The influence of sulfur and nitrogen on cadmium tolerance in *Euglena gracilis*: an RNA-sequencing investigation

A thesis submitted to the Committee of Graduate Studies

in partial fulfillment of the requirements for the degree of

Master of Science

in the faculty of Arts and Science

TRENT UNIVERSITY

Peterborough, Ontario, Canada

Environmental and Life Sciences M.Sc. Graduate Program

September 2022

ABSTRACT

The Influence of nitrogen and sulfur on cadmium tolerance in *Euglena gracilis*: an RNA-Sequencing investigation

Victoria Kennedy

Heavy metal pollution threatens human and ecosystem health. *E. gracilis* was investigated for its potential use in bioremediation due to its tolerance for heavy metals and ability to sequester them from the environment. *E. gracilis* can remove metals by producing metal binding compounds enriched in sulfur and nitrogen. In this thesis, *E. gracilis* cultures that were pretreated with elevated levels of sulfur or nitrogen had increased tolerance to $CdCl_2$ compared to non-pretreated cultures. RNA-sequencing revealed that both pretreatments led to transcript level changes and that exposure to $CdCl_2$ led to further transcript level changes. Gene ontology (GO) enrichment analysis reflected changes in nitrogen and sulfur metabolism as well as physiological processes related to metal binding. The data from this thesis revealed important transcription level changes that occur when *E. gracilis* is challenged with $CdCl_2$ and helps us understand how organisms adapt to heavy metal pollution in the environment.

KEYWORDS: *Euglena gracilis, Cadmium, bioremediation, metal-binding, RNA-Sequencing, GO-enrichment*

Acknowledgements

I would first like to thank my supervisor Dr. Barry Saville, who's willingness to take on new projects afforded me the opportunity to complete this work. His knowledge and financial support helped guide me through the switch from forensic science with an emphasis on chemistry to molecular biology and bioinformatics. Along with Dr. Saville, I would also like to thank my committee members, Dr. Michael Donaldson, Dr. Neil Emery, and Dr. Scott Farrow for their guidance during committee meetings and words of encouragement. I would especially like to thank Dr. Michael Donaldson for lending his expertise in bioinformatics, without it I probably would have smashed my computer.

Thank you to my peers in the Saville Lab including Amanda Seto, Emma Kaszecki and Neha Rhampertab for your support during weekly lab meetings, and for being there for every triumph and trial of my research. I would especially like to thank Emilee Storfie, who trained me during my early days as an undergrad, without her help, I would not be the researcher I am today. To Dr. Colleen Doyle, thank you for your constant optimism and for always being around when I need to troubleshoot. I cannot end this section without giving a special acknowledgement to Monique Lariviere for always being there with suggestions and coffee whenever I spiraled.

Finally, I would like to thank my support system: Solomon, Ashley, Ciara and Mackenzie and Anita for always being supportive and encouraging (special thanks to Ashley and Anita for always making sure I was fed when I would come home from lab late). To my family, thank you for your love and support throughout my long journey here at Trent, even if you wished I was closer to home. I also want to give a quick shoutout to both Starbucks and tequila for helping me get through the long days in lab and writing.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

°C	degrees Celsius
3’	three-prime
5’	five-prime
As	arsenic
BLAST	basic local alignment search tool
BLASTX	basic local alignment search tool (nucleotide to protein query)
Cd	cadmium
$CdCl_2$	Cd chloride
cDNA	complementary deoxyribonucleic acid
Co	cobalt
Cr	chromium
Cu	copper
DEPC	diethylpyrocarbonate
dH_2O	deionized water
DNA	deoxyribonucleic acid
dT	oligo (dT)16
EtOH	ethanol
Fe	iron
g	gram (s)
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
hrs	hours
L	litre (s)
LN_2	liquid nitrogen
M	molar
MAM	modified acid media
mins	minutes
µg	microgram (s)
µL	microlitre (s)
mL	millilitre (s)
Mn	maganese
n	sample size
ng	nanogram (s)
Ni	nickel
NH_4NO_3	ammonium nitrate
P	p-value (probability)
Pb	lead
PCs	phytochelatins
PCR	polymerase chain reaction
RNA-Seq	RNA sequencing
RT	room temperature
s	second (s)
SO_4	sulfate
Zinc	Zn

CHAPTER ONE: GENERAL INTRODUCTION

In 2017, 90% of the global population had an improved drinking water source that was at least a 30 min round trip from their homes, while 785 million people continue to lack this resource (World Health Organization, 2022). Access to clean water can become more challenging in instances where water pollution is prevalent. Water can become polluted through sewage leakages and from industrial wastewater (Owa, 2013). Industrial wastewater includes waters that are generated from the production of goods such as batteries, fertilizers, and mining operations. All of which, are common sources of wastewater pollution (Borba et al., 2006; Fu & Wang, 2011).

Mine runoffs as well as mine tailing ponds can directly pollute water sources as well as produce aerosolized pollution and soil pollution that can be carried to neighboring water sources (Khademi et al., 2018). In one study of the processes that lead to pollution, the dust in abandoned mine tailing ponds and wind eroded particles showed high concentrations of Cd, Zn, Pb and Mn (Khademi et al., 2018). The industrial wastewater produced by mine tailing ponds as well as the aerosolization of the pollutants, can impact agriculture (Gabarrón et al., 2018). Agriculture soils found near mine tailing ponds such as La Unión and Mazarrón in Spain were enriched in As, Cd, Pb, Zn and iron (Fe) because of their proximity to the mine tailing ponds (Gabarrón et al., 2018).

Since mine tailing ponds have the potential to pollute both soil and groundwater, they can have devastating impacts on both human and ecosystem health. In the Gadoon Amazaii industrial estate in Swabi Pakistan, industrial wastewater as well as groundwater

were investigated for their physicochemical properties, the concentrations of toxic elements, and the risk to human health (Muhammad et al., 2021). The pollution from wastewater irrigation contained materials that posed carcinogenic and non-carcinogenic dermal health risks (Muhammad et al., 2021). Along with the risks to human health, industrial wastewater is known to impact the health of the surrounding ecosystems (Wright et al., 2015) The contamination of neighboring bodies of water by wastewater from mine tailing ponds is found to be detrimental to the biodiversity of aquatic ecosystems (Wright et. al., 2015). In Redbank Creek and Bargo River (New South Wales), the mine tailing ponds caused a reduction of dissolved oxygen levels, and a 4 X and 20 X increase in the Zn and Ni concentration respectively. Both changes contributed to a decrease in biodiversity (Wright et al., 2015).

Currently, methods exist to clean up the pollution caused by mine tailing ponds. These methods include chemical precipitation, filtration systems, resin-based ion exchange, flotations, coagulation, and the use of activated carbons (Fu & Wang, 2011). While these methods are currently accepted, they are expensive and require a substantial amount of energy (Volesky, 2001). Another drawback of these older methods is their overall ineffectiveness at cleaning pollutants from industrial wastewater. It was found that on average these methods remove less than 100 mg/L of heavy metals from water which is insufficient to tackle the amount of heavy metal contaminants from mine tailing ponds (Volesky, 2001). Due to these drawbacks, research has begun to focus on removing these toxic metals using bioremediation.

Bioremediation is the process of using living organisms to remove contaminants and pollutants from the environment (Boopathy, 2000). Fungi are one group of organisms known for their role in bioremediation. Two examples include *Beaveria bassina* and *Rhodotorula mucilaginosa* (Purchase et al., 2009). When grown in environments with elevated Zn and Pb, *B. bassina* absorbs up to 0.64% (~200 mg/L) of the available Zn and up to 8.44% (~100 mg/L) of Pb, while *R. mucilaginosa* absorb 2.05% (~200 mg/L) of the total Zn and 16.55% (~ 40 mg/L) of the total Pb in the media. Both *B. bassina* and *R. mucilaginosa* show no difference in metal tolerance at decreased temperatures (Purchase et al., 2009). Other fungal species such as *Aspergillus niger*, *Aspergillus foetidus* and *Penicillium simplicissimum* have also demonstrated a tolerance to heavy metals including molybdenum (Mo), vanadium (V) as well as manganese (Mn) (Anahid et al., 2011). While *Aspergillus foetidus* has a lower tolerance to metals such as Ni, Co, and Zn, all three fungi showed an elevated tolerance to Mo and V (Anahid et al., 2011). The tolerance of some species of fungi makes them suitable as agents to use in the bioremediation of metal polluted environments.

While bioremediation commonly uses microbes, the cultivation of specific plants can also be used for bioremediation efforts since the rhizosphere creates an environment that is favorable for microbial proliferation (Stout & Nüsslein, 2010). Phytoremediation can be accomplished through the exploitation of the plant-endophyte relationship (Weyens et al., 2009). Some heavy metals including arsenic can be absorbed in the leaves of the wildflower *Ballota hirsute*, which is native to the western Mediterranean (Gabarrón et al., 2018). Soils and plants were collected from the La Unión and Mazarrón mining districts in Spain, and results showed that the mine forests and the agricultural soils were both

contaminated with As, Cd, Pb, Zn , and Cu from mine waste (Gabarrón et al., 2018). The metal(oids) accumulated in the leaves by *B. hirsuta* differed between La Unión and Mazarrón, with Fe being the most accumulated in La Unión and As being the most accumulated in Mazarrón (Gabarrón et al., 2018). The ability of *B. hirsuta* to absorb these metals, make it a tool in bioremediation efforts similar to fungi.

Along with fungi and plants, *E. gracilis* has emerged as a potential organism that can be used for bioremediation. *E. gracilis* has the potential to absorb high concentrations of heavy metals including Cu, Pb, Hg and Cd (Winters et al., 2017, Devars et al., 2000, (Avilés et al., 2003). It was initially found that after being exposed to 100 μM of HgCl, *E. gracilis* cells were able to uptake HgCl from growth media and retain it (Devars et al., 2000). More recently, toxicity assays showed that the EC_{50} values for *E. gracilis* cells were 39.2 mg/L for Cu and 27.4 mg/L for Ni. The sorption experiments performed in tandem with the toxicity assay showed that the bioaccumulation of both Cu and Ni occurred within 10-30 mins before plateauing in single metal systems (Winters et al., 2017). While plants such as *Ballota hirsute* absorbed approximately 1.7 mg/kg of Ni and 12 mg/kg of Cu, the removal of Ni by *E. gracilis* was found to increase at higher concentrations (5-110 mg/L) (Winters et al., 2017). This indicates that *E. gracilis* might be more efficient at absorbing specific metals compared to other organisms.

While the above--mentioned organisms can absorb metals naturally, metal absorption can be performed in other species using enhanced pretreatments. Rakhshaee *et. al.*, collected samples of *Lemna minor* (common duckweed) from a lake in northern Iran and pretreated the samples with various concentrations of hydrochloric acid (HCl), sodium

hydroxide (NaOH), calcium chloride ($CaCl_2$), magnesium chloride ($MgCl_2$) and sodium chloride (NaCl) at different pH values. (Rakhshaee et al., 2009). After pretreatments, *L. minor was* able to remove various concentrations of chromium (Cr) III and IV and Cu II. This was due to the increase in $-COO^-$ groups from the pH changes and the activation with chloride (Rakhshaee et al., 2009). Likewise, *E. gracilis* also had an increased tolerance to heavy metals after pretreatment. Pretreatment of *E. gracilis* cells with $HgCl_2$ resulted in cultures with a higher tolerance to $CdCl_2$ when compared to controls. This indicated that pretreatment with HgCl enhanced the tolerance of *E. gracilis* to $CdCl_2$ (Avilés et al., 2003).

Current methods of heavy industrial wastewater treatment such as filtration, coagulation, chemical precipitation, and ion exchange are not economically or environmentally sustainable (Fu & Wang, 2011, Volesky, 2011). As such, bioremediation is emerging as a new method for industrial wastewater cleanup. *E. gracilis*, like other organisms, can remove heavy metals such as Hg and Cd from its environment. This ability can be enhanced through pretreating the cells. *E. gracilis* has the added benefit of being able to tolerate and absorb multiple heavy metals in a range of otherwise harsh environments, making it a superior candidate for bioremediation compared to plants and fungi (Winters et al., 2017). Past studies into the removal of heavy metals by *Euglena gracilis* also found that once in the cell, metals such as mercury do not then get expelled from the cell back into the media (Devars et., al., 2000). Understanding the mechanisms of this tolerance will expand our knowledge of how *E. gracilis* can be used in bioremediation efforts to mitigate the harmful effects of heavy metal pollution.

There are two objectives to my research. Objective 1: Assess the tolerance of *E. gracilis* to $CdCl_2$ after pretreating the cells separately with either sulfur or nitrogen. Objective 2: identify gene functions that are associated with altered transcript levels in response to both the pretreatment with either sulfur or nitrogen, and the exposure to $CdCl_2$ To accomplish Objective 1, I pretreated axenic *E. gracilis* cultures with either 60 mM SO_4 or 75 mM of nitrogen in the form of NH_4^+ and NO_3^-, both of which are higher concentrations than what is currently found in most unpolluted environments. I then exposed both pretreated cultures and cells grown in Modified Acid Media (MAM) to various concentrations of $CdCl_2$ and counted the living cells remaining in the culture, using Trypan Blue to distinguish between living and dead cells. To complete Objective 2, I used RNA-sequencing, differential gene expression analysis, and GO term analysis to identify gene functions associated with altered transcript levels. The insight gained from this research will contribute to the understanding of methods to enhance heavy metal tolerance of *E. gracilis* and genes that are differentially expressed as a result of both the pretreatments and exposure to $CdCl_2$.

CHAPTER TWO: LITERATURE REVIEW

Cadmium exists naturally in the environment; however, mine tailing ponds can be significant sources of Cd pollution that can cause negative impacts on human and ecosystem health (Muhammad et al., 2021; Wright et al., 2015). Bioremediation methods can be used to remove Cd and other metals from mine tailing ponds (Volesky, 2001). Some microorganism and plants have a tolerance to Cd and can be used in these bioremediation efforts. To help augment the tolerance and bioremediation capabilities of some plants, nitrogen and sulfur have been used (Liang et al., 2016; Yotsova et al., 2020). Increased exposure to both sulfur and nitrogen have been linked to protecting photosynthesis in plants through maintaining the Asa-GSH cycle and mitigating the reduction of chlorophyll and carotenoids respectively (Liang et al., 2016; Yotsova et al., 2020). While sulfur must be added to plants, some microalgae grow in aquatic ecosystems which are rich in sulfur and are able to modulate their sulfur absorption to maintain metabolic pathways (Giordano et al., 2005). When grown in nitrogen-rich conditions, microalgae growth increased with carotenoid and chlorophyll content (World Health Organization, 2022). This demonstrates that both nitrogen and sulfur could play a role in increasing the metal tolerance of algae, as they do in plants. *E. gracilis* is a freshwater plankton that has a known tolerance to Cd that can be enhanced through the pretreatment of cells with HgCl (Avilés et al., 2003; Devars et al., 2000). Metabolomics and transcriptomics done on *E. gracilis* cells identified metal-binding compounds that were enriched with nitrogen and sulfur (Mangal et al., 2022). Some of these compounds shared similarities to thiol compounds which are known to bind metals (Mangal et al., 2022, Mangal & Guéguen, 2015). Along with the identified compounds, specific proteins have been found that are related to metal tolerance and

binding in *E. gracilis* (Khatiwada et al., 2020). My research objectives were to assess the tolerance of *E. gracilis* to $CdCl_2$ after pretreating cells separately with sulfur or nitrogen, and to identify gene candidates that may explain any pretreatment effects.

2.1 Cadmium in the Environment

Cadmium is found within the Earth's crust at a concentration of about 0.18 ppm and exists in the environments as a divalent cation bound with other elements such as chloride ($CdCl_2$) (Bernhoft, 2013; Mastalerz & Drobniak, 2007). In humans, Cd can cause tissue injury through the creation of oxidative stress within the cells (Matović et al., 2011). It can also cause epigenetic changes and the inhibition or upregulation of specific transport pathways (Wang et al., 2012). Cd toxicity in plants causes a plethora of issues including interfering with the uptake of important minerals like calcium, magnesium, phosphorous and potassium (Das et al., 1997). Cd is also known to interfere with nitrate reductase, causing a reduction in the ability of a plant to absorb nitrogen (Hernández et al., 1997).Within plants, Cd can induce the excessive production of reactive oxygen species (ROS) which can damage the plant's ability to photosynthesize (Mobin & Khan, 2007).

An environment with a Cd concentration between 0.32 µM and 1 µM is considered moderately polluted (Sanità Di Toppi & Gabbrielli, 1999). Some sources of Cd pollution include mine tailings, and the contamination of soil and neighbouring water sources by mine tailings (Zhao et. al., 2002). In southern China, approximately 50% of drinking water comes from surface water that can have Cd concentrations up to 100 µg/L (0.89 µM) (Zhao et al., 2002). While water can be directly contaminated by industrial wastewaters mixing with contaminated ground water, mine tailing ponds are also a significant source of both

water and soil contamination (Boussen et al., 2013). The erosion of mine tailing ponds acts as a dispersion method whereby wind can distribute aerosolized particles to surrounding areas and introduces the potential for increased environmental impact since the pollution cannot be contained in one area (Boussen et al., 2013)

Plants have some Cd defense mechanisms (Pinto et al., 2004). When plants such as sorghum and maize were exposed to various concentrations of Cd, a difference in organic acid exudation patterns were observed (Pinto et al., 2004). For example, sorghum showed enhanced malate exudation whereas maize showed increased citrate exudation indicating that malate and citrate are a part of their overall tolerance mechanisms, respectively (Pinto, Simo˜es and Mota, 2008).

