

Agarose (Sigma, cat. no. A9539-500)	0.60	g
10xFA Buffer	5	ml
37 % Formaldehyde	900	µl
Milli-Q water	q.s	
Ethidium bromide	0.5	µl

Agarose was dissolved by heating the solution in a microwave oven. The solution was then allowed to cool for a couple of minutes before adding the ethidium bromide. After preparation, the gel was left for equilibration in the running buffer for at least 30 minutes.

10 x FA-buffer (1000 ml)

MOPS	200	mM
Na-acetate x3H_2O	50	mM
EDTA	10	mM

Adjust to pH 7.0 with NaOH, store at ambient temperature.

50xTris-acetate EDTA (TAE) buffer (1000 ml)

Tris base	242	g
Glacial acetic acid	57.1	ml
0.5 M EDTA pH 8.0	100	ml
Milli-Q water	q.s	

The solution was sterilized by autoclaving and stored at ambient temperature. Working solution was diluted to 1xTAE.

Running buffer for FA-gel (500 ml)

10 x FA-buffer	50	ml
37 % Formaldehyde	10	ml
Milli-Q water	q.s	

A2. PRIMERS

A2.1 PRIMERS FOR RT-qPCR

Table A 1 Primers for RT-qPCR

Gene	Sequence	Orientation
Cyp19A1	5'-CAACGGTCATCCTACAATACGA-3'	Forward
	3'-CGATGTCCCAATGCGTTTCAAG-5'	Reverse
Cyp51A1	5'-TCGATCCAACCCGGTTTGCG-3'	Forward
	3'-GTTCCAGGCAAACTTGCGAC-3'	Reverse
Cyp51B1+2	5'-GCGTGCAAATTCCGATCAGTCA-3'	Forward
	3'-TCCTTATGCAGACACTCAACCT-5'	Reverse
Cyp86A2	5'-TCCCATGCAAGTCACATTGAGG-3'	Forward
	3'-GTTACAGTAGTATCCACCAGTC-5'	Reverse
Cyp86A3	5'-GACATGATGGCACCCTTATACC-3'	Forward
	3'-TCTGCAAGGCCTGCTATCAACG-5'	Reverse
Cyp86A6	5'-GGACACAATGGAATGTACGACT-3'	Forward
	3'-CTATTGTCGCAACCGTACCTGA-5'	Reverse
Cyp97A1	5'-GCCATTCCGCCTGAAGAAGTTG-3'	Forward
	3'-CATCCGTCTTTGGCAATGGTTC-5'	Reverse
Cyp97B1	5'-AACGATCATCCCCAGCATTTGG-3'	Forward
	3'-TCTTCACGATCATGTGCAATCC-5'	Reverse
SrCyp10	5'-CGTATCGTCAACAATCCCAACT-3'	Forward
	3'-TCAATGAGACTCCAGGAAATGC-5'	Reverse
SrCyp12_1	5'-GCGTGAAGTTGGAGATGAACC-3'	Forward
	3'-ACAGTTCCTTCCTCCAGCCGA-5'	Reverse
SrCyp12_2	5'-TGGATACCCTGATGACTTTGCTG-3'	Forward
	3'-CTTTGAAGCAACGTTCCTGAACT-5'	Reverse
SrCyp13_1	5'-GACTTGTTGGCCAGTTTGATTGA-3'	Forward
	3'-CAAGCTTCTGGGAGAACTACAG-5'	Reverse
SrCyp13_2	5'-AGAAGTCTATCGAAGAGGTGTC-3'	Forward
	3'-TGATACTCGTGGTGTCATATCC-5'	Reverse

SrCyp16	5’-AATGATACAGGAACTGTCGGCT-3’	Forward
	3’-TCAACCATCCACAAGAAGGGTC-5’	Reverse
SrCyp17	5’-GTTCATTCCATAGTGAAACTGC-3’	Forward
	3’-TGTCACCTTCGAGCTTTGATGA-5’	Reverse
SrCyp18	5’-GGTCATGGGGTTCGAAAGTGTC-3’	Forward
	3’-GGTTGGGGGCAAGTTCAAACTG-3’	Reverse
SrCyp20	5’-CCAGCAAGCTAGTGCGTGTAGA-3’	Forward
	3’-CAGCTACCGGTAACGCATGGAT-5’	Reverse
SrCyp21	5’-TAGCTGTCATGCAAAGTTGAGC-3’	Forward
	3’-ACTCGCCATACAAGAGGATACA-5’	Reverse
SrCyp22	5’-GCTGGATTTACCGTCTTACCAC-3’	Forward
	3’-GCGCTTACGGAGGTTCGATTCA-5’	Reverse
SrCyp23	5’-CAGATGTTTGGTGATCAAACTG-3’	Forward
	3’-AACCTAGCTTGTTCATCTCGCA-5’	Reverse
SrCyp28	5’-AGGTGTTCCTTGCCATATTCGC-3’	Forward
	3’-ATTGTGCTAGGGTTCCACTTGA-5’	Reverse
SrCyp44	5’-TCTTCGAGGTAGGCAGCAACT-3’	Forward
	3’-GATCCCAGTACAGCAATCAAG-5’	Reverse
SrCyp46	5’-GACACTGTTAATGCAAGGGACT-3’	Forward
	3’-GCAAAGCCCAAATACTGAGTTG-5’	Reverse
SrCyp52	5’-CCGGAAAGCTAGTCAACGATCC-3’	Forward
	3’-CTTGACTTCGTTGGCAATGAAC-5’	Reverse
SrCyp54	5’-TCATGCATCCGGTGGAAGATGT-3’	Forward
	3’-GTTTCGTACTACACCGTATCCA-5’	Reverse
SrCyp70	5’-CAAAGCATTCAATGGATGGAGC-3’	Forward
	3’-CCCAACTCGTCAACAGGAATCT-5’	Reverse
SrCyp73	5’-CTTGGAAATTCCCGTGGATCAC-3’	Forward
	3’-ACATGATTGCCAACCACAGTTC-5’	Reverse

A2.2 PRIMERS FOR PCR

Gene	Sequence	Orientation
Cyp 19A1_1	5'-CAAACGGAAGCTGTTTGTTGAG-3'	Forward
	3'-AGTAAGCCGTAACCGATGAAGT-5'	Reverse
Cyp19A1_2	5'-CGGCTTTCATAGATGCGATTGT-3'	Forward
	3'-ATGACCCTTTAAGCTGGATACG-5'	Reverse

A3. LINREG RESULTS

Table A 2 LinReg results for the reference gene *qSrVps35* in RT-qPCR.

Sample	Individual PCR-efficiency	Mean PCR-efficiency	Cq-value
Const_0h_ctrl_1 *qSrVps35*	1.964	1.939	28.490
Const_0h_ctrl_2 *qSrVps35*	1.892	1.939	28.014
Const_0h_ctrl_3 *qSrVps35*	1.935	1.939	27.078
Const_0h_ctrl_4 *qSrVps35*	1.943	1.939	27.078
Const_2h_1 *qSrVps35*	1.919	1.939	28.017
Const_2h_2 *qSrVps35*	1.945	1.939	27.367
Const_2h_3 *qSrVps35*	1.930	1.939	28.127
Const_2h_4 *qSrVps35*	1.929	1.939	27.160
Const_24h_1 *qSrVps35*	1.950	1.939	26.977
Const_24h_2 *qSrVps35*	1.953	1.939	27.204
Const_24h_3 *qSrVps35*	1.916	1.939	27.366
Const_24h_4 *qSrVps35*	1.979	1.939	27.346
D/n_1 *qSrVps35*	1.960	1.939	27.411
D/n_2 *qSrVps35*	1.964	1.939	27.234
D/n_3 *qSrVps35*	1.920	1.939	26.985
D/n_4 *qSrVps35*	1.921	1.939	26.965
qSrVps35 NTC	1.000	0.000	0.000
qSrVps35 NTC	1.932	0.000	45.277

Table A 3 LinReg results for the reference gene *qSrMFS* in RT-qPCR.

Sample	Individual PCR-efficiency	Mean PCR-efficiency	Cq-value
Const_0h_1 *qSrMFS*	2.026	1.998	32.880
Const_0h_2 *qSrMFS*	1.995	1.998	32.265
Const_0h_3 *qSrMFS*	2.041	1.998	31.480
Const_0h_4 *qSrMFS*	1.996	1.998	30.458
Const_2h_1 *qSrMFS*	1.984	1.998	32.405
Const_2h_2 *qSrMFS*	2.048	1.998	31.839
Const_2h_3 *qSrMFS*	2.037	1.998	31.820
Const_2h_4 *qSrMFS*	1.999	1.998	30.506
Const_24h_1 *qSrMFS*	1.888	1.998	30.100
Const_24h_2 *qSrMFS*	1.961	1.998	30.254
Const_24h_3 *qSrMFS*	2.017	1.998	32.803
Const_24h_4 *qSrMFS*	1.978	1.998	31.087
D/n_1 *qSrMFS*	1.954	1.998	31.086
D/n_2 *qSrMFS*	1.971	1.998	32.032
D/n_3 *qSrMFS*	2.011	1.998	30.598
D/n_4 *qSrMFS*	1.930	1.998	31.103
qSrMFS NTC	1.000	0.000	0.000
qSrMFS NTC	1.906	0.000	43.624

Table A 4 LinReg results for the control without reverse transcriptase in RT-qPCR. The primerset for the target gene *Cyp86A3* is used as this does not contain any introns.

Sample	Individual PCR-efficiency	Mean PCR-efficiency	Cq-value
Const_0h_ctrl_1 *Cyp86A3*-RT	1.000	0.000	0.000
Const_0h_ctrl_2 *Cyp86A3*-RT	1.000	0.000	0.000
Const_0h_ctrl_3 *Cyp86A3*-RT	1.960	0.000	41.131
Const_0h_ctrl_4 *Cyp86A3*-RT	1.000	0.000	0.000
Const_2h_1 *Cyp86A3*-RT	1.000	0.000	0.000
Const_2h_2 *Cyp86A3*-RT	1.000	0.000	0.000
Const_2h_3 *Cyp86A3*-RT	1.000	0.000	0.000
Const_2h_4 *Cyp86A3*-RT	1.000	0.000	0.000
Const_24h_1 *Cyp86A3*-RT	1.000	0.000	0.000
Const_24h_2 *Cyp86A3*-RT	1.000	0.000	0.000
Const_24h_3 *Cyp86A3*-RT	1.000	0.000	0.000
Const_24h_4 *Cyp86A3*-RT	1.000	0.000	0.000
D/n_1 *Cyp86A3*-RT	1.000	0.000	0.000
D/n_2 *Cyp86A3*-RT	1.000	0.000	0.000
D/n_3 *Cyp86A3*-RT	1.000	0.000	0.000
D/n_4 *Cyp86A3*-RT	1.000	0.000	0.000

Table A 5 LinReg results for the target gene *Cyp19A1* in RT-qPCR

Sample	Individual PCR-efficiency	Mean PCR-efficiency	Cq-value
Const_0h_ctrl_1 *Cyp19A1*	1.927	1.948	34.010
Const_0h_ctrl_2 *Cyp19A1*	2.012	1.948	33.735
Const_0h_ctrl_3 *Cyp19A1*	2.009	1.948	33.676
Const_0h_ctrl_4 *Cyp19A1*	1.927	1.948	33.125
Const_2h_1 *Cyp19A1*	1.941	1.948	34.200
Const_2h_2 *Cyp19A1*	1.977	1.948	35.520
Const_2h_3 *Cyp19A1*	1.965	1.948	35.241
Const_2h_4 *Cyp19A1*	1.932	1.948	35.083
Const_24h_1 *Cyp19A1*	1.929	1.948	33.003
Const_24h_2 *Cyp19A1*	1.947	1.948	34.201
Const_24h_3 *Cyp19A1*	1.950	1.948	32.805
Const_24h_4 *Cyp19A1*	1.944	1.948	32.149
D/n_1 *Cyp19A1*	1.958	1.948	33.621
D/n_2 *Cyp19A1*	1.968	1.948	34.547
D/n_3 *Cyp19A1*	1.924	1.948	34.069
D/n_4 *Cyp19A1*	1.977	1.948	32.432
Cyp19A1 NTC	1.000	0.000	0.000
Cyp19A1 NTC	1.000	0.000	0.000

Table A 6 LinReg results for the target gene *Cyp51A1* in RT-qPCR

Sample	Individual PCR-efficiency	Mean PCR-efficiency	Cq-value
Const_0h_ctrl_1 *Cyp51A1*	1.906	1.893	28.709
Const_0h_ctrl_2 *Cyp51A1*	1.906	1.893	27.944
Const_0h_ctrl_3 *Cyp51A1*	1.926	1.893	27.079
Const_0h_ctrl_4 *Cyp51A1*	1.863	1.893	26.867
Const_2h_1 *Cyp51A1*	1.954	1.893	27.346
Const_2h_2 *Cyp51A1*	1.888	1.893	26.608
Const_2h_3 *Cyp51A1*	1.896	1.893	28.147
Const_2h_4 *Cyp51A1*	1.870	1.893	26.976
Const_24h_1 *Cyp51A1*	1.934	1.893	26.552
Const_24h_2 *Cyp51A1*	1.894	1.893	26.640
Const_24h_3 *Cyp51A1*	1.904	1.893	26.488
Const_24h_4 *Cyp51A1*	1.856	1.893	26.917
D/n_1 *Cyp51A1*	1.934	1.893	26.463
D/n_2 *Cyp51A1*	1.877	1.893	26.459
D/n_3 *Cyp51A1*	1.867	1.893	26.609
D/n_4 *Cyp51A1*	1.871	1.893	25.939
Cyp51A1 NTC	1.000	0.000	0.000
Cyp51A1 NTC	1.951	0.000	39.651

Table A 7 LinReg results for the target gene *Cyp51B1+2* in RT-qPCR

Sample	Individual PCR-efficiency	Mean PCR-efficiency	Cq-value
Const_0h_ctrl_1 *Cyp51B1+2*	1.899	1.916	25.069
Const_0h_ctrl_2 *Cyp51B1+2*	1.922	1.916	24.200
Const_0h_ctrl_3 *Cyp51B1+2*	1.948	1.916	23.571
Const_0h_ctrl_4 *Cyp51B1+2*	1.915	1.916	23.545
Const_2h_1 *Cyp51B1+2*	1.854	1.916	25.990
Const_2h_2 *Cyp51B1+2*	1.938	1.916	25.455
Const_2h_3 *Cyp51B1+2*	1.889	1.916	25.942
Const_2h_4 *Cyp51B1+2*	1.933	1.916	25.327
Const_24h_1 *Cyp51B1+2*	1.920	1.916	25.223
Const_24h_2 *Cyp51B1+2*	1.913	1.916	25.052
Const_24h_3 *Cyp51B1+2*	1.958	1.916	24.667
Const_24h_4 *Cyp51B1+2*	1.936	1.916	25.561
D/n_1 *Cyp51B1+2*	1.892	1.916	23.546
D/n_2 *Cyp51B1+2*	1.891	1.916	23.020
D/n_3 *Cyp51B1+2*	1.888	1.916	22.949
D/n_4 *Cyp51B1+2*	1.901	1.916	23.006
Cyp51B1+2 NTC	1.916	0.000	44.758
Cyp51B1+2 NTC	1.000	0.000	0.000

Table A 8 LinReg results for the target gene *Cyp86A2* in RT-qPCR

Sample	Individual PCR-efficiency	Mean PCR-efficiency	Cq-value
Const_0h_ctrl_1 *Cyp86A2*	1.900	1.898	28.048
Const_0h_ctrl_2 *Cyp86A2*	1.880	1.898	26.893
Const_0h_ctrl_3 *Cyp86A2*	1.901	1.898	26.793
Const_0h_ctrl_4 *Cyp86A2*	1.893	1.898	26.462
Const_2h_1 *Cyp86A2*	1.982	1.898	28.246
Const_2h_2 *Cyp86A2*	1.947	1.898	29.315
Const_2h_3 *Cyp86A2*	1.882	1.898	29.909
Const_2h_4 *Cyp86A2*	1.899	1.898	28.800
Const_24h_1 *Cyp86A2*	1.932	1.898	27.384
Const_24h_2 *Cyp86A2*	1.961	1.898	27.646
Const_24h_3 *Cyp86A2*	1.921	1.898	27.144
Const_24h_4 *Cyp86A2*	1.943	1.898	27.245
D/n_1 *Cyp86A2*	1.893	1.898	26.126
D/n_2 *Cyp86A2*	1.856	1.898	25.248
D/n_3 *Cyp86A2*	1.876	1.898	25.003
D/n_4 *Cyp86A2*	1.849	1.898	25.031
Cyp86A2 NTC	1.000	0.000	0.000
Cyp86A2 NTC	1.000	0.000	0.000

Table A 9 LinReg results for the target gene *Cyp86A3* in RT-qPCR

Sample	Individual PCR-efficiency	Mean PCR-efficiency	Cq-value
D/n_1 *Cyp86A3*	1.876	1.882	30.210
D/n_2 *Cyp86A3*	1.875	1.882	31.097
D/n_3 *Cyp86A3*	1.873	1.882	31.215
D/n_4 *Cyp86A3*	1.881	1.882	29.572
Const_0h_1 *Cyp86A3*	1.862	1.882	32.944
Const_0h_2 *Cyp86A3*	1.884	1.882	32.574
Const_0h_3 *Cyp86A3*	1.885	1.882	32.057
Const_0h_4 *Cyp86A3*	1.882	1.882	31.467
Const_2h_1 *Cyp86A3*	1.893	1.882	33.707
Const_2h_2 *Cyp86A3*	1.890	1.882	34.293
Const_2h_3 *Cyp86A3*	1.869	1.882	34.904
Const_2h_4 *Cyp86A3*	1.910	1.882	34.346
Const_24h_1 *Cyp86A3*	1.887	1.882	32.320
Const_24h_2 *Cyp86A3*	1.873	1.882	33.168
Const_24h_3 *Cyp86A3*	1.883	1.882	32.925
Const_24h_4 *Cyp86A3*	1.883	1.882	32.311
Cyp86A3 NTC	1.000	0.000	0.000
Cyp86A3 NTC	1.000	0.000	0.000

Table A 10 LinReg results for the target gene *Cyp86A6* in RT-qPCR

Sample	Individual PCR-efficiency	Mean PCR-efficiency	Cq-value
Const_0h_ctrl_1 *Cyp86A6*	1.950	1.926	26.387
Const_0h_ctrl_2 *Cyp86A6*	1.943	1.926	26.403
Const_0h_ctrl_3 *Cyp86A6*	1.929	1.926	26.446
Const_0h_ctrl_4 *Cyp86A6*	1.922	1.926	26.114
Const_2h_1 *Cyp86A6*	1.900	1.926	24.978
Const_2h_2 *Cyp86A6*	1.923	1.926	25.818
Const_2h_3 *Cyp86A6*	1.954	1.926	26.622
Const_2h_4 *Cyp86A6*	1.917	1.926	25.986
Const_24h_1 *Cyp86A6*	1.915	1.926	25.079
Const_24h_2 *Cyp86A6*	1.931	1.926	26.166
Const_24h_3 *Cyp86A6*	1.917	1.926	26.127
Const_24h_4 *Cyp86A6*	1.942	1.926	25.774
D/n_1 *Cyp86A6*	1.911	1.926	25.256
D/n_2 *Cyp86A6*	1.917	1.926	24.395
D/n_3 *Cyp86A6*	1.919	1.926	24.755
D/n_4 *Cyp86A6*	1.928	1.926	25.210
Cyp86A6 NTC	1.937	0.000	42.665
Cyp86A6 NTC	1.924	0.000	41.948

Table A 11 LinReg results for the target gene *Cyp97A1* in RT-qPCR

Sample	Individual PCR-efficiency	Mean PCR-efficiency	Cq-value
Const_0h_ctrl_1 *Cyp97A1*	1.983	1.966	24.618
Const_0h_ctrl_2 *Cyp97A1*	1.961	1.966	24.403
Const_0h_ctrl_3 *Cyp97A1*	2.023	1.966	23.716
Const_0h_ctrl_4 *Cyp97A1*	1.978	1.966	22.754
Const_2h_1 *Cyp97A1*	1.946	1.966	24.659
Const_2h_2 *Cyp97A1*	1.925	1.966	23.975
Const_2h_3 *Cyp97A1*	1.948	1.966	24.361
Const_2h_4 *Cyp97A1*	1.966	1.966	23.755
Const_24h_1 *Cyp97A1*	1.947	1.966	24.927
Const_24h_2 *Cyp97A1*	2.002	1.966	26.231
Const_24h_3 *Cyp97A1*	1.999	1.966	25.526
Const_24h_4 *Cyp97A1*	1.997	1.966	24.869
D/n_1 *Cyp97A1*	1.946	1.966	26.520
D/n_2 *Cyp97A1*	1.987	1.966	25.676
D/n_3 *Cyp97A1*	1.959	1.966	26.180
D/n_4 *Cyp97A1*	1.953	1.966	26.341
Cyp97A1 NTC	1.000	0.000	0.000
Cyp97A1 NTC	1.000	0.000	0.000

Table A 12 LinReg results for the target gene *Cyp97B1* in RT-qPCR

Sample	Individual PCR-efficiency	Mean PCR-efficiency	Cq-value
Const_0h_ctrl_1 *Cyp97B1*	1.892	1.891	25.370
Const_0h_ctrl_2 *Cyp97B1*	1.890	1.891	25.364
Const_0h_ctrl_3 *Cyp97B1*	1.898	1.891	24.328
Const_0h_ctrl_4 *Cyp97B1*	1.866	1.891	24.035
Const_2h_1 *Cyp97B1*	1.912	1.891	26.686
Const_2h_2 *Cyp97B1*	1.913	1.891	25.716
Const_2h_3 *Cyp97B1*	1.862	1.891	26.088
Const_2h_4 *Cyp97B1*	1.896	1.891	25.336
Const_24h_1 *Cyp97B1*	1.869	1.891	24.029
Const_24h_2 *Cyp97B1*	1.911	1.891	24.276
Const_24h_3 *Cyp97B1*	1.904	1.891	23.400
Const_24h_4 *Cyp97B1*	1.921	1.891	23.489
D/n_1 *Cyp97B1*	1.879	1.891	28.421
D/n_2 *Cyp97B1*	1.874	1.891	30.191
D/n_3 *Cyp97B1*	1.889	1.891	29.410
D/n_4 *Cyp97B1*	1.873	1.891	27.992
Cyp97B1 NTC	1.846	1.870	40.892
Cyp97B1 NTC	1.852	1.870	40.779

Table A 13 LinReg results for the target gene *SrCyp10* in RT-qPCR

Sample	Individual PCR-efficiency	Mean PCR-efficiency	Cq-value
Const_0h_ctrl_1 *SrCyp10*	1.816	1.818	28.223
Const_0h_ctrl_2 *SrCyp10*	1.799	1.818	26.978
Const_0h_ctrl_3 *SrCyp10*	1.818	1.818	25.874
Const_0h_ctrl_4 *SrCyp10*	1.825	1.818	26.647
Const_2h_1 *SrCyp10*	1.839	1.818	30.738
Const_2h_2 *SrCyp10*	1.803	1.818	30.044
Const_2h_3 *SrCyp10*	1.807	1.818	31.141
Const_2h_4 *SrCyp10*	1.830	1.818	30.323
Const_24h_1 *SrCyp10*	1.819	1.818	27.802
Const_24h_2 *SrCyp10*	1.826	1.818	28.894
Const_24h_3 *SrCyp10*	1.821	1.818	28.813
Const_24h_4 *SrCyp10*	1.799	1.818	29.526
D/n_1 *SrCyp10*	1.801	1.818	27.260
D/n_2 *SrCyp10*	1.824	1.818	27.622
D/n_3 *SrCyp10*	1.829	1.818	26.644
D/n_4 *SrCyp10*	1.824	1.818	26.444
SrCyp10 NTC	1.838	0.000	39.670
SrCyp10 NTC	1.761	0.000	45.918

Table A 14 LinReg results for the target gene *SrCyp12* in RT-qPCR. Two primersets are included for this gene.

Sample	Individual PCR-efficiency	Mean PCR-efficiency	Cq-value
Const_0h_ctrl_1 *SrCyp12_1*	1.915	1.926	28.219
Const_0h_ctrl_2 *SrCyp12_1*	1.806	1.926	27.336
Const_0h_ctrl_3 *SrCyp12_1*	1.927	1.926	25.987
Const_0h_ctrl_4 *SrCyp12_1*	1.941	1.926	25.844
Const_2h_1 *SrCyp12_1*	1.914	1.926	27.054
Const_2h_2 *SrCyp12_1*	1.924	1.926	26.993
Const_2h_3 *SrCyp12_1*	1.928	1.926	27.341
Const_2h_4 *SrCyp12_1*	1.951	1.926	26.469
Const_24h_1 *SrCyp12_1*	1.935	1.926	24.030
Const_24h_2 *SrCyp12_1*	1.953	1.926	23.758
Const_24h_3 *SrCyp12_1*	1.933	1.926	23.448
Const_24h_4 *SrCyp12_1*	1.930	1.926	23.595
D/n_1 *SrCyp12_1*	1.930	1.926	25.578
D/n_2 *SrCyp12_1*	1.899	1.926	26.039
D/n_3 *SrCyp12_1*	1.898	1.926	26.219
D/n_4 *SrCyp12_1*	1.915	1.926	24.621
SrCyp12_1 NTC	1.984	0.000	39.575
SrCyp12_1 NTC	1.000	0.000	0.000
Const_0h_ctrl_1 *SrCyp12_2*	1.966	1.940	30.350
Const_0h_ctrl_2 *SrCyp12_2*	1.911	1.940	28.947
Const_0h_ctrl_3 *SrCyp12_2*	1.929	1.940	27.753
Const_0h_ctrl_4 *SrCyp12_2*	1.933	1.940	27.425
Const_2h_1 *SrCyp12_2*	1.908	1.940	29.032
Const_2h_2 *SrCyp12_2*	1.911	1.940	29.067
Const_2h_3 *SrCyp12_2*	1.947	1.940	29.556
Const_2h_4 *SrCyp12_2*	1.966	1.940	28.505
Const_24h_1 *SrCyp12_2*	1.990	1.940	25.767
Const_24h_2 *SrCyp12_2*	2.001	1.940	25.715
Const_24h_3 *SrCyp12_2*	1.949	1.940	25.596
Const_24h_4 *SrCyp12_2*	1.986	1.940	25.634
D/n_1 *SrCyp12_2*	1.940	1.940	27.583
D/n_2 *SrCyp12_2*	1.932	1.940	28.339
D/n_3 *SrCyp12_2*	1.905	1.940	28.019
D/n_4 *SrCyp12_2*	1.975	1.940	26.524
SrCyp12_2 NTC	1.994	0.000	39.649
SrCyp12_2 NTC	1.000	0.000	0.000

Table A 15 LinReg results for the target gene *SrCyp13* in RT-qPCR. Two primersets are included for this gene.

Sample	Individual PCR-efficiency	Mean PCR-efficiency	Cq-value
Const_0h_ctrl_1 *SrCyp13_1*	1.849	1.889	24.473
Const_0h_ctrl_2 *SrCyp13_1*	1.873	1.889	24.269
Const_0h_ctrl_3 *SrCyp13_1*	1.855	1.889	23.726
Const_0h_ctrl_4 *SrCyp13_1*	1.892	1.889	23.675
Const_2h_1 *SrCyp13_1*	1.854	1.889	24.064
Const_2h_2 *SrCyp13_1*	1.913	1.889	24.101
Const_2h_3 *SrCyp13_1*	1.912	1.889	24.680
Const_2h_4 *SrCyp13_1*	1.888	1.889	24.148
Const_24h_1 *SrCyp13_1*	1.890	1.889	24.201
Const_24h_2 *SrCyp13_1*	1.916	1.889	24.624
Const_24h_3 *SrCyp13_1*	1.912	1.889	24.152
Const_24h_4 *SrCyp13_1*	1.900	1.889	24.351
D/n_1 *SrCyp13_1*	1.894	1.889	24.324
D/n_2 *SrCyp13_1*	1.913	1.889	23.372
D/n_3 *SrCyp13_1*	1.896	1.889	24.191
D/n_4 *SrCyp13_1*	1.863	1.889	23.950
SrCyp13_1 NTC	1.000	0.000	0.000
SrCyp13_1 NTC	1.000	0.000	0.000
Const_0h_ctrl_1 *SrCyp13_2*	1.813	1.811	28.531
Const_0h_ctrl_2 *SrCyp13_2*	1.806	1.811	28.075
Const_0h_ctrl_3 *SrCyp13_2*	1.834	1.811	26.853
Const_0h_ctrl_4 *SrCyp13_2*	1.790	1.811	27.367
Const_2h_1 *SrCyp13_2*	1.800	1.811	27.965
Const_2h_2 *SrCyp13_2*	1.806	1.811	27.575
Const_2h_3 *SrCyp13_2*	1.828	1.811	28.591
Const_2h_4 *SrCyp13_2*	1.814	1.811	27.401
Const_24h_1 *SrCyp13_2*	1.815	1.811	27.219
Const_24h_2 *SrCyp13_2*	1.817	1.811	28.615
Const_24h_3 *SrCyp13_2*	1.818	1.811	28.278
Const_24h_4 *SrCyp13_2*	1.812	1.811	28.291
D/n_1 *SrCyp13_2*	1.819	1.811	28.237
D/n_2 *SrCyp13_2*	1.789	1.811	27.485
D/n_3 *SrCyp13_2*	1.819	1.811	28.121
D/n_4 *SrCyp13_2*	1.795	1.811	27.851
SrCyp13_2 NTC	1.000	0.000	0.000
SrCyp13_2 NTC	1.000	0.000	0.000

Table A 16 LinReg results for the target gene *SrCyp16* in RT-qPCR. A possible contamination is seen in one NTC, but is not considered important as the Cq-value is significant higher than the samples'.

Sample	Individual PCR-efficiency	Mean PCR-efficiency	Cq-value
Const_0h_ctrl_1 *SrCyp16*	1.921	1.913	27.634
Const_0h_ctrl_2 *SrCyp16*	1.921	1.913	27.529
Const_0h_ctrl_3 *SrCyp16*	1.924	1.913	27.444
Const_0h_ctrl_4 *SrCyp16*	1.891	1.913	25.917
Const_2h_1 *SrCyp16*	1.888	1.913	26.994
Const_2h_2 *SrCyp16*	1.935	1.913	26.659
Const_2h_3 *SrCyp16*	1.902	1.913	26.990
Const_2h_4 *SrCyp16*	1.894	1.913	26.433
Const_24h_1 *SrCyp16*	1.895	1.913	27.201
Const_24h_2 *SrCyp16*	1.912	1.913	27.152
Const_24h_3 *SrCyp16*	1.924	1.913	26.597
Const_24h_4 *SrCyp16*	1.935	1.913	26.659
D/n_1 *SrCyp16*	1.907	1.913	26.355
D/n_2 *SrCyp16*	1.919	1.913	26.140
D/n_3 *SrCyp16*	1.917	1.913	26.795
D/n_4 *SrCyp16*	1.923	1.913	27.426
SrCyp16 NTC	1.000	1.876	0.000
SrCyp16 NTC	1.876	1.876	35.912

Table A 17 LinReg results for the target gene *SrCyp17* in RT-qPCR

Sample	Individual PCR-efficiency	Mean PCR-efficiency	Cq-value
Const_0h_ctrl_1 *SrCyp17*	1.888	1.903	33.231
Const_0h_ctrl_2 *SrCyp17*	1.911	1.903	33.040
Const_0h_ctrl_3 *SrCyp17*	1.908	1.903	31.928
Const_0h_ctrl_4 *SrCyp17*	1.915	1.903	31.256
Const_2h_1 *SrCyp17*	1.880	1.903	34.721
Const_2h_2 *SrCyp17*	1.899	1.903	34.044
Const_2h_3 *SrCyp17*	1.915	1.903	34.426
Const_2h_4 *SrCyp17*	1.907	1.903	33.841
Const_24h_1 *SrCyp17*	1.906	1.903	30.440
Const_24h_2 *SrCyp17*	1.912	1.903	31.042
Const_24h_3 *SrCyp17*	1.911	1.903	30.486
Const_24h_4 *SrCyp17*	1.903	1.903	30.359
D/n_1 *SrCyp17*	1.879	1.903	30.594
D/n_2 *SrCyp17*	1.895	1.903	31.232
D/n_3 *SrCyp17*	1.918	1.903	31.169
D/n_4 *SrCyp17*	1.897	1.903	29.914
SrCyp17 NTC	1.000	0.000	0.000
SrCyp17 NTC	1.000	0.000	0.000

Table A 18 LinReg results for the target gene *SrCyp18* in RT-qPCR

Sample	Individual PCR-efficiency	Mean PCR-efficiency	Cq-value
Const_0h_ctrl_1 *SrCyp18*	1.976	1.975	29.823
Const_0h_ctrl_2 *SrCyp18*	1.961	1.975	30.107
Const_0h_ctrl_3 *SrCyp18*	1.977	1.975	29.171
Const_0h_ctrl_4 *SrCyp18*	1.963	1.975	28.175
Const_2h_1 *SrCyp18*	1.977	1.975	30.712
Const_2h_2 *SrCyp18*	1.986	1.975	30.692
Const_2h_3 *SrCyp18*	1.963	1.975	30.776
Const_2h_4 *SrCyp18*	1.980	1.975	30.362
Const_24h_1 *SrCyp18*	1.999	1.975	30.451
Const_24h_2 *SrCyp18*	1.962	1.975	31.220
Const_24h_3 *SrCyp18*	1.962	1.975	31.026
Const_24h_4 *SrCyp18*	2.010	1.975	30.489
D/n_1 *SrCyp18*	1.991	1.975	26.689
D/n_2 *SrCyp18*	1.932	1.975	27.041
D/n_3 *SrCyp18*	2.011	1.975	27.645
D/n_4 *SrCyp18*	1.949	1.975	26.120
SrCyp18 NTC	1.000	0.000	0.000
SrCyp18 NTC	1.960	0.000	39.976

Table A 19 LinReg results for the target gene *SrCyp20* in RT-qPCR

Sample	Individual PCR-efficiency	Mean PCR-efficiency	Cq-value
Const_0h_ctrl_1 *SrCyp20*	1.915	1.922	31.118
Const_0h_ctrl_2 *SrCyp20*	1.931	1.922	31.246
Const_0h_ctrl_3 *SrCyp20*	1.913	1.922	30.759
Const_0h_ctrl_4 *SrCyp20*	1.928	1.922	29.852
Const_2h_1 *SrCyp20*	1.921	1.922	31.316
Const_2h_2 *SrCyp20*	1.924	1.922	31.077
Const_2h_3 *SrCyp20*	1.911	1.922	31.076
Const_2h_4 *SrCyp20*	1.915	1.922	30.819
Const_24h_1 *SrCyp20*	1.933	1.922	28.325
Const_24h_2 *SrCyp20*	1.908	1.922	27.993
Const_24h_3 *SrCyp20*	1.939	1.922	27.579
Const_24h_4 *SrCyp20*	1.920	1.922	28.017
D/n_1 *SrCyp20*	1.932	1.922	28.692
D/n_2 *SrCyp20*	1.909	1.922	27.425
D/n_3 *SrCyp20*	1.913	1.922	28.217
D/n_4 *SrCyp20*	1.935	1.922	29.085
SrCyp20 NTC	1.970	0.000	40.464
SrCyp20 NTC	1.986	0.000	39.828

Table A 20 LinReg results for the target gene *SrCyp21* in RT-qPCR

Sample	Individual PCR-efficiency	Mean PCR-efficiency	Cq-value
Const_0h_ctrl_1 *SrCyp21*	2.022	1.960	34.813
Const_0h_ctrl_2 *SrCyp21*	1.958	1.960	35.161
Const_0h_ctrl_3 *SrCyp21*	1.941	1.960	34.371
Const_0h_ctrl_4 *SrCyp21*	1.987	1.960	34.578
Const_2h_1 *SrCyp21*	2.026	1.960	31.661
Const_2h_2 *SrCyp21*	2.060	1.960	31.740
Const_2h_3 *SrCyp21*	2.018	1.960	31.557
Const_2h_4 *SrCyp21*	2.039	1.960	31.732
Const_24h_1 *SrCyp21*	1.938	1.960	29.015
Const_24h_2 *SrCyp21*	1.970	1.960	28.340
Const_24h_3 *SrCyp21*	1.935	1.960	28.036
Const_24h_4 *SrCyp21*	1.962	1.960	28.932
D/n_1 *SrCyp21*	1.940	1.960	32.985
D/n_2 *SrCyp21*	1.990	1.960	32.513
D/n_3 *SrCyp21*	1.978	1.960	32.451
D/n_4 *SrCyp21*	1.960	1.960	32.339
SrCyp21 NTC	1.890	2.029	44.400
SrCyp21 NTC	1.000	2.029	0.000

Table A 21 LinReg results for the target gene *SrCyp22* in RT-qPCR

Sample	Individual PCR-efficiency	Mean PCR-efficiency	Cq-value
Const_0h_ctrl_1 *SrCyp22*	2.039	1.994	25.746
Const_0h_ctrl_2 *SrCyp22*	2.060	1.994	25.619
Const_0h_ctrl_3 *SrCyp22*	1.997	1.994	25.334
Const_0h_ctrl_4 *SrCyp22*	1.976	1.994	24.347
Const_2h_1 *SrCyp22*	1.995	1.994	30.296
Const_2h_2 *SrCyp22*	2.036	1.994	29.602
Const_2h_3 *SrCyp22*	1.975	1.994	30.176
Const_2h_4 *SrCyp22*	1.953	1.994	29.007
Const_24h_1 *SrCyp22*	1.992	1.994	25.882
Const_24h_2 *SrCyp22*	1.960	1.994	25.995
Const_24h_3 *SrCyp22*	1.986	1.994	25.364
Const_24h_4 *SrCyp22*	2.042	1.994	24.721
D/n_1 *SrCyp22*	2.003	1.994	27.612
D/n_2 *SrCyp22*	2.023	1.994	28.817
D/n_3 *SrCyp22*	1.970	1.994	28.238
D/n_4 *SrCyp22*	1.965	1.994	27.104
SrCyp22 NTC	1.000	0.000	0.000
SrCyp22 NTC	1.000	0.000	0.000

Table A 22 LinReg results for the target gene *SrCyp23* in RT-qPCR

Sample	Individual PCR-efficiency	Mean PCR-efficiency	Cq-value
Const_0h_ctrl_1 *SrCyp23*	2.104	2.086	32.763
Const_0h_ctrl_2 *SrCyp23*	2.065	2.086	32.154
Const_0h_ctrl_3 *SrCyp23*	1.909	2.086	31.590
Const_0h_ctrl_4 *SrCyp23*	2.175	2.086	31.385
Const_2h_1 *SrCyp23*	2.043	2.086	30.784
Const_2h_2 *SrCyp23*	2.073	2.086	31.177
Const_2h_3 *SrCyp23*	2.029	2.086	31.003
Const_2h_4 *SrCyp23*	2.081	2.086	31.298
Const_24h_1 *SrCyp23*	2.096	2.086	29.327
Const_24h_2 *SrCyp23*	2.091	2.086	29.176
Const_24h_3 *SrCyp23*	2.161	2.086	28.485
Const_24h_4 *SrCyp23*	2.133	2.086	28.728
D/n_1 *SrCyp23*	1.988	2.086	28.188
D/n_2 *SrCyp23*	2.058	2.086	28.744
D/n_3 *SrCyp23*	2.118	2.086	29.527
D/n_4 *SrCyp23*	2.090	2.086	27.860
SrCyp23 NTC	2.066	2.083	35.101
SrCyp23 NTC	2.101	2.083	35.422

Table A 23 LinReg results for the target gene *SrCyp28* in RT-qPCR

Sample	Individual PCR-efficiency	Mean PCR-efficiency	Cq-value
Const_0h_ctrl_1 *SrCyp28*	1.982	1.948	30.616
Const_0h_ctrl_2 *SrCyp28*	1.974	1.948	30.623
Const_0h_ctrl_3 *SrCyp28*	1.912	1.948	29.981
Const_0h_ctrl_4 *SrCyp28*	1.898	1.948	28.917
Const_2h_1 *SrCyp28*	1.910	1.948	30.991
Const_2h_2 *SrCyp28*	1.978	1.948	30.632
Const_2h_3 *SrCyp28*	1.917	1.948	31.057
Const_2h_4 *SrCyp28*	1.928	1.948	30.117
Const_24h_1 *SrCyp28*	1.971	1.948	27.571
Const_24h_2 *SrCyp28*	1.923	1.948	28.026
Const_24h_3 *SrCyp28*	1.950	1.948	27.462
Const_24h_4 *SrCyp28*	1.968	1.948	27.828
D/n_1 *SrCyp28*	1.968	1.948	29.711
D/n_2 *SrCyp28*	1.915	1.948	29.906
D/n_3 *SrCyp28*	1.960	1.948	29.631
D/n_4 *SrCyp28*	1.957	1.948	29.556
SrCyp28 NTC	1.971	1.983	39.534
SrCyp28 NTC	1.888	1.983	44.475

Table A 24 LinReg results for the target gene *SrCyp44* in RT-qPCR

Sample	Individual PCR-efficiency	Mean PCR-efficiency	Cq-value
Const_0h_ctrl_1 *SrCyp44*	1.789	1.801	32.033
Const_0h_ctrl_2 *SrCyp44*	1.793	1.801	31.484
Const_0h_ctrl_3 *SrCyp44*	1.812	1.801	31.378
Const_0h_ctrl_4 *SrCyp44*	1.808	1.801	30.392
Const_2h_1 *SrCyp44*	1.797	1.801	33.530
Const_2h_2 *SrCyp44*	1.744	1.801	34.762
Const_2h_3 *SrCyp44*	1.801	1.801	35.126
Const_2h_4 *SrCyp44*	1.798	1.801	34.423
Const_24h_1 *SrCyp44*	1.802	1.801	30.808
Const_24h_2 *SrCyp44*	1.806	1.801	31.611
Const_24h_3 *SrCyp44*	1.809	1.801	30.936
Const_24h_4 *SrCyp44*	1.815	1.801	30.248
D/n_1 *SrCyp44*	1.806	1.801	29.641
D/n_2 *SrCyp44*	1.786	1.801	31.048
D/n_3 *SrCyp44*	1.781	1.801	31.991
D/n_4 *SrCyp44*	1.815	1.801	29.669
SrCyp44 NTC	1.000	0.000	0.000
SrCyp44 NTC	1.000	0.000	0.000

Table A 25 LinReg results for the target gene *SrCyp46* in RT-qPCR

Sample	Individual PCR-efficiency	Mean PCR-efficiency	Cq-value
Const_0h_ctrl_1 *SrCyp46*	1.903	1.943	32.092
Const_0h_ctrl_2 *SrCyp46*	1.980	1.943	31.768
Const_0h_ctrl_3 *SrCyp46*	1.932	1.943	31.370
Const_0h_ctrl_4 *SrCyp46*	1.966	1.943	30.762
Const_2h_1 *SrCyp46*	1.954	1.943	32.583
Const_2h_2 *SrCyp46*	1.936	1.943	33.792
Const_2h_3 *SrCyp46*	1.927	1.943	34.158
Const_2h_4 *SrCyp46*	1.935	1.943	34.219
Const_24h_1 *SrCyp46*	1.956	1.943	32.333
Const_24h_2 *SrCyp46*	1.956	1.943	31.584
Const_24h_3 *SrCyp46*	1.952	1.943	31.383
Const_24h_4 *SrCyp46*	1.979	1.943	32.520
D/n_1 *SrCyp46*	1.927	1.943	30.410
D/n_2 *SrCyp46*	1.955	1.943	30.621
D/n_3 *SrCyp46*	1.914	1.943	32.070
D/n_4 *SrCyp46*	1.913	1.943	30.181
SrCyp46 NTC	1.000	0.000	0.000
SrCyp46 NTC	1.000	0.000	0.000

Table A 26 LinReg results for the target gene *SrCyp52* in RT-qPCR

Sample	Individual PCR-efficiency	Mean PCR-efficiency	Cq-value
Const_0h_ctrl_1 *SrCyp52*	1.972	1.950	32.374
Const_0h_ctrl_2 *SrCyp52*	1.982	1.950	31.744
Const_0h_ctrl_3 *SrCyp52*	1.994	1.950	30.764
Const_0h_ctrl_4 *SrCyp52*	1.958	1.950	30.719
Const_2h_1 *SrCyp52*	1.914	1.950	32.929
Const_2h_2 *SrCyp52*	1.973	1.950	31.735
Const_2h_3 *SrCyp52*	1.925	1.950	33.006
Const_2h_4 *SrCyp52*	1.926	1.950	31.002
Const_24h_1 *SrCyp52*	1.919	1.950	30.142
Const_24h_2 *SrCyp52*	1.908	1.950	31.127
Const_24h_3 *SrCyp52*	1.951	1.950	31.336
Const_24h_4 *SrCyp52*	1.996	1.950	30.715
D/n_1 *SrCyp52*	1.934	1.950	30.390
D/n_2 *SrCyp52*	2.019	1.950	29.782
D/n_3 *SrCyp52*	1.925	1.950	30.338
D/n_4 *SrCyp52*	1.978	1.950	30.622
SrCyp52 NTC	1.995	0.000	39.671
SrCyp52 NTC	1.924	0.000	43.728

Table A 27 LinReg results for the target gene *SrCyp54* in RT-qPCR

Sample	Individual PCR-efficiency	Mean PCR-efficiency	Cq-value
Const_0h_ctrl_1 *SrCyp54*	1.921	1.932	33.019
Const_0h_ctrl_2 *SrCyp54*	1.912	1.932	31.035
Const_0h_ctrl_3 *SrCyp54*	1.983	1.932	30.548
Const_0h_ctrl_4 *SrCyp54*	1.935	1.932	31.166
Const_2h_1 *SrCyp54*	1.900	1.932	30.910
Const_2h_2 *SrCyp54*	1.928	1.932	32.034
Const_2h_3 *SrCyp54*	1.974	1.932	31.482
Const_2h_4 *SrCyp54*	1.871	1.932	31.053
Const_24h_1 *SrCyp54*	1.992	1.932	32.646
Const_24h_2 *SrCyp54*	1.951	1.932	34.292
Const_24h_3 *SrCyp54*	1.929	1.932	32.944
Const_24h_4 *SrCyp54*	1.991	1.932	32.233
D/n_1 *SrCyp54*	1.945	1.932	29.613
D/n_2 *SrCyp54*	2.002	1.932	28.678
D/n_3 *SrCyp54*	1.934	1.932	28.660
D/n_4 *SrCyp54*	1.920	1.932	29.008
SrCyp54 NTC	1.970	0.000	43.235
SrCyp54 NTC	1.989	0.000	42.842

Table A 28 LinReg results for the target gene *SrCyp70* in RT-qPCR

Sample	Individual PCR-efficiency	Mean PCR-efficiency	Cq-value
Const_0h_ctrl_1 *SrCyp70*	1.904	1.926	31.109
Const_0h_ctrl_2 *SrCyp70*	1.923	1.926	31.523
Const_0h_ctrl_3 *SrCyp70*	1.929	1.926	30.357
Const_0h_ctrl_4 *SrCyp70*	1.918	1.926	29.437
Const_2h_1 *SrCyp70*	1.939	1.926	30.663
Const_2h_2 *SrCyp70*	1.899	1.926	30.172
Const_2h_3 *SrCyp70*	1.927	1.926	30.340
Const_2h_4 *SrCyp70*	1.943	1.926	30.618
Const_24h_1 *SrCyp70*	1.898	1.926	28.047
Const_24h_2 *SrCyp70*	1.905	1.926	28.066
Const_24h_3 *SrCyp70*	1.949	1.926	27.650
Const_24h_4 *SrCyp70*	1.965	1.926	27.625
D/n_1 *SrCyp70*	1.924	1.926	29.572
D/n_2 *SrCyp70*	1.910	1.926	31.305
D/n_3 *SrCyp70*	1.939	1.926	31.651
D/n_4 *SrCyp70*	1.939	1.926	28.532
SrCyp70 NTC	1.968	1.994	39.489
SrCyp70 NTC	1.991	1.994	39.735

Table A 29 LinReg results for the target gene *SrCyp73* in RT-qPCR

Sample	Individual PCR-efficiency	Mean PCR-efficiency	Cq-value
Const_0h_ctrl_1 *SrCyp73*	1.930	1.949	32.160
Const_0h_ctrl_2 *SrCyp73*	1.924	1.949	32.170
Const_0h_ctrl_3 *SrCyp73*	1.990	1.949	30.724
Const_0h_ctrl_4 *SrCyp73*	1.922	1.949	30.020
Const_2h_1 *SrCyp73*	1.954	1.949	31.738
Const_2h_2 *SrCyp73*	1.987	1.949	30.764
Const_2h_3 *SrCyp73*	1.878	1.949	31.949
Const_2h_4 *SrCyp73*	1.927	1.949	30.095
Const_24h_1 *SrCyp73*	1.942	1.949	29.213
Const_24h_2 *SrCyp73*	1.998	1.949	29.760
Const_24h_3 *SrCyp73*	1.967	1.949	29.383
Const_24h_4 *SrCyp73*	1.993	1.949	29.593
D/n_1 *SrCyp73*	1.959	1.949	29.581
D/n_2 *SrCyp73*	1.956	1.949	30.473
D/n_3 *SrCyp73*	1.906	1.949	29.989
D/n_4 *SrCyp73*	1.934	1.949	29.386
SrCyp73 NTC	1.000	0.000	0.000
SrCyp73 NTC	1.000	0.000	0.000

A4. MELTING TEMPERATURES

Table A 30 Melting temperatures (°C) for the reference gene *qSrVps35* in RT-qPCR.

Name	Tm1	Tm2
Const_0h_ctrl_1 *qSrVps35*	82.60	
Const_0h_ctrl_2 *qSrVps35*	82.45	
Const_0h_ctrl_3 *qSrVps35*	82.53	
Const_0h_ctrl_4 *qSrVps35*	82.70	
Const_2h_1 *qSrVps35*	82.58	
Const_2h_2 *qSrVps35*	82.62	
Const_2h_3 *qSrVps35*	82.66	
Const_2h_4 *qSrVps35*	82.77	
Const_24h_1 *qSrVps35*	82.84	
Const_24h_2 *qSrVps35*	82.73	
Const_24h_3 *qSrVps35*	82.55	
Const_24h_4 *qSrVps35*	82.66	
D/n_1 *qSrVps35*	82.65	
D/n_2 *qSrVps35*	82.55	
D/n_3 *qSrVps35*	82.53	
D/n_4 *qSrVps35*	82.54	
qSrVps35 NTC	74.59	
qSrVps35 NTC	82.75	

Table A 31 Melting temperatures (°C) for the reference gene *qSrMFS* in RT-qPCR.