2.2 Mine Tailings as a source of heavy metal pollution

As previously mentioned, mine tailing ponds are a significant source of heavy metal contamination. However, their impact on the environment comes not only from their constitutive metals, but also from the rate at which metals leach into the environment (Nikolaidis et al., 2010; Kim & Jung, 2003). Two examples of this are in the old tailing ponds from mines in both Greece and Korea. Samples from the Agios Philippos mine in Greece had a concentration of Cd and As in the soil around the tailing ponds of 174 mg/kg and 241 mg/kg, respectively (Nikolaidis et al., 2010). Water samples from the same tailing ponds had no detectable levels of As but did have an average Cd concentration of ~4.08 μg/L (0.0362 μM) with the highest concentration being 11.2 μg/L (0.0993 μM). The leaching rates of metals was found to be 0.14 - 1.7 % for arsenic and Cd, respectively, in mine tailing ponds in Korea (Kim & Jung, 2004). The following research indicated that

tailing ponds are not the only sources for heavy metal contamination. Rather that the areas surrounding mine tailing ponds are also contaminated and pose risks to ecosystems. The danger not only coming from the high concentration of Cd, but also from the rate at which it leached into the environment.

2.3 Role of nitrogen and sulfur in cadmium tolerance

While Cd pollution from sources such as mine tailing ponds poses a risk to surrounding ecosystems, there are some plants that have a tolerance to Cd (Liang et al., 2016). For example, *Brassica chinensis* L. (pakchoi) plants treated with sulfur showed an increased tolerance to Cd by achieving an increased overall growth compared to untreated plants (Liang et al., 2016). The mechanism of tolerance is linked to a reduction of Cd translocation from roots to shoots, which is accomplished through an increased production of non-protein, thiol containing compounds such as phytochelatins and GSH, that can sequester Cd to the plant's storage vacuoles (Liang et al., 2016). The addition of sulfur also stabilizes the Asa-GSH cycle, which is crucial for maintaining the balance of reactive oxygen species (ROS) that can negatively impact photosynthesis when overproduced (Liang et al., 2016).

The effect that Cd has on plants such as *Triticun aestivum*, has been linked to the use of different nitrogen supplies (Yotsova et al., 2020). While 10 mM NO_3^- was found to be optimal for the photosynthetic performance within wheat, increasing the nitrogen supply to 20 mM NO_3^- reduced the oxidative stress and growth inhibition typically experienced as a result of exposure to Cd (Yotsova et al., 2020). By increasing the nitrogen availability to 20 mM NO_3^-, the concentration of Cd in the roots and shoots of the wheat cultivates also

increased. Exposure to Cd causes a reduction of chlorophyll and carotenoids. This reduction hinders photosynthesis however, the presence of elevated nitrogen levels mitigated this impact (Yotsova et al., 2020).

2.4 Sulfur and Nitrogen in algae

Sulfur is essential in metabolic and catalytic activities within living cells. One such example is in photosynthesis where sulfur is essential in light and carbon reactions (Norici et al., 2005). The average sulfur assimilation by marine phytoplankton over the course of a year is 1320 pg (Giordano et al., 2005). Due to the high concentrations of sulfur that is found in oceans, phytoplankton live in an environment with supraoptimal sulfur concentrations (Giordano, 1994). This excess sulfur may require changes to ensure that metabolic pathways can be maintained. Therefore, algae must have a method of modulating their sulfur absorption and metabolism (Giordano et al., 2005).

Pollution of freshwater ecosystems can have an impact on the sulfur availability since sulfate has to compete with other anions such as selenate and molybdate for common transporters (Ramaiah & Shanmugasundaram, 1962). While this is a more common issue for terrestrial ecosystems, if the area is heavily polluted, then competition for common transporters can lower sulfur availability that is available for metabolic processes in algae (Wheeler et al., 1982) In some instances, heavy metal contamination can mimic the characteristics of sulfur deficiency. One example of this is pollution with Cd for which, even in an environment of sufficient sulfate concentrations, the Cd can inhibit photosynthesis (Adhikari et al., 2018; Ramaiah & Shanmugasundaram, 1962). In maize, genes responded to both sulfur limitation and Cd exposure, indicating that the stress

response to Cd could be like that of sulfur deprivation (Adhikari et al., 2018). In cases of sulfur deficiency, cell division can stop, and algae cells can begin to accumulate large amounts of starch and experience a decreased photosynthetic rate (Kolber et al., 1988). When sulfur-deprived *Chlamydomonous reinhardtii* was examined, there was a 10-fold increase in cellular starch after 24 hrs (Zhang et al., 2001). Sulfur deprivation can also be responsible for adjustments to the photosynthetic capacity of green algae, which is caused by the downregulation of photosynthetic reactions in the stroma (Melis et al., 2000). This may indicate that sulfate metabolites are responsible for the tolerance of microalgae to heavy metals by protecting photosynthesis after Cd exposure. Nitrogen limitation has been used as a strategy to enhance algal lipid production, where nitrogen repletion results in algal biomass that is rich in proteins (Tossavainen et al., 2019). Studies on *Euglena gracilis* fatty acid, protein, chlorophyll a, and carotenoid concentrations indicate that, under nitrogen limitation, the total lipid content increased in ageing cultures. However, when nitrogen limitation was prolonged, there was a higher proportion of saturated fatty acids (Tossavainen et. al., 2019). With a higher nitrogen concentration, *Euglena* showed enhanced growth, protein, carotenoid, chlorophyll a production, and remained stable during stationary growth (Tossavainen et. al., 2019). This increase contrasts with what is seen after exposure to Cd, where chlorophyll and carotenoid are in a deficit, indicating the potential benefit of nitrogen as a defense to Cd exposure in algae (Yotsova et al., 2020).

In environments where there is excess loading of nitrogen, the trophic structure of lakes can shift from an macrophyte-dominated state to an algal dominated turbid state (Olsen et. al., 2015). With macrophytes such as *Potamogeton lucens* and *Cabomba caroliniana* their overall abundance declined in the low nitrogen and high nitrogen

treatments while completely disappearing by the following summer in the high nitrogen treatment (Olsen et al., 2015). The overall abundance of the microalga increased during the winter and then decreased during the summer with no obvious impact of the nitrogen treatments.

2.5 *Euglena gracilis*

Euglena gracilis is a freshwater, single-celled eukaryote that can be cultured with ease (Rodríguez-Zavala et al., 2007). Cells can grow up to 100 μm in length and possess a photoreceptor for light perception (O'Neill et. al, 2015). *E. gracilis* is part of the group of flagellates known as Euglenozoa (Bicudo et al., 2016), and this monophyletic group contains Euglenida, Kinetoplastidea and Diplonemea and Symbiontida.

E. gracilis possesses several unique and interesting characteristics. It can grow under photoautotrophic, heterotrophic and mixotrophic conditions (Rodriguez-Couto et al., 2019). Using Cryo-EM, the structure of the cytosolic ribosome was solved and found to possess a large ribosomal subunit that contains 14 discrete rRNA fragments (Matzov et al., 2020). These fragments are assembled into the canonical ribosome structure non-covalently and are stabilized through a network of post-transcriptional modifications that spread beyond the catalytic core (Matzov et. al., 2015). *E. gracilis*, unlike other organisms, can survive in a variety of harsh environments including low pH (pH 2.5-7), which is typical of mine runoff locations where *E. gracilis* is typically found (Winters et al., 2017).

E. gracilis has the capacity to remove heavy metals such as Hg from its growth media (Devars et al., 2000). When exposed to HgCl in culture, the concentration of HgCl in the growth media decreased over time while the concentration within the cell increased (Devars et al., 2000). *E. gracilis* also tolerates and absorbs Cu, Ni, and Cd at better than most plants (Avilés et al., 2003; Winters et al., 2017). Interestingly, Euglena's tolerance for heavy metals can be enhanced by using specific pretreatments. For example, a pretreatment of HgCl prior to exposing the cultures to Cd increased the organism's overall tolerance to Cd and increased the amount of Cd it could remove from the growth media (Avilés et al., 2003). This is indicative that, like some plants, the environment of *E. gracilis* can be altered to increase metal tolerance.

2.6 Sulfur & Nitrogen in *Euglena gracilis*

E. gracilis can metabolize sulfur in the form of sulfate. This is in part due to the sulfate activating system that is present on the outside of the inner mitochondrial membrane (Saidha et al., 1988). The mitochondria of *E. gracilis* has necessary enzymes for sulphate activation and reduction. The mitochondria also provide the necessary ATP for sulfate activation through coupled phosphorylation (Saidha et al., 1988). While the mitochondria contain the necessary enzymes for sulfate activation and reduction, the chloroplast is devoid of this function. However, they do contain a thylakoid sulpholipid indicating that *E. gracilis* mitochondria must convert the sulphates into intermediates that can be used by the chloroplast for sulpholipid formation (Saidha et. al., 1988).

E. gracilis has been grown in sulfate sufficient, sulfate deficient and in cultures with Cd exposure or cysteine overload, to examine its impact on sulfate uptake (García-García et al., 2012). Kinetic analysis done at the stationary growth phase identified three saturable components. The Km value suggested the presence of one high affinity (*Eg*PMST1) and two low affinity (*Eg*PMST2 and *Eg*PMST3) sulfate transporters (García-García et. al., 2012). This indicated that the sulfate transporters had increased activity in sulfate deficient environments and in environments with Cd exposure, (García-García et. al., 2012). This could indicate that the response of *E. gracilis* is similar in sulfur deficient environments and environments with heavy metals.

E. gracilis is unable to use urea, nitrate or nitrite as a suitable nitrogen source. As such, ammonium is the supplied nitrogen source in cultures (Richter et al., 2015). While *E. gracilis* is unable to use nitrate and nitrites, past work has shown that *E. gracilis* has grown in environments for which amino acids were the sole nitrogen source. The growth kinetics of *E. gracilis were studied* in the presence of amino acids such as glycine, glutamine, glutamic acid, leucine, and threonine; as along with the uptake of these amino acids into the cells (Richter et. al., 2015). The cell growth of cultures grown with glycine and glutamine was comparable to that of the cells grown with ammonium as the nitrogen source. While absorption of the amino acids was generally comparable to that of ammonium, the consumption of glutamate was poor (Richter et. al., 2015). Overall, while ammonium is the current nitrogen source used to culture *Euglena gracilis* in the lab, amino acids can act as an alternative nitrogen source. This would allow for culturing *E. gracilis* without the decomposition of ammonium to ammonia (Richter et. al., 2015).

2.7 Metal binding compounds

It is known that certain plants, fungi, and algae possess the ability to tolerate and remove heavy metals from their environment. Some fungi, including *Aspergillus niger*, have been able to sequester metal ions such as Cd, Cu (II), Zn, Ni (II), and Co from concentrated solutions (Akthar & Mohan, 1995). With the fungus *Flagellospora curta*, high levels of thiols were found after exposure to copper, Cd, or Zn (Guimarães-Soares et al., 2006). The high number of thiols allowed for *F. curta* to bind more copper and Cd when compared to *Fontanospora fusiramosa* (Guimarães-Soaresa et al., 2006). With other types of organisms such as *E. gracilis*, it was found that when cells are exposed to Cd, approximately 60% of Cd is found in the chloroplast (Mendoza-Cózatl & Moreno-Sánchez, 2005). When studying the absorption kinetics in the *E. gracilis* chloroplast the initial uptake rate was resolved in two different components with one being hyperbolic and saturable and the other linear and non-saturable. The uptake of Cd was unaffected by either metabolic inhibitors or illumination, suggesting that a cation-diffusion facilitator protein was involved (Mendoza-Cózatl & Moreno-Sánchez, 2005). A correlation was found between Cd retention in the chloroplast and the amount of internal sulfur compounds. Chloroplasts that were isolated from cells grown in 50 μM $CdCl_2$ contained 4.4 X more thiol-compounds and sulfide when compared with control chloroplasts, and they retained 6 times more Cd (Mendoza-Cózatl & Moreno-Sánchez, 2005).

In 2018, metabolomic analysis into *E. gracilis* identified specific metal binding compounds (Mangal et al., 2022). There were six Hg binding compounds found in *E. gracilis* that were patented along with the methods used to remove them from *E. gracilis*.

Five of these compounds contain nitrogen. Most of the metal binding metabolites that were identified in *E. gracilis* have been classified as dissolved polyphenols or lignin monomoers (Mangal et al., 2022). Some of the identified monomers were modified when sulfur or nitrogen groups, or a combination of the two was added.

2.8 Genes related to metal tolerance

Compounds such as thiols and phytochelatins have been extensively studied in relation to metal binding in various organisms. Along with these compounds, there are various proteins that have been identified as playing a role in an organisms' ability to tolerate heavy metals such as Cd. These include chemotaxis and heat shock proteins.

Chemotaxis is the mechanism that allows bacteria to rapidly respond to changes in the chemical composition of their environment. This allows for cells to move from unfavorable conditions to more favorable ones (Bren & Eisenbach, 2000). Bacteria with a tolerance to metal had a downregulation of chemotaxis proteins. In the specific case of Cd, when *Caulobacter crescentus* was exposed to $CdSO_4$, proteomics showed that there was a downregulation in chemotaxis proteins. The downregulation of chemotaxis proteins was also seen in other metal tolerant bacteria such as *Acidithiobacillus caldus* when the metal tolerant bacteria was exposed to higher concentrations of Cu^{2+} (Feng et al., 2020). In bacteria, the change in expression of chemotaxis proteins is typically accompanied with changes in expression of flagellar motor switch proteins (Yung et al., 2014). Flagellar motor switch proteins make up what is known as the "switch complex". This complex is responsible for the direction of flagellar rotation, impacting the overall motility of the cell (Paul et al., 2011). In bacteria, chemotaxis regulators-transmit various chemoreceptor

signals to the flagellar motor CheY component which starts the cascade causing the change in rotational sense which is required for chemotaxis (Sarkar et al., 2010). Organisms, such as bacteria, lose their motility to ensure their survival when facing environmental stressors. (Sharma et al., 2019). When *Klebsiella pneumoniae* was exposed to meropenem stress, proteomics revealed that there was a downregulation of proteins related to flagellar formation and subsequent function including flagellar motor switch protein FliG.

Within plants, heat shock proteins protect photosynthesis during stresses such as exposure to heavy metals. When plants were exposed to heavy metals, the content of smHSPs increases in the chloroplast, allowing the plant to continue to carry out photosynthesis (Heckathorn et al., 2004). Within *E. gracilis* heat shock proteins are known to be upregulated when the protist is exposed to Cd. These findings are supported through proteomics where HSPs were found to be upregulated in plants that were stressed with heavy metals, and they were also produced by plants that had been exposed to Cd (Barque et al., 1996).

2.9 *Euglena gracilis* genome, transcriptome, and proteome analysis

The *E. gracilis* nuclear genome was estimated to be upwards of 500 megabases in size Ebenezer et al., 2019). *E. gracilis* is known to support multiple splicing pathways which include both conventional and non-conventional *cis*- and *trans*-splicing (Tessier et al., 1991). To better understand the biology of *E. gracilis*, a combination of genomics, transcriptomics and proteomics research is being conducted (Ebenezer et. al., 2019).

Using Roche 454 sequencing technology, the haploid genome size of Euglena was estimated to be over 450 Mb in size (Ebenezer et al., 2019). Sequencing the genome using

Illumina paired-end sequencing and supplementing this with PacBio sequencing data did not improve the quality of the assembly significantly. The authors indicated that the poor results with PacBio could be related to low coverage (Ebenezer et. al., 2019). The final assembly had an extremely low N50 value of 955. The latest attempt to assemble a genome was hampered due to the heterozygosity, size, and the frequency of sequences with low complexity within the dataset used. (Ebenezer et. al., 2019).

Along with attempts to sequence and annotate the *E. gracilis* genome, transcriptome assemblies have been completed as early as 2015. In light grown cells, transcript analysis identified 22,814 predicted protein encoding genes, while 26,738 were identified in the dark grown cells (O'Neill et al., 2015). Other efforts were able to assemble 22 Mb of coding sequences, with 26,479 transfrags that had an FPKM value greater than one, indicating that it was potentially expressed. Out of those sequences, about 40 % had an assignable function based on their sequence similarity to SWISS-PROT (Yoshida et. al. 2016). In another study, sequencing and assembling an *E. gracilis* transcriptome using in-house and publicly available data found over 36,000 unique coding sequences. Approximately 50 % of all the transcripts were annotated using UniRef, while over 12,000 were associated with a GO term (Ebenezer et al., 2019). BUSCO was used to assess the quality of the newly generated transcriptome, and to attempt to improve the quality by combining sequences generated from three separate studies (Ebenezer et. al., 2019). Individually, the number of complete BUSCOs was similar amongst the three studies, with the assembly done by O'Neill et al. showing 240 complete BUSCOs compared to that of the 262 and 255 complete BUSCOs from the transcriptome assembly performed by Yoshida et. al., and Ebenezer et. al., respectively (O'Neil et. al., 2015, Yoshida et. al., 2016,

Ebenezer et. al., 2019). When combining the three datasets to form one single assembly, the number of complete BUSCOs increased to 274 (Ebenezer et. al., 2019). More recent transcriptome assemblies have been done separate transcriptome was assembled using trasnfrags from *E gracilis* cultures grown in the presence and absence of mercury. The resulting transcriptome had 99% (299 complete BUSCOs) orthologs from the *eukaryote* OrthoDB v9 (Mangal et.2022).

Gene ontology (GO)-term analysis was conducted on *E. gracilis* after exposure to Cd and identified enrichment of GO-terms related to response to stress, metabolic process, iron sulfur clustering, metal ion binding, transmembrane transport, and catalytic activity (Khatiwada et al., 2020). Specifically, there was an increase in thiol-containing proteins such as glutaredoxin. Despite the increase in thiol-containing compounds, exposure of *E. gracilis* to heavy metals alone did not bring a significant increase in the abundance of sulfur transporters, even if the activity of high and low affinity sulfate transporters is increased both during sulfate starvation as well as exposure to Cd (Khatiwada et al., 2020). Along with thiol-containing compounds, a major facilitator superfamily (MFS) transporter was increased after exposure to Cd and has the known function of transporting metals in and out of cells. An identified heavy metal pump, P1 B ATPase showed increased expression when *E. gracilis* was exposed to Cd and is known for maintain metal homeostasis by extruding transition metals from the cytoplasmic area and translocating metal ions across the cellular membrane (Khatiwada et al., 2020). The transporter that had the highest abundance in response to Cd was identified as TrkA transporter. This transmembrane transporter is responsible for cation transportation such as potassium transport and sodium/sulfate symport. Specific stress responses were identified including glutathione

transferase which aids in detoxification and protects cells against ROS death and cysteine desulferase which, in turn, enables iron-sulfur clustering through by catalyzing the conversion of L-cysteine to L-alanine (Khatiwada et al., 2020). Most recently, in 2022, RNA-Sequencing was used to assess transcript level changes in *E. gracilis* cells expose to Hg. Subsequent GO-term analysis identified enriched GO-terms related to amide biosynthetic and metabolic processes, peptide biosynthetic and metabolic process, photosynthesis (light harvesting, light harvesting IN Photosystem, light reaction), translation, and purine ribonucleotide triphosphate biosynthetic and metabolic process (Mangal et. al., 2022).

2.10 Research objectives

In the past, methods of removing heavy metal contaminants from industrial wastewater have predominantly focused on filtration, chemical precipitation, resin-based ion exchange and coagulation of activated carbons (Fu & Wang, 2011). These methods are expensive, relatively ineffective and have the potential of introducing new chemicals into the already polluted waters (Wang & Cheng, 2009, Volesky, 2001). Due to these factors, the need for more economically and environmentally feasible options of wastewater cleanup are needed. This makes research into methods of bioremediation crucial to address industrial wastewater pollution.

One such organism that is being studied for its potential role in bioremediation is *E. gracilis*. It is currently known that *E. gracilis* can absorb high concentrations of toxic metals including Hg, Cu, Ni, and Cd (Devars et. al., 2000, Winters *et. al*, 2017). Along

with being able to absorb these metals, the cells are shown to have some tolerance to metals such as Cd which can also be enhance through pretreating the cells (Aviles *et. al*, 2003). Recently it was found that *Euglena gracilis* cells can produce specific metal binding, sulfur and nitrogen enriched compounds that allow the uptake of metals such as Hg and Cd (Winters et al., 2021, Mangal et al., 2022).

Based on the work by Winters et. al., 2019 and Mangal et al., 2022, I hypothesized that a pretreatment of *E. gracilis* with either sulfur or nitrogen would enhance the organism's tolerance to $CdCl_2$. I also hypothesized that there will be genes with altered transcript levels as a result of either the pretreatments or the pretreatments and exposure to $CdCl_2$. There are two objectives to my research. Objective 1: Assess the tolerance of *E. gracilis* to $CdCl_2$ after pretreating the cells separately with either sulfur or nitrogen. Objective 2: identify gene functions that are associated with altered transcript levels in response to both the pretreatment with either sulfur or nitrogen, and the exposure to $CdCl_2$ To accomplish Objective 1, I pretreated axenic *E. gracilis* cultures with either 60 mM SO_4 or 75 mM of nitrogen in the form of NH_4^+ and NO_3^-. I then exposed both pretreated cultures and cells grown in MAM to various concentrations of $CdCl_2$ and counted the living cells remaining in the culture, using Trypan Blue to distinguish between living and dead cells. To complete Objective 2, I used RNA-sequencing, differential gene expression analysis, and GO-term analysis to identify gene functions associated with altered transcript levels. The insight gained from this research will contribute to the understanding of methods to enhance heavy metal tolerance of *E. gracilis* and potential genes that are differentially expressed as a result of both the pretreatments and exposure to $CdCl_2$.

CHAPTER THREE: METHODS

3.1 *Euglena gracilis* cultivation

Field samples of *E. gracilis gracilis* were collected by the Canadian Phycological Culture Centre (CPCC) located at the University of Waterloo. The axenic strain of *E. gracilis* (CPCC95) was grown photoautotrophically at 22 ^{o}C while shaking at 80 rpm with a 12hr:12hr light/dark photoperiod and a light intensity of 2,691 µMol/m^2/h in a Conviron CMP5090 environmental chamber. Cultures were grown in Modified Acid Medium (MAM) which included: 0.5g/L of $(NH_4)_2SO_4$, and $MgSO_4*7H_2O$, 0.3g/L of KH_2PO_4, 0.01 g/L of $CaCl_2*_2H_2O$, 0.03 g/L of NaCl, 0.01 g/L of $Na_2EDTA*2H_2O$, 0.00498 g/L of $FeSO_4*7H_2O$ with 0.00000843 g/L of H_2SO_4. The media also included a Trace Metal solution containing a final concentration of: 2.86 µg/L of H_3BO_3, 1.81 µg/L of $MnCl_2*4H_2O$, 0.22 µg/L of $ZnSO_4*7H_2O$, 0.39 µg/L of $Na_2MoO_4*2H_2O$, 0.079 µg/L of $CuSO_4*5H_2O$ and 0.0494 µg/L of $Co(NO_3)_2*6H_2O$ prepared to a pH of 4.3. After the media was autoclaved, a vitamin stock was added which included a final concentration of 10 µg/L of Vitamin B12, 0.001 µg/L of Biotin and 0.002 g/L of Thiamine HCl.

3.2 Sulfur and Nitrogen Pretreatment of CPCC 95 *Euglena gracilis*

The concentrations of both sulfur and nitrogen pretreatments were determined by increasing the total concentration of SO_4 and NH_4NO_3 within the media so that the total concentration of SO_4 in MAM was 60 mM for the sulfur pretreatment and the total ion concentration for NH_4 and NO_3 was 76 mM for the nitrogen pretreatment. A stock culture

of CPCC 95 *E. gracilis* was used to make three technical replicates of a control group grown in MAM, a sulfur enriched group grown in MAM + 13.85 g/L of $MgSO_4*7H_2O$, and a nitrogen enriched group grown in MAM + 5.85 g/L of NH_4NO_3. Pretreatments were carried out for 44 days. During this time, the cells were counted with a hemocytometer to track growth and the entire volume of the culture was centrifuged at 3783 x *g* for 10 mins and the cells were resuspended in fresh media every 96 hrs.

3.3 Cadmium Exposure of Pretreated and Non-Pretreated CPCC 95 *Euglena gracilis* cultures

Pretreated and non-pretreated cultures were inoculated in Modified Acid Media (MAM) and MAM with 25 µM Cd chloride ($CdCl_2$) to a final concentration of 200,000 cells/mL. Cd was added to MAM from a sterile 100 mM stock solution prior to inoculation. Cell count and viability was recorded every 96 hrs for a period 8 days using Trypan blue staining. Briefly, 100 mL of culture was stained with 0.32% Trypan Blue and a 10 µL aliquot was place on a hemocytometer (Hausser Scientific) and the cells were observed at 10 X magnification using a Zeiss AXIO microscope. Living and dead cells were visualized and counted in each of the eight quadrants with each field representing 0.1 µL of cells. Counts were repeated five times with a fresh 10 µL of culture being placed on the hemocytometer each time. A t-test was preformed to determine statistical differences in the amount of living cells in the nitrogen and sulfur pretreated *Euglena gracilis* cultures exposed to Cd compared to non-pretreated control cultures exposed to Cd. After counting, the remaining cultures were washed with sterile dH_2O and used to inoculated fresh MAM

or MAM+ $CdCl_2$ to a concentration of 200,000 cells/mL in a 40 mL cultures. After the final day of growth, total culture volume was centrifuged at 5,250 x *g* for six mins and the supernatant was removed. Cell pellets were frozen using LN_2 and stored at -80 ^oC until RNA isolation.

3.4 RNA Isolation and DNase 1 Treatment

Cells from CPCC 95 *E. gra*cilis were removed from -80 ^oC and resuspended in 2 mL of TRIzol. Samples were incubated at RT for five mins before being transferred to MP Biomedical Lysing Matrix C tubes. Cells were lysed using a MP Biomedical Fast Prep-24, and subsequently shaken at 6.5 m/s for 45 s. Cells were incubated on ice for one minute before being shaken again at 6.5 m/s for 45 s and incubated on ice for one minute. Samples were incubated at RT for five minutes before centrifugation at 12,200 x *g* for 10 mins at 4 ^oC. Supernatant was transferred to a new 1.5 mL microtube tube and centrifuged again at 4^oC at 12,200 x *g* for 10 mins. To prepare for the phase separation step, Invitrogen Phasemaker tubes were centrifuged at 16,000 x *g* for 30 s, followed by addition of supernatant. 200 µL of chloroform was added to each tube, and the sample was gently inverted 20 X. Samples were incubated at RT for three mins before being centrifuged at 12,200 x *g* for 15 mins at 4 ^oC. The supernatant from each sample was transferred to a new 1.5 mL microcentrifuge tube and 200 µL of chloroform was added, and the tubes were inverted 20 X and the samples were centrifuged at 12,200 x *g* for 10 mins at 4 ^oC. The supernatant from each sample was transferred to a final microcentrifuge tube and the tube was labeled for storage. The RNA was precipitated using 250 µL of 0.8M disodium

citrate/1.2 M NaCl and 250 µL isopropanol. The samples were mixed by inversion 20 times and incubated at RT for 10 mins. The samples were incubated at -20 ºC for 20 mins and centrifuged for 10 mins at 4 ºC and 12,200 x *g*. The supernatant was removed, leaving the RNA pellet which was then washed with 1 mL of 75 % ethanol (EtOH) and centrifuged for five mins at 7,500 x *g* and 4 ºC. EtOH was removed, and the tubes containing the RNA pellet were left open at RT until samples were dry. The pellet was redissolved in 60 µL of diethyl-pyrocarbonate treated (DEPC-treated) water and incubated at RT before being incubated at 65 ºC for 10 mins. After the final incubation, 2 µL of each sample was loaded onto a Thermo Scientific NanoDrop 8000 to measure RNA concentration.

To remove DNA from the samples, aliquots containing 15 µg of RNA were diluted with DEPC-treated water to a final volume of 120 µL. Each sample was treated with 15 µL of DNase 1 enzyme and 15 µL DNase Buffer and incubated at 37 ºC for 30 mins. Once incubated, 3 µL of 250 mM EDTA was added to each sample and DNase 1 was inactivated at 75 ºC for 10 mins. RNA was precipitated using 37.5 µL of 0.8 M disodium citrate/1.2M NaCl and 75 µL of isopropanol. The samples were inverted 20 times and incubated at RT for 10 mins before incubation at -20 ºC for 20 mins. The samples were centrifuged at 12,200 x *g* for 15 mins at 4 ºC. The supernatant was removed, and the pellet was centrifuged for 10 s at maximum speed using the short function on an Eppendorf Centrifuge 5430 and the remaining supernatant was removed. Pellets were washed with 1mL of 75% EtOH and centrifuged at 7,500 x *g* at 4ºC for five mins. The EtOH was removed, and the wash step was repeated. After the final wash, EtOH was removed, and the pellet was left to air dry until the pellet appeared dry. The pellet was resuspended in 50 µL of Ambion Nuclease-Free Water and incubated at RT for five mins before being incubated at 65 ºC for

five mins and storing on ice. The RNA concentration for each sample was determined using a Nanodrop 8000 (Thermo Scientific, Massachusetts US) and samples were stored at -80 ^{o}C for future use. Each sample was screened for gDNA contamination by PCR amplification of *gapdh*. Each reaction was carried out in a 25 µL final volume containing 1.5 µL of DNase treated RNA and 23.5 µL of master mix which contained filter sterilized water, 1x Dreamtaq Buffer, 0.625 U Dreamtaq DNA Polymerase and 200 nM of forward and reverse *gapdh* primers *E. gracilis gapdh*-F (5' GGTCTGATGACCACCATCCAT3') and *E. gracilis gapdh*-R, (5' CGACGACACGGTTGGAGTAT 3'). The *gapdh* PCR was run in a GeneAmp PCR System 9700 from Applied Biosystems under the following conditions: 95 ^{o}C for 10 mins (95 ^{o}C 30 s, 65 ^{o}C 30 s, 72 ^{o}C 1 min) x 40 cycles, 72 ^{o}C for 10 mins followed by 4 ^{o}C hold. 5 µL of 6x loading dye was added to the 25 µL PCR product and 5 µL of the PCR product was loaded onto a 1.2% agarose gel and ran for 60 mins at 90V.

3.5 Quality Assessment of DNase-Treated RNA

The quality of each sample was assessed by visualizing RNA following electrophoretic separation on a 1.5% BPTE agarose gel following glyoxal denaturation. DNase-treated RNA and a ssRNA ladder (New England BioLabs, Whitby) were incubated at 65 ^{o}C for five mins and 3 µL were combined with 12 µL of a stock reaction mixture of glyoxal prepared using the protocol outlined by (Sambrook & Russell, 2001). The total 15 µL of each sample was loaded onto the 1.5% agarose gel and the gel was run at 100 V for 60

mins, then dropped to 80 V and run for another 15 mins to allow for the smaller fragments to separate and migrate further down the gel without running off.

3.6 cDNA library preparation and RNA-Sequencing

22 µL aliquots of selected DNase-treated RNA samples were sent to the Centre for Applied Genomics at the Hospital for Sick Children (Toronto, Canada) where they underwent further quality assessment, cDNA library preparation and RNA-sequencing. The quality of each RNA sample was assessed using a BioAnalyzer (Agilent Technologies). The results from the BioAalyzer were assessed and the 6 highest quality biological replicates from pretreated and non-pretreated *E. gracilis* cultures grown in 0 µM and 25 µM of $CdCl_2$ were selected for sequencing. The quality of the biological replicates was determined based on the RNA Integrity Number (RIN) and the DV200% with an emphasis being placed on the DV200%. Selected RNA samples underwent Poly(A) enrichment using oligo dT-beads and the subsequent cDNA libraries were prepared using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England BioLabs). Once prepped, the 36 barcoded libraries were pooled and sequenced on two lanes of the NovaSeq (Illumina Inc) system producing paired-end reads that were 150 bp long. The samples where the cultures were grown in the absence of a pretreatment were labeled as "Ctrl or C" or labeled "S" for sulfur pretreated cultures and "N" for nitrogen pretreated cultures; followed by either a letter to indicate that the sample was not challenged with $CdCl_2$ or 25 to indicate that the cells had been exposed to 25 µM of $CdCl_2$ (Table S1).

3.7 Sequence Quality Assessment, Trimming

Fastq files for the 36 sequenced libaries were downloaded using the TCAG data portal command line tool. Sequence quality was assessed using FastQC (v0.11.9) as outlined in the literature (Andrews et al., 2010). The generated html files were reviewed to assess Phred scores across all the generated reads for each file. Once the minimum quality was assessed, sequences were trimmed using Trimmomatic (v0.39) with the following specifications: Illuminaclip:TruSeq3-PE-2.fa:2:30:10 sliding window:4:5 leading:5 tailing:5 minlen:25 (Bolger et al., 2014).

3.8 *De novo* transcriptome assembly

I used the trimmed sequences in the formation of a single *de novo* transcriptome assembly of *E. gracilis* using Trinity (v2.12.0) set to 'strand-specific mode (RF)' (Grabherr et al., 2013). The completeness of the assembly was assessed using BUSCO (v5.2.2) and the 'eukaryota_odb9' dataset (Waterhouse et al., 2018) . Trimmed sequences were aligned to the assembled transcriptome using Bowtie2 (v2.4.2) with the 'SS_lib_type RF' feature (Langmead &Salzberg, 2012). Finally, RSEM (v1.3.3) was used to estimate the transfrag abundance for each library based on the settings used by Li & Dewey, 2011.

3.9 Differential gene expression and GO-term analysis

DeSeq2 (v1.30.0) was implemented using the SarTools (v1.7.4.0) program to assess the changes in transcript levels (Love et al., 2014, Varet et. al., 2016). Individual 1 on 1 comparisons were completed to assess genes that were differentially expressed based on the pretreatment of the cultures, and their exposure to $CdCl_2$ (Table S1). The analysis was also used to assess the changes in transcript levels of the pretreated *Euglena gracilis* cultures compared to non-pretreated *Euglena gracilis* cells. The statistical model "treatment" was used for the comparison with a false discovery rate of (FDR) <0.05. DeSeq2 was run with the settings: cooksCutoff = True, independentFiltering = True, alpha = 0.05 to set the threshold for significance, pAdjustMethod = BH (Benjamini/Hochberg p-value adjustment method), typeTrans = rlog, and locfunc = median to estimate size factors. The genes that were identified to have transcript level changes according to DESeq2 were characterized using NCBI blast 2.11.0+ blastx to search the NCBI protein databases as well as SWISS-PROT databases (accessed January 2022) using an E-value of 1E-5 to identify genes encoding proteins with similar sequences. The *Arabidopsis thaliana, Chlamydomonas reinhardtii, Synechocystis sp., Homo sapiens and Trypanosoma brucei* databases on Ensembl 90 were searched using blastx to characterize genes that were further analyzed for gene ontology (GO) enrichment analysis (accessed February 2022). The genes listed were run in the PANTHER gene ontology database (http://pantherdb.org/index.jsp) on the default settings to test for statistical over-representation of the identified GO terms.