Name	Tm1	Tm2
Const_0h_1 *qSrMFS*	83.69	
Const_0h_2 *qSrMFS*	83.57	
Const_0h_3 *qSrMFS*	83.66	
Const_0h_4 *qSrMFS*	83.76	
Const_2h_1 *qSrMFS*	83.73	
Const_2h_2 *qSrMFS*	83.73	
Const_2h_3 *qSrMFS*	83.73	
Const_2h_4 *qSrMFS*	83.84	
Const_24h_1 *qSrMFS*	83.79	
Const_24h_2 *qSrMFS*	83.82	
Const_24h_3 *qSrMFS*	84.18	
Const_24h_4 *qSrMFS*	83.86	
D/n_1 *qSrMFS*	83.71	
D/n_2 *qSrMFS*	83.65	
D/n_3 *qSrMFS*	83.66	
D/n_4 *qSrMFS*	83.67	
qSrMFS NTC	72.70	
qSrMFS NTC	77.80	

Table A 32 Melting temperatures (°C) for the control without reverse transcriptase in RT-qPCR. The primerset for the target gene *Cyp86A3* is used as this does not contain any introns.

Name	Tm1	Tm2
Const_0h_1 *Cyp86A3* -RT	73,11	
Const_0h_2 *Cyp86A3* -RT	76,29	
Const_0h_3 *Cyp86A3* -RT	78,92	
Const_0h_4 *Cyp86A3* -RT	78,28	
Const_2h_1 *Cyp86A3* -RT	76,56	
Const_2h_2 *Cyp86A3* -RT		
Const_2h_3 *Cyp86A3* -RT	76,52	
Const_2h_4 *Cyp86A3* -RT		
Const_24h_1 *Cyp86A3* -RT	77,64	
Const_24h_2 *Cyp86A3* -RT	76,24	
Const_24h_3 *Cyp86A3* -RT	76,15	
Const_24h_4 *Cyp86A3* -RT		
D/n_1 *Cyp86A3* -RT		
D/n_2 *Cyp86A3* -RT		
D/n_3 *Cyp86A3* -RT	71,09	
D/n_4 *Cyp86A3* -RT	79,44	

Table A 33 Melting temperatures (°C) for the target gene *Cyp19A1* in RT-qPCR.

Name	Tm1	Tm2
Const_0h_ctrl_1 *Cyp19A1*	79.82	
Const_0h_ctrl_2 *Cyp19A1*	79.83	
Const_0h_ctrl_3 *Cyp19A1*	79.89	
Const_0h_ctrl_4 *Cyp19A1*	79.82	
Const_2h_1 *Cyp19A1*	79.88	
Const_2h_2 *Cyp19A1*	79.89	
Const_2h_3 *Cyp19A1*	79.88	
Const_2h_4 *Cyp19A1*	79.99	
Const_24h_1 *Cyp19A1*	79.96	
Const_24h_2 *Cyp19A1*	79.90	
Const_24h_3 *Cyp19A1*	80.23	
Const_24h_4 *Cyp19A1*	79.93	
D/n_1 *Cyp19A1*	79.69	
D/n_2 *Cyp19A1*	79.83	
D/n_3 *Cyp19A1*	79.85	
D/n_4 *Cyp19A1*	79.89	
Cyp19A1 NTC		
Cyp19A1 NTC	70.63	

Table A 34 Melting temperatures (°C) for the target gene *Cyp51A1* in RT-qPCR.

Name	Tm1	Tm2
Const_0h_ctrl_1 *Cyp51A1*	86.07	
Const_0h_ctrl_2 *Cyp51A1*	86.08	
Const_0h_ctrl_3 *Cyp51A1*	86.11	
Const_0h_ctrl_4 *Cyp51A1*	86.17	
Const_2h_1 *Cyp51A1*	86.13	
Const_2h_2 *Cyp51A1*	86.12	
Const_2h_3 *Cyp51A1*	86.36	
Const_2h_4 *Cyp51A1*	86.28	
Const_24h_1 *Cyp51A1*	86.35	
Const_24h_2 *Cyp51A1*	86.18	
Const_24h_3 *Cyp51A1*	86.27	
Const_24h_4 *Cyp51A1*	86.25	
D/n_1 *Cyp51A1*	85.99	
D/n_2 *Cyp51A1*	86.08	
D/n_3 *Cyp51A1*	86.06	
D/n_4 *Cyp51A1*	86.12	
Cyp51A1 NTC	70.38	82.01
Cyp51A1 NTC	81.03	

Table A 35 Melting temperatures (°C) for the target gene *Cyp51B1+2* in RT-qPCR.

Name	Tm1	Tm2
Const_0h_ctrl_1 *Cyp51B1+2*	84.16	
Const_0h_ctrl_2 *Cyp51B1+2*	84.13	
Const_0h_ctrl_3 *Cyp51B1+2*	84.25	
Const_0h_ctrl_4 *Cyp51B1+2*	84.15	
Const_2h_1 *Cyp51B1+2*	83.90	
Const_2h_2 *Cyp51B1+2*	84.11	
Const_2h_3 *Cyp51B1+2*	84.11	
Const_2h_4 *Cyp51B1+2*	84.22	
Const_24h_1 *Cyp51B1+2*	84.41	
Const_24h_2 *Cyp51B1+2*	84.16	
Const_24h_3 *Cyp51B1+2*	84.42	
Const_24h_4 *Cyp51B1+2*	84.07	
D/n_1 *Cyp51B1+2*	84.09	
D/n_2 *Cyp51B1+2*	84.11	
D/n_3 *Cyp51B1+2*	84.07	
D/n_4 *Cyp51B1+2*	84.01	
Cyp51B1+2 NTC	79.54	
Cyp51B1+2 NTC	71.96	79.64

Table A 36 Melting temperatures (°C) for the target gene *Cyp86A2* in RT-qPCR.

Name	Tm1	Tm2
Const_0h_ctrl_1 *Cyp86A2*	79.78	
Const_0h_ctrl_2 *Cyp86A2*	79.75	
Const_0h_ctrl_3 *Cyp86A2*	79.92	
Const_0h_ctrl_4 *Cyp86A2*	79.85	
Const_2h_1 *Cyp86A2*	79.84	
Const_2h_2 *Cyp86A2*	79.89	
Const_2h_3 *Cyp86A2*	79.92	
Const_2h_4 *Cyp86A2*	80.04	
Const_24h_1 *Cyp86A2*	80.08	
Const_24h_2 *Cyp86A2*	79.94	
Const_24h_3 *Cyp86A2*	79.95	
Const_24h_4 *Cyp86A2*	79.84	
D/n_1 *Cyp86A2*	79.68	
D/n_2 *Cyp86A2*	79.75	
D/n_3 *Cyp86A2*	79.80	
D/n_4 *Cyp86A2*	79.76	
Cyp86A2 NTC	69.96	78.94
Cyp86A2 NTC	70.79	

Table A 37 Melting temperatures (°C) for the target gene *Cyp86A3* in RT-qPCR.

Name	Tm1	Tm2
Const_0h_ctrl_1 *Cyp86A3*	83.64	
Const_0h_ctrl_2 *Cyp86A3*	83.74	
Const_0h_ctrl_3 *Cyp86A3*	83.88	
Const_0h_ctrl_4 *Cyp86A3*	83.70	
Const_2h_1 *Cyp86A3*	83.76	
Const_2h_2 *Cyp86A3*	83.80	
Const_2h_3 *Cyp86A3*	83.82	
Const_2h_4 *Cyp86A3*	83.94	
Const_24h_1 *Cyp86A3*	84.03	
Const_24h_2 *Cyp86A3*	83.84	
Const_24h_3 *Cyp86A3*	84.09	
Const_24h_4 *Cyp86A3*	83.81	
D/n_1 *Cyp86A3*	83.66	
D/n_2 *Cyp86A3*	83.70	
D/n_3 *Cyp86A3*	83.78	
D/n_4 *Cyp86A3*	83.66	
Cyp86A3 NTC	71.02	
Cyp86A3 NTC	77.73	

Table A 38 Melting temperatures (°C) for the target gene *Cyp86A6* in RT-qPCR.

Name	Tm1	Tm2
Const_0h_ctrl_1 *Cyp86A6*	81.62	
Const_0h_ctrl_2 *Cyp86A6*	81.58	
Const_0h_ctrl_3 *Cyp86A6*	81.43	
Const_0h_ctrl_4 *Cyp86A6*	81.54	
Const_2h_1 *Cyp86A6*	81.65	
Const_2h_2 *Cyp86A6*	81.69	
Const_2h_3 *Cyp86A6*	81.71	
Const_2h_4 *Cyp86A6*	81.83	
Const_24h_1 *Cyp86A6*	81.86	
Const_24h_2 *Cyp86A6*	81.81	
Const_24h_3 *Cyp86A6*	81.79	
Const_24h_4 *Cyp86A6*	81.78	
D/n_1 *Cyp86A6*	81.49	
D/n_2 *Cyp86A6*	81.56	
D/n_3 *Cyp86A6*	81.60	
D/n_4 *Cyp86A6*	81.66	
Cyp86A6 NTC	79.31	
Cyp86A6 NTC	70.03	78.10

Table A 39 Melting temperatures (°C) for the target gene *Cyp97A1* in RT-qPCR.

Name	Tm1	Tm2
Const_0h_ctrl_1 *Cyp97A1*	82.84	
Const_0h_ctrl_2 *Cyp97A1*	82.99	
Const_0h_ctrl_3 *Cyp97A1*	82.99	
Const_0h_ctrl_4 *Cyp97A1*	83.06	
Const_2h_1 *Cyp97A1*	83.11	
Const_2h_2 *Cyp97A1*	83.17	
Const_2h_3 *Cyp97A1*	83.17	
Const_2h_4 *Cyp97A1*	83.25	
Const_24h_1 *Cyp97A1*	83.35	
Const_24h_2 *Cyp97A1*	83.48	
Const_24h_3 *Cyp97A1*	83.25	
Const_24h_4 *Cyp97A1*	83.12	
D/n_1 *Cyp97A1*	82.92	
D/n_2 *Cyp97A1*	82.88	
D/n_3 *Cyp97A1*	82.81	
D/n_4 *Cyp97A1*	82.99	
Cyp97A1 NTC	75.74	
Cyp97A1 NTC	70.30	81.94

Table A 40 Melting temperatures (°C) for the target gene *Cyp97B1* in RT-qPCR.

Name	Tm1	Tm2
Const_0h_ctrl_1 *Cyp97B1*	84.29	
Const_0h_ctrl_2 *Cyp97B1*	84.22	
Const_0h_ctrl_3 *Cyp97B1*	84.45	
Const_0h_ctrl_4 *Cyp97B1*	84.32	
Const_2h_1 *Cyp97B1*	84.35	
Const_2h_2 *Cyp97B1*	84.35	
Const_2h_3 *Cyp97B1*	84.33	
Const_2h_4 *Cyp97B1*	84.53	
Const_24h_1 *Cyp97B1*	84.58	
Const_24h_2 *Cyp97B1*	84.42	
Const_24h_3 *Cyp97B1*	84.65	
Const_24h_4 *Cyp97B1*	84.36	
D/n_1 *Cyp97B1*	84.26	
D/n_2 *Cyp97B1*	84.28	
D/n_3 *Cyp97B1*	84.33	
D/n_4 *Cyp97B1*	84.29	
Cyp97B1 NTC	77.84	
Cyp97B1 NTC	77.74	

Table A 41 Melting temperatures (°C) for the target gene *SrCyp10* in RT-qPCR.

Name	Tm1	Tm2
Const_0h_ctrl_1 *SrCyp10*	85.66	
Const_0h_ctrl_2 *SrCyp10*	85.74	
Const_0h_ctrl_3 *SrCyp10*	85.74	
Const_0h_ctrl_4 *SrCyp10*	85.74	
Const_2h_1 *SrCyp10*	85.88	
Const_2h_2 *SrCyp10*	85.91	
Const_2h_3 *SrCyp10*	85.97	
Const_2h_4 *SrCyp10*	86.10	
Const_24h_1 *SrCyp10*	86.11	
Const_24h_2 *SrCyp10*	85.99	
Const_24h_3 *SrCyp10*	86.03	
Const_24h_4 *SrCyp10*	85.98	
D/n_1 *SrCyp10*	85.69	
D/n_2 *SrCyp10*	85.65	
D/n_3 *SrCyp10*	85.71	
D/n_4 *SrCyp10*	85.79	
SrCyp10 NTC	85.98	
SrCyp10 NTC	86.49	

Table A 42 Melting temperatures (°C) for the target gene *SrCyp12* in RT-qPCR. Two primersets are included for this gene.

Name	Tm1	Tm2
Const_0h_ctrl_1 *SrCyp12_1*	86.59	
Const_0h_ctrl_2 *SrCyp12_1*	86.32	
Const_0h_ctrl_3 *SrCyp12_1*	86.77	
Const_0h_ctrl_4 *SrCyp12_1*	86.65	
Const_2h_1 *SrCyp12_1*	86.65	
Const_2h_2 *SrCyp12_1*	86.68	
Const_2h_3 *SrCyp12_1*	86.67	
Const_2h_4 *SrCyp12_1*	86.79	
Const_24h_1 *SrCyp12_1*	86.87	
Const_24h_2 *SrCyp12_1*	86.73	
Const_24h_3 *SrCyp12_1*	86.99	
Const_24h_4 *SrCyp12_1*	86.62	
D/n_1 *SrCyp12_1*	86.61	
D/n_2 *SrCyp12_1*	86.66	
D/n_3 *SrCyp12_1*	86.58	
D/n_4 *SrCyp12_1*	86.63	
SrCyp12_1 NTC	81.86	
SrCyp12_1 NTC	75.07	
Const_0h_ctrl_1 *SrCyp12_2*	83.73	
Const_0h_ctrl_2 *SrCyp12_2*	83.73	
Const_0h_ctrl_3 *SrCyp12_2*	83.69	
Const_0h_ctrl_4 *SrCyp12_2*	83.84	
Const_2h_1 *SrCyp12_2*	83.73	
Const_2h_2 *SrCyp12_2*	83.70	
Const_2h_3 *SrCyp12_2*	83.74	
Const_2h_4 *SrCyp12_2*	83.76	
Const_24h_1 *SrCyp12_2*	83.96	
Const_24h_2 *SrCyp12_2*	83.72	
Const_24h_3 *SrCyp12_2*	83.82	
Const_24h_4 *SrCyp12_2*	83.78	
D/n_1 *SrCyp12_2*	83.72	
D/n_2 *SrCyp12_2*	83.68	
D/n_3 *SrCyp12_2*	83.60	
D/n_4 *SrCyp12_2*	83.71	
SrCyp12_2 NTC	80.27	
SrCyp12_2 NTC	75.25	

Table A 43 Melting temperatures (°C) for the target gene *SrCyp13* in RT-qPCR. Two primersets are included for this gene.

Name	Tm1	Tm2
Const_0h_ctrl_1 *SrCyp13_1*	83.44	
Const_0h_ctrl_2 *SrCyp13_1*	83.55	
Const_0h_ctrl_3 *SrCyp13_1*	83.48	
Const_0h_ctrl_4 *SrCyp13_1*	83.42	
Const_2h_1 *SrCyp13_1*	83.61	
Const_2h_2 *SrCyp13_1*	83.60	
Const_2h_3 *SrCyp13_1*	83.63	
Const_2h_4 *SrCyp13_1*	83.77	
Const_24h_1 *SrCyp13_1*	83.53	
Const_24h_2 *SrCyp13_1*	83.66	
Const_24h_3 *SrCyp13_1*	83.67	
Const_24h_4 *SrCyp13_1*	83.46	
D/n_1 *SrCyp13_1*	83.63	
D/n_2 *SrCyp13_1*	83.46	
D/n_3 *SrCyp13_1*	83.39	
D/n_4 *SrCyp13_1*	83.55	
SrCyp13_1 NTC		
SrCyp13_1 NTC		
Const_0h_ctrl_1 *SrCyp13_2*	81.77	
Const_0h_ctrl_2 *SrCyp13_2*	81.68	
Const_0h_ctrl_3 *SrCyp13_2*	81.97	
Const_0h_ctrl_4 *SrCyp13_2*	81.82	
Const_2h_1 *SrCyp13_2*	81.82	
Const_2h_2 *SrCyp13_2*	81.88	
Const_2h_3 *SrCyp13_2*	81.78	
Const_2h_4 *SrCyp13_2*	81.98	
Const_24h_1 *SrCyp13_2*	82.09	
Const_24h_2 *SrCyp13_2*	82.18	
Const_24h_3 *SrCyp13_2*	81.82	
Const_24h_4 *SrCyp13_2*	81.95	
D/n_1 *SrCyp13_2*	81.75	
D/n_2 *SrCyp13_2*	81.81	
D/n_3 *SrCyp13_2*	81.67	
D/n_4 *SrCyp13_2*	81.86	
SrCyp13_2 NTC		
SrCyp13_2 NTC		

Table A 44 Melting temperatures (°C) for the target gene *SrCyp16* in RT-qPCR.

Name	Tm1	Tm2
Const_0h_ctrl_1 *SrCyp16*	86.52	
Const_0h_ctrl_2 *SrCyp16*	86.52	
Const_0h_ctrl_3 *SrCyp16*	86.53	
Const_0h_ctrl_4 *SrCyp16*	86.66	
Const_2h_1 *SrCyp16*	86.62	
Const_2h_2 *SrCyp16*	86.60	
Const_2h_3 *SrCyp16*	86.65	
Const_2h_4 *SrCyp16*	86.74	
Const_24h_1 *SrCyp16*	86.79	
Const_24h_2 *SrCyp16*	86.75	
Const_24h_3 *SrCyp16*	86.74	
Const_24h_4 *SrCyp16*	86.71	
D/n_1 *SrCyp16*	86.55	
D/n_2 *SrCyp16*	86.51	
D/n_3 *SrCyp16*	86.53	
D/n_4 *SrCyp16*	86.62	
SrCyp16 NTC		
SrCyp16 NTC	86.56	

Table A 45 Melting temperatures (°C) for the target gene *SrCyp17* in RT-qPCR.

Name	Tm1	Tm2
Const_0h_ctrl_1 *SrCyp17*	80.62	
Const_0h_ctrl_2 *SrCyp17*	80.72	
Const_0h_ctrl_3 *SrCyp17*	80.68	
Const_0h_ctrl_4 *SrCyp17*	80.70	
Const_2h_1 *SrCyp17*	80.82	
Const_2h_2 *SrCyp17*	80.84	
Const_2h_3 *SrCyp17*	80.80	
Const_2h_4 *SrCyp17*	80.98	
Const_24h_1 *SrCyp17*	80.98	
Const_24h_2 *SrCyp17*	80.91	
Const_24h_3 *SrCyp17*	80.68	
Const_24h_4 *SrCyp17*	80.92	
D/n_1 *SrCyp17*	80.62	
D/n_2 *SrCyp17*	80.49	
D/n_3 *SrCyp17*	80.66	
D/n_4 *SrCyp17*	80.70	
SrCyp17 NTC	69.93	
SrCyp17 NTC	69.93	

Table A 46 Melting temperatures (°C) for the target gene *SrCyp18* in RT-qPCR.

Name	Tm1	Tm2
Const_0h_ctrl_1 *SrCyp18*	82.25	
Const_0h_ctrl_2 *SrCyp18*	82.37	
Const_0h_ctrl_3 *SrCyp18*	82.32	
Const_0h_ctrl_4 *SrCyp18*	82.47	
Const_2h_1 *SrCyp18*	82.64	
Const_2h_2 *SrCyp18*	82.53	
Const_2h_3 *SrCyp18*	82.52	
Const_2h_4 *SrCyp18*	82.70	
Const_24h_1 *SrCyp18*	82.69	
Const_24h_2 *SrCyp18*	82.55	
Const_24h_3 *SrCyp18*	82.60	
Const_24h_4 *SrCyp18*	82.36	
D/n_1 *SrCyp18*	82.28	
D/n_2 *SrCyp18*	82.30	
D/n_3 *SrCyp18*	82.33	
D/n_4 *SrCyp18*	82.45	
SrCyp18 NTC	70.99	
SrCyp18 NTC	81.80	

Table A 47 Melting temperatures (°C) for the target gene *SrCyp20* in RT-qPCR.

Name	Tm1	Tm2
Const_0h_ctrl_1 *SrCyp20*	84.10	
Const_0h_ctrl_2 *SrCyp20*	84.07	
Const_0h_ctrl_3 *SrCyp20*	84.14	
Const_0h_ctrl_4 *SrCyp20*	84.11	
Const_2h_1 *SrCyp20*	84.20	
Const_2h_2 *SrCyp20*	84.16	
Const_2h_3 *SrCyp20*	84.18	
Const_2h_4 *SrCyp20*	84.35	
Const_24h_1 *SrCyp20*	84.12	
Const_24h_2 *SrCyp20*	84.26	
Const_24h_3 *SrCyp20*	84.16	
Const_24h_4 *SrCyp20*	84.25	
D/n_1 *SrCyp20*	84.29	
D/n_2 *SrCyp20*	84.13	
D/n_3 *SrCyp20*	84.12	
D/n_4 *SrCyp20*	84.18	
SrCyp20 NTC	80.33	
SrCyp20 NTC	79.70	

Table A 48 Melting temperatures (°C) for the target gene *SrCyp21* in RT-qPCR.

Name	Tm1	Tm2
Const_0h_ctrl_1 *SrCyp21*	81.21	
Const_0h_ctrl_2 *SrCyp21*	81.18	
Const_0h_ctrl_3 *SrCyp21*	81.24	
Const_0h_ctrl_4 *SrCyp21*	81.27	
Const_2h_1 *SrCyp21*	81.23	
Const_2h_2 *SrCyp21*	81.27	
Const_2h_3 *SrCyp21*	81.27	
Const_2h_4 *SrCyp21*	81.42	
Const_24h_1 *SrCyp21*	81.48	
Const_24h_2 *SrCyp21*	81.32	
Const_24h_3 *SrCyp21*	81.50	
Const_24h_4 *SrCyp21*	81.24	
D/n_1 *SrCyp21*	81.15	
D/n_2 *SrCyp21*	81.14	
D/n_3 *SrCyp21*	81.14	
D/n_4 *SrCyp21*	81.14	
SrCyp21 NTC	78.64	
SrCyp21 NTC	74.28	

Table A 49 Melting temperatures (°C) for the target gene *SrCyp22* in RT-qPCR.

Name	Tm1	Tm2
Const_0h_ctrl_1 *SrCyp22*	83.39	
Const_0h_ctrl_2 *SrCyp22*	83.34	
Const_0h_ctrl_3 *SrCyp22*	83.37	
Const_0h_ctrl_4 *SrCyp22*	83.45	
Const_2h_1 *SrCyp22*	83.48	
Const_2h_2 *SrCyp22*	83.48	
Const_2h_3 *SrCyp22*	83.49	
Const_2h_4 *SrCyp22*	83.55	
Const_24h_1 *SrCyp22*	83.66	
Const_24h_2 *SrCyp22*	83.57	
Const_24h_3 *SrCyp22*	83.56	
Const_24h_4 *SrCyp22*	83.53	
D/n_1 *SrCyp22*	83.35	
D/n_2 *SrCyp22*	83.36	
D/n_3 *SrCyp22*	83.38	
D/n_4 *SrCyp22*	83.46	
SrCyp22 NTC	75.00	
SrCyp22 NTC	73.53	

Table A 50 Melting temperatures (°C) for the target gene *SrCyp23* in RT-qPCR.

Name	Tm1	Tm2
Const_0h_ctrl_1 *SrCyp23*	79.17	
Const_0h_ctrl_2 *SrCyp23*	79.39	
Const_0h_ctrl_3 *SrCyp23*	79.28	
Const_0h_ctrl_4 *SrCyp23*	79.19	
Const_2h_1 *SrCyp23*	79.36	
Const_2h_2 *SrCyp23*	79.36	
Const_2h_3 *SrCyp23*	79.38	
Const_2h_4 *SrCyp23*	79.42	
Const_24h_1 *SrCyp23*	79.29	
Const_24h_2 *SrCyp23*	79.37	
Const_24h_3 *SrCyp23*	79.39	
Const_24h_4 *SrCyp23*	79.23	
D/n_1 *SrCyp23*	79.36	
D/n_2 *SrCyp23*	79.19	
D/n_3 *SrCyp23*	79.19	
D/n_4 *SrCyp23*	79.27	
SrCyp23 NTC	77.03	
SrCyp23 NTC	76.87	

Table A 51 Melting temperatures (°C) for the target gene *SrCyp28* in RT-qPCR.

Name	Tm1	Tm2
Const_0h_ctrl_1 *SrCyp28*	82.87	
Const_0h_ctrl_2 *SrCyp28*	82.72	
Const_0h_ctrl_3 *SrCyp28*	82.84	
Const_0h_ctrl_4 *SrCyp28*	82.88	
Const_2h_1 *SrCyp28*	82.86	
Const_2h_2 *SrCyp28*	82.92	
Const_2h_3 *SrCyp28*	82.91	
Const_2h_4 *SrCyp28*	83.07	
Const_24h_1 *SrCyp28*	83.10	
Const_24h_2 *SrCyp28*	82.93	
Const_24h_3 *SrCyp28*	83.00	
Const_24h_4 *SrCyp28*	82.98	
D/n_1 *SrCyp28*	82.81	
D/n_2 *SrCyp28*	82.83	
D/n_3 *SrCyp28*	82.84	
D/n_4 *SrCyp28*	82.89	
SrCyp28 NTC	78.67	
SrCyp28 NTC	80.49	

Table A 52 Melting temperatures (°C) for the target gene *SrCyp44* in RT-qPCR.

Name	Tm1	Tm2
Const_0h_ctrl_1 *SrCyp44*	84.88	
Const_0h_ctrl_2 *SrCyp44*	84.93	
Const_0h_ctrl_3 *SrCyp44*	84.88	
Const_0h_ctrl_4 *SrCyp44*	84.88	
Const_2h_1 *SrCyp44*	84.86	
Const_2h_2 *SrCyp44*	84.84	
Const_2h_3 *SrCyp44*	84.76	
Const_2h_4 *SrCyp44*	85.10	
Const_24h_1 *SrCyp44*	84.83	
Const_24h_2 *SrCyp44*	84.89	
Const_24h_3 *SrCyp44*	84.89	
Const_24h_4 *SrCyp44*	84.81	
D/n_1 *SrCyp44*	85.10	
D/n_2 *SrCyp44*	84.88	
D/n_3 *SrCyp44*	85.05	
D/n_4 *SrCyp44*	84.85	
SrCyp44 NTC		
SrCyp44 NTC	71.14	78.42

Table A 53 Melting temperatures (°C) for the target gene *SrCyp46* in RT-qPCR.

Name	Tm1	Tm2
Const_0h_ctrl_1 *SrCyp46*	82.83	
Const_0h_ctrl_2 *SrCyp46*	82.94	
Const_0h_ctrl_3 *SrCyp46*	82.93	
Const_0h_ctrl_4 *SrCyp46*	82.91	
Const_2h_1 *SrCyp46*	83.12	
Const_2h_2 *SrCyp46*	83.14	
Const_2h_3 *SrCyp46*	83.14	
Const_2h_4 *SrCyp46*	83.30	
Const_24h_1 *SrCyp46*	83.31	
Const_24h_2 *SrCyp46*	83.18	
Const_24h_3 *SrCyp46*	83.21	
Const_24h_4 *SrCyp46*	83.03	
D/n_1 *SrCyp46*	82.89	
D/n_2 *SrCyp46*	82.90	
D/n_3 *SrCyp46*	82.93	
D/n_4 *SrCyp46*	82.96	
SrCyp46 NTC	72.21	
SrCyp46 NTC	75.04	

Table A 54 Melting temperatures (°C) for the target gene *SrCyp52* in RT-qPCR.

Name	Tm1	Tm2
Const_0h_ctrl_1 *SrCyp52*	84.50	
Const_0h_ctrl_2 *SrCyp52*	84.52	
Const_0h_ctrl_3 *SrCyp52*	84.48	
Const_0h_ctrl_4 *SrCyp52*	84.53	
Const_2h_1 *SrCyp52*	84.62	
Const_2h_2 *SrCyp52*	84.68	
Const_2h_3 *SrCyp52*	84.66	
Const_2h_4 *SrCyp52*	84.77	
Const_24h_1 *SrCyp52*	84.57	
Const_24h_2 *SrCyp52*	84.49	
Const_24h_3 *SrCyp52*	84.52	
Const_24h_4 *SrCyp52*	84.59	
D/n_1 *SrCyp52*	84.42	
D/n_2 *SrCyp52*	84.39	
D/n_3 *SrCyp52*	84.40	
D/n_4 *SrCyp52*	84.30	
SrCyp52 NTC	78.80	
SrCyp52 NTC	79.97	

Table A 55 Melting temperatures (°C) for the target gene *SrCyp54* in RT-qPCR.

Name	Tm1	Tm2
Const_0h_ctrl_1 *SrCyp54*	80.47	
Const_0h_ctrl_2 *SrCyp54*	80.00	
Const_0h_ctrl_3 *SrCyp54*	80.51	
Const_0h_ctrl_4 *SrCyp54*	80.44	
Const_2h_1 *SrCyp54*	80.68	
Const_2h_2 *SrCyp54*	80.62	
Const_2h_3 *SrCyp54*	80.55	
Const_2h_4 *SrCyp54*	80.58	
Const_24h_1 *SrCyp54*	80.51	
Const_24h_2 *SrCyp54*	80.46	
Const_24h_3 *SrCyp54*	80.52	
Const_24h_4 *SrCyp54*	80.61	
D/n_1 *SrCyp54*	80.37	
D/n_2 *SrCyp54*	80.33	
D/n_3 *SrCyp54*	80.29	
D/n_4 *SrCyp54*	80.42	
SrCyp54 NTC	78.76	
SrCyp54 NTC	78.91	

Table A 56 Melting temperatures (°C) for the target gene *SrCyp70* in RT-qPCR.

Name	Tm1	Tm2
Const_0h_ctrl_1 *SrCyp70*	80.87	
Const_0h_ctrl_2 *SrCyp70*	80.71	
Const_0h_ctrl_3 *SrCyp70*	80.71	
Const_0h_ctrl_4 *SrCyp70*	81.02	
Const_2h_1 *SrCyp70*	80.96	
Const_2h_2 *SrCyp70*	80.92	
Const_2h_3 *SrCyp70*	81.01	
Const_2h_4 *SrCyp70*	81.08	
Const_24h_1 *SrCyp70*	81.18	
Const_24h_2 *SrCyp70*	81.12	
Const_24h_3 *SrCyp70*	81.13	
Const_24h_4 *SrCyp70*	81.09	
D/n_1 *SrCyp70*	80.83	
D/n_2 *SrCyp70*	80.70	
D/n_3 *SrCyp70*	80.72	
D/n_4 *SrCyp70*	80.97	
SrCyp70 NTC	77.80	
SrCyp70 NTC	77.44	

Table A 57 Melting temperatures (°C) for the target gene *SrCyp73* in RT-qPCR.

Name	Tm1	Tm2
Const_0h_ctrl_1 *SrCyp73*	83.89	
Const_0h_ctrl_2 *SrCyp73*	83.91	
Const_0h_ctrl_3 *SrCyp73*	84.02	
Const_0h_ctrl_4 *SrCyp73*	84.06	
Const_2h_1 *SrCyp73*	84.26	
Const_2h_2 *SrCyp73*	84.24	
Const_2h_3 *SrCyp73*	84.18	
Const_2h_4 *SrCyp73*	84.35	
Const_24h_1 *SrCyp73*	84.35	
Const_24h_2 *SrCyp73*	84.12	
Const_24h_3 *SrCyp73*	84.34	
Const_24h_4 *SrCyp73*	84.13	
D/n_1 *SrCyp73*	84.03	
D/n_2 *SrCyp73*	83.91	
D/n_3 *SrCyp73*	84.03	
D/n_4 *SrCyp73*	84.09	
SrCyp73 NTC	73.81	
SrCyp73 NTC	74.04	

A5. BAR CHARTS

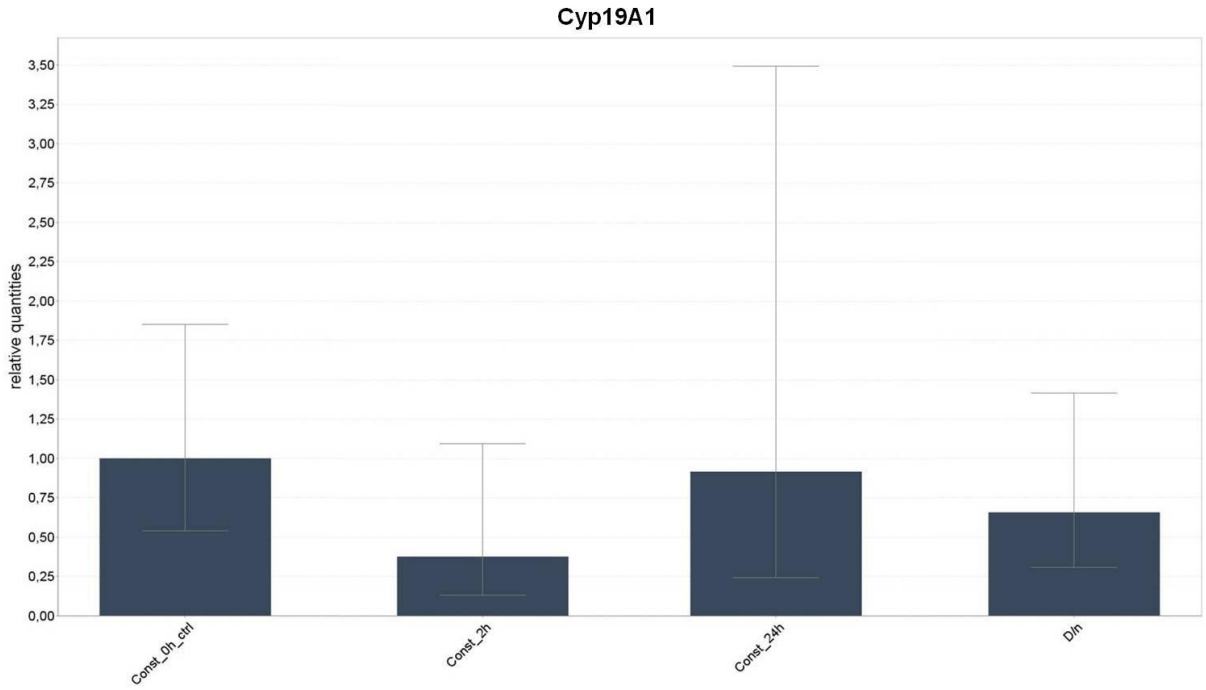


Figure A 1 Bar chart for the target gene *Cyp19A1.* The x-axis shows the experimental conditions, and the expression ratio is given on the y-axis, with the control sample set to 1.000.

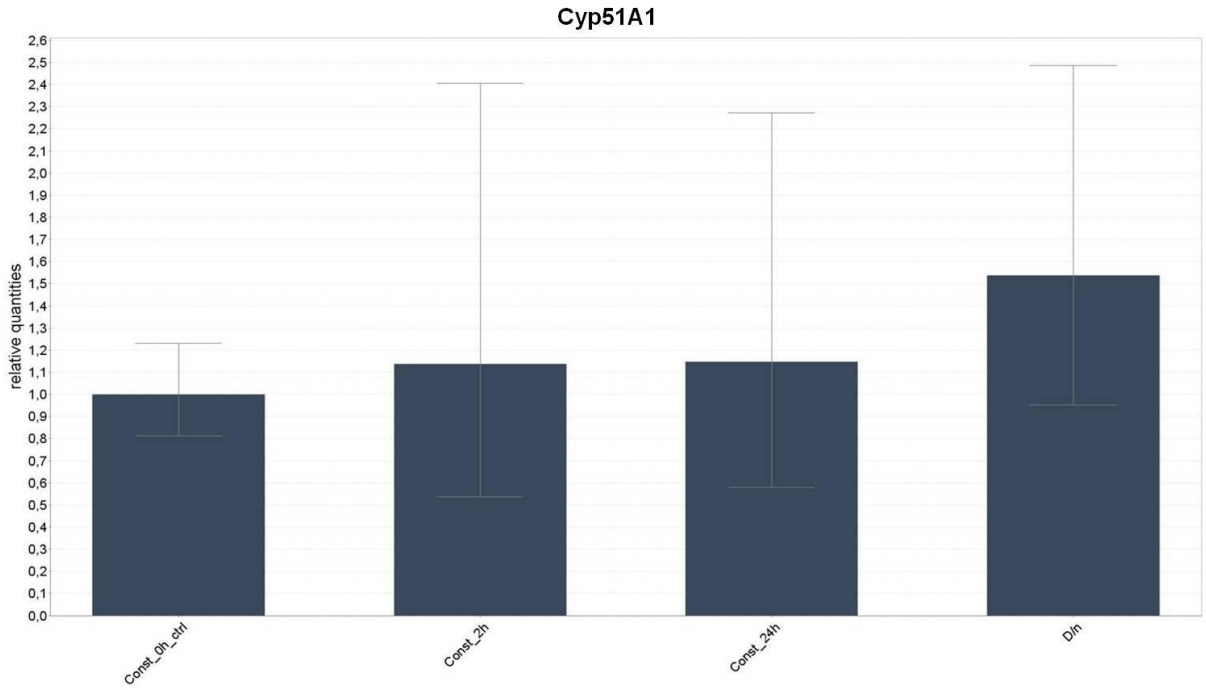


Figure A 2 Bar chart for the target gene *Cyp51A1*. The x-axis shows the experimental conditions, and the expression ratio is given on the y-axis, with the control sample set to 1.000.

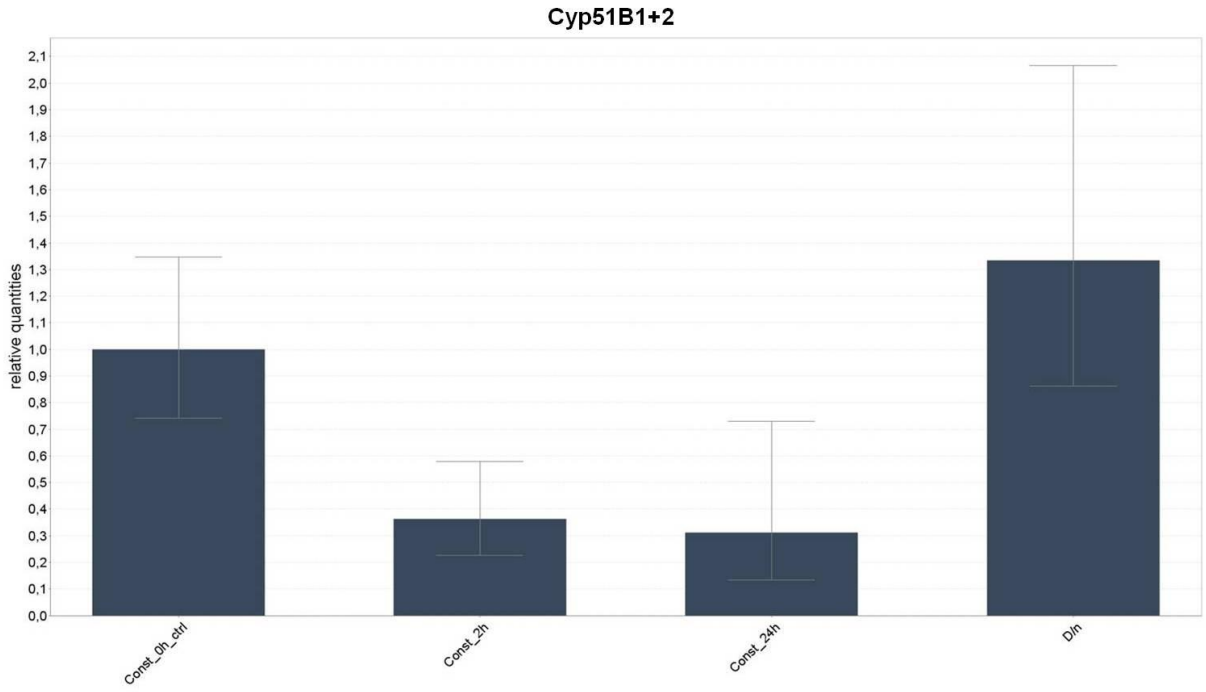


Figure A 3 Bar chart for the target genes *Cyp51B1* and *Cyp51B2*. *Cyp51B1* and *Cyp51B2* were analyzed with one primer pair recognizing both genes. The x-axis shows the experimental conditions, and the expression ratio is given on the y-axis, with the control sample set to 1.000.

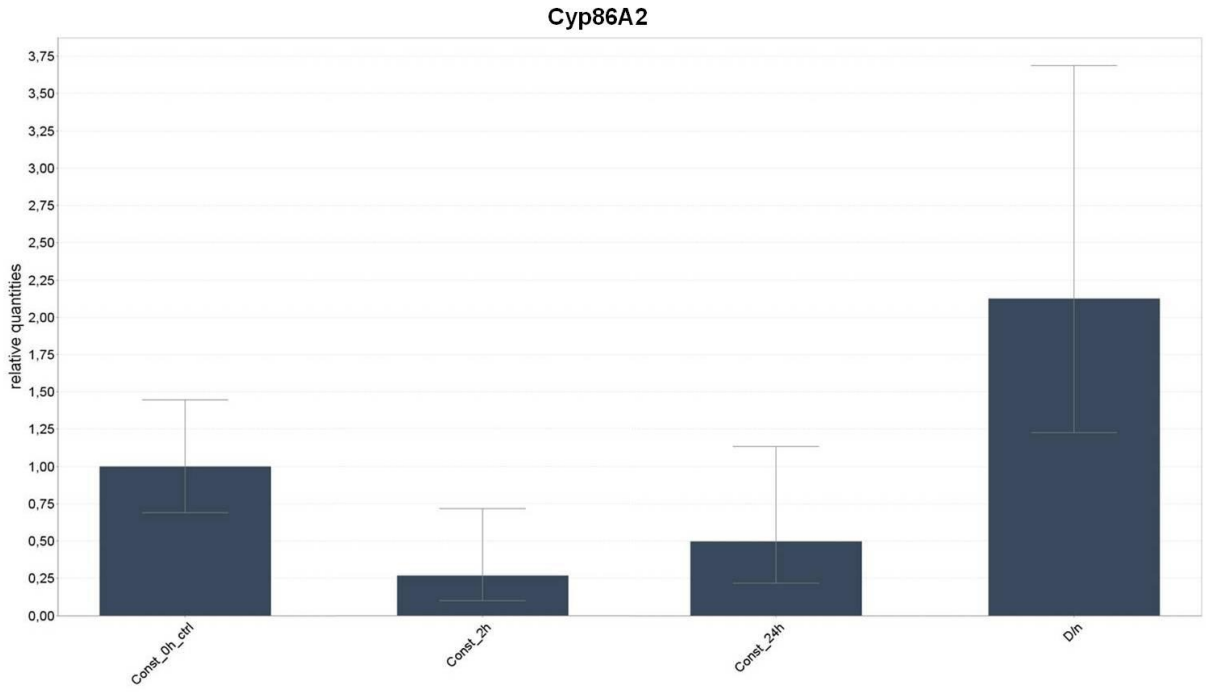


Figure A 4 Bar chart for the target gene *Cyp86A2*. The x-axis shows the experimental conditions, and the expression ratio is given on the y-axis, with the control sample set to 1.000.

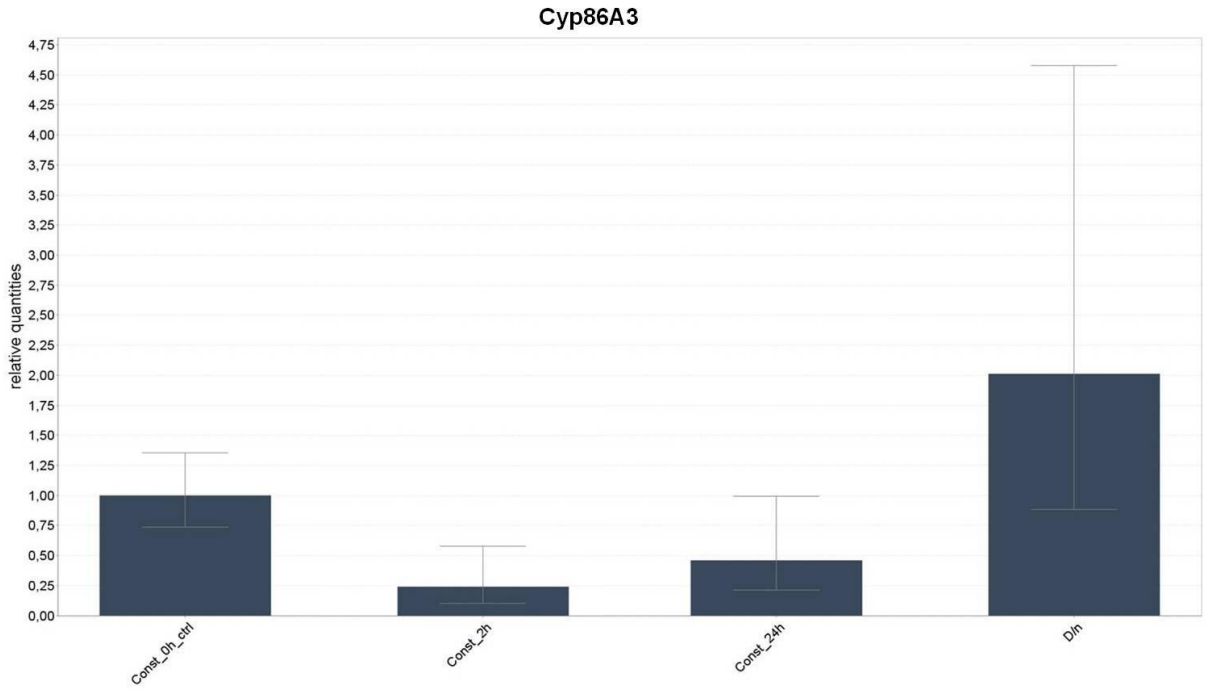


Figure A 5 Bar chart for the target gene *Cyp86A3*. The x-axis shows the experimental conditions, and the expression ratio is given on the y-axis, with the control sample set to 1.000.

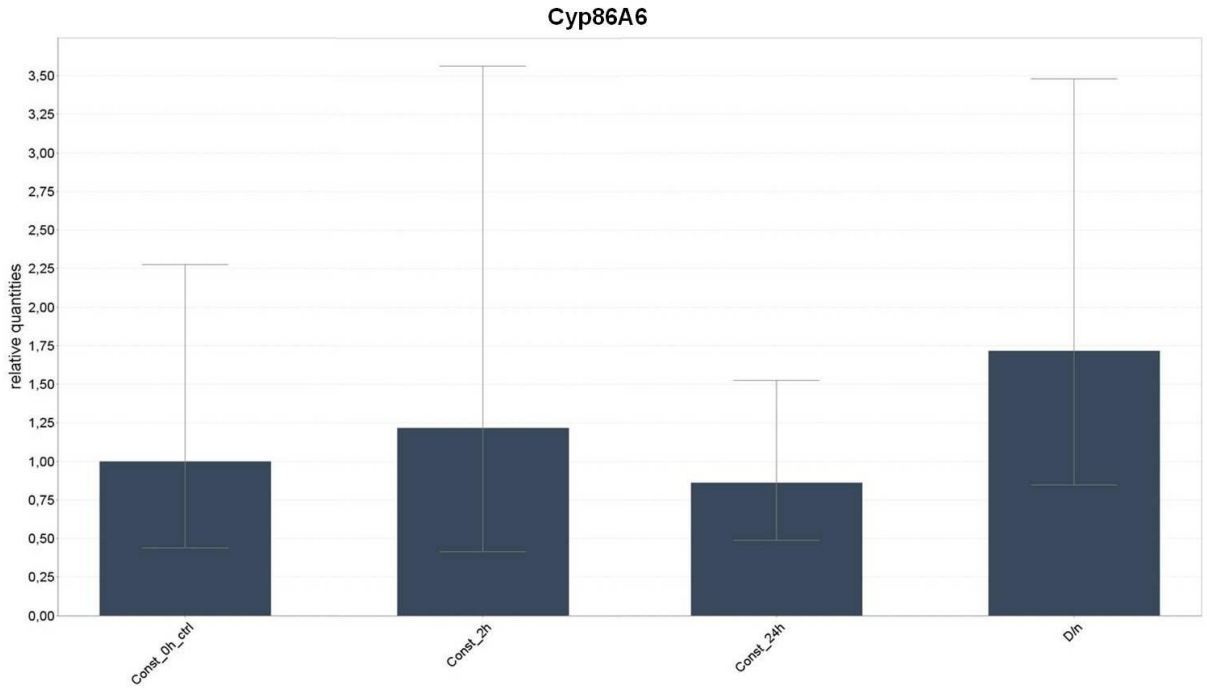


Figure A 6 Bar chart for the target gene *Cyp86A6*. The x-axis shows the experimental conditions, and the expression ratio is given on the y-axis, with the control sample set to 1.000.

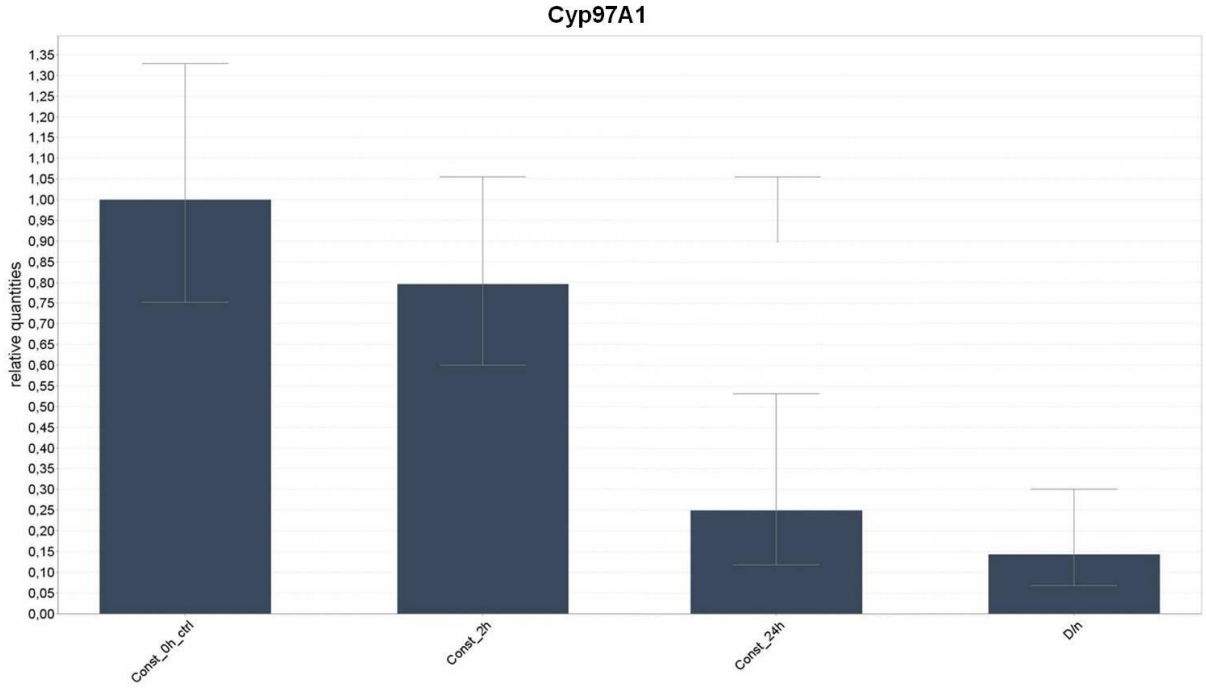


Figure A 7 Bar chart for the target gene *Cyp97A1*. The x-axis shows the experimental conditions, and the expression ratio is given on the y-axis, with the control sample set to 1.000.