CHAPTER FOUR: RESULTS

Work done by Aviles et al., 2003 determined that a pretreatment of *E. gracilis* with HgCl enhanced its tolerance to Cd and Mangal et al., 2022 identified that specific metal binding compounds were produced by *E. gracilis* that were enriched with sulfur and nitrogen respectively. The objectives of this thesis included 1) microscopic observations of *E. gracilis* cultures grown in MAM + 60 mM total SO_4 using $MgSO_4*7H_2O$ (herein referred to as sulfur pretreated), MAM + 75 mM total nitrogen using NH_4NO_3 (herein referred to as nitrogen pretreated), and *E. gracilis* cells grown only in MAM (herein referred to as "non-pretreated") and then exposed to various concentrations of $CdCl_2$ to determine if there were changes in the tolerance of the respective cultures to Cd, and 2) RNA-Seq and DGE and GO term analyses to identify gene candidates with that were differentially expressed as a result of both the pretreatments and exposure to $CdCl_2$.

4.1 Exposure to increased concentrations of $MgSO_4*7H_2O$ and NH_4NO_3 demonstrated the ability of *Euglena gracilis* to survive changing environments

Cultures grown in MAM supplemented with $MgSO_4*7H_2O$ with a total SO_4 concentration of 30 mM, 60 mM or 120 mM (Figure 1 A-D) showed no visible morphological changes eight days post inoculation when compared cells grown in MAM without supplemental SO_4. In each condition, cells remained green and maintained their shape and overall motility. When cells were grown for 10 days in MAM with NH_4NO_3 with a final concentration of NO_3 of 8, 38, 76 or 151 mM there was a notable difference in cell morphology in the 76 and 151 mM NO_3 treatments (Figure 1-E-H). Growth in 76 mM and

151 mM NH_4NO_3 led to a change in the cell morphology after 10 days whereby the majority of cells became spherical. At 151 mM total nitrogen, cells also became brown in colour. This indicated that, unlike SO_4 treatments, cells became stressed with higher concentrations of NH_4NO_3. This is thought to be because *E. gracilis* is only able to metabolize nitrogen in the form of NH_4^+, which would prevent it from using the nitrate in the media (Richter et. al., 2015).

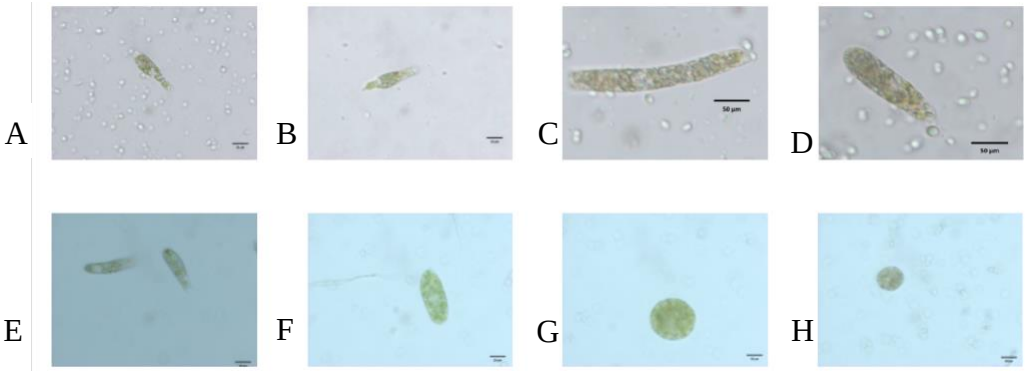


Figure 1: *Euglena gracilis* grown in MAM with A) 6 mM SO_4, B) 30 mM SO_4, C) 60 mM SO_4, D) 120 mM SO_4 after 8 days of growth and E) 8 mM NO_3 F) 38 mM NO_3 G) 76 mM NO_3, and H) 151 mM NO_3 after 10 days of growth. Photos were taken at 40x magnification. Scale Bar, 50 μ**m**.

For there to be a suitable number of cells to inoculate into $CdCl_2$, the target concentration of cells for transfer to $CdCl_2$ and subsequent RNA isolation was 300,000 cells/mL. *Euglena gracilis* attained this concentration after 28 days in non-pretreated conditions, 36 days in sulfur pretreated conditions, and 45 days in nitrogen pretreated conditions (Figure 2). While there was a trend showing slower growth in cultures pretreated with nitrogen compared to cultures that were not pretreated and cultures that were pretreated with sulfur, the difference was not consistently statistically significant (p

> 0.05) using a one-way ANOVA test. There were isolated days where the difference in cells/mL were statistically significant, but overall, the difference was not significant.

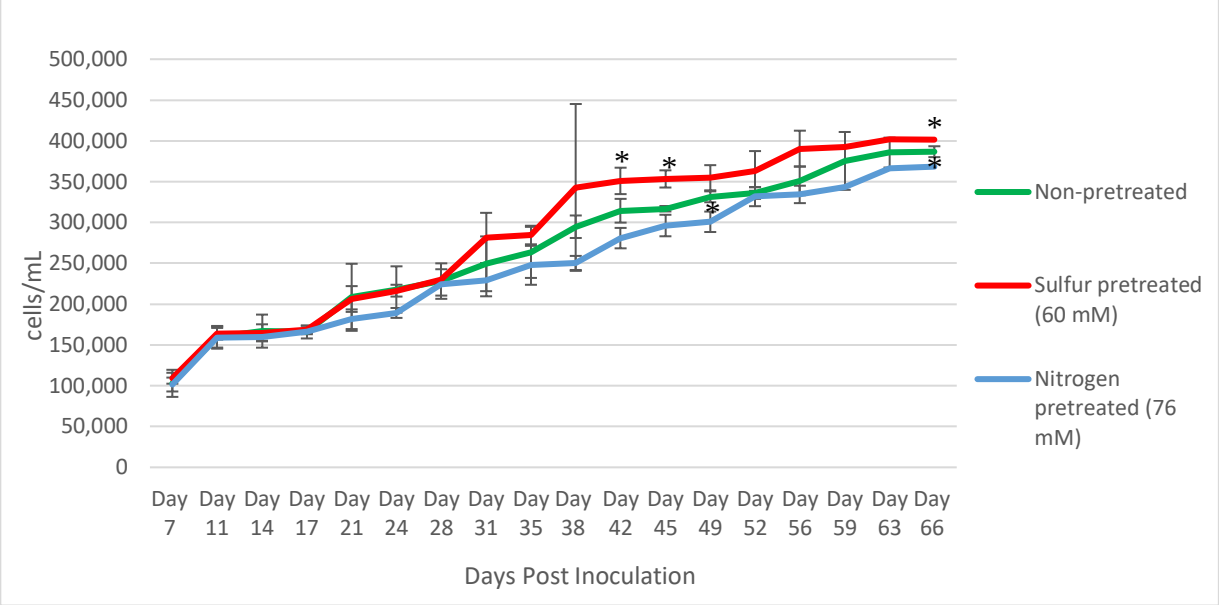


Figure 2: Growth of non-pretreated, sulfur pretreated and nitrogen pretreated *Euglena gracilis* cultures over the course of 66 days. Non-pretreated, sulfur pretreated and nitrogen pretreated *E. gracilis* cultures were grown at 22ºC while shaking at 80 rpm with a 12:12 light/dark photoperiod and a light intensity of 2691 µMol/m^2/h in a Conviron CMP5090 environmental chamber for 66 days. Error bars indicate the standard deviation between three biological replicates (n=3). * Indicates differences in cells/mL that are statistically significant using a one-way ANOVA ($p < 0.05$).

4.2 *Euglena gracilis* cultures pretreated with either excess sulfur or nitrogen had an increased tolerance to Cd

When exposed to increasing concentrations of $CdCl_2$, *E. gracilis* cultures that were grown without a pretreatment showed a statistically lower concentration of viable cells ($p < 0.05$)

when compared to cultures that were not exposed to $CdCl_2$ (Figure 3 A). Over the course of 20 days, exposure to $CdCl_2$ decreased the number of viable cells in the culture. Overall, cultures that were not pretreated did not show tolerance to $CdCl_2$ at concentrations higher than 10 µM. However, the sulfur and nitrogen pretreated cultures showed an increased tolerance ($p < 0.05$) in the number of viable cells at increasing concentrations of $CdCl_2$ over a prolonged period (Figure 3 B & C). Based on these results, it was clear that 25 µM $CdCl_2$ was a high enough concentration to cause a stress response in the *E. gracilis* cultures without depleting the number of viable cells in the non-pretreated cultures to a point where sufficient RNA could not be isolated.

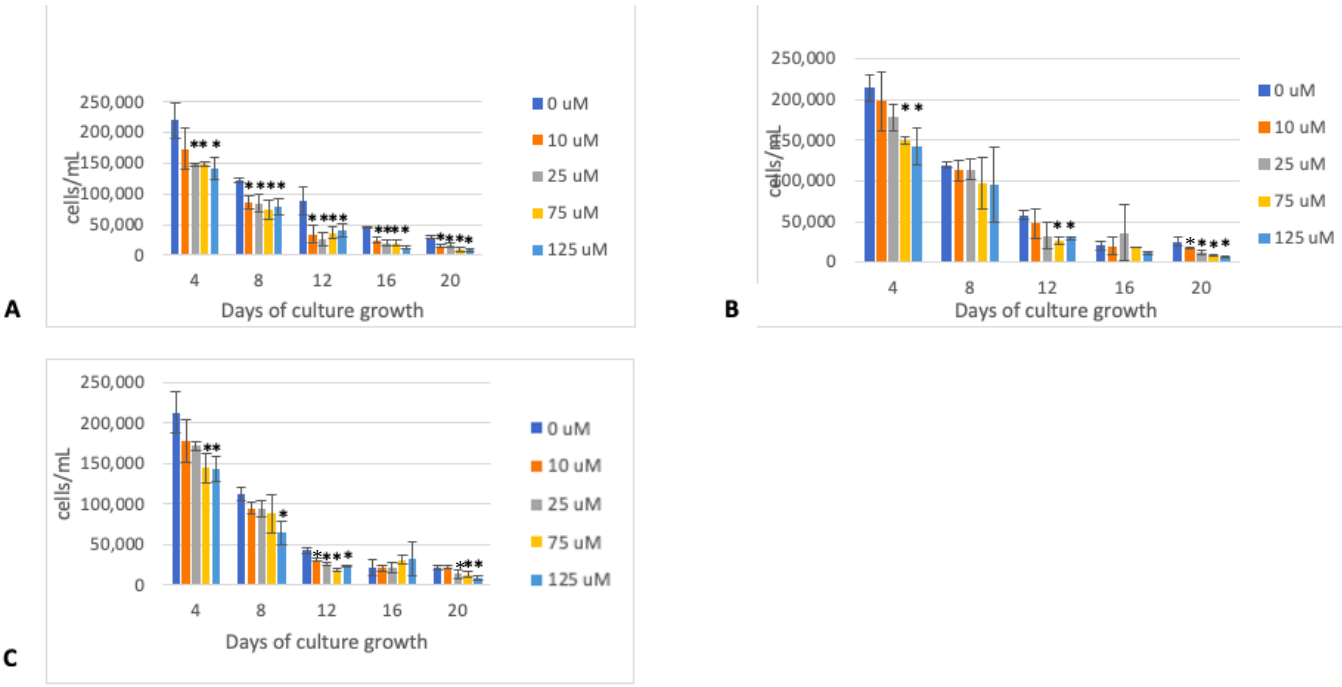


Figure 3: *Euglena gracilis* cells that were pretreated with either increased sulfur or nitrogen showed an increased tolerance to various concentrations of $CdCl_2$. Bars represent the concentration of viable cells following growth after transfer to the MAM containing the concentrations of $CdCl_2$ indicated in the legend are presented. A) Non-pretreated cultures B) Sulfur pretreated cultures. C) Nitrogen pretreated cultures. Error bars representing the standard deviation between biological replicates (n=3) with 5 assessments of concentrations per time point. * Indicates statistical difference between the sulfur and nitrogen pretreated cultures when compared to the non-pretreated cultures as determined by a one-way ANOVA test ($p < 0.05$).

At a concentration of 25 µM $CdCl_2$, the cultures for which the cells had been pretreated with either increased sulfur or nitrogen had a significantly higher number of cells

($p < 0.05$) when compared to the cultures that were pre-grown in MAM (Figure 4). Even after 8 days the average number of cells in the non-pretreated cultures was just below 250,000 cells/mL compared to the sulfur and nitrogen pretreated cultures, which both averaged above 300,000 cells/mL. This indicated that while *E. gracilis* had some baseline tolerance to 25 µM $CdCl_2$, the overall tolerance was improved in the cells that were pretreated with nutrients.

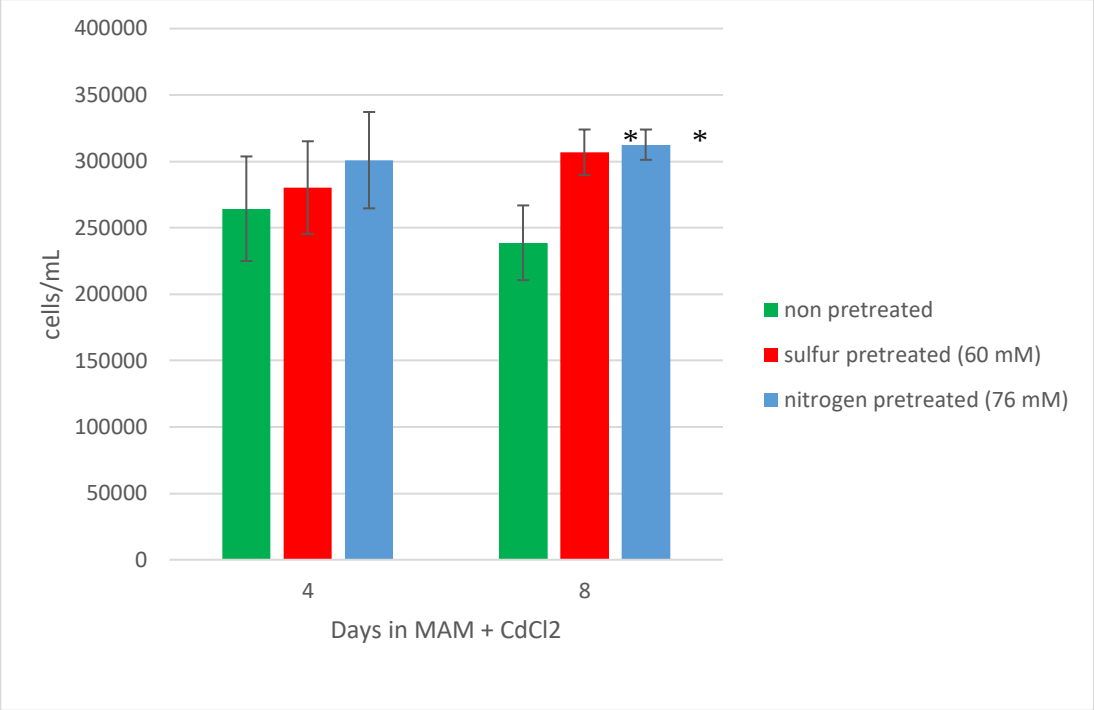


Figure 4: *Euglena gracilis* cultures pretreated with increased levels of sulfur or nitrogen showed an increased tolerance to 25 µM $CdCl_2$. The number of viable cells was measured 4 days and 8 days post inoculation. Bars depict the number of viable cells/mL. * Indicates statistical difference between the sulfur and nitrogen pretreated cultures when compared to the non-pretreated cultures as determined by a one-way ANOVA test ($p < 0.05$) across seven biological replicates. Error bars represent the standard deviation between the cells/mL of the seven biological replicates (n=7).

Trypan Blue (Sigma, Mississauga) was used to distinguish between living and dead cells, allowing viable cells to be counted (Figure 5). After 4 days, both pretreated and non-pretreated cultures showed viable cells in cultures without $CdCl_2$ (Table 1). However, in the presence of 25 µM $CdCl_2$, cultures pretreated with either sulfur or nitrogen had no dead cells while cultures that were not pretreated began to show signs of cell death. After 8 days, the cultures that were pretreated with sulfur prior to their exposure to $CdCl_2$ continued to show no indication of cell death with Trypan Blue, whereas the cells that were pretreated with increased nitrogen did show some cell death but not to the extent of the cells grown without a pretreatment (Table 1).

Table 1: Average number of living and dead cells in the non-pretreated, sulfur pretreated and nitrogen pretreated *Euglena gracilis* cultures after 4 days and 8 days in 25 μM $CdCl_2$.