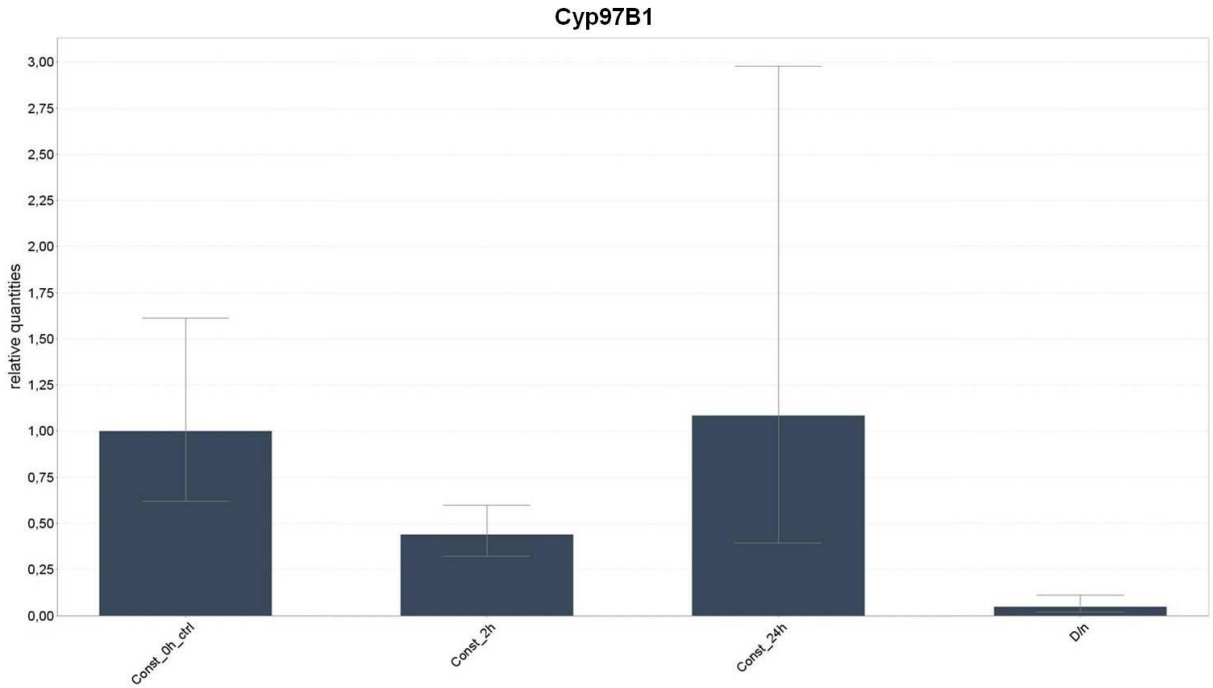


Figure A 8 Bar chart for the target gene *Cyp97B1*. The x-axis shows the experimental conditions, and the expression ratio is given on the y-axis, with the control sample set to 1.000.

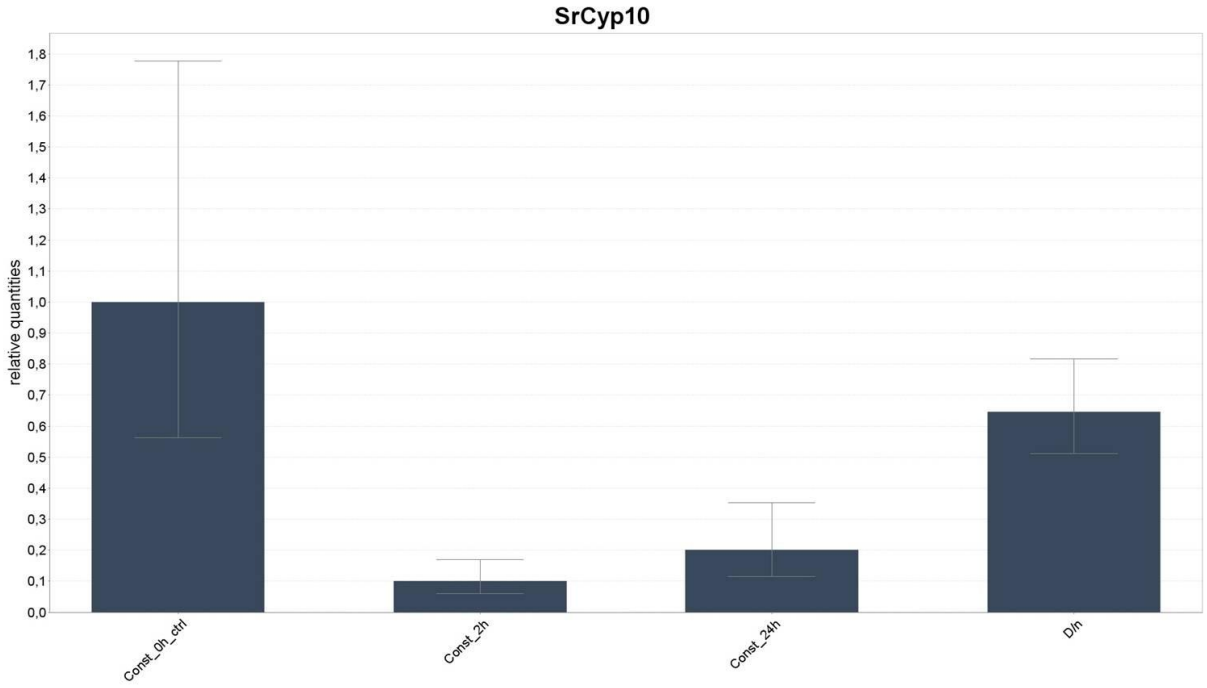


Figure A 9 Bar chart for the target gene *SrCyp10*. The x-axis shows the experimental conditions, and the expression ratio is given on the y-axis, with the control sample set to 1.000.

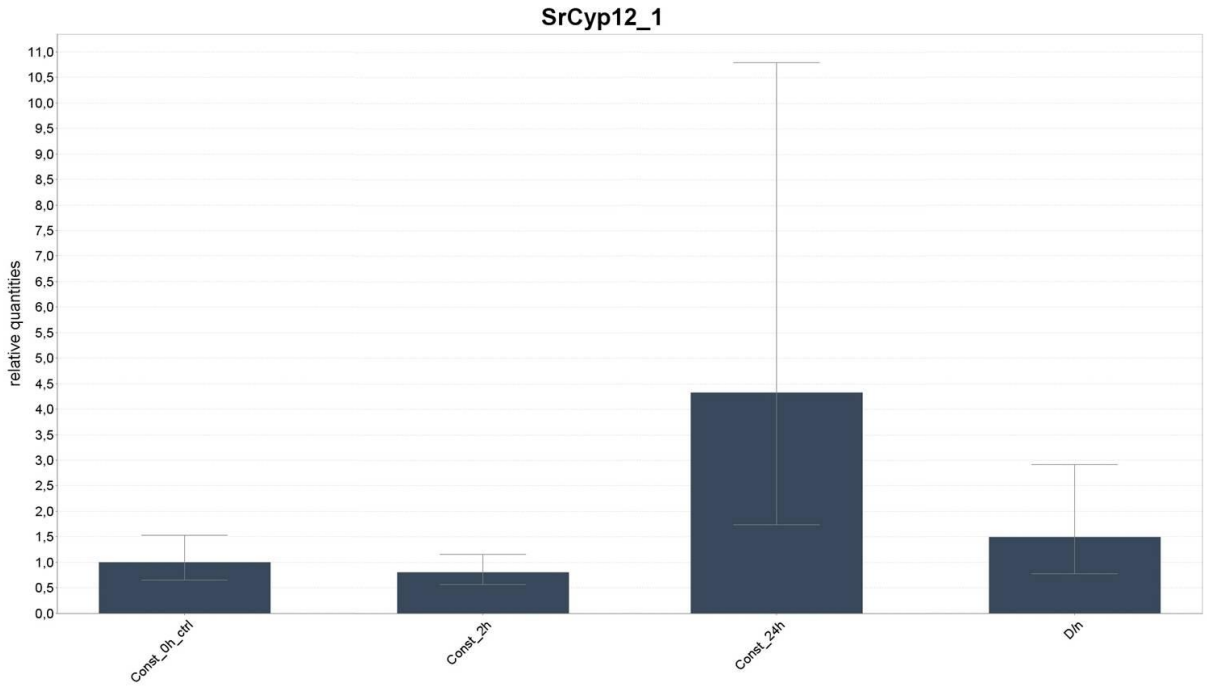


Figure A 10 Bar chart for the target gene *SrCyp12_1*. The x-axis shows the experimental conditions, and the expression ratio is given on the y-axis, with the control sample set to 1.000.

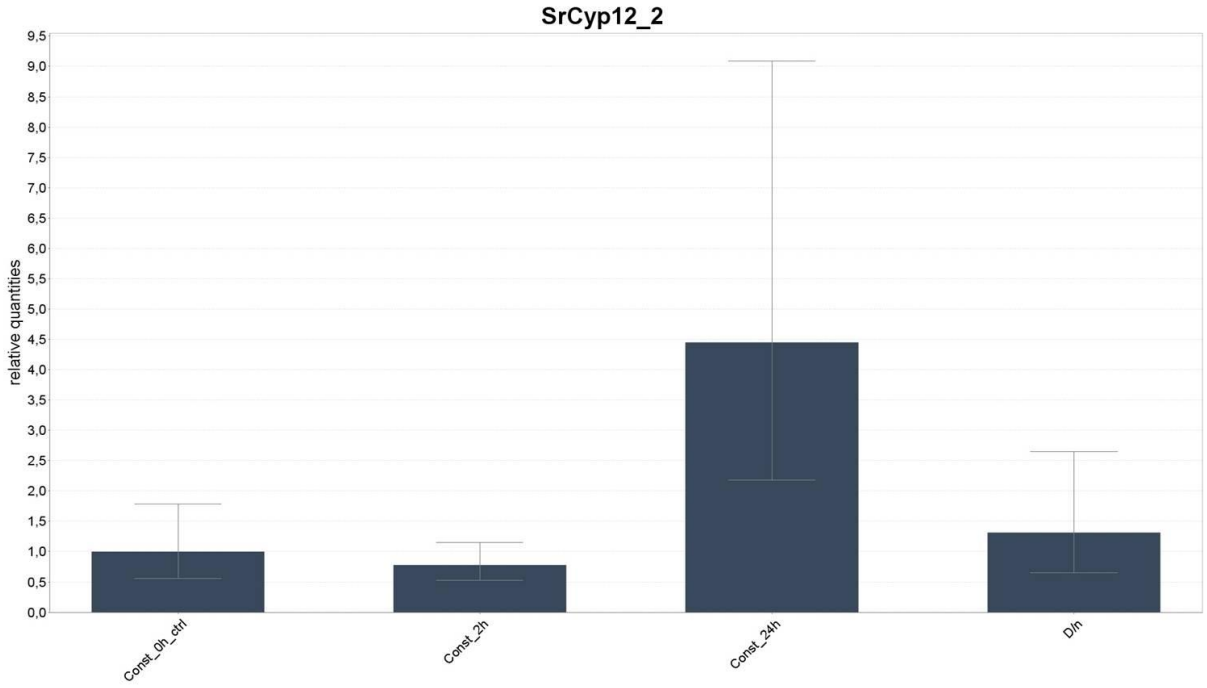


Figure A 11 Bar chart for the target gene *SrCyp12_2*. The x-axis shows the experimental conditions, and the expression ratio is given on the y-axis, with the control sample set to 1.000.

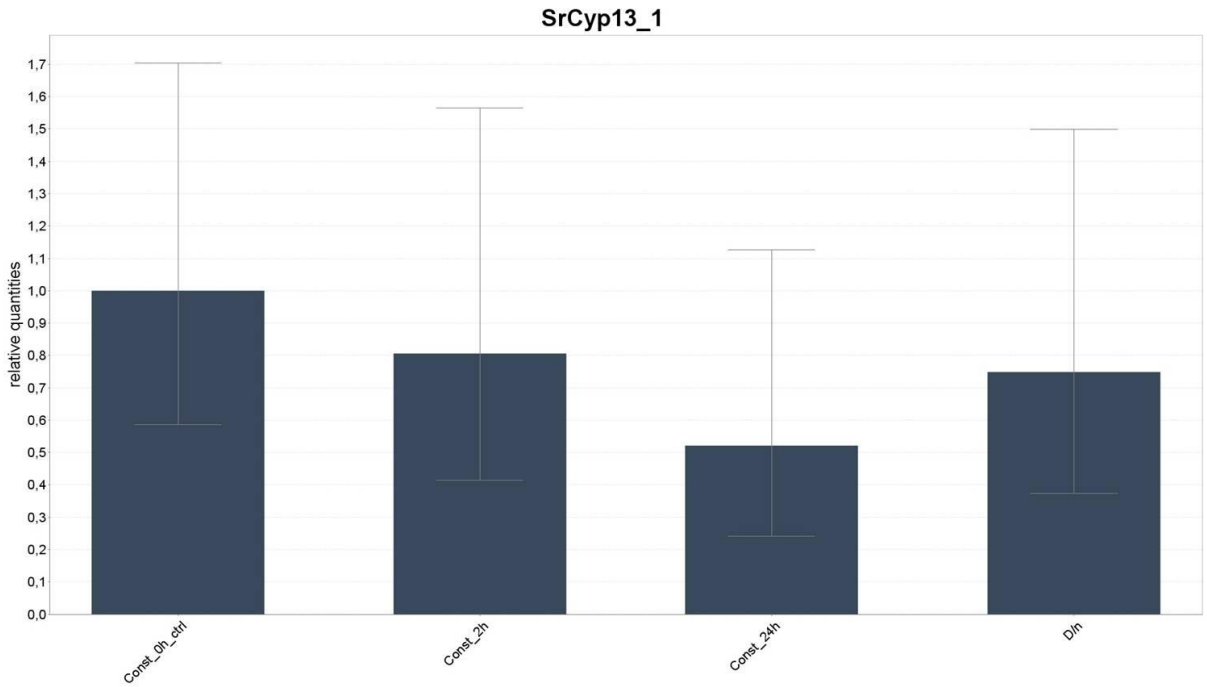


Figure A 12 Bar chart for the target gene *SrCyp13_1*. The x-axis shows the experimental conditions, and the expression ratio is given on the y-axis, with the control sample set to 1.000.

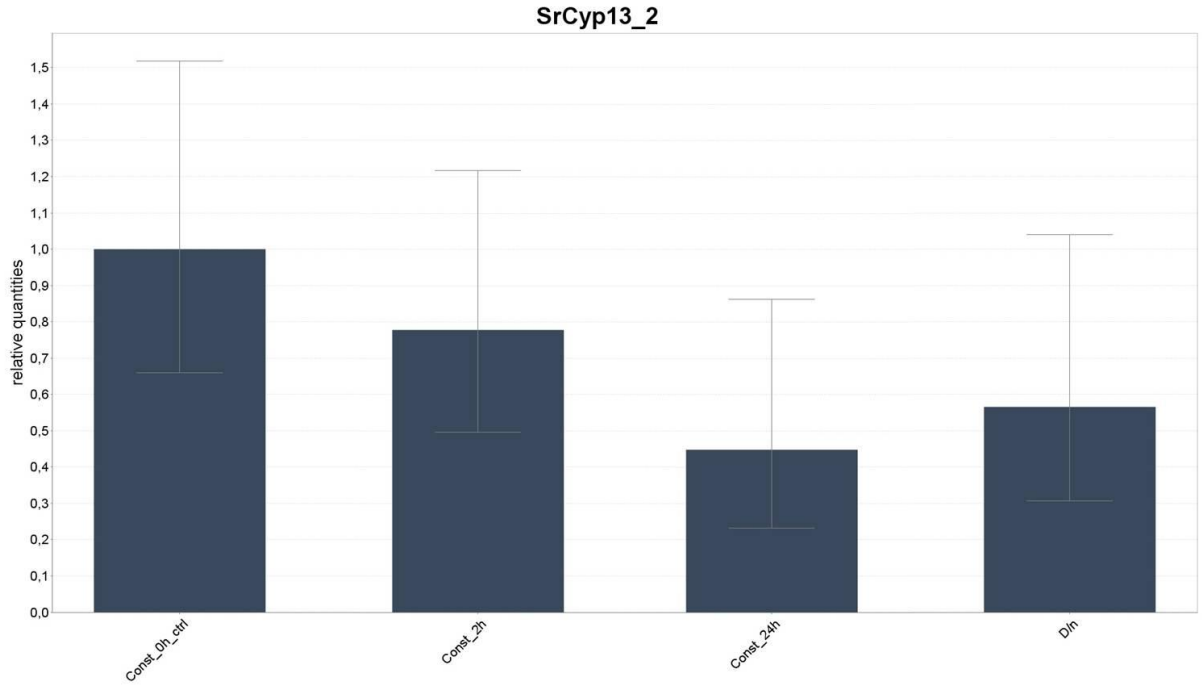


Figure A 13 Bar chart for the target gene *SrCyp13_2*. The x-axis shows the experimental conditions, and the expression ratio is given on the y-axis, with the control sample set to 1.000.

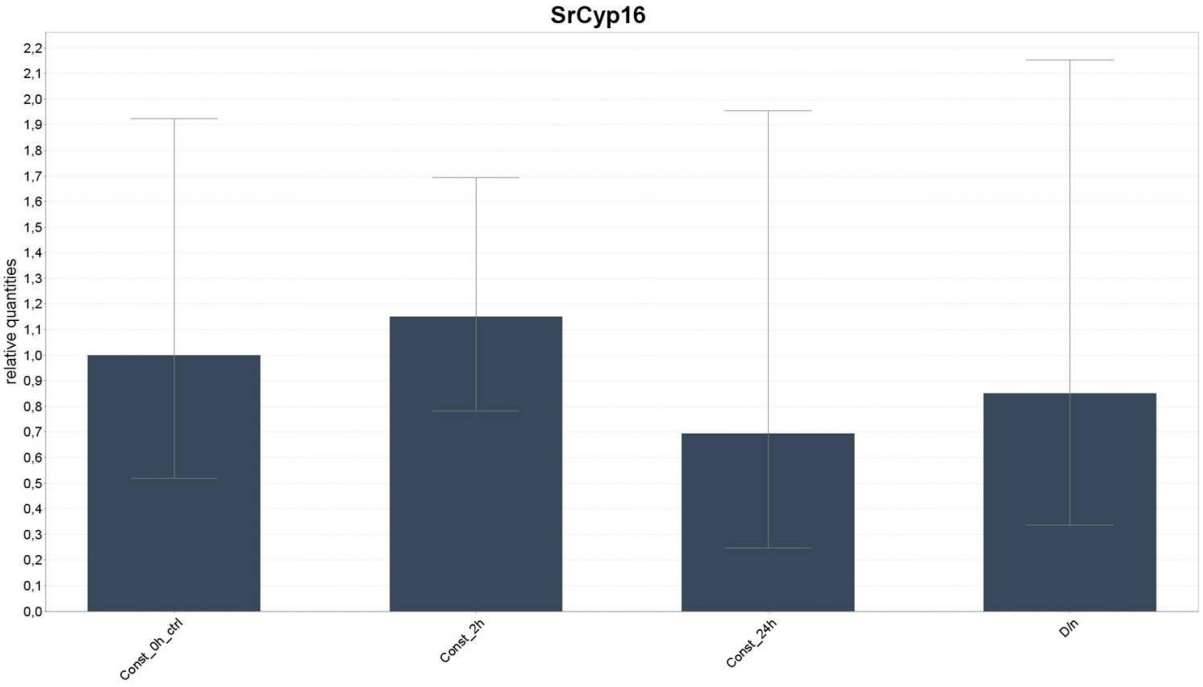


Figure A 14 Bar chart for the target gene *SrCyp16*. The x-axis shows the experimental conditions, and the expression ratio is given on the y-axis, with the control sample set to 1.000.

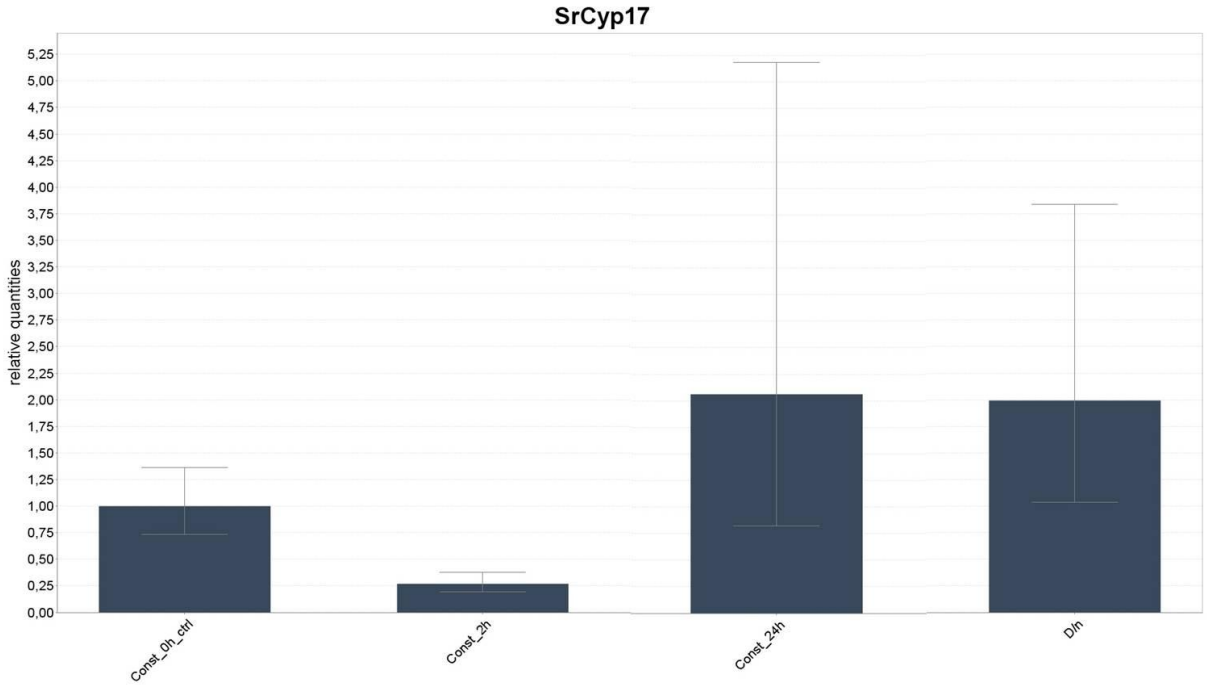


Figure A 15 Bar chart for the target gene *SrCyp17*. The x-axis shows the experimental conditions, and the expression ratio is given on the y-axis, with the control sample set to 1.000.

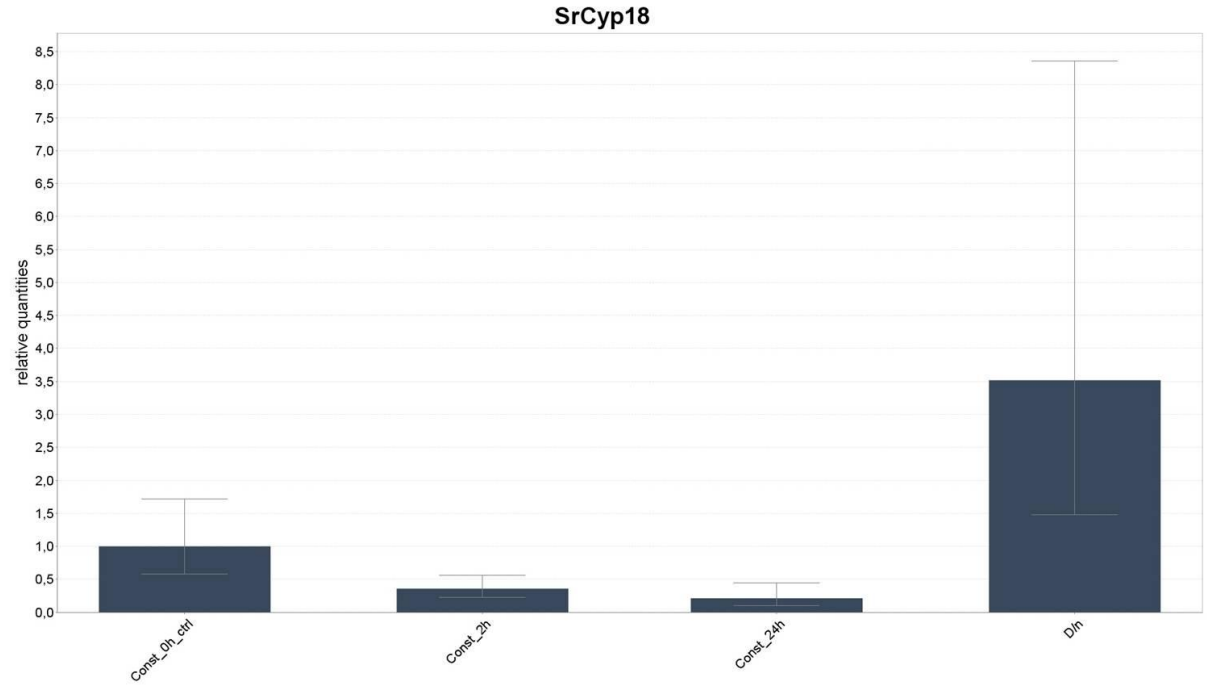


Figure A 16 Bar chart for the target gene *SrCyp18*. The x-axis shows the experimental conditions, and the expression ratio is given on the y-axis, with the control sample set to 1.000.