	Day 4			Day 8		
Treatment	Living cells (cells/mL)	Dead cells (cells/mL)	% Of dead cells	Living cells (cells/mL)	Dead cells (cells/mL)	% Of dead cells
Non-pretreated	264,330	1,785	0.68 %	238,690	3,273	1.35 %
Sulfur pretreated	280,666	0	0 %	306,785	0	0 %
Nitrogen pretreated	300,892	0	0 %	312,559	892	0.28 %

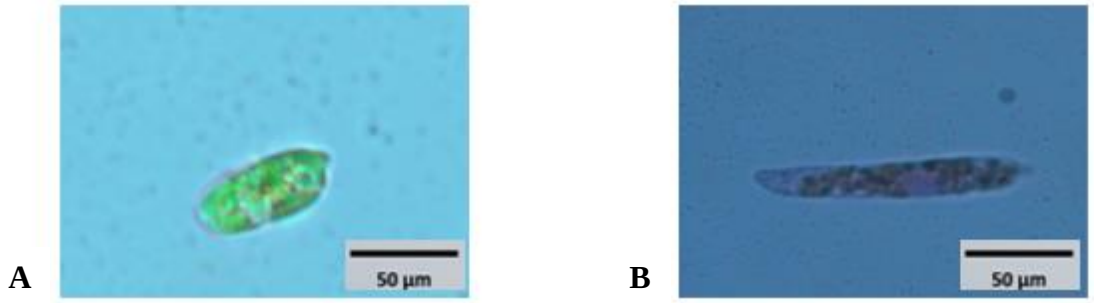


Figure 5: Trypan Blue staining revels protective effect of pretreatment after 8 days of $CdCl_2$ exposure. A) Living cells were stained with 0.32% Trypan Blue after 8 days of growth in 25 µM $CdCl_2$., B) Dead cells were stained with 0.32% Trypan Blue after 8 days of growth in 25 µM $CdCl_2$. Blue staining indicated the cells were not viable. Scale bar, 50 µm.

4.3 *Euglena gracilis* RNA-sequencing and *de novo* transcriptome assembly

RNA-Sequencing was carried out on six different sample types: non-pretreated cultures (n=6), non-pretreated cultures with 25 µM $CdCl_2$ (n=6), sulfur pretreated cultures (n=6), sulfur pretreated cultures exposed to 25 µM $CdCl_2$ (n=6), nitrogen pretreated cultures (n=6), and nitrogen pretreated cultures exposed to 25 µM $CdCl_2$. This resulted in an average of 23.3 million raw paired end reads per library (Table 2). After the adaptor sequences and poor-quality sequences were removed, there was an average of 19.5 million trimmed paired end reads per library.

Table 2: Average total number of raw and trimmed reads across 6 biological replicates of each RNA-Seq library.

Library	Number of Raw reads	Number of Trimmed Reads
Control	22,474,149	18,867,533
Control + 25 µM $CdCl_2$	22,938,670	19,041,189
SO_4	21,179,800	17,495,371
SO_4 + 25 µM $CdCl_2$	21,329,961	17,848,454
NH_4NO_3	26,431,615	22,355,220
NH_4NO_3 + 25 µM $CdCl_2$	25,430,061	21,699,217

For the *de novo* assembly, a combined total of 117.3 million trimmed paired-end reads were used. Once completed, the transfrag assembly represented 838,537 transcripts with an N50 value of 892 and these transcripts were from 459,533 *Euglena gracilis* genes. BUSCO analysis comparing *E. gracilis* transfrag assembly to the *eukaryota* Orthodb v10 orthologs showed 82.8% of orthologs within the database were found in the *E. gracilis de novo* assembly. An average of 23.2 million trimmed paired-end reads were aligned to the *de novo* transcriptome assembly across all libraries (Table 3).

Table 3: Average total number of reads mapped to de novo assembly using bowtie2 (v2.4.2).

Library	Number of Reads Mapped
Control	33,479,732
Control + 25 µM $CdCl_2$	33,516,640
SO_4	28,305,301
SO_4 + 25 µM $CdCl_2$	31,105,360
NH_4NO_3	39,336,238
NH_4NO_3 + 25 µM $CdCl_2$	38,067,499

4.4 Differential gene expression analysis comparing samples based on pretreatment and exposure to $CdCl_2$

For the differential gene expression (DGE) analysis, cluster dendrograms were created and principal component analysis was carried out (Figures S2 A-N). DESeq2 was used to find genes that were differentially expressed when comparing both pretreated and non-pretreated *E. gracilis* that were exposed to 0 µM and 25 µM $CdCl_2$ (Figures S3 A-G). The following sections present the results of the clustering analyses as well as the results of DGE analysis. Overall, DGE analysis of the non-pretreated, sulfur pretreated and nitrogen pretreated *E. gracilis* cultures identified 185, 136 and 311 unique transcripts respectively

(Figure 6). The results of the DGE analysis indicated that the response of *E. gracilis* to $CdCl_2$ is varied depending on the pretreatment used. While the transcripts that did not overlap are of high significance, the overlap of the 10 transcripts differentially expressed in all treatments and the 22 that specifically overlapped between the sulfur and nitrogen pretreated cultures could be a potential avenue of future work once more is known about the genome of *E. gracilis*.

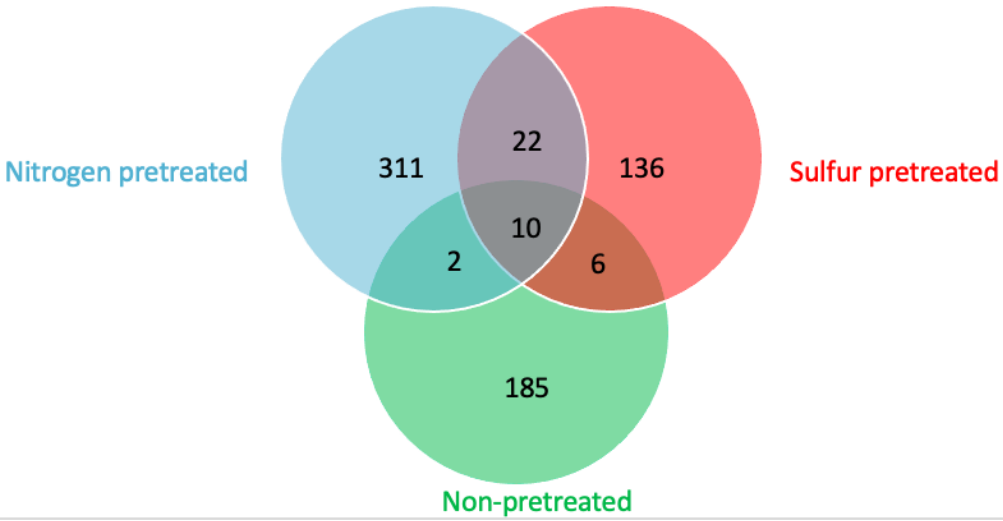


Figure 6: Venn diagram of Differentially expressed genes between A) non-pretreated *Euglena gracilis* cultures (green), nitrogen pretreated *Euglena gracilis* cultures (blue) and sulfur pretreated *Euglena gracilis* cultures (red).

4.4.1 Differential gene expression analysis on non-pretreated Euglena gracilis cultures after exposure to $CdCl_2$

When non-pretreated *E. gracilis* cells were compared between the presence and absence of $CdCl_2$, a total of 203 differentially expressed genes were identified with 115 being upregulated (Figure S3 A). The clustering analysis showed no separation of the samples (Figure S2 A-B). Blastx was used to characterize the genes identified by DESeq2 and 102/203 genes were found to have similarities to sequences with the non-redundant protein database and SWISS-PROT protein databases (Table S3-S4). Transcripts showed sequence similarity to *A. thaliana* (12), *H. sapiens* (12), *C. reinhardtii* (10), *Synechocystis sp.* (11) and *T. brucei* (8) (Table S5-S9).

The differentially expressed transcripts that were identified in both the NCBI non-redundant and SWISS-PROT protein databases included: chemotaxis proteins, ferredoxin proteins as well as flagellar motor switch proteins. The transcript identified as a chemotaxis protein was upregulated according to the DGE analysis while the ferredoxin family protein and flagellar motor switch protein were found to be downregulated in the non-pretreated *E. gracilis* cells exposed to $CdCl_2$.

The lists resulting from the BLAST searches to the *A thaliana*, *C. reinhardtii*, *Synechocystis*, *H. sapiens* and *T. brucei* databases, were used to search the PANTHER databases for GO term enrichment including biological process, cellular component, and molecular function. When looking at the GO-terms for the *E. gracilis* samples in the presence and absence of $CdCl_2$, GO term enrichment was seen for proteins related to the positive regulation of proteasomal protein catabolic processes and phosphotransferase activity (Table 4) in the *C. reinhardtii* and *A. thaliana* databases respectively. Upregulated

proteins that were related to both these GO-terms were identified as a WD_REPEATS_REGION domain containing protein and 26s proteasome regulatory subunit.

4.4.2 Differential gene expression analysis of sulfur pretreated Euglena gracilis cultures both in the presence and absence of $CdCl_2$

Between samples of *E. gracilis* cells that were pretreated with sulfur and non-pretreated cultures in the presence of $CdCl_2$, 180 transcripts were found to be differentially expressed, 62 of which were upregulated (Figure S3 B). The clustering analysis showed no separation of the samples (Figure S2 C-D). Out of the 180 transcripts, 54 showed sequence similarities within the non-redundant and SWISS-PROT protein databases from NCBI (Table S10-S11). When compared to species specific databases, 11 were identified in the *A. thaliana* database, 9 in the *H. sapiens* database, 13 in the *C. reinhardtii* database, 5 in *Synechocystis sp* database and 9 were identified in the *T. brucei* database (Table S12-16).

When comparing transcripts from sulfur treated *E. gracilis* cells in the presence and absence of $CdCl_2$, 174 were differentially expressed, 104 of which were upregulated (Figure S3 C). The clustering analysis showed no separation of the samples (Figure S2 E-F). Out of the 174 differentially expressed transcripts, 104 were identified using the NCBI non-redundant and SWISS-PROT protein databases (Table S17-S18). When the same 174 transcripts were searched in the Ensembl protein databases, 31 showed a similarity to sequences in the *A. thaliana* database, 26 had similarities to the *H. sapiens* database, 27 had similarities to the *C. reinhardtii* database, eight showed similarities to proteins in the *Synechocystis sp.* database, and 28 were found to have sequence similarities with proteins in the *T. brucei* databases (Table S19-S23).

E. gracilis cultures that were pretreated with sulfur, compared to non-pretreated *E. gracilis* cells, yielded 157 transcripts that were differentially expressed, 62 of which were upregulated (Figure S3 D) while the clustering analysis showed no separation of the samples (Figure S2 G-H). Out of 157 genes that were differentially expressed, 34 were identified in the NCBI non-redundant and SWISS-PROT protein databases (Table S24-S25). Using the Ensembl protein databases, 11 transcripts showed a similarity to sequences in the *A. thaliana* database, five showed similarities to proteins in the *H. sapiens* database, nine showed sequence similarities to proteins in the *C. reinhardtii* database, seven showed sequence similarities to proteins in the *Synechocystis sp.* database, and six show sequence similarities to proteins in the *T. brucei* database (Table S26-S30).

The differentially expressed transcripts of sulfur pretreated cultures exposed to Cd compared to cells grown in MAM prior to Cd exposure when searched against the NCBI non-redundant protein and SWISS-PROT protein databases, were identified as having sequence similarities to chemotaxis proteins, heat shock proteins and ABC transporter proteins. The chemotaxis protein was also identified when comparing differentially expressed transcripts in *E. gracilis* cells grown only in MAM in the presence and absence of $CdCl_2$, except when the cells were pretreated with sulfur, the identified transcript was downregulated. The transcript identified as encoding an ABC transporter was also downregulated while the transcript for the gene encoding a heat shock protein was upregulated. A heat shock protein transcript was also identified as upregulated when differential gene expression analysis was done comparing sulfur pretreated *E. gracilis* cells in the presence and absence of $CdCl_2$. Along with the heat shock protein, a dehydration responsive protein was also upregulated. Another downregulated transcript encoded a

potential serine/threonine kinase protein. When comparing the sulfur pretreated *E. gracilis* cultures to non-pretreated cultures, the transcripts that were differentially expressed included various dehydrogenase enzymes, reverse transcriptase and intron maturase, pyruvate formate and a variety of hypothetical proteins.

While the comparison of sulfur pretreated cultures and non-pretreated cultures yielded no GO-term enrichment, in the presence of Cd there was GO-term enrichment for cysteine metabolic process as well as catalytic activity. Cysteine metabolic process was identified as enriched across the *T. brucei* database when comparing sulfur pretreated *E. gracilis* cells in the presence of $CdCl_2$ to non-pretreated cultures exposed to $CdCl_2$ (Table 4). Catalytic activity was identified within sulfur pretreated *E. gracilis* cells in the presence and absence of $CdCl_2$ (Table 4).

4.3 Differential gene expression analysis of nitrogen pretreated Euglena gracilis cultures in the presence and absence of $CdCl_2$

When comparing nitrogen pretreated cultures and the non-pretreated cultures, both in the presence of $CdCl_2$, a total of 293 transcripts were identified as being differentially expressed with DESeq2 with 104 of them being upregulated (Figure S3 E). The clustering analysis showed no separation of the samples (Figure S2 I-J). Out of the 293 transcripts, a total of 166 genes were identified as having sequence similarity to genes in the non-redundant and SWISS-PROT protein databases on NCBI (Table S31-S32). When compared to the databases for *A. thaliana,* 36 transcripts showed protein sequence similarities, 33 transcripts showed sequence similarities to proteins in the *H. sapiens* database, 28 showed sequence similarities to proteins in the *C. reinhardtii* database, 28

showed sequence similarities to proteins in the *Synechocystis sp.*, and 29 transcripts had sequence similarities to proteins in the *T. brucei* database (Table S33-37).

Nitrogen pretreated *E. gracilis* cells grown in the presence and absence of $CdCl_2$ were compared and 345 transcripts were found to be differentially expressed while 204 were upregulated with a > 2 log2 fold change (Figure S3 F). The clustering analysis showed no separation of the samples (Figure S2 K-L). Of the 345 sequences that were searched in the blastx non-redundant and SWISS-PROT protein databases, a total of 203 had sequence similarity with proteins in those databases (Table S38-S39). Out of the 345 transcripts, 40 were identified as having sequence similarities to proteins in the *Arabidopsis thaliana* database, 28 were identified in the *Homo sapiens* protein database and 39 were identified in the *Chlamydomonas reinhardtii*, 23 in *Synechocystis sp.* and 34 in *Trypanosoma brucei* (Table S40-S44).

When comparing nitrogen pretreated cultures to non-pretreated cultures, a total of 279 transcripts were differentially expressed, 130 of which were upregulated (Figure S3 G) while the clustering analysis showed no separation of the samples (Figure S2 M-N). When searching the NCBI non-redundant and SWISS-PROT protein database, 59 of the 279 were identified to have sequence similarities (Table S45-S46). When the same 279 transcripts were searched for in specific protein databases, 17 were identified in the *A. thaliana* database, 15 were identified in the *H. sapiens* database, 16 were identified in the *C. reinhardtii* database, 7 in the *Synechocystis sp.* database and 13 in the *T. brucei* protein database (Table S47-S51).

The NCBI non-redundant and SWISS-PROT protein databases were searched to identify the transcripts that were differentially expressed. Similar to the results of the sulfur pretreated cultures, the identified proteins were different than those differentially expressed in the non-pretreated *E. gracilis* cultures. With nitrogen pretreated samples that were exposed to $CdCl_2$ compared to non-pretreated *E. gracilis* cultures exposed to $CdCl_2$, chemotaxis proteins, heat shock proteins, ABC transporter proteins and serine/threonine kinase identified based on their sequence similarities to the transcripts that were differentially expressed based on the DGE analysis. Like the comparison between the sulfur pretreated cultures exposed to $CdCl_2$ and the non-pretreated cultures exposed to $CdCl_2$, transcripts for genes encoding chemotaxis proteins and ABC transporter proteins were downregulated. The nitrogen pretreated cultures exposed to $CdCl_2$ vs non-pretreated *E. gracilis* cultures exposed to $CdCl_2$ also saw the upregulation of transcripts for genes encoding heat shock proteins and the downregulation of transcripts of genes encoding serine/threonine kinase. GO-term analysis identified the enrichment of GO-terms relating to biological processes such as translational elongation and transmembrane transport in *A. thaliana*, cellular components including plasmodesma and vacuoles in *A. thaliana* and molecular function including ABC-type transporter activity as well as Nucleoside triphosphatase activity and ATP binding within the *C. reinhardtii* database (Table 4).

Transcripts were found to be differentially expressed between nitrogen pretreated cultures that were exposed to 0 µM and 25 µM $CdCl_2$. These included transcripts for genes encoding enzymes related to amino acid biosynthesis such as, threonine dehydrogenase (downregulated) and 2-isopropylmalate synthase (upregulated). Other upregulated transcripts for genes encoding cold shock proteins and ATP-dependent chaperone protein,

a polyamine ABC transporter ATP binding protein and a leucine rich repeat containing protein were identified. The GO-term analysis identified GO-term enrichment for the GO Molecular function annotation sets including hydrolase activity acting on acid anhydrides, iron-sulfur cluster binding, primary active transmembrane transport activity, ATP-dependent activity and ATP binding, metal ion binding and catalytic activity in the *C. reinhardtii* database (Table 4).