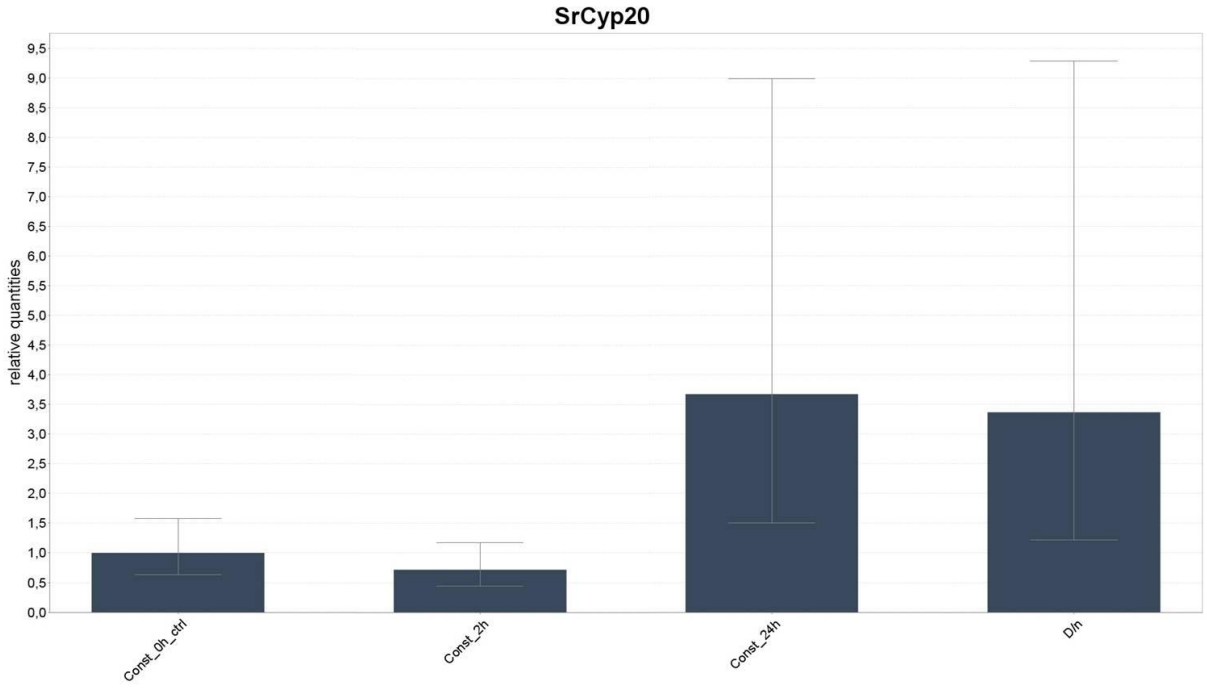


Figure A 17 Bar chart for the target gene *SrCyp20*. The x-axis shows the experimental conditions, and the expression ratio is given on the y-axis, with the control sample set to 1.000.

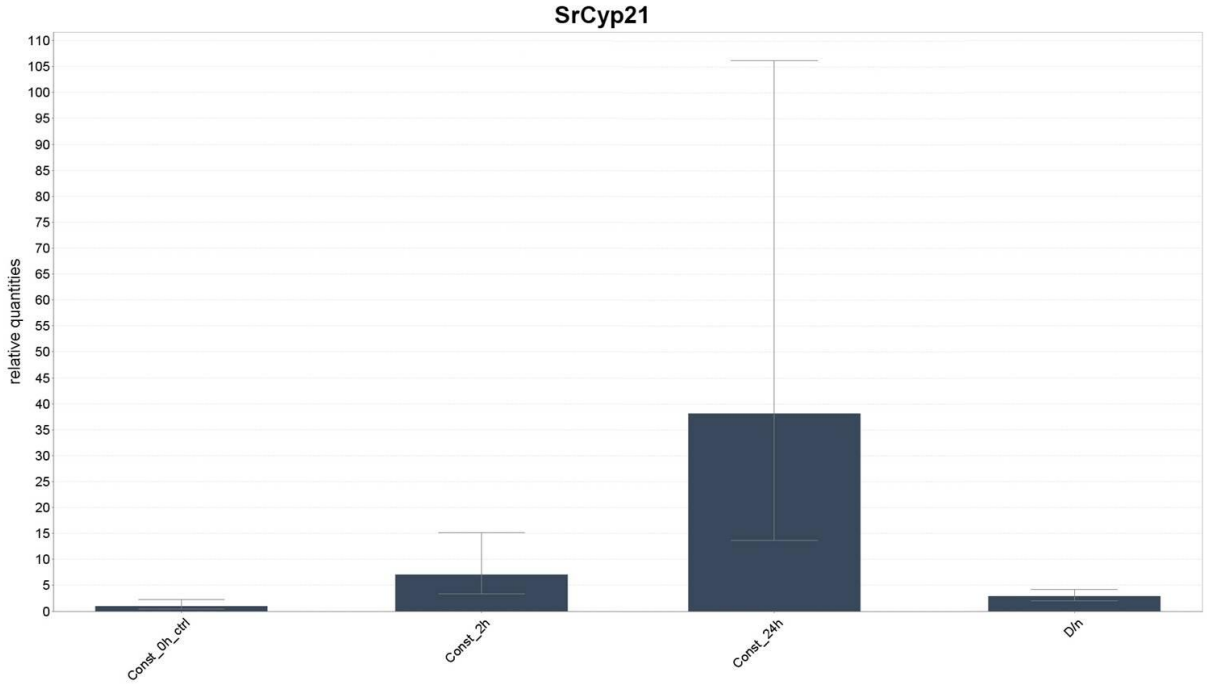


Figure A 18 Bar chart for the target gene *SrCyp21*. The x-axis shows the experimental conditions, and the expression ratio is given on the y-axis, with the control sample set to 1.000.

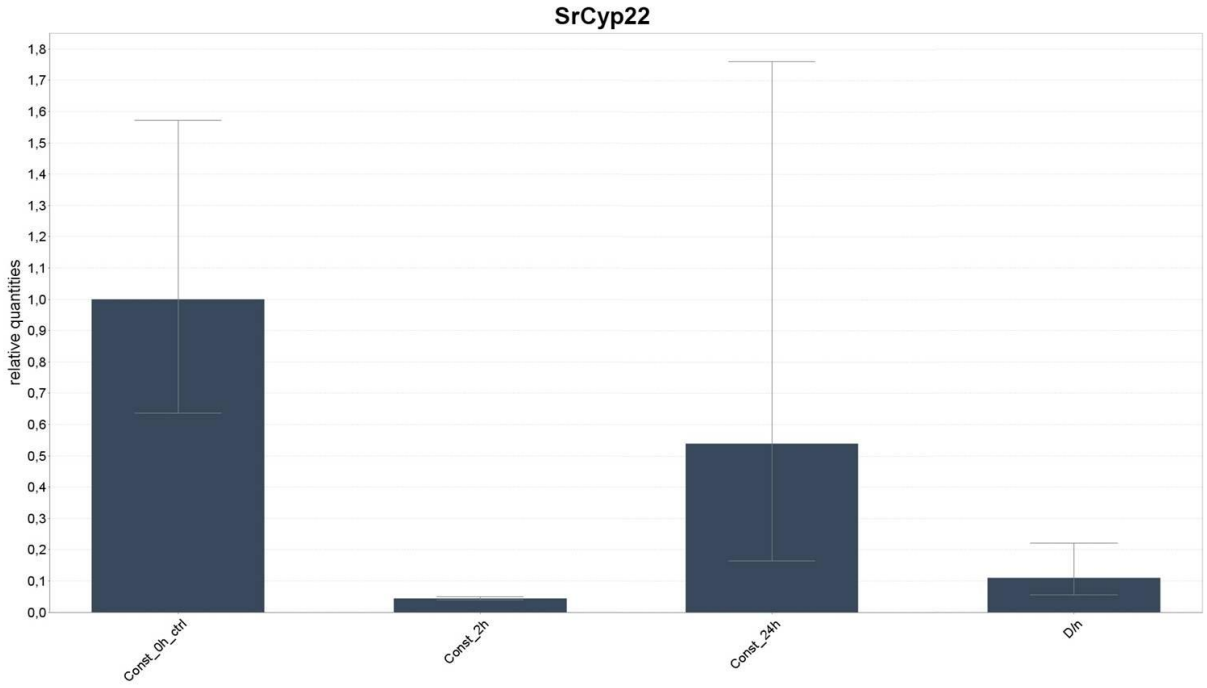


Figure A 19 Bar chart for the target gene *SrCyp22*. The x-axis shows the experimental conditions, and the expression ratio is given on the y-axis, with the control sample set to 1.000.

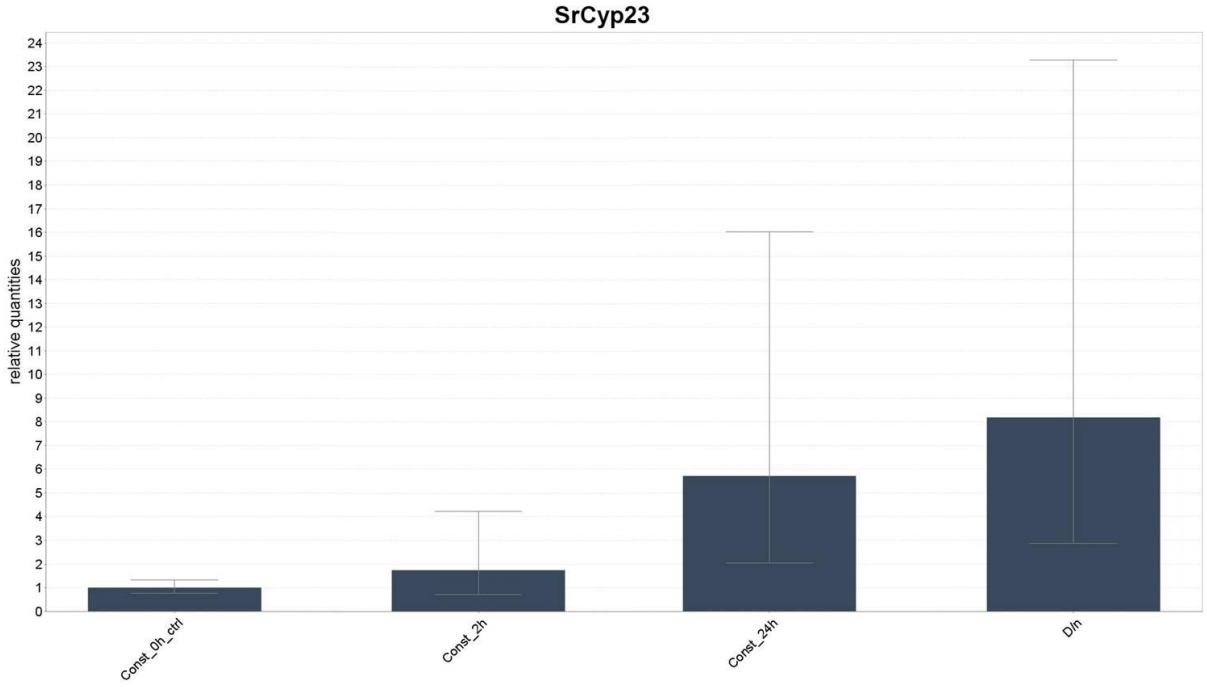


Figure A 20 Bar chart for the target gene *SrCyp23*. The x-axis shows the experimental conditions, and the expression ratio is given on the y-axis, with the control sample set to 1.000.

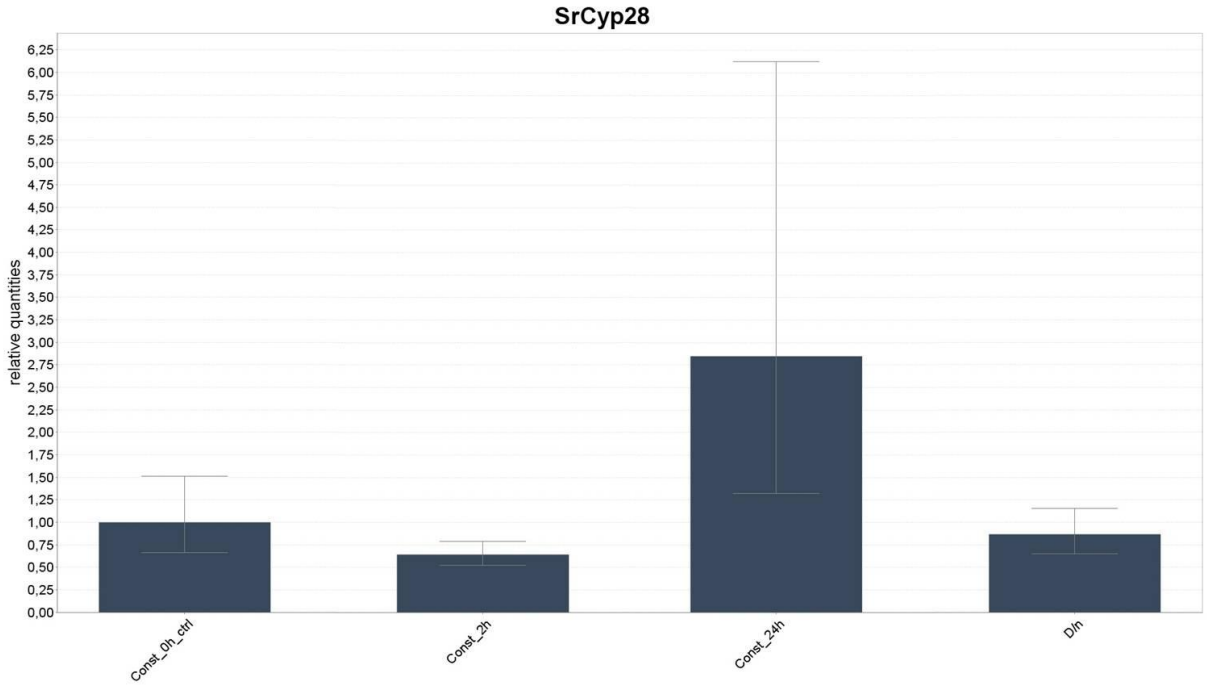


Figure A 21 Bar chart for the target gene *SrCyp28*. The x-axis shows the experimental conditions, and the expression ratio is given on the y-axis, with the control sample set to 1.000.

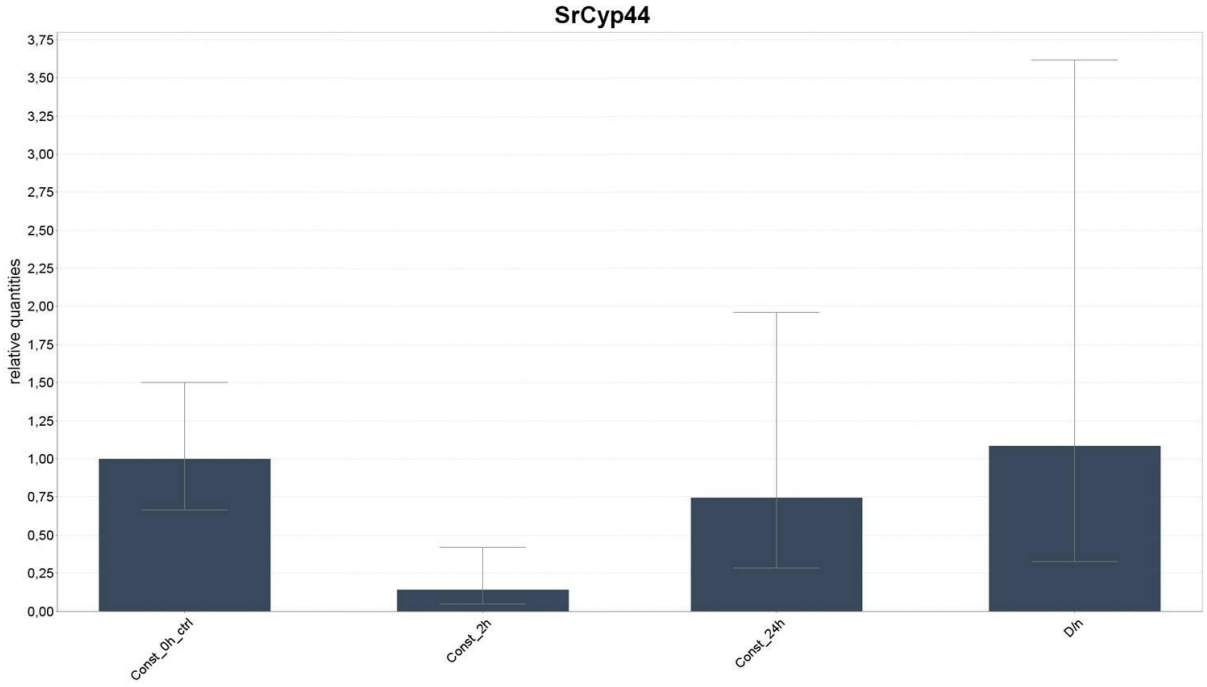


Figure A 22 Bar chart for the target gene *SrCyp44*. The x-axis shows the experimental conditions, and the expression ratio is given on the y-axis, with the control sample set to 1.000.

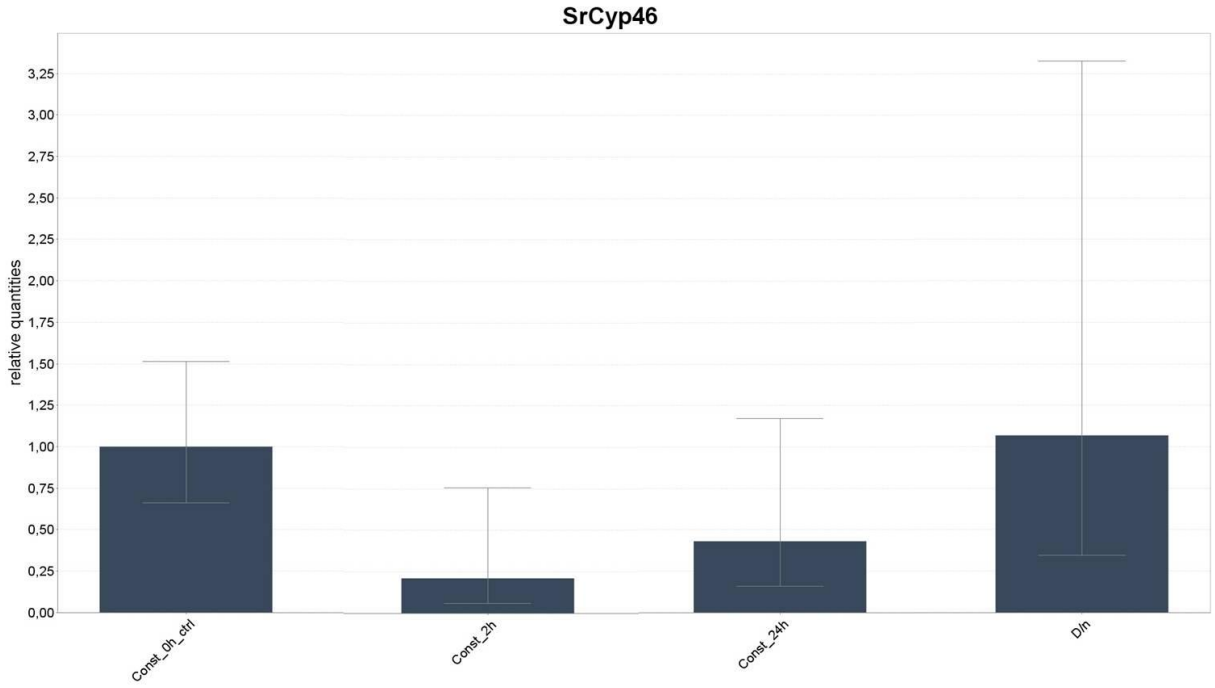


Figure A 23 Bar chart for the target gene *SrCyp46*. The x-axis shows the experimental conditions, and the expression ratio is given on the y-axis, with the control sample set to 1.000.

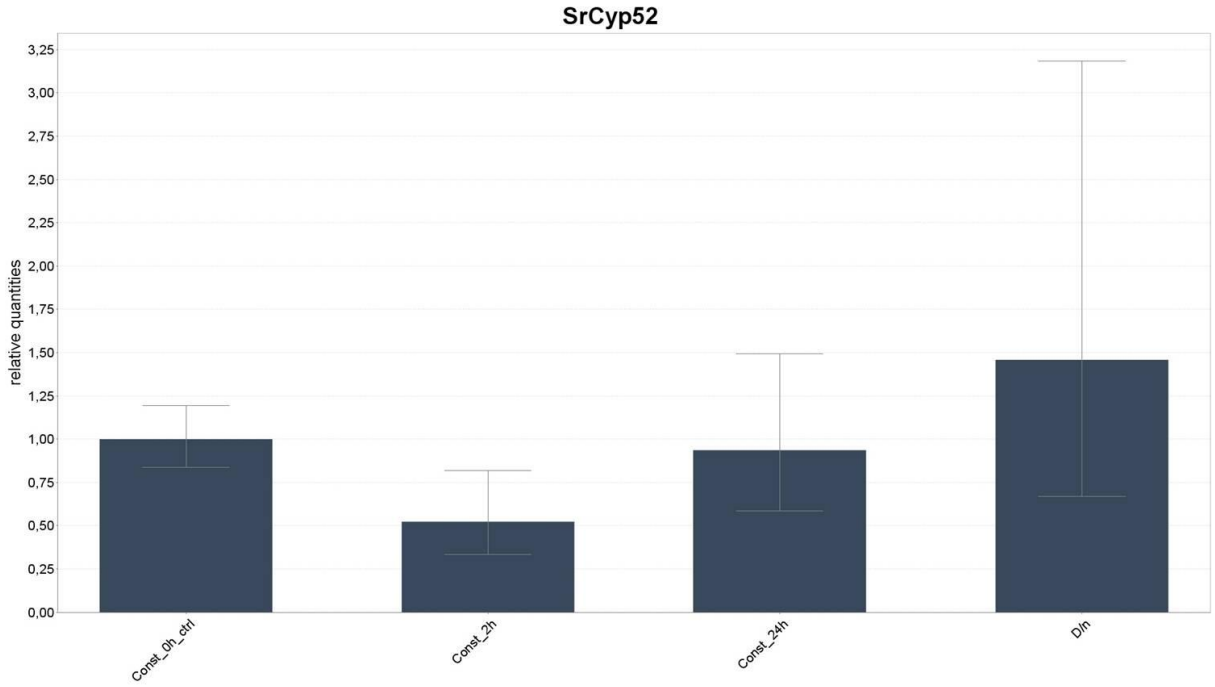


Figure A 24 Bar chart for the target gene *SrCyp52*. The x-axis shows the experimental conditions, and the expression ratio is given on the y-axis, with the control sample set to 1.000.

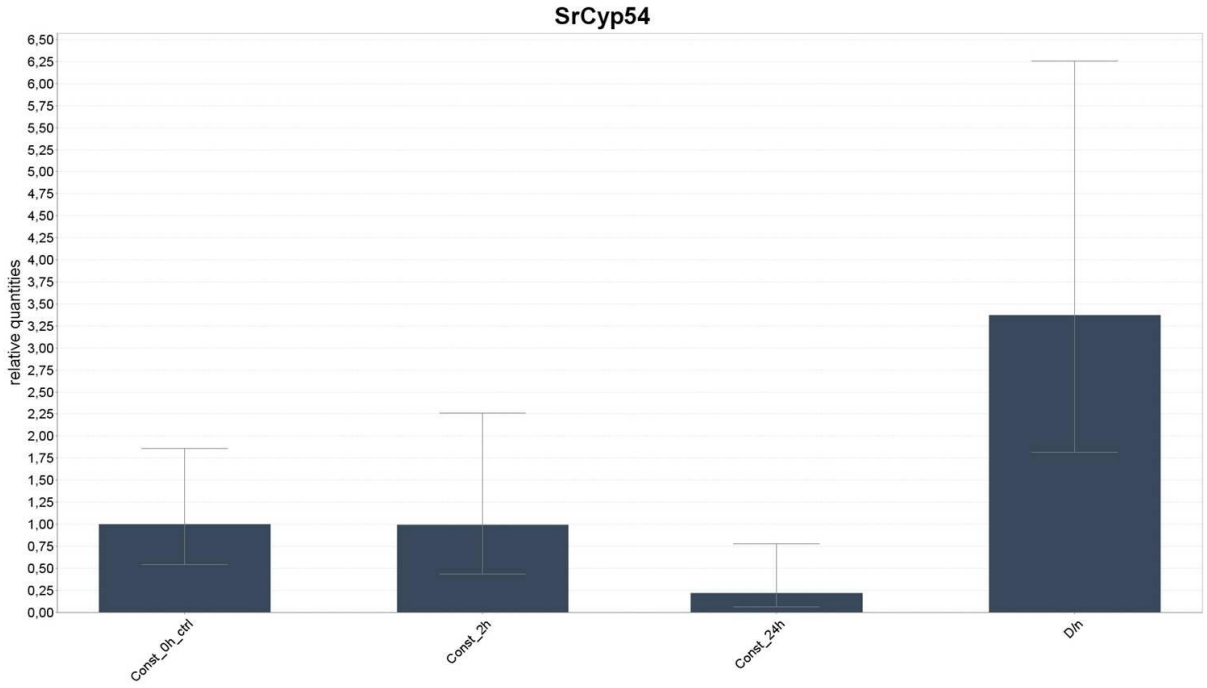


Figure A 25 Bar chart for the target gene *SrCyp54*. The x-axis shows the experimental conditions, and the expression ratio is given on the y-axis, with the control sample set to 1.000.

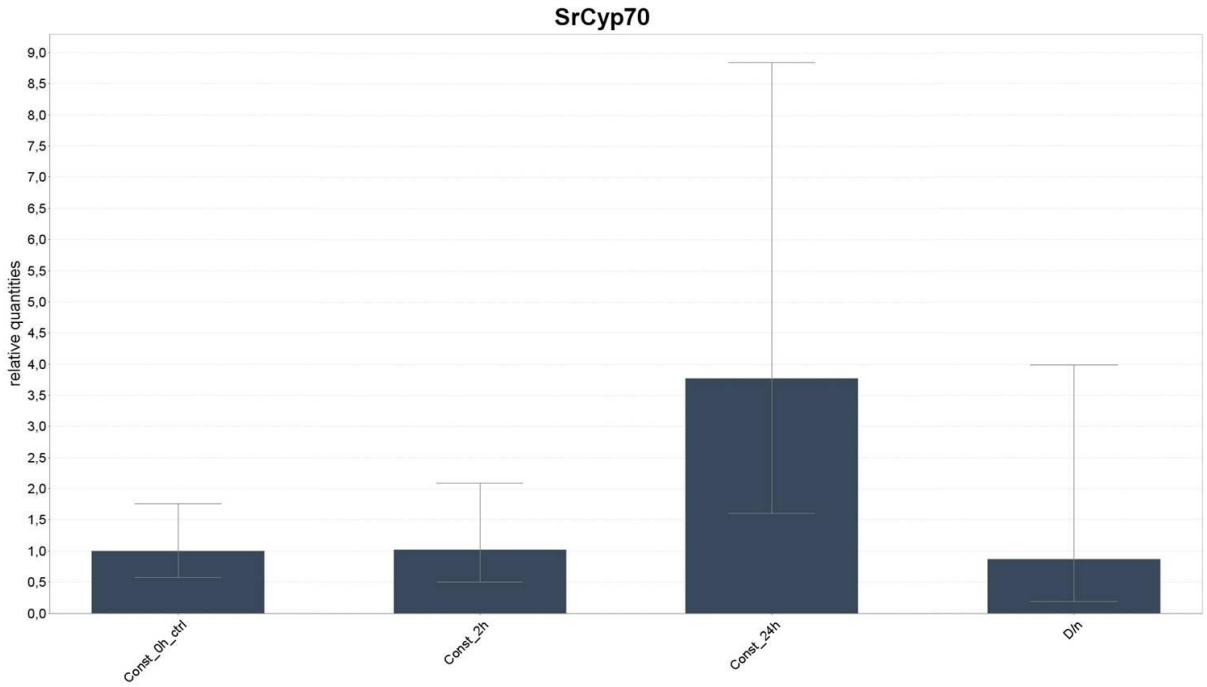


Figure A 26 Bar chart for the target gene *SrCyp70*. The x-axis shows the experimental conditions, and the expression ratio is given on the y-axis, with the control sample set to 1.000.

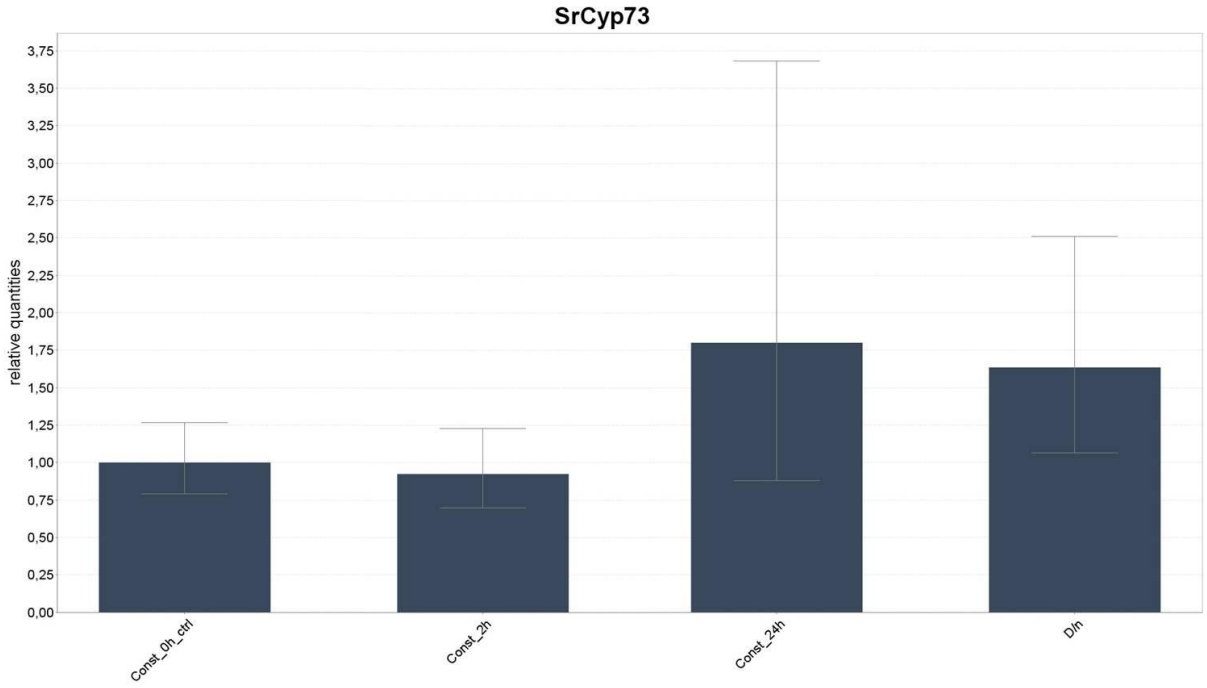


Figure A 27 Bar chart for the target gene *SrCyp73*. The x-axis shows the experimental conditions, and the expression ratio is given on the y-axis, with the control sample set to 1.000.

A6. PREDICTION OF PROTEIN LOCALIZATION

Table A 58 Results from prediction of protein localization. Indicated for each tool is the network used (plant/animal, plant/non-plant and mammalian). The RC-score given for results from TargetP is an indication of the reliability of the result, with 1 as the most reliable, and 5 as the least reliable. An NN-score of >0.500 from SecretomeP could indicate a secreted protein, in a non-classical pathway if a signal protein is not predicted. The CYPs marked with a * are still not complete.

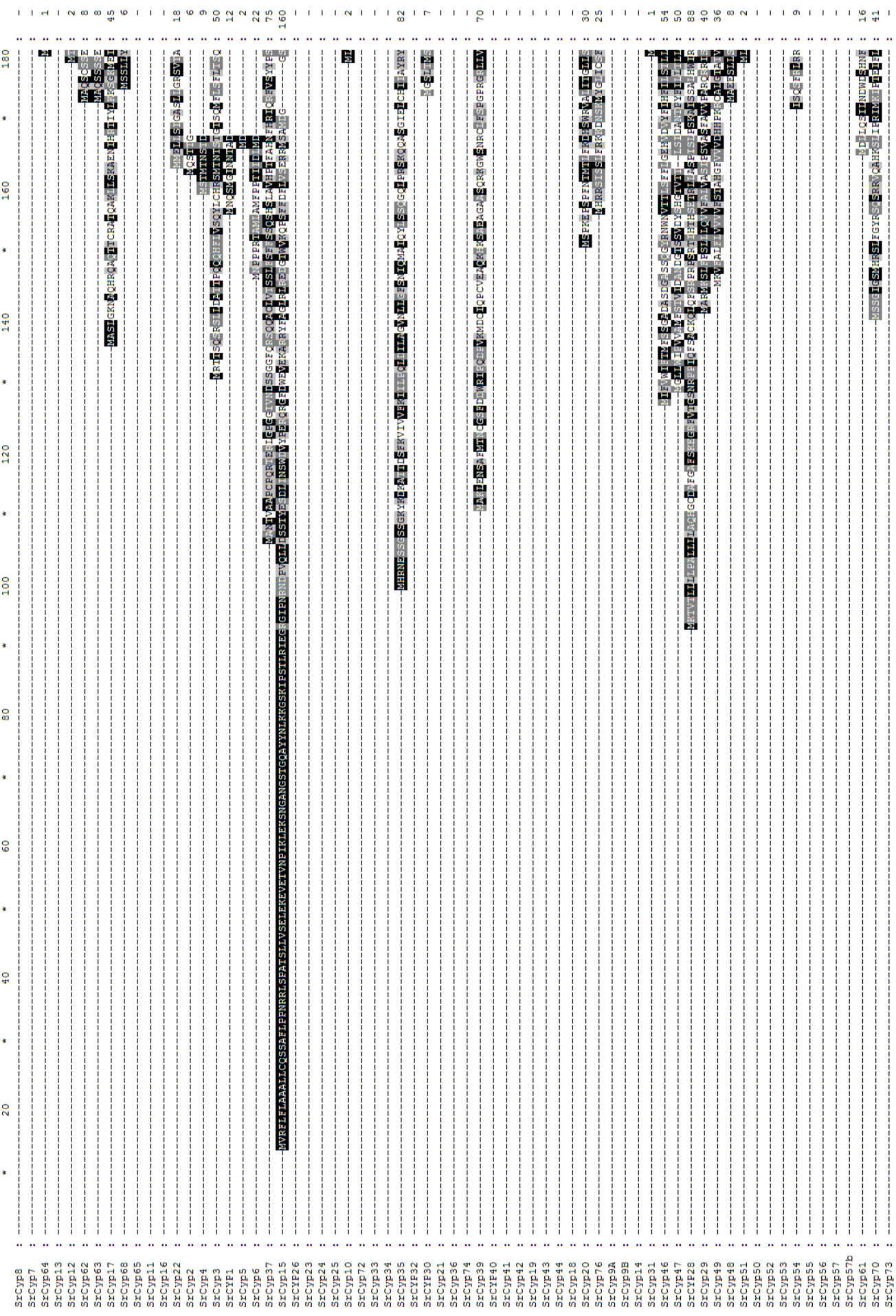
	WoLF PSORT		TargetP		SecretomeP
CYP	Plant	Animal	Plant	Non-plant	Mammalian
Cyp19A1	ER	Extracellular	Secreted RC: 3	Secreted RC: 4	NN: 0.474
Cyp51A1	Chloroplast	Membrane-bound ER	Other RC: 3	Other RC: 5	NN: 0.616
Cyp51B1	Plasmodial	Membrane-bound ER	Secreted RC: 5	Secreted RC: 1	NN: 0.716 Signal peptide predicted
Cyp51B2	Plasmodial	Membrane-bound ER	Secreted RC: 4	Secreted RC: 1	NN: 0.716 Signal peptide predicted
Cyp86A1	Cytoplasmic	Membrane-bound, ER	Other RC: 5	Secreted RC: 3	NN: 0.513
Cyp86A2	Vacuolar	Membrane-bound, ER	Secreted RC: 5	Secreted RC: 5	NN: 0.525
Cyp86A3	Chloroplast	Membrane-bound, ER	Other RC: 3	Other RC: 5	NN: 0.565
Cyp86A4	Chloroplast	Mitochondrial	Mitochondrial RC: 4	Mitochondrial RC: 2	NN: 0.505
Cyp86A5	Vacuolar	Membrane-bound, ER	Secreted RC: 1	Secreted RC: 1	NN: 0.570
Cyp86A6	Chloroplast	Membrane-bound ER	Other RC: 5	Secreted RC: 1	NN: 0.654
Cyp86B1	Nuclear	Cytoplasmic	Other RC: 5	Other RC: 5	NN: 0.672
Cyp97A1	Chloroplast	Mitochondrial	Mitochondrial RC: 4	Secreted RC: 2	NN: 0.676 Signal peptide predicted
Cyp97B1	Chloroplast	Cytoplasmic	Other RC: 1	Other RC: 1	NN: 0.655
SrCyp7	Chloroplast	Membrane-bound ER	Secreted RC: 3	Secreted RC: 3	NN: 0.348
SrCyp8	Extracellular	Peroxisomal or mitochondrial	Other RC: 3	Other RC: 2	NN: 0.532
SrCyp10	Cytoplasmic	Cytoplasmic	Other RC: 3	Other RC: 5	NN: 0.407

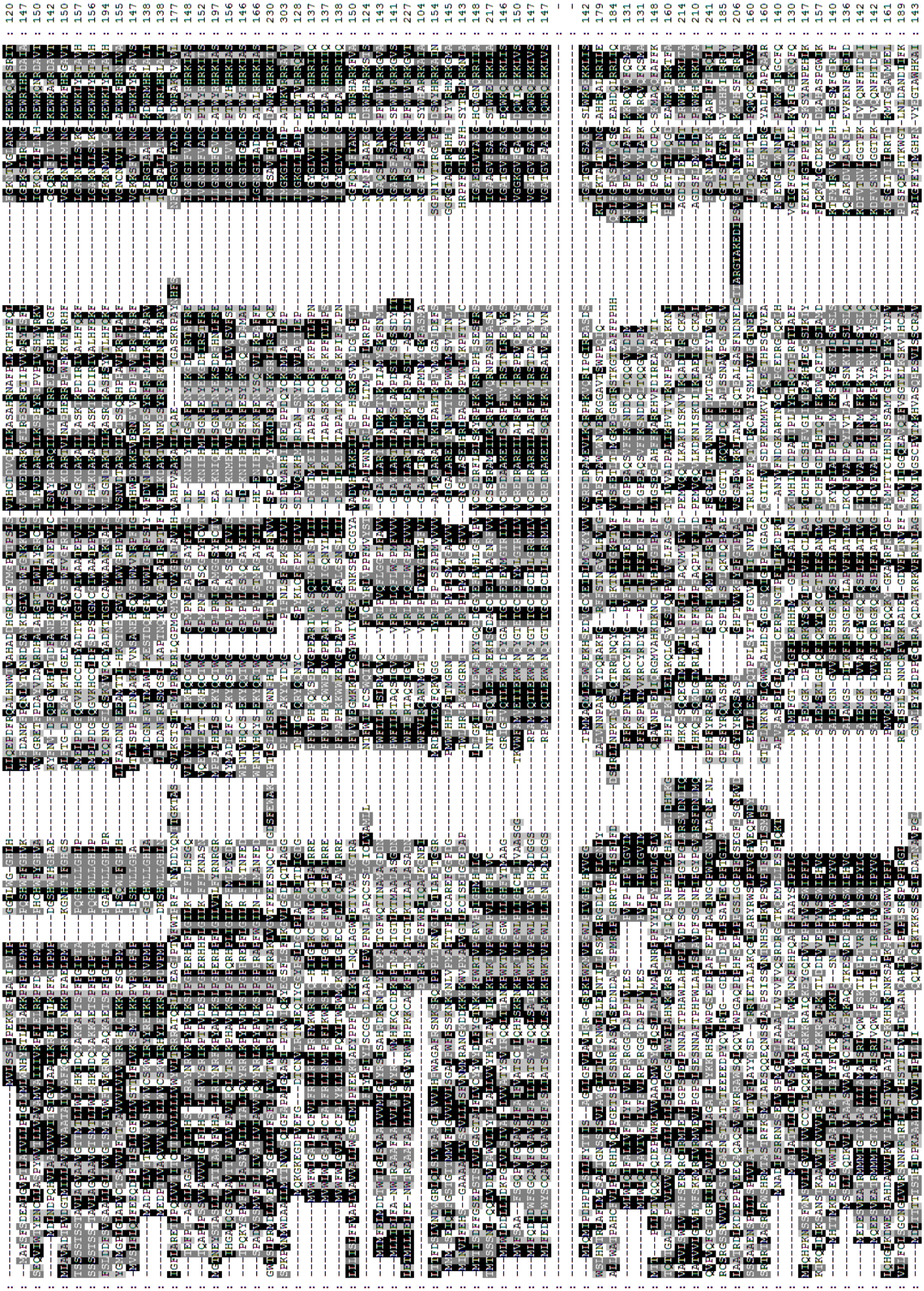
SrCyp11	Chloroplast	Membrane-bound ER	Secreted RC: 4	Secreted RC: 1	NN: 0.484 Signal peptide predicted
SrCyp12	Plasmodial	Membrane-bound ER	Secreted RC: 2	Secreted RC: 3	NN: 0.450 Signal peptide predicted
SrCyp13	ER	Membrane-bound ER	Secreted RC: 2	Secreted RC: 2	NN: 0.548 Signal peptide predicted
SrCyp16	Chloroplast	Membrane-bound ER	Secreted RC: 4	Secreted RC: 1	NN: 0.370 Signal peptide predicted
SrCyp17	Mitochondrial	Mitochondrial	Mitochondrial RC: 3	Mitochondrial RC: 3	NN: 0.492
SrCyp18	Nuclear	Membrane-bound ER	Secreted RC: 5	Secreted RC: 3	NN: 0.521 Signal peptide predicted
SrCyp19	ER	Membrane-bound ER	Other RC: 5	Other RC: 4	NN: 0.475
SrCyp20	Chloroplast	Cytoplasmic	Other RC: 4	Secreted RC: 5	NN: 0.474
SrCyp21	ER	Membrane-bound ER	Secreted RC: 2	Secreted RC: 1	NN: 0.434 Signal peptide predicted
SrCyp22	Chloroplast	Extracellular	Other RC: 3	Secreted RC: 3	NN: 0.490 Signal peptide predicted
SrCyp23	Vacuolar	Membrane-bound ER	Mitochondrial mTP: 0.746	Secreted SP: 0.913	NN: 0.344 Signal peptide predicted
SrCyp24	Chloroplast	Membrane-bound ER	Mitochondrial RC: 3	Secreted RC: 3	NN: 0.500 Signal peptide predicted
SrCyp25	Chloroplast	Membrane-bound ER	Secreted RC: 5	Secreted RC: 1	NN: 0.703 Signal peptide predicted
SrCyp28	Plasmodial	Plasmodial	Secreted RC: 5	Secreted RC: 1	NN: 0.729 Signal peptide predicted
SrCyp29	Chloroplast	Membrane-bound ER	Other RC: 2	Secreted RC: 1	NN: 0.722 Signal peptide predicted
SrCyp30	Plasmodial	Membrane-bound ER	Secreted RC: 5	Other RC: 5	NN: 0.528
SrCyp31	Plasmodial	Membrane-bound ER	Secreted RC: 5	Secreted RC: 4	NN: 0.762

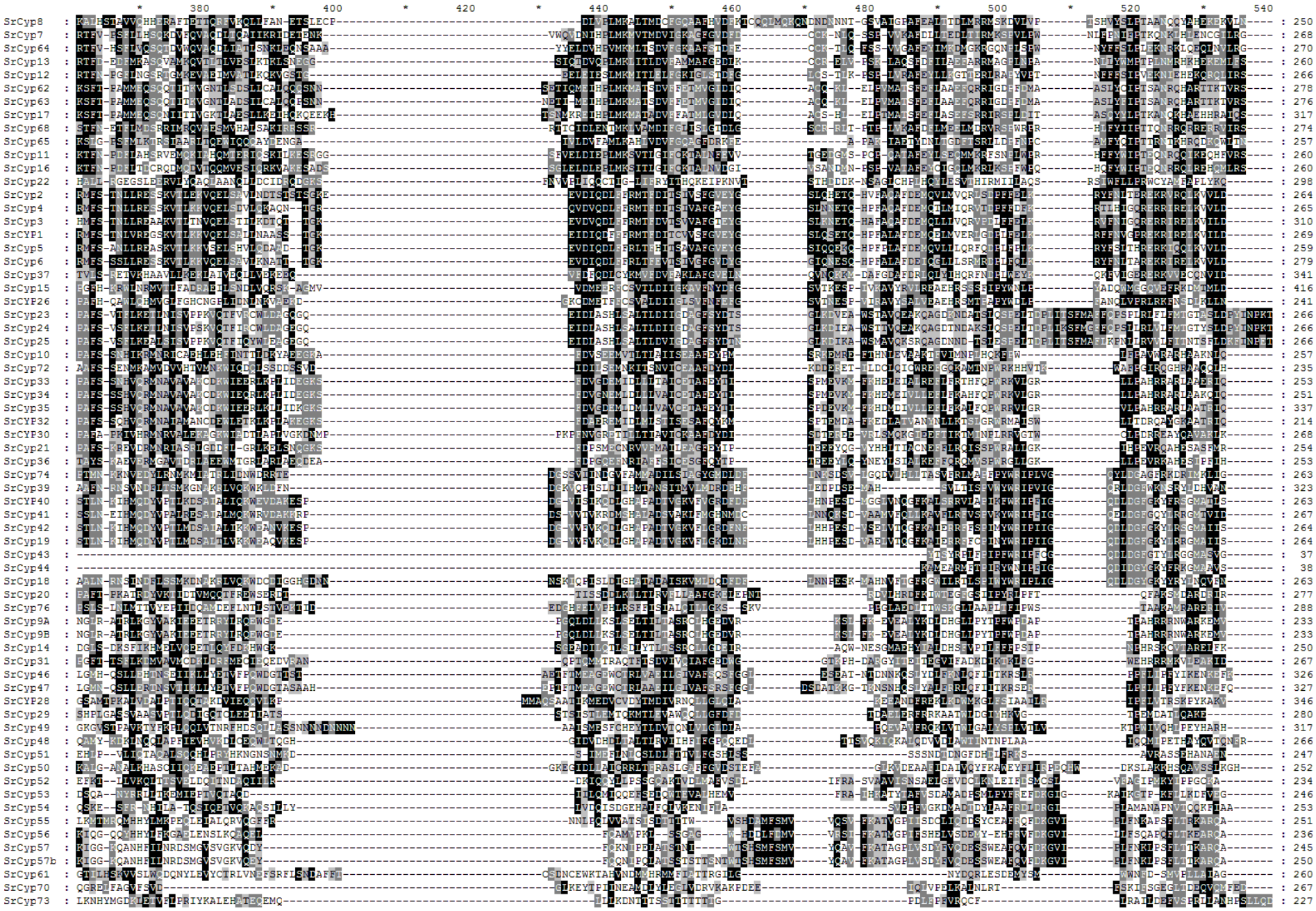
SrCyp32*	Cytoplasmic	Cytoplasmic	Other RC: 3	Other RC: 1	NN: 0.509
SrCyp33	Plasmodial	Membrane-bound ER	Other RC: 5	Secreted RC: 2	NN: 0.690 Signal peptide predicted
SrCyp34	Cytoplasmic	Extracellular	Secreted RC: 3	Secreted RC: 2	NN: 0.508 Signal peptide predicted
SrCyp35	Nuclear	Membrane-bound ER	Other RC: 1	Other RC: 4	NN: 0.267 Signal peptide predicted
SrCyp36	Nuclear	Membrane-bound ER	Secreted RC: 2	Secreted RC: 2	NN: 0.188 Signal peptide predicted
SrCyp39	Cytoplasmic	Cytoplasmic	Other RC: 3	Other RC: 3	NN: 0.696
SrCyp40	ER	Membrane-bound ER	Secreted RC: 3	Secreted RC: 1	NN: 0.554 Signal peptide predicted
SrCyp41	Plasmodial	Plasmodial	Secreted RC: 1	Secreted RC: 3	NN: 0.648 Signal peptide predicted
SrCyp42	ER	Membrane-bound ER	Secreted RC: 1	Secreted RC: 1	NN: 0.514 Signal peptide predicted
SrCyp43*	Mitochondrial	Mitochondrial	Mitochondrial RC: 4	Other RC: 5	NN: 0.497
SrCyp44*	Chloroplast	Cytoplasmic	Other RC: 3	Other RC: 2	NN: 0.383
SrCyp46	Nuclear	Membrane-bound ER	Other RC: 4	Secreted RC: 2	NN: 0.849 Signal peptide predicted
SrCyp47	Extracellular	Extracellular	Secreted RC: 5	Secreted RC: 1	NN: 0.756 Signal peptide predicted
SrCyp48	Chloroplast	Extracellular	Other RC: 5	Other RC: 5	NN: 0.201 Odds: 0.404
SrCyp49	Vacuolar	Membrane-bound ER	Secreted RC: 2	Secreted RC: 1	NN: 0.732 Signal peptide predicted
SrCyp51	Mitochondrial	Mitochondrial	Mitochondrial RC: 4	Mitochondrial RC: 1	NN: 0.584
SrCyp52	Mitochondrial	Nuclear	Mitochondrial RC: 4	Mitochondrial RC: 1	NN: 0.584