When comparing nitrogen pretreated cultures to non-pretreated cultures, the NCBI non redundant and SWISS-PROT database identified some of the differentially expressed transcripts as having sequence similarities to genes that encode heat shock proteins and DHA2 family efflux transporter permease subunit proteins. Unlike when nitrogen pretreated cultures were exposed to Cd, the heat shock proteins were downregulated. GO-term analysis on these cell types identified GO-term enrichment of the biological process of response to heat as well as molecular functions such as unfolded protein binding, small molecule binding, heterocyclic compound binding and organic cyclic compounds binding in *C. reinhardtii* (Table 4).

Table 4: GO-Term analysis results for non-pretreated, sulfur pretreated and nitrogen pretreated *Euglena gracilis* cultures.

GO-Term	Annotation Set	Fold Enrichment	P-Value	Database
Non-pretreated with and without $CdCl_2$				
positive regulation of proteasomal protein catabolic process (1901800)	Biological Process	>100	4.38E-5	*Chlamydomonas reinhardtii*
Phosphotransferase activity carboxyl group as acceptor (GO:0016774)	Molecular Function	>100	8.90 E-6	*Arabidopsis thaliana*
Sulfur pretreated and non-pretreated in $CdCl_2$				
Cysteine metabolic process (GO:0006534)	Biological Process	> 100	9.79 E-6	*Trypanosoma brucei*
Sulfur pretreated grown in the presence and absence of $CdCl_2$				
Catalytic activity (GO:0003824)	Molecular Function	3.13	1.42 E-6	*Arabidopsis thaliana*
Nitrogen pretreated and non-pretreated in $CdCl_2$				
Translational elongation (GO:0006414)	Biological Process	>100	4.00E-8	*Arabidopsis thaliana*

Transmembrane transport (GO:0055085)	Biological Process	11.07	3.35E-7	*Arabidopsis thaliana*
Plasmodesma (GO:0009506)	Cellular component	7.88	8.53E-5	*Arabidopsis thaliana*
Vacuole (GO:0005773)	Cellular Component	7.23	2.18E-5	*Arabidopsis thaliana*
ABC-type transporter activity (GO:0140359)	Molecular Function	40.65	6.28E-5	*Chlamydomonas reinhardtii*
Nucleoside-triphosphatase activity (GO:0017111)	Molecular Function	10.87	8.56E-5	*Chlamydomonas reinhardtii*
ATP binding (GO:0005524)	Molecular Function	4.68	6.35E-6	*Chlamydomonas reinhardtii*
Nitrogen pretreated grown in the presence and absence of $CdCl_2$				
Hydrolase activity acting on acid anhydrides	Biological Process	17.52	9.28E-7	*Arabidopsis thaliana*

(GO:0016818)				
Iron-sulfur cluster binding (GO:0051536)	Molecular Function	23.88	2.46E-5	*Chlamydomonas reinhardtii*
Primary active transmembrane transporter activity (GO:0015399)	Molecular Function	20.72	4.29E-4	*Chlamydomonas reinhardtii*
ATP-dependent activity (GO:0140657)	Molecular Function	7.56	4.54E-4	*Chlamydomonas reinhardtii*
ATP binding (GO:0005524)	Molecular Function	5.20	8.33E-6	*Chlamydomonas reinhardtii*
Metal ion binding (GO:0046872)	Molecular Function	4.25	3.53E-4	*Chlamydomonas reinhardtii*
Catalytic activity (GO:0003824)	Molecular Function	2.44	3.49E-5	*Chlamydomonas reinhardtii*

Nitrogen pretreated and non-pretreated cultures				
Response to heat (GO:0009408)	Biological Process	32.29	5.94E-6	*Arabidopsis thaliana*
Unfolded protein binding (GO:0051082)	Molecular Function	>100	2.11E-6	*Homo sapiens*
Small molecule binding (GO:0036094)	Molecular Function	5.16	4.35E-5	*Chlamydomonas reinhardtii*
Heterocyclic compound binding (GO:1901363)	Molecular Function	4.05	4.60E-5	*Chlamydomonas reinhardtii*
Organic cyclic compound binding (GO:0097159)	Molecular Function	4.04	4.70E-6	*Chlamydomonas reinhardtii*

CHAPTER FIVE: DISCUSSION

E. gracilis has demonstrated an overall tolerance to heavy metals such as Hg and Cd (Devars et al., 2000, Avilés et al., 2003). This tolerance is accompanied by an ability for *E. gracilis* to uptake these and store them in their cells (Devars et al., 2000). This natural tolerance can be enhanced using pretreatments with compounds such as Hg (Avilés et al., 2003). Studies investigating how *E. gracilis* can tolerate and remove heavy metals led to the discovery of specific metal binding compounds that are produced within the cell and are enriched with both sulfur and nitrogen (Winters et al., 2021, Mangal et al., 2022). In this thesis, the impact of sulfur or nitrogen pretreatment on the tolerance of *E. gracilis* cultures to $CdCl_2$ was explored by focusing on comparing the pretreated cultures to non-pretreated cultures in the presence and absence of $CdCl_2$. This was done using growth assays to monitor the viability of cells, as well as using RNA-seq and subsequent GO term analysis to identify gene functions associated with altered transcript levels. Growth assays demonstrated that, at concentrations of 60 mM and 75 mM of sulfur and nitrogen respectively, there were no immediate or long-term changes to cell morphology metal cultures. When pretreated and non-pretreated *E. gracilis* cells were exposed to increasing concentrations of $CdCl_2$, sulfur and nitrogen had a positive impact on tolerance. This was demonstrated by a significantly higher number of living cells after exposure to $CdCl_2$ when compared to non-pretreated cultures exposed to the same concentration of $CdCl_2$. RNA-Seq analysis revealed several differentially expressed transcripts when comparing pretreated cultures to non-pretreated cultures both in the presence and absence of $CdCl_2$. Using GO term analysis, there was GO term enrichment when comparing the sulfur and nitrogen pretreated cultures to non-pretreated cultures, both in the presence and absence of

$CdCl_2$. This investigation into the use of a sulfur and nitrogen pretreatment to enhance *E. gracilis* cell's tolerance to $CdCl_2$ expanded our understanding into how *E. gracilis* binds heavy metals by revealing important transcript level changes that occur when pretreateed *E. gracilis* cultures were challenged with $CdCl_2$. The results from this study can be used to potentially enhance *E. gracilis*' metal binding capabilities, improving its bioremediation potential.

5.1 Exposure to increased concentrations of $MgSO_4*7H_2O$ and NH_4NO_3 demonstrated the ability of *Euglena gracilis* to survive changing environments

The first objective of this thesis was to assess the impact of a sulfur and nitrogen pretreatment on the $CdCl_2$ tolerance of *E. gracilis* cultures. Growth assays were done to determine if increasing sulfur or nitrogen concentrations in the media would have a significant impact on *E. gracilis* cells. To increase the sulfur concentration, excess $MgSO_4*7H_2O$ was used. This compound was selected because *E. gracilis* can uptake sulfate and, when under stresses such as SO_4 deficiency or Cd^{2+} exposure, it is sulfate transporters that show increased activity (García-García et al., 2012). Our results support the use of $MgSO_4*7H_2O$ as the sulfur pretreatment since, even at a concentration of 120 mM SO_4, the cultures continued to grow at comparable rates to the non-pretreated cultures. Sulfur pretreated cultures had no consistent significant changes in growth compared to the non-pretreated cultures over the course of 66 days except for on three isolated incidents on da. This indicated that *E. gracilis* was able to grow in environments with increased sulfur and that it was able to tolerate the excess sulfate provided by $MgSO_4*7H_2O$.

On the other hand, *E. gracilis* is only able to efficiently metabolize nitrogen in the form of NH_4^+ and does not metabolize nitrate, nitrite, or urea (Richter et al., 2015). In this study, when grown at the higher concentration of NH_4NO_3 tested (76 mM and 151 mM), *E. gracilis* cells began to appear visibly stressed after 10 days of growth, with those grown in 151 mM NH_4NO_3 changing color from green to brown. Long term, nitrogen treated cells showed a trend of reduced growth compared to those that were either sulfur treated or non-pretreated. However, this difference was not statistically significant ($p > 0.05$) except for on the final day of the growth assay. This could indicate that while 75 mM NH_4NO_3 was not harmful to the cultures, there is a possibility that the trend of reduced growth that was seen was because of an inability of *E. gracilis* to utilize nitrate and could only metabolize NH_4^+. It was also possible that the concentration of NH_4NO_3 that was used for the nitrogen pretreatement was too high that it caused the trend of reduced growth in the culture.

5.2 Cultures pretreated with either excess sulfur or nitrogen had an increased tolerance to various $CdCl_2$ concentrations compared to cells grown in MAM

Growth assays were conducted with sulfur and nitrogen pretreated *E. gracilis* cultures that were exposed to 0 µM, 10 µM, 25 µM, 75 µM and 125 µM of $CdCl_2$. The methods used were similar to those of Avilés et al., (2003), except cell counts and transfers into fresh media were completed every 96 hrs. The results indicated that non-pretreated *E. gracilis* cultures had a moderate tolerance to 10 µM $CdCl_2$ since there was no statistical difference in the number of viable cells compared to the cultures exposed to 0 µM $CdCl_2$. Both the

sulfur and nitrogen pretreated cultures showed no statistically significant difference at 10 µM and 25 µM $CdCl_2$ when compared to pretreated cultures not exposed to 0 µM. This trend continued over the course of 8 days. The results were consistent with those of Mendoza-Cozatl et al., 2002 and showed that non pretreated *E. gracilis* cells had a moderate tolerance to $CdCl_2$. The results of the pretreated cultures, indicate that, similar to the results found by Aviles et al.,2003, *E. gracilis* can be induced to have a higher tolerance to Cd using a sulfur and nitrogen pretreatment.

5.3 *Euglena gracilis de novo* transcriptome assembly using Trinity

Currently, *E. gracilis* does not have a sequenced and annotated genome that can be used as a reference for transcript level analysis. The main reason for this has been difficulty in sequencing and annotating the genome. Recent attempts to assemble the genome have been less successful due to the heterozygosity, size, and the frequency of sequences with low complexity (Ebenezer et al., 2019). Here, we report a *de novo* transfrag assembly using trimmed paired-end transcript sequences. In the past, *de novo* transcript assemblies have been generated for *E. gracilis* and the quality of the assembly has been determined using BUSCO. Our assembly searched a total of 255 BUSCO groups and found 211 complete BUSCOs (82.8%). These results are similar to other *E. gracilis* assemblies, where: O'Neill et al., 2015 found 267 complete BUSCOs (88.1%), Yoshida et al., 2016 found 262 complete BUSCOs (86.4%), Ebenezer et al., 2019 had 255 complete BUSCOs (84.1%) and Mangal et al., 2022 had 299 complete BUSCOs (99%). The difference between our study and others is the version of database that was used. Our assessment used the

eukaryota Orthodb v10 database was used while previous studies used the *eukaryota* Orthodb v9 database for analysis - so there was a change in the number of BUSCO groups that were searched with v9 searching 303 BUSCO groups and v10 only searching 255.

5.4 Increased *Euglena gracilis* tolerance to $CdCl_2$ can be linked to the use of a sulfur pretreatment

The second objective of this thesis was to identify genes with altered transcript levels because of both the pretreatment with sulfur and nitrogen, and the exposure to $CdCl_2$. Comparing the differentially expressed transcripts identified with the sulfur pretreated cultures and the non-pretreated cultures, there were differences in the transcripts that were differentially expressed. These genes were identified as having sequence similarities to those that encode proteins related to transmembrane transport, stress response and enzymes related to cysteine metabolism.

5.41 Transmembrane transport

When the transcript level changes for non-pretreated *E. gracilis* cells that were exposed to $CdCl_2$ were compared to those of non-pretreated cultures grown only in MAM, there were no transmembrane transport proteins identified when the sequences were searched against the NCBI non-redundant and SWISS-PROT protein databases. This contrasted with what was identified in sulfur pretreated samples exposed to $CdCl_2$. The NCBI non-redundant and SWISS-PROT protein databases identified one downregulated transcript as having sequence similarity to an ABC-transporter protein. This is indicative that in the presence

of Cd, a pretreatment with sulfur impacts areas of transmembrane transport that is not seen in the non-pretreated cultures.

ABC transporters are ATP-binding cassette transporters that are responsible for ATP-powered translocation of many substrates across membranes (Song et al., 2014). They are part of a superfamily of importer and exporter membrane proteins (Locher, 2016). Other microorganisms such as *Thermus thermophilus* can transport metal ions across the cell membrane (Mandal et. al., 2019). The ions are first sequestered by substrate-binding proteins before being transferred to transmembrane domains for future transport (Mandal et al., 2019). In organisms such as yeast, once Cd forms complexes with PCs and glutathione, it is then transported into the vacuole of the cells by ABC transporters (Bovet et al., 2005). There is evidence indicating that ABC transporters are involved in heavy metal resistance in eukaryotic cells. ABC transporters can confer heavy metal tolerance in organisms such as *Saccharomyces cerevisae* and *Schizosaccharomyces pombe* (Song et al., 2014). In mammalian cells, ABC transporter proteins ABCG2 and ABCB4 were downregulated in the placenta of rats exposed to Cd (Liu et al., 2016). The results obtained by Lui et al., 2016 led to the hypothesis that these ABC transport proteins are responsible for the regulation of Cd through the placenta.

An ABC transporter protein was downregulated in sulfur pretreated *E. gracilis* exposed to $CdCl_2$ when compared to non-pretreated *E. gracilis* cells exposed to $CdCl_2$.This may indicate that the ABC transporter protein regulates the flow of $CdCl_2$ in an *E. gracilis* cell. Since this was identified in the sulfur pretreated cultures, it could also indicate that the sulfur pretreatment, through decreased expression of a potential ABC transporter, can help regulate the flow of $CdCl_2$ through the cell, thereby increasing tolerance.

5.4.2 Stress response

Along with ABC-transporter proteins, genes encoding various stress response proteins were differentially expressed in *E. gracilis* cultures exposed to $CdCl_2$, many of which were different between sulfur pretreated cultures and non-pretreated cultures. The proteins identified by NCBI's non-redundant and SWISS-PROT database included chemotaxis proteins and heat shock proteins both of which have been identified as stress response proteins in bacteria and plants respectively (Barque et al., 1996; Berleman & Bauer, 2005; Heckathorn et al., 2004).

Within bacteria, chemotaxis proteins are responsible for controlling chemotactic responses (Berleman & Bauer, 2005). Chemotaxis proteins are differentially expressed when metal tolerant bacteria are exposed to metals such as Cd (Yung et al., 2014). In the presence of $CdSO_4$, chemotaxis proteins were downregulated in *C. crescentus* (Yung et al., 2014). In this study, genes encoding potential chemotaxis proteins were differentially expressed when comparing both sulfur pretreated and non-pretreated *E. gracilis* cells in the presence of $CdCl_2$. When non-pretreated *E. gracilis* cultures were exposed to $CdCl_2$ the transcript for the gene encoding a potential chemotaxis protein was upregulated. In contrast, the sulfur pretreated *E. gracilis* cultures, had downregulated genes encoding a potential chemotaxis protein compared to non-pretreated cultures exposed to $CdCl_2$. This could indicate that, because of the sulfur pretreatment, there was an increased tolerance to $CdCl_2$ which was seen at the transcript level with a potential chemotaxis protein being downregulated.

In bacteria, the change in expression of chemotaxis proteins is also accompanied by changes in expression of genes encoding flagellar motor switch proteins (Yung et al.,

2014). Flagellar motor switch proteins make up what is known as the "switch complex". This complex is responsible for the direction of flagellar rotation, impacting the overall motility of the cell (Paul et al., 2011). In bacteria, chemotaxis regulators transmit various chemoreceptor signals to the flagellar motor CheY component which starts the cascade causing the change in rotational sense which is required for chemotaxis (Sarkar et al., 2010). Organisms, such as bacteria, will lose their motility to ensure their survival when facing environmental stressors. (Sharma et al., 2019). When *Klebsiella pneumoniae* was exposed to meropenem stress, proteomics revealed that there was a downregulation of proteins related to flagellar formation and its subsequent functioning, including flagellar motor switch protein FliG. In this study, when non pretreated *E. gracilis* cultures were exposed to $CdCl_2$, there was a downregulation of genes encoding potential flagellar motor switch proteins. This is indicative of the need for *E. gracilis* to modify its motility as a response to the stress of $CdCl_2$ in its environment like bacteria. In this study, the downregulation of transcripts that were identified as having sequence similarities to chemotaxis proteins could indicate that, rather than needing to leave an unfavorable environment, pretreated *E. gracilis* cultures had an increased tolerance, allowing them to remain in the environment with $CdCl_2$.

Genes encoding putative heat shock proteins, were identified in *E. gracilis* cells that were pretreated with sulfur prior to $CdCl_2$ exposure. In plants, small heat shock proteins work to protect photosynthesis and electron transport during a variety of stressors including heavy metal stress in species such as *Zea mays* (Heckathorn et al., 2004) In *E. gracilis*, exposure to Cd caused an overexpression of heat shock proteins, specifically hsp90, hsp70, hsp55 and hsp40 which was also confirmed through proteomic analysis

(Barque et al., 1996, Khatiwada et al., 2020). In this study, sulfur pretreated *E. gracilis* cultures showed an upregulation in transcripts that were identified as heat shock proteins based on sequence similarities to *E. gracilis* heat shock proteins in the NCBI non-redundant protein database. This could indicate that *E. gracilis* uses heat shock proteins to protect photosynthesis and electron transport when under heavy metal stress. This trend was not seen in the non-pretreated cultures, further indicating that the pretreatment with sulfur enhanced the tolerance of *E. gracilis* to $CdCl_2$ since these proteins have been hypothesized to allow for *E. gracilis* to be tolerant to metals (Barque et al., 1996).