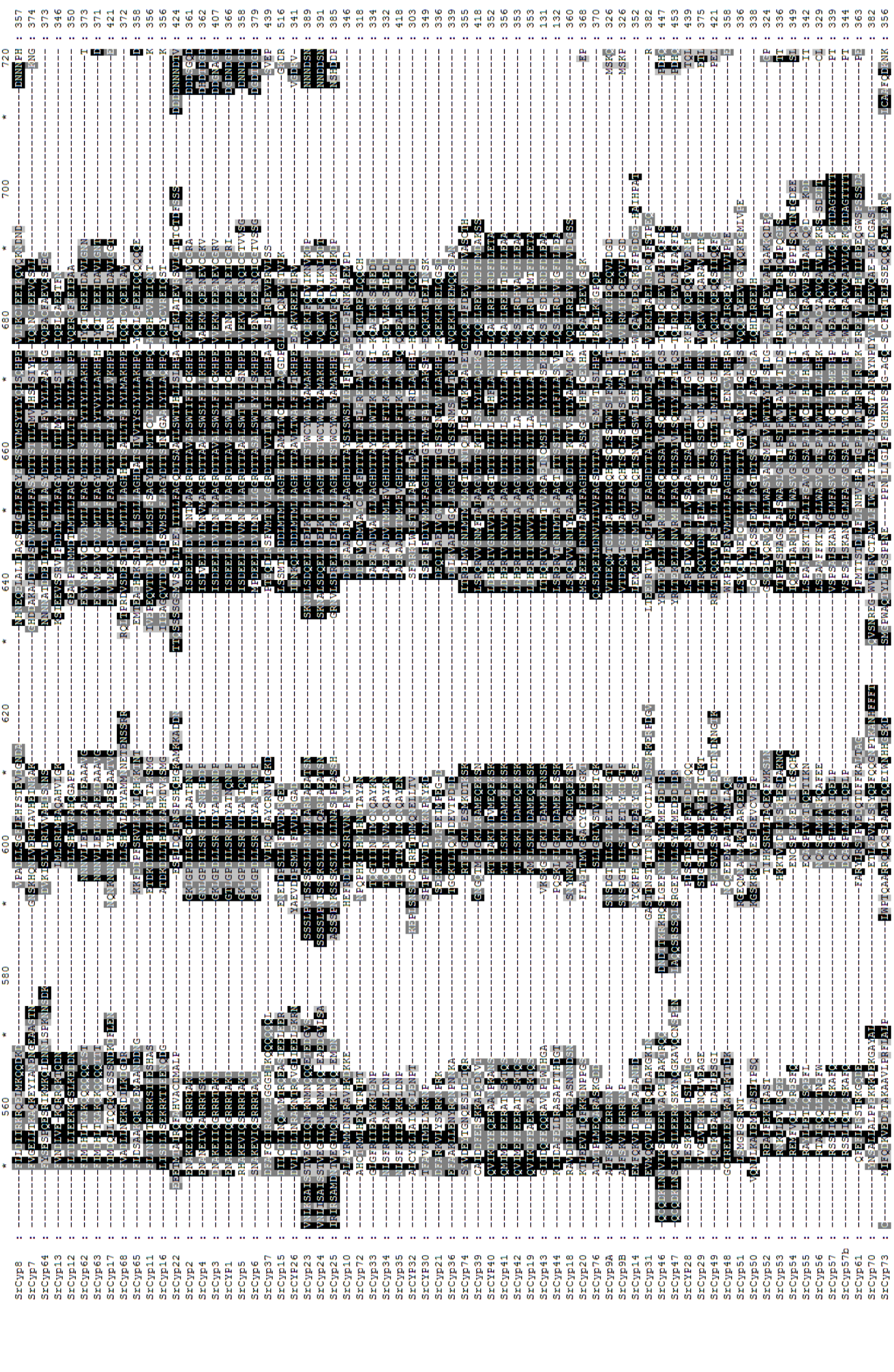
SrCyp53	Extracellular	Extracellular	Secreted RC: 1	Secreted RC: 1	NN: 0.509 Signal peptide predicted
SrCyp54	Mitochondrial	Mitochondrial	Mitochondrial RC: 4	Other RC: 5	NN: 0.528
SrCyp55	Chloroplast	Extracellular	Other RC: 4	Secreted RC: 1	NN: 0.550 Signal peptide predicted
SrCyp56	ER	ER	Secreted RC: 2	Secreted RC: 3	NN: 0.635
SrCyp57	Chloroplast	Membrane-bound ER	Secreted RC: 4	Secreted RC: 3	NN: 0.597 Signal peptide predicted
SrCyp57b	ER	Extracellular	Secreted RC: 3	Secreted RC: 3	NN: 0.521 Signal peptide predicted
SrCyp61	Nuclear	Cytoplasmic	Secreted RC: 5	Secreted RC: 2	NN: 0.609
SrCyp62	ER	Extracellular	Chloroplast RC: 4	Other RC: 4	NN: 0.550
SrCyp63	ER	Extracellular	Chloroplast RC: 4	Mitochondrial RC: 5	NN: 0.489
SrCyp64	ER	Extracellular	Secreted RC: 3	Secreted RC: 2	NN: 0.541
SrCyp65	ER	Membrane-bound ER	Secreted RC: 2	Secreted RC: 2	NN: 0.400 Signal peptide predicted
SrCyp67	Cytoplasmic	Membrane-bound ER	Other RC: 5	Secreted RC: 4	NN: 0.596
SrCyp68	ER	ER	Secreted RC: 2	Secreted RC: 1	NN: 0.318 Signal peptide predicted
SrCyp70	Chloroplast	Mitochondrial	Mitochondrial RC: 4	Mitochondrial RC: 3	NN: 0.436
SrCyp72	ER	Extracellular	Other RC: 3	Secreted RC: 1	NN: 0.539 Signal peptide predicted
SrCyp73	Nuclear	Membrane-bound ER	Other RC: 3	Other RC: 5	NN: 0.422
SrCyp74	ER	Membrane-bound ER	Secreted RC: 4	Secreted RC: 5	NN: 0.772
SrCyp76	Chloroplast	Mitochondrial	Mitochondrial RC: 4	Other RC: 5	NN: 0.624 Signal peptide predicted

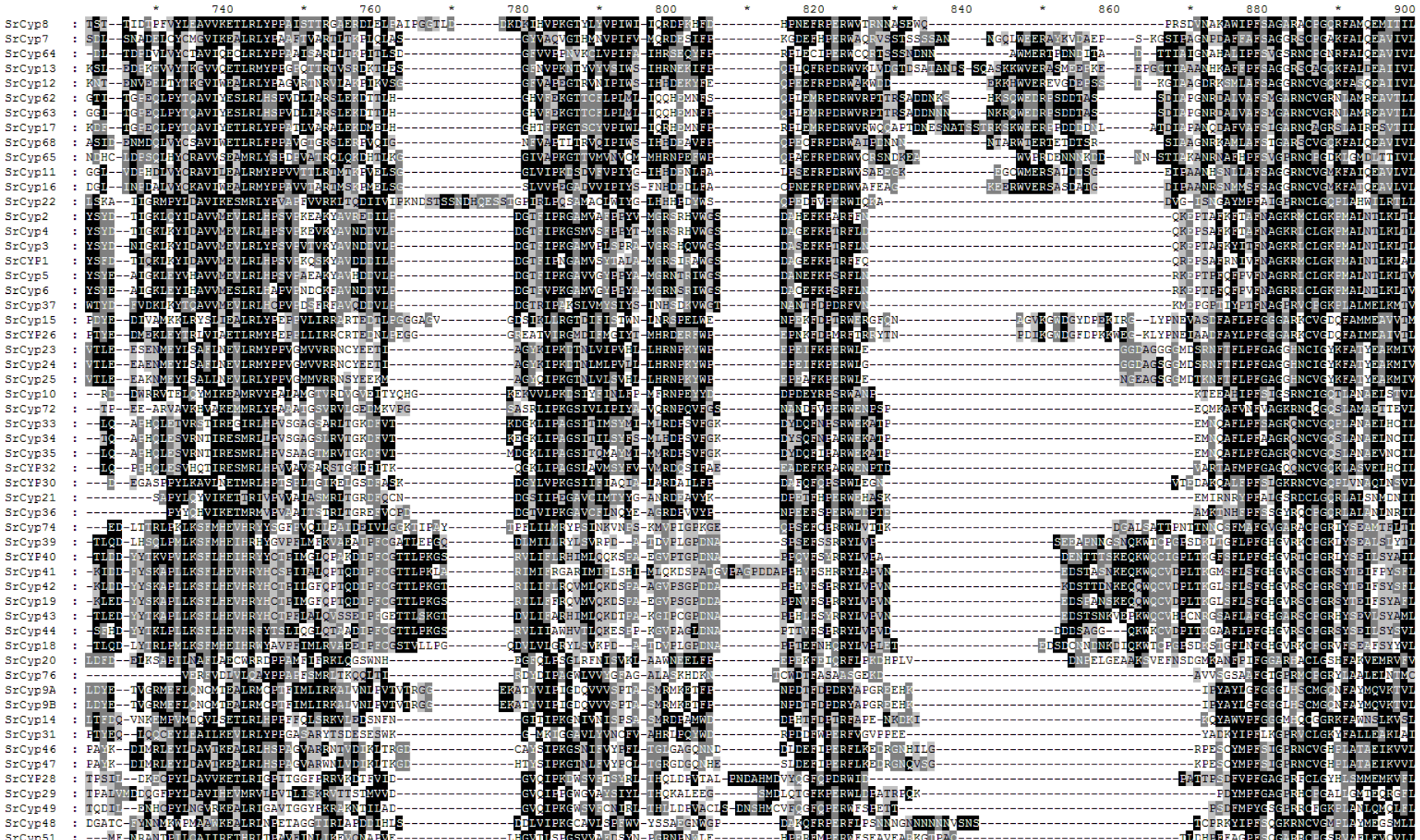
A7. GENEDOC ALIGNMENT










SrCyp8
SrCyp7
SrCyp64
SrCyp13
SrCyp12
SrCyp62
SrCyp63
SrCyp17
SrCyp68
SrCyp65
SrCyp11
SrCyp16
SrCyp22
SrCyp2
SrCyp4
SrCyp3
SrCYP1
SrCyp5
SrCyp6
SrCyp37
SrCyp15
SrCYP26
SrCyp23
SrCyp24
SrCyp25
SrCyp10
SrCyp72
SrCyp33
SrCyp34
SrCyp35
SrCYP32
SrCYP30
SrCyp21
SrCyp36
SrCyp74
SrCyp39
SrCYP40
SrCyp41
SrCyp42
SrCyp19
SrCyp43
SrCyp44
SrCyp18
SrCyp20
SrCyp76
SrCyp9A
SrCyp9B
SrCyp14
SrCyp31
SrCyp46
SrCyp47
SrCYP28
SrCyp29
SrCyp49
SrCyp48
SrCyp51
SrCyp50
SrCyp52
SrCyp53
SrCyp54
SrCyp55
SrCyp56
SrCyp57
SrCyp57b
SrCyp61
SrCyp70
SrCyp73

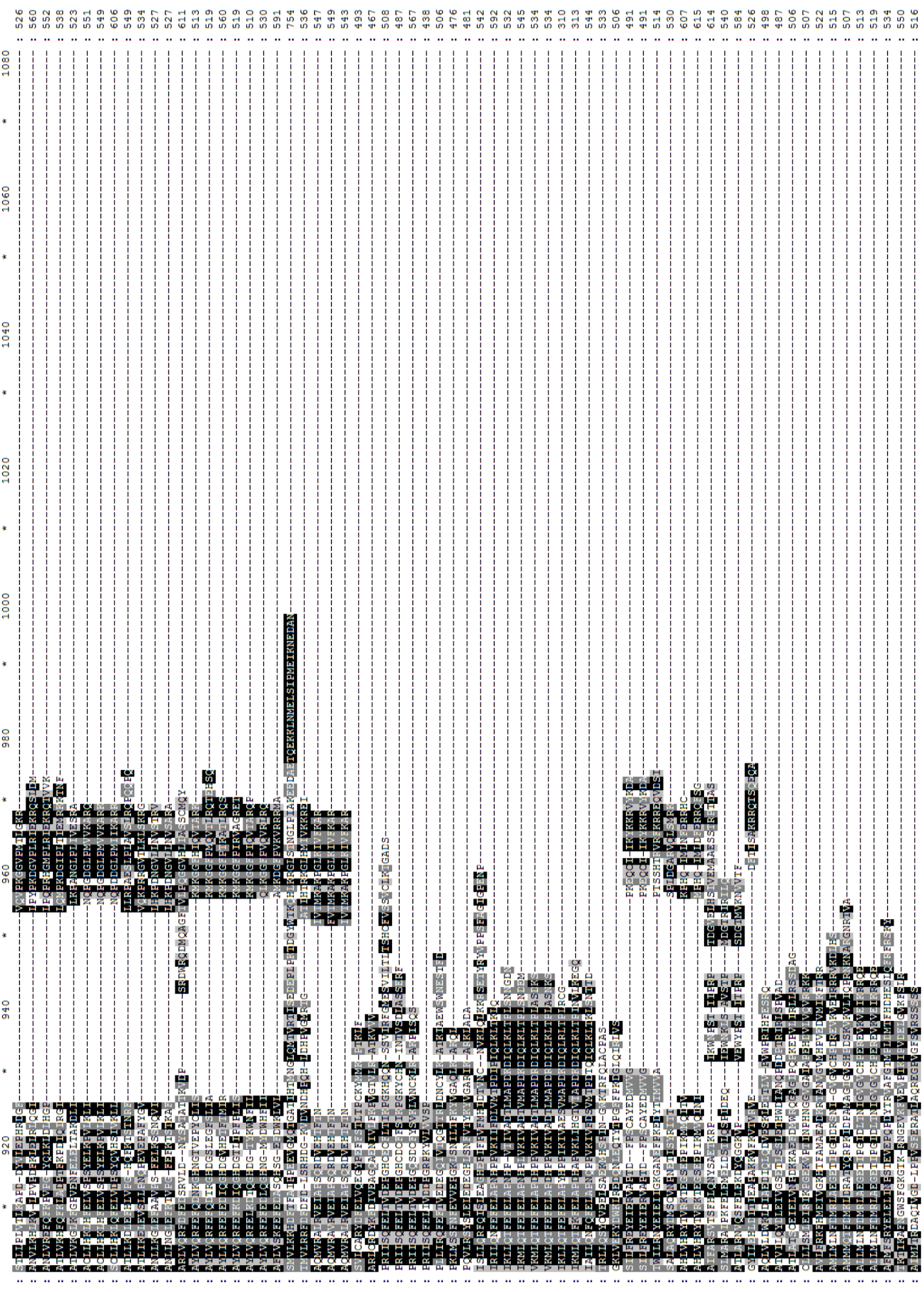


Figure A 28 GeneDoc alignment of the CYPs in *S. robusta*. The original alignment contains related proteins from other organisms, but is too large to include.

A8.AMINO ACID SIMILARITY

The CYPs that are still not complete are marked by a * in the tables below.

Table A 59 Amino acid similarity in the CYP86 family.

	Cyp86A2	Cyp86A4	Cyp86A3	Cyp86A5	Cyp86A6	Cyp86B1	
Cyp86A1	71 %	63 %	67 %	64 %	59 %	29 %	% Exact
	84 %	76 %	81 %	79 %	76 %	47 %	%Score
	2 %	9 %	2 %	2 %	4 %	18 %	%Gap
Cyp86A2		63 %	63 %	61 %	58 %	30 %	% Exact
		75 %	80 %	77 %	76 %	48 %	%Score
		8 %	1 %	3 %	4 %	18 %	%Gap
Cyp86A4			61 %	57 %	56 %	31 %	% Exact
			73 %	71 %	71 %	48 %	%Score
			7 %	10 %	6 %	10 %	%Gap
Cyp86A3				58 %	58 %	28 %	% Exact
				75 %	76 %	47 %	%Score
				2 %	2 %	16 %	%Gap
Cyp86A5					73 %	29 %	% Exact
					83 %	49 %	%Score
					4 %	18 %	%Gap
Cyp86A6						29 %	% Exact
						48 %	%Score
						15 %	%Gap

Table A 60 Amino acid similarity in the CYP51 family.

	Cyp51B2	Cyp51A1	
Cyp51B1	99 %	28 %	% Exact
	99 %	48 %	%Score
	0 %	10 %	%Gap
Cyp51B2		28 %	% Exact
		48 %	%Score
		10 %	%Gap

Table A 61 Amino acid similarity in the CYP97 family.

	Cyp97B1	
Cyp97A1	34 %	% Exact
	45 %	%Score
	30 %	%Gap

Table A 62 Amino acid similarity in Families I and II

	SrCyp72	SrCyp33	SrCyp34	SrCyp35	SrCyp32*	SrCyp30	SrCyp21	SrCyp36	
SrCyp10	22 %	27 %	25 %	21 %	23 %	24 %	21 %	24 %	% Exact
	38 %	47 %	46 %	40 %	41 %	43 %	41 %	41 %	%Score
	11 %	10 %	7 %	18 %	15 %	9 %	7 %	7 %	%Gap
SrCyp72		22 %	25 %	21 %	19 %	24 %	20 %	22 %	% Exact
		39 %	39 %	35 %	37 %	38 %	37 %	38 %	%Score
		12 %	9 %	21 %	10 %	13 %	10 %	9 %	%Gap
SrCyp33			70 %	57 %	42 %	29 %	30 %	27 %	% Exact
			81 %	67 %	56 %	47 %	48 %	48 %	%Score
			4 %	19 %	16 %	9 %	9 %	9 %	%Gap
SrCyp34				68 %	41 %	29 %	27 %	27 %	% Exact
				75 %	56 %	48 %	48 %	48 %	%Score
				16 %	13 %	7 %	6 %	6 %	%Gap
SrCyp35					36 %	27 %	25 %	23 %	% Exact
					48 %	43 %	43 %	43 %	%Score
					25 %	18 %	18 %	18 %	%Gap
SrCyp32*						23 %	23 %	23 %	% Exact
						39 %	41 %	41 %	%Score
						18 %	14 %	15 %	%Gap
SrCyp30							24 %	26 %	% Exact
							44 %	45 %	%Score
							8 %	9 %	%Gap
SrCyp21								47 %	% Exact
								66 %	%Score
								3 %	%Gap

Table A 63 Amino acid similarity in Family III.

	SrCyp24	SrCyp25	
SrCyp23	91 %	76 %	% Exact
	94 %	90 %	%Score
	0 %	1 %	%Gap
SrCyp24		76 %	% Exact
		88 %	%Score
		1 %	%Gap

Table A 64 Amino acid similarity in Family IV

	SrCyp29	SrCyp49*	
SrCyp28	24 %	34 %	% Exact
	41 %	50 %	%Score
	17 %	18 %	%Gap
SrCyp29		27 %	% Exact
		43 %	%Score
		12 %	%Gap

Table A 65 Amino acid similarity in Families V and VI.

	SrCyp46	SrCyp47	
SrCyp31	13 %	13 %	% Exact
	27 %	27 %	%Score
	17 %	18 %	%Gap
SrCyp46		72 %	% Exact
		81 %	%Score
		2 %	%Gap

Table A 66 Amino acid similarity in Families VII and VIII

	SrCyp76	
SrCyp20	15 %	% Exact
	32 %	%Score
	12 %	%Gap

Table A 67 Amino acid similarity in Families IX, X, XI and XII.

	SrCyp53	SrCyp54	SrCyp55	SrCyp56	SrCyp57	SrCyp57b	SrCyp61	
SrCyp52	23 %	18 %	20 %	21 %	21 %	21 %	15 %	% Exact
	39 %	36 %	38 %	38 %	37 %	37 %	28 %	%Score
	7 %	12 %	10 %	12 %	11 %	10 %	18 %	%Gap
SrCyp53		27 %	22 %	21 %	23 %	23 %	11 %	% Exact
		47 %	40 %	37 %	40 %	40 %	24 %	%Score
		8 %	9 %	11 %	10 %	9 %	16 %	%Gap
SrCyp54			24 %	21 %	25 %	25 %	10 %	% Exact
			40 %	39 %	41 %	42 %	27 %	%Score
			11 %	13 %	12 %	11 %	12 %	%Gap
SrCyp55				44 %	51 %	51 %	12 %	% Exact
				62 %	68 %	67 %	26 %	%Score
				5 %	6 %	7 %	17 %	%Gap
SrCyp56					40 %	41 %	12 %	% Exact
					60 %	60 %	24 %	%Score
					6 %	7 %	19 %	%Gap
SrCyp57						95 %	12 %	% Exact
						97 %	28 %	%Score
						1 %	19 %	%Gap
SrCyp57b							12 %	% Exact
							28 %	%Score
							18 %	%Gap

Table A 68 Amino acid similarity in Families XIII and XIV.

	SrCyp7	SrCyp64	SrCyp13	SrCyp12	SrCyp62	SrCyp63	SrCyp17	SrCyp68	SrCyp65	SrCyp11	SrCyp16	SrCyp67	
SrCyp8	21 %	22 %	24 %	20 %	22 %	22 %	22 %	21 %	22 %	22 %	22 %	20 %	% Exact
	39 %	39 %	41 %	39 %	40 %	40 %	37 %	39 %	39 %	40 %	40 %	36 %	%Score
	18 %	18 %	18 %	19 %	17 %	17 %	24 %	20 %	16 %	15 %	15 %	18 %	%Gap
SrCyp7		38 %	33 %	32 %	27 %	27 %	26 %	30 %	27 %	31 %	31 %	28 %	% Exact
		57 %	52 %	52 %	47 %	47 %	46 %	51 %	48 %	52 %	51 %	48 %	%Score
		4 %	6 %	8 %	7 %	8 %	12 %	9 %	7 %	7 %	7 %	8 %	%Gap
SrCyp64			33 %	36 %	27 %	28 %	27 %	30 %	27 %	31 %	31 %	28 %	% Exact
			50 %	55 %	48 %	48 %	44 %	52 %	49 %	51 %	51 %	47 %	%Score
			7 %	7 %	8 %	8 %	13 %	8 %	6 %	7 %	8 %	8 %	%Gap
SrCyp13				39 %	28 %	28 %	27 %	33 %	29 %	31 %	33 %	29 %	% Exact
				57 %	47 %	47 %	44 %	52 %	50 %	54 %	54 %	50 %	%Score
				6 %	8 %	8 %	15 %	8 %	8 %	7 %	7 %	8 %	%Gap
SrCyp12					27 %	27 %	26 %	41 %	27 %	38 %	40 %	33 %	% Exact
					46 %	45 %	45 %	59 %	49 %	58 %	61 %	51 %	%Score
					6 %	6 %	14 %	6 %	7 %	7 %	6 %	6 %	%Gap
SrCyp62						93 %	56 %	29 %	24 %	27 %	27 %	26 %	% Exact
						96 %	70 %	48 %	42 %	47 %	47 %	45 %	%Score
						0 %	9 %	7 %	7 %	8 %	8 %	8 %	%Gap
SrCyp63							55 %	29 %	25 %	28 %	27 %	26 %	% Exact
							70 %	48 %	43 %	47 %	47 %	45 %	%Score
							9 %	7 %	7 %	8 %	8 %	8 %	%Gap
SrCyp17								26 %	22 %	25 %	25 %	23 %	% Exact
								43 %	41 %	44 %	44 %	40 %	%Score
								15 %	14 %	15 %	15 %	16 %	%Gap
SrCyp68									27 %	37 %	38 %	33 %	% Exact
									47 %	55 %	57 %	49 %	%Score
									9 %	7 %	7 %	7 %	%Gap
SrCyp65										29 %	29 %	28 %	% Exact
										50 %	48 %	46 %	%Score
										7 %	7 %	7 %	%Gap
SrCyp11											67 %	37 %	% Exact
											81 %	56 %	%Score
											0 %	4 %	%Gap
SrCyp16												43 %	% Exact
												58 %	%Score
												4 %	%Gap

Table A 69 Amino acid similarity in Families XV and XVI.

	SrCyp39	SrCyp40	SrCyp41	SrCyp42	SrCyp19	SrCyp43*	SrCyp44*	SrCyp18	
SrCyp74	26 %	30 %	26 %	28 %	28 %	17 %	17 %	29 %	% Exact
	43 %	48 %	46 %	48 %	49 %	28 %	27 %	49 %	%Score
	19 %	7 %	9 %	7 %	7 %	47 %	46 %	9 %	%Gap
SrCyp39		38 %	34 %	37 %	37 %	23 %	23 %	53 %	% Exact
		53 %	49 %	54 %	54 %	32 %	31 %	66 %	%Score
		17 %	18 %	17 %	17 %	50 %	50 %	17 %	%Gap
SrCyp40			58 %	68 %	67 %	34 %	35 %	46 %	% Exact
			75 %	84 %	83 %	44 %	45 %	63 %	%Score
			3 %	1 %	1 %	43 %	43 %	4 %	%Gap
SrCyp41				70 %	67 %	31 %	31 %	40 %	% Exact
				82 %	81 %	41 %	40 %	59 %	%Score
				2 %	2 %	44 %	44 %	6 %	%Gap
SrCyp42					90 %	35 %	36 %	44 %	% Exact
					95 %	45 %	44 %	64 %	%Score
					0 %	43 %	43 %	4 %	%Gap
SrCyp19						34 %	35 %	44 %	% Exact
						44 %	45 %	64 %	%Score
						43 %	43 %	4 %	%Gap
SrCyp43*							55 %	26 %	% Exact
							71 %	38 %	%Score
							2 %	43 %	%Gap
SrCyp44*								27 %	% Exact
								38 %	%Score
								43 %	%Gap