The upregulation of a transcript that was identified to have sequence similarities to heat shock proteins, is indicative of the influence of a sulfur pretreatment on *E. gracilis'* tolerance to $CdCl_2$. Similar to the findings of Barque et al., 1996 and Khatiwada et al., 2020, there was an upregulation of a potential heat shock protein in the sulfur pretreated *E. gracilis* cultures from this thesis, when the pretreated cultures were exposed to $CdCl_2$. This trend was not seen in the non-pretreated cultures, further indicating that the pretreatment with sulfur enhanced the tolerance of *E. gracilis* to $CdCl_2$ since these proteins have been hypothesized to allow for *E. gracilis* to be tolerant to metals.

5.4.3 Metabolic changes

Using RNA isolated from sulfur pretreated *E. gracilis* cultures that were exposed to $CdCl_2$, GO term analysis identified an enrichment of genes related to cysteine metabolic process. This was unique to the sulfur pretreated *E. gracilis* cultures exposed to $CdCl_2$. In protists, the sulfur assimilation pathway leads to cysteine biosynthesis and involves sulfate transporters as well as enzymes including cystathionine beta-synthase and cystathionine

gamma-lyase (Mendoza-Cózatl et al., 2005). These pathways can be altered by sulfur deficiency and heavy metal exposures (Mendoza-Cózatl et al., 2005).

In *E. gracilis*, metabolomic analysis identified various metabolites related to sulfur and sulfur metabolism when *E. gracilis* was exposed to Hg (Mangal et al., 2022). These included sulfate, cysteine, and glutathione. The metabolites are important chelators for metals like Cd. Cysteine is an important precursor of glutathione biosynthesis which works to mitigate the oxidative stress induced by Cd (Zhang et al., 2005) . When sulfur pretreated cultures of *E. gracilis* were exposed to $CdCl_2$, there was an increase in transcripts having sequence similarities to proteins found within the cysteine metabolic process including in cystathionine gamma-lyase and cystathionine beta-synthase. Those specific enzymes catalyze the metabolism of cysteine which is used to synthesize glutathione (Mangal et al., 2022). Transcript level analysis also identified significantly enriched genes related to amide biosynthesis and metabolism such specifically amino acids such as cysteine, a known precursor to glutathione (Mangal et al., 2022). These results are indicative that pretreatment of *E. gracilis* with sulfur modified and increased cysteine metabolism and could have downstream impacts on the synthesis of glutathione and other known phytochelatins.

5.5 Increased Cd tolerance can be linked to pretreatment of *Euglena gracilis* with nitrogen

The second objective of this thesis was to identify genes that had altered transcript levels because of both the pretreatment with sulfur and nitrogen, and the exposure to $CdCl_2$. The

pretreatment of *E. gracilis* cultures with nitrogen lead to changes in transcript levels. These transcripts were for genes encoding transmembrane transport proteins, stress response proteins, metal binding proteins and enzymes for amino acid biosynthesis.

5.5.1 Transmembrane transport

Nitrogen pretreated *E. gracilis* cultures had downregulated transcripts that had sequence similarities with an ABC-transporter protein. This result was similar to what was observed with the sulfur pretreated cultures. This contrasted with non-pretreated cultures that were exposed to $CdCl_2$ where there was no differential expression of genes encoding proteins related to transmembrane transport. This indicates that the use of a nitrogen pretreatment increased the overall tolerance of $CdCl_2$ by regulating the transport of $CdCl_2$ throughout the cell, similarly to what has been seen across other organisms (Liu et al., 2016; Song et al., 2014).

5.5.2 Stress response

The use of a nitrogen pretreatment prior to the exposure to $CdCl_2$ resulted in differentially expressed transcripts having sequence similarities to chemotaxis proteins and heat shock proteins. When nitrogen pretreated cultures were compared to non-pretreated cultures, there was GO-term enrichment for the GO-term related to response to heat.

The impact of the nitrogen pretreatment on overall metal tolerance was similar to that of the sulfur pretreated *E. gracilis* cultures when exposed to $CdCl_2$. When nitrogen pretreated cultures that were exposed to $CdCl_2$ were compared to the non-pretreated *E. gracilis* cultures that were exposed to $CdCl_2$, genes encoding a potential chemotaxis protein were downregulated. As was found with the sulfur pretreated cultures, the downregulation

of the transcript encoding a potential chemotaxis protein could indicate a higher tolerance of *E. gracilis* to Cd since it is an identified stress response protein in bacteria (Bren & Eisenbach, 2000; Sarkar et al., 2010).

Other transcripts that were differentially expressed when nitrogen pretreated cultures were exposed to $CdCl_2$ included transcripts with sequence similarities to heat shock proteins in the NCBI non-redundant protein database. The data in this study shows that nitrogen pretreated *E. gracilis* cultures that were exposed to $CdCl_2$ had a downregulation of heat shock proteins when nitrogen pretreated cultures that were exposed to $CdCl_2$ were compared to non-pretreated cultures that were exposed to $CdCl_2$. The downregulation of genes encoding potential heat shock protein was also seen in nitrogen pretreated *E. gracilis* cultures when compared back to non-pretreated *E. gracilis* cultures when neither were exposed to $CdCl_2$. As was seen with the sulfur pretreated *E. gracilis* cultures, heat shock proteins are related to stress response in plants, including heavy metal stress response. These results could indicate that the stress response mechanisms are different on the nitrogen pretreated cultures even with the addition of $CdCl_2$ since unlike with the sulfur pretreated cultures, there was already a downregulation of a transcript identified as a potential heat shock protein.

5.5.3 Amino acid biosynthesis

GO term analysis of the nitrogen pretreated cultures exposed to cadmium did not identify enriched genes related to amino acid biosynthesis. However, there were two enzymes coded by differentially expressed transcripts that were relevant to this process: threonine dehydratase and 2-isopropylmalate synthase, that were identified when searching transcript

sequences in the NCBI non-redundant and SWISS-PROT protein databases. Threonine dehydratase has been downregulated in fungi as a response to $CdCl_2$ exposure, specifically in the fungus *Psxillud involutus*. Cd can impact amino acid metabolism to suppress the TCA cycle (Ding et al., 2020). The downregulation of threonine dehydratase was thought to be the result of the redirection of carbon skeletons from ASP to the TCA cycle (Jacob et al., 2004). 2-isopropylmalate synthase is responsible for amino acid biosynthesis in plants and nitrogen metabolism (Wang et al., 2018). In plants, Isopropylmalate converts acetylCoA and a-ketoisovalerate into a-isopropylmalate which is essential in leucine biosynthesis (de Kraker et al., 2007). Aminotransferase class I/II-fold pyridoxal phosphate-dependent enzyme was another enzyme that was differentially expressed based on its sequence similarity to one of the differentially expressed transcripts in this thesis. These enzymes are differentially expressed as a response to various stressors including furfural and acetic acid (Shabbir et al., 2021).

In this study, nitrogen pretreated cultures that were exposed to $CdCl_2$ showed a downregulation of the transcript that shared sequence similarities to the enzyme threonine synthase, while the transcript that had sequence similarities to the enzyme 2-isopropylmalate synthase was upregulated. Since both are responsible for amino acid biosynthesis and only threonine dehydratase was downregulated, this could indicate a shift the mechanism for amino acid biosynthesis due to the use of the nitrogen pretreatment. The shift could be from the use of threonine dehydratase as the primary catalyst amino acid biosynthesis, to utilizing isopropylmalate-synthase. In nitrogen pretreated cultures without $CdCl_2$, the differential expression of a potential aminotransferase class I/II-fold pyridoxal phosphate-dependent enzyme could be related to stress response however, the exact link

between the use of a nitrogen pretreatment and this enzyme are not discussed in current literature.

5.5.4 Metal ion binding

GO-term analysis of nitrogen pretreated *E. gracilis* cultures exposed to $CdCl_2$ identified the enriched GO-terms: “related to metal cluster binding” and “metal ion binding as being enriched”. Within these GO-terms, specific genes that were identified based on the transcript sequence similarity to proteins in the *C. reinhardtti* database. The genes included calmodulin, phospholipid transporting ATPase, L-ascorbate peroxidase, and serine/threonine kinase.

Calmodulin is a calcium (Ca) binding protein. Due to the chemical similarities of Ca and Cd, it has been hypothesized that Cd can bind to the binding sites on calmodulin instead of Ca (Choong et al., 2014). In wheat, growth was affected by the interactions between lanthanum (La) and Ca and Cd and the interplay of all three (Yang et al., 2021). It was in the face of calcium deficiency, that La was stimulated to mitigate Cd toxicity. These results suggested La, Ca and Cd had shared binding sites in calmodulin (Yang et al., 2021). Inhibiting calmodulin caused a decrease in intracellular Cd and a decrease in GSH and PCs levels in *Ulva compressa* (Riechers et al., 2021). Proteomics analysis on *E. gracilis* cultures identified calmodulin as one of the enriched GO-terms (Khatiwada et al., 2020). Within the phospholipid transporting ATPase family of enzymes, specific ATPases exist for transporting metals such as copper transporting ATPase (MERAKLI et al., 2021). Plants such as *Brassica nigra* L., when exposed to Cu, showed an upregulation of not only copper-transporting ATPase, but also of a putative phospholipid-transporting ATPase 5

protein. This specific transporter gene was also differentially expressed when *Triticum aestivum* cultivates were exposed to both drought and salinity stress (Dugasa et al., 2021). The findings in this study are supported by the current literature and indicate that calmodulin can potentially bind Cd and be the reason that *E. gracilis* can remove Cd and that transporter genes such as phospholipid-transporting ATPase 5 could be linked to the metal tolerance and binding capabilities of *E. gracilis*.

Two other enzymes that were identified in the metal binding GO-term were ascorbate peroxidase and serine/threonine kinases. Ascorbate peroxidase is a ROS scavenger protein, similar to that of glutathione S-transferase. These enzymes are known to play a key role in response to metal-stress in plants (Merakli et al., 2021). With *B. nigra* L., L-ascorbate peroxidase was upregulated after exposure to copper as a response to the metal-related oxidative stress. (Merakli et al., 2021). Ascorbate peroxidase can scavenge ROS allowing plants such as *Lactuca* sativa to absorb Cd(Kolahi et al., 2020). Serine/threonine kinases was identified as one of the genes in the *C. reinhardtti* database that was under the GO-term "metal binding". This group of kinases functions as one part of the stress response cascade in plants (Muhammad et al., 2019). In one study, the overexpression of *SlMAPK3* in tomatoes increased Cd tolerance which was indicated through the decrease of malondialdehyde and hydrogen peroxide accumulation (Muhammad et al., 2019). In *C. reinhardtii*, the *Sac3* gene also regulates the response to sulfur limitation which is known to have similar effects on plants as heavy metal contamination (Davies et al., 1999, Adhikari et al., 2018). Within *E. gracilis*, proteomic analysis indicated an increase in serine/threonine kinases expression and was classified as metal binding proteins based on the proteomic analysis (Khatiwada et al., 2020). In this

study, a downregulated transcript was identified as L-ascorbate peroxidase based on its sequence similarity to the enzyme in the NCBI non-redundant protein database. This could indicate that while ascorbate peroxidase can play a role in scavenging reactive oxygen species, the use of the nitrogen pretreatment prior to $CdCl_2$ did not enhance the binding of $CdCl_2$ by these means and the increased tolerance was controlled by other mechanisms. Nitrogen pretreated *E. gracilis* cultures showed that a potential serine/threonine kinase was differentially expressed. These results could be indicative of a cascade being started by serine/threonine kinase that is related to the tolerance and binding of $CdCl_2$, similar to its mechanism in plants.

5.6 Future directions

Mangal et al., (2022) identified metabolites that were related to tolerance of *E. gracilis* to Hg. To compare the results of a sulfur and nitrogen pretreatment on $CdCl_2$ tolerance, a similar metabolomic analysis can be conducted. This would allow for comparisons at the metabolite level to identify similar and unique metabolites among pretreatments as well as the use of $CdCl_2$ compared to mercury. Along with metabolomics, mass spectrometry of both the sulfur and nitrogen pretreated *E. gracilis* cultures can quantify the increased $CdCl_2$ tolerance. Mass spectrometry (ICP-MS) can be used to quantify the concentration of $CdCl_2$ that remained in the media, and the concentration in the cell. Based on the RNA-Sequencing results, the next step in terms of transcript level data would be to confirm the RNA-Sequencing results by selecting transcripts that were identified as having sequence similarity to transmembrane transport, stress response, cysteine metabolic process and

amino acid biosynthesis as well as the transcripts that were identified as potential metal ion binding proteins using GO-term analysis. Once selected, RT-qPCR can be performed by designing primers for these sequences to confirm the results of the RNA-seq analysis (Khatiwada et al., 2020).

The sulfur pretreatment in this thesis increased the total sulfate concentration to 60 mM through the addition of excess $MgSO_4*7H_2O$ prior to exposure to 25 µM $CdCl_2$. This research can be expanded by increasing both the concentration of $MgSO_4*7H_2O$ used for the pretreatment, and by increasing the $CdCl_2$ concentration that the cultures are exposed to. Increasing both concentrations is feasible since the results of this thesis show that *E. gracilis* can tolerate up to 120 mM excess $MgSO_4*7H_2O$ and that sulfur pretreated cultures had a tolerance up to 75 µM $CdCl_2$. By increasing both the concentration of the sulfur pretreatment, and the $CdCl_2$ concentration that the pretreated cultures were exposed to, there is a chance that more sulfate-based proteins would be identified in the subsequent RNA-Seq and GO-term analysis. This would allow for a closer comparison to the work done by Mangal et al., 2022.

The connection between sulfur and increased Cd tolerance has been studied in various plants (Liang et al., 2016). The work in this thesis indicates that a pretreatment with sulfur increases the tolerance of *E. gracilis* to $CdCl_2$ similar to the impact it has on several plants. It has yet to be determined if the use of a sulfur pretreatment works specifically for increasing $CdCl_2$ tolerance, or if its impact can be seen when looking at other metals such as Pb, Cr, or As. In *Arabidopsis*, the use of hydrogen sulfide enhanced its tolerance to Cr (VI) by inducing cysteine accumulation (Fang et al., 2016). Hydrogen sulfide also improves the removal of silver from wastewater by *Phanerochaete*

chrysosporium (Huang et al., 2019). This data is indicative of two things 1) that sulfate is not the only source of sulfur that has been linked to an organism's increased tolerance to heavy metals and 2) that the benefits of sulfur on heavy metal tolerance is not reserved exclusively for Cd. This indicates that, while $MgSO_4*7H_2O$ was chosen for the sulfur pretreatment in this case because it is already in the growth media; thus, it is worth exploring other sources of sulfur such as hydrogen sulfide for sulfur pretreatments.

The work in this thesis can be expanded upon to determine effects of exposure to higher concentrations of $CdCl_2$. Similar to the work done by Avilés et al., 2003, where 100 µM of $CdCl_2$ was used, the nitrogen pretreated cultures showed a relative tolerance of upwards of 75 µM $CdCl_2$ based on the data in this study (Figure 4 C). Increasing the concentration of $CdCl_2$ could lead to different genes being upregulated or having the same genes being upregulated, just to a greater extent.

The potentially negative impact of NH_4NO_3 on growth as well as its impact on $CdCl_2$ tolerance could be explored further. The source of the nitrogen would have to be modified to either ammonium phosphate or amino acids. The use of amino acids as the nitrogen source for the pretreatment can also be hypothesized to increase the metal tolerance and metal binding since an amino acid, such as cysteine, is a precursor to glutathione, a known metal binding compound (Mangal et al, 2022, Khatiwada et al., 2020).

While the data from this thesis has opened more avenues for studying *E. gracilis* and its potential use in bioremediation, future studies into the transcript level changes are yet hindered by the lack of a completed and annotated genome sequence. Currently

researchers are working on creating a reference genome sequence for *E. gracilis* (Ebenezer et al., 2019). However, obtaining an annotated genome for *E. gracilis* has proven to be difficult due to its large size and repetitive sequence (Ebenezer et. al., 2019). With an annotated genome comes improved databases which are also needed for *E. gracilis*. In this thesis, GO-term analysis was completed using Ensembl databases of *A. thaliana*, *H. sapiens, C. reinhardtii*, *Synechocistis sp.*, and *T. brucei*. While these encompass organisms that are either similar to *E. gracilis* in some form or are well sequenced in the case of the *H. sapiens* database; the lack of a comprehensive database for *E. gracilis* limits one's ability to identify genes that are differentially expressed. Current efforts to obtain an annotated genome and to build databases for *E. gracilis* should be expanded to enhance future gene expression analysis, yielding more insight into *E. gracilis* and its ability to tolerate and bind metals. Combining the work done in this thesis with an annotated genome and better *E. gracilis* databases could lead to an ability to enhance its tolerance and binding of heavy metals. Understanding what genes showed altered transcript levels, combined with good sequence databases and a genome could allow for molecular manipulation of *E. gracilis*. In the future, this could allow for *E. gracilis* to be enhanced and then used as a bioremediation tool in areas such as mine tailing ponds.

CHAPTER SIX: CONCLUSION

As mining industries continue to expand, so does the pollution from their industrial wastewater. Mine tailing ponds are polluted with heavy metals such as Cd, which can have devasting consequences for both human and ecosystem health. To combat these negative effects, new methods in bioremediation are being investigated. Organisms such as *E. gracilis*, are being studied to find more viable options for cleaning industrial wastewaters. In this study, I examined the influence of sulfur and nitrogen on the Cd tolerance of *E. gracilis* and identified gene functions that were associated with changes in transcript levels.

Euglena gracilis has a good tolerance to heavy metals such as Cd, and this tolerance can be enhanced through the use of pretreatments (Avilés et al., 2003). It was hypothesized that this tolerance to metal is due to specific metal binding compounds that have been identified in *E. gracilis* and are found to be enriched with sulfur and nitrogen (Winters et al., 2021). I pretreated *E. gracilis* cells with either 60 mM sulfate using $MgSO_4*7H_2O$ or 76 mM of nitrogen using NH_4NO_3. Microscopy indicated that, at these concentrations, the cells remained viable and growth assays indicated no change in the growth of cells because of these pretreatments. The *E. gracilis* cultures that were either pretreated with sulfur or nitrogen exhibited a higher tolerance to $CdCl_2$ when compared to cultures that were only grown in MAM. The tolerance was determined through the increase in viable cells in the sulfur and nitrogen pretreated cultures when compared to cultures grown in MAM. This tolerance was maintained over a course of 8 days at a concentration of 25 µM of $CdCl_2$. Growth assays indicated that pretreated cultures could tolerate a higher concentration of $CdCl_2$ however, the number of viable cells grown in the non-pretreated cultures were depleting which would have negatively impacted future transcriptomic analysis. The

ability of *E. gracilis* to use sulfur and nitrogen in its media to enhance its tolerance to Cd builds upon previous work by Avilés et al., 2003 and Winters et al., 2021 and provides a base for the use of *E. gracilis* as a tool for bioremediation.

Using RNA-seq and subsequent GO term analysis, we were able to identify gene functions that are associated with altered transcript levels. The results indicated that the pretreatment with either sulfur or nitrogen as well as the exposure to $CdCl_2$ caused transcript level changes. In non-pretreated *E. gracilis* cultures, the exposure to $CdCl_2$ showed an upregulation of transcripts that were identified as having sequence similarities to chemotaxis proteins in the NCBI non-redundant protein database with GO-term analysis showing GO-term enrichment for positive regulation of proteasomal protein catabolic process and phosphotransferase activity carboxyl group as acceptor. The use of a sulfur and nitrogen pretreatment showed differential expression of transcripts that were identified as having sequence similarities to stress response proteins including chemotaxis proteins as well as heat shock proteins. The sulfur and nitrogen pretreated cultures also had downregulated transcripts that were identified as having a sequence similarity to the ABC transporter proteins. Along with stress response proteins, sulfur pretreated cultures identified enzymes related to cysteine metabolism that were differentially expressed after exposure of sulfur pretreated cultures to $CdCl_2$ with the GO-term cysteine metabolic process being enriched. In contrast, enzymes related to amino acid biosynthesis were identified such as threonine dehydratase and 2-isopropylmalate synthase which are responsible for the synthesis of amino acids such as leucine (de Kraker et al., 2007). Specifically with the nitrogen pretreated cultures, GO-term analysis also identified GO terms related to metal binding and metal ion binding. These included specific proteins such

as calmodulin, phospholipid-transporting ATPase, ascorbate peroxidase, serine/threonine kinase and chaperone proteins. These results indicated that the use of either a sulfur or a nitrogen pretreatment enhanced the tolerance of *E. gracilis* to $CdCl_2$. This was seen with altered transcript levels for genes potentially related to transmembrane transport, stress response, metabolic processes and metal and metal ion binding. These results are indicative that the tolerance of *E. gracilis* to $CdCl_2$ is impacted by multiple factors rather than just one specific mechanism.

Overall, the results from this thesis suggest that metal tolerance in *E. gracilis* is a dynamic process, that is potentially facilitated by a variety of pathways. The knowledge gained from studying the influence of sulfur and nitrogen on the tolerance of *E. gracilis* to $CdCl_2$ contributes to the understanding of the mechanisms that allow *E. gracilis* to tolerate and bind heavy metals. The knowledge gained from this study is not limited to *E. gracilis* but can potentially be used to investigate the metal tolerance of other algae and even other species of Euglena like *Euglena mutabilis*. The results from this thesis expand our knowledge of the metal binding process in *E.gracilis* , which could make this organism a better candidate for bioremediation in polluted wastewater such as mine tailing ponds.

REFERENCES

Adhikari, S., Ghosh, S., Azahar, I., Adhikari, A., Shaw, A. K., Konar, S., Roy, S., & Hossain, Z. (2018). Sulfate improves cadmium tolerance by limiting cadmium accumulation, modulation of sulfur metabolism and antioxidant defense system in maize. *Environmental and Experimental Botany*, *153*, 143–162. https://doi.org/10.1016/J.ENVEXPBOT.2018.05.008

Akthar, M. N., & Mohan, P. M. (1995). Bioremediation of toxic metal ions from polluted lake waters and industrial effluents by fungal biosorbent. In *Source: Current Science* (Vol. 69, Issue 12). https://about.jstor.org/terms

Anahid, S., Yaghmaei, S., & Ghobadinejad, Z. (2011). Heavy metal tolerance of fungi. *Scientia Iranica*, *18*(3), 502–508. https://doi.org/10.1016/J.SCIENT.2011.05.015

Avilés, C., Loza-Tavera, H., Terry, N., & Moreno-Sánchez, R. (2003). Mercury pretreatment selects an enhanced cadmium-accumulating phenotype in Euglena gracilis. *Archives of Microbiology*, *180*(1), 1–10. https://doi.org/10.1007/s00203-003-0547-2

Barque, J. P., Abahamid, A., Chacun, H., & Bonaly, J. (1996). Different Heat-Shock Proteins Are Constitutively Overexpressed in Cadmium and Pentachlorophenol AdaptedEuglena gracilisCells. *Biochemical and Biophysical Research Communications*, *223*(1), 7–11. https://doi.org/10.1006/BBRC.1996.0837

Berleman, J. E., & Bauer, C. E. (2005). Involvement of a Che-like signal transduction cascade in regulating cyst cell development in Rhodospirillum centenum. *Molecular Microbiology*, *56*(6), 1457–1466. https://doi.org/10.1111/j.1365-2958.2005.04646.x

Bernhoft, R. A. (2013). Cadmium Toxicity and Treatment. *The Scientific World Journal*, *2013*. https://doi.org/10.1155/2013/394652

Bolger, A. M., Lohse, M., & Usadel, B. (2014). *Genome analysis Trimmomatic: a flexible trimmer for Illumina sequence data*. *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Boopathy, R. (2000). Factors limiting bioremediation technologies. *Bioresource Technology*, *74*(1), 63–67. https://doi.org/10.1016/S0960-8524(99)00144-3

Borba, C. E., Guirardello, R., Silva, E. A., Veit, M. T., & Tavares, C. R. G. (2006). Removal of nickel(II) ions from aqueous solution by biosorption in a fixed bed column: Experimental and theoretical breakthrough curves. *Biochemical Engineering Journal*, *30*(2), 184–191. https://doi.org/10.1016/J.BEJ.2006.04.001

Boussen, S., Soubrand, M., Bril, H., Ouerfelli, K., & Abdeljaouad, S. (2013). Transfer of lead, zinc and cadmium from mine tailings to wheat (Triticum aestivum) in carbonated Mediterranean (Northern Tunisia) soils. *Geoderma*, *192*(1), 227–236. https://doi.org/10.1016/J.GEODERMA.2012.08.029

Bovet, L., Feller, U., & Martinoia, E. (2005). Possible involvement of plant ABC transporters in cadmium detoxification: A cDNA sub-microarray approach. *Environment International*, *31*(2), 263–267. https://doi.org/10.1016/j.envint.2004.10.011

Bren, A., & Eisenbach, M. (2000). How Signals Are Heard during Bacterial Chemotaxis: Protein-Protein Interactions in Sensory Signal Propagation. In *JOURNAL OF BACTERIOLOGY* (Vol. 182, Issue 24).

Choong, G., Liu, Y., & Templeton, D. M. (2014). Interplay of calcium and cadmium in mediating cadmium toxicity. *Chemico-Biological Interactions*, *211*(1), 54–65. https://doi.org/10.1016/J.CBI.2014.01.007

Das, P., Samantaray, S., & Rout, G. R. (1997). Studies on cadmium toxicity in plants: A review. *Environmental Pollution*, *98*(1), 29–36. https://doi.org/10.1016/S0269-7491(97)00110-3

Davies, J. P., Yildiz, F. H., & Grossman, A. R. (1999). Sac3, an Snf1-like Serine/Threonine Kinase That Positively and Negatively Regulates the Responses of Chlamydomonas to Sulfur Limitation. In *The Plant Cell* (Vol. 11). www.plantcell.org

de Kraker, J.-W., Luck, K., Textor, S., Tokuhisa, J. G., & Gershenzon, J. (n.d.). *Two Arabidopsis Genes (IPMS1 and IPMS2) Encode Isopropylmalate Synthase, the Branchpoint Step in the Biosynthesis of Leucine 1[W][OA]*. https://doi.org/10.1104/pp.106.085555

Devars, S., Avilés, C., Cervantes, C., & Moreno-Sánchez, R. (2000). Mercury uptake and removal by euglena gracilis. *Archives of Microbiology*, *174*(3), 175–180. https://doi.org/10.1007/s002030000193

Ding, N., Wang, L., Kang, Y., Luo, K., Zeng, D., Man, Y. B., Zhang, Q., Zeng, L., Luo, J., & Jiang, F. (2020). The comparison of transcriptomic response of green microalga Chlorella sorokiniana exposure to environmentally relevant concentration of cadmium(II) and 4-n-nonylphenol. *Environmental Geochemistry and Health*, *42*(9), 2881–2894. https://doi.org/10.1007/s10653-020-00526-1

Dugasa, M. T., Feng, · Xue, Wang, N.-H., Wang, J., & Feibo Wu, ·. (2021). Comparative transcriptome and tolerance mechanism analysis in the two contrasting wheat (Triticum aestivum L.) cultivars in response to drought and salinity stresses. *Plant Growth Regulation*, *94*, 101–114. https://doi.org/10.1007/s10725-021-00699-4

Ebenezer, T. E., Zoltner, M., Burrell, A., Nenarokova, A., Novák Vanclová, A. M. G., Prasad, B., Soukal, P., Santana-Molina, C., O'neill, E., Nankissoor, N. N., Vadakedath, N., Daiker, V., Obado, S., Silva-Pereira, S., Jackson, A. P., Devos, D. P., Lukeš, J., Lebert, M., Vaughan, S., … Field, M. C. (n.d.). *Transcriptome, proteome and draft genome of Euglena gracilis*. https://doi.org/10.1186/s12915-019-0626-8

Fang, H., Liu, Z., Jin, Z., Zhang, L., Liu, D., & Pei, Y. (2016). An emphasis of hydrogen sulfide-cysteine cycle on enhancing the tolerance to chromium stress in Arabidopsis. *Environmental Pollution*, *213*, 870–877. https://doi.org/10.1016/J.ENVPOL.2016.03.035

Feng, S., Hou, S., Cui, Y., Tong, Y., & Yang, H. (2020). Metabolic transcriptional analysis on copper tolerance in moderate thermophilic bioleaching microorganism Acidithiobacillus caldus. *Journal of Industrial Microbiology and Biotechnology*, *47*(1), 21–33. https://doi.org/10.1007/s10295-019-02247-6

Fu, F., & Wang, Q. (2011). Removal of heavy metal ions from wastewaters: A review. *Journal of Environmental Management*, *92*(3), 407–418. https://doi.org/10.1016/J.JENVMAN.2010.11.011

Gabarrón, M., Faz, A., Martínez-Martínez, S., & Acosta, J. A. (2018). Change in metals and arsenic distribution in soil and their bioavailability beside old tailing ponds. *Journal of Environmental Management*, *212*, 292–300. https://doi.org/10.1016/J.JENVMAN.2018.02.010

García-García, J. D., Olin-Sandoval, V., Saavedra, E., Girard, L., Hernández, G., & Moreno-Sánchez, R. (2012). Sulfate uptake in photosynthetic Euglena gracilis. Mechanisms of regulation and contribution to cysteine homeostasis. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1820*(10), 1567–1575. https://doi.org/10.1016/J.BBAGEN.2012.05.002

Giordano, M., Norici, A., & Hell, R. (2005). Sulfur and Phytoplankton: Acquisition, Metabolism and Impact on the Environment. In *Source: The New Phytologist* (Vol. 166, Issue 2).

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Author, N. B. (2013). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data HHS Public Access Author manuscript. *Nat Biotechnol*, *29*(7), 644–652. https://doi.org/10.1038/nbt.1883

Guimarães-Soares, L., Felícia, H., João Bebianno, M., & Cássio, F. (2006). Metal-binding proteins and peptides in the aquatic fungi Fontanospora fusiramosa and Flagellospora curta

exposed to severe metal stress. *Science of The Total Environment*, *372*(1), 148–156. https://doi.org/10.1016/J.SCITOTENV.2006.09.017

Heckathorn, S. A., Mueller, J. K., Laguidice, S., Zhu, B., Barrett, T., Blair, B., & Dong, Y. (2004). Chloroplast small heat‐shock proteins protect photosynthesis during heavy metal stress <link href="#FN1">1</link>; Chloroplast small heat‐shock proteins protect photosynthesis during heavy metal stress <link href="#FN1">1</link>. In *American Journal of Botany* (Vol. 91, Issue 9). Wollgiehn and Neumann. https://doi.org/10.3732/ajb.91.9.1312

Hernández, L. E., Gárate, A., & Carpena-Ruiz, R. (1997). Effects of cadmium on the uptake, distribution and assimilation of nitrate in Pisum sativum. In *Plant and Soil* (Vol. 189). Kluwer Academic Publishers.

Huang, Z., He, K., Song, Z., Zeng, G., Chen, A., Yuan, L., Li, H., & Chen, G. (2019). Alleviation of heavy metal and silver nanoparticle toxicity and enhancement of their removal by hydrogen sulfide in Phanerochaete chrysosporium. *Chemosphere*, *224*, 554–561. https://doi.org/10.1016/J.CHEMOSPHERE.2019.02.190

Jacob, C., Courbot, M., Martin, F., Brun, A., & Chalot, M. (2004). Transcriptomic responses to cadmium in the ectomycorrhizal fungus Paxillus involutus. *FEBS Letters*, *576*(3), 423–427. https://doi.org/10.1016/J.FEBSLET.2004.09.028

Khademi, H., Abbaspour, A., Martínez-Martínez, S., Gabarrón, M., Shahrokh, V., Faz, A., & Acosta, J. A. (2018). Provenance and environmental risk of windblown materials from mine tailing ponds, Murcia, Spain. *Environmental Pollution*, *241*, 432–440. https://doi.org/10.1016/J.ENVPOL.2018.05.084

Khatiwada, B., Hasan, M. T., Sun, A., Kamath, K. S., Mirzaei, M., Sunna, A., & Nevalainen, H. (2020). Proteomic response of Euglena gracilis to heavy metal exposure – Identification of key proteins involved in heavy metal tolerance and accumulation. *Algal Research*, *45*, 101764. https://doi.org/10.1016/J.ALGAL.2019.101764

Kim, M.-J., & Jung, Y. (2004). Vertical Distribution and Mobility of Arsenic and Heavy Metals in and around Mine Tailings of an Abandoned Mine. *Journal of Environmental Science and Health, Part A*, *39*(1), 203–222. https://doi.org/10.1081/ESE-120027379

Kolahi, M., Mohajel Kazemi, E., Yazdi, M., & Goldson-Barnaby, A. (2020). Oxidative stress induced by cadmium in lettuce (Lactuca sativa Linn.): Oxidative stress indicators and prediction of their genes. *Plant Physiology and Biochemistry*, *146*, 71–89. https://doi.org/10.1016/J.PLAPHY.2019.10.032

Kolber, Z., Zehr, J., & Falkowski, P. (1988). Effects of Growth Irradiance and Nitrogen Limitation on Photosynthetic Energy Conversion in Photosystem I1l. In *Plant Physiol* (Vol. 88). https://academic.oup.com/plphys/article/88/3/923/6083407

Kwiatowski, J. M., Sáez, A. G., Karnkowska, A., De, C. E., Bicudo, M., & Menezes, M. (2016). Phylogeny and Classification of Euglenophyceae: A Brief Review. *Phylogeny and Classification of Euglenophyceae: A Brief Review. Front. Ecol. Evol*, *4*, 17. https://doi.org/10.3389/fevo.2016.00017

Li, B., & Dewey, C. N. (2011). *RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome*. https://doi.org/10.1186/1471-2105-12-323

Liang, T., Ding, H., Wang, G., Kang, J., Pang, H., & Lv, J. (2016). Sulfur decreases cadmium translocation and enhances cadmium tolerance by promoting sulfur assimilation and glutathione metabolism in Brassica chinensis L. *Ecotoxicology and Environmental Safety*, *124*, 129–137. https://doi.org/10.1016/J.ECOENV.2015.10.011

Liu, L., Zhou, L., Hu, S., Zhou, S., Deng, Y., Dong, M., Huang, J., Zeng, Y., Chen, X., Zhao, N., Li, H., & Ding, Z. (2016). Down-regulation of ABCG2 and ABCB4 transporters in the placenta of rats exposed to cadmium. *Oncotarget*, *7*(25), 38154–38163. https://doi.org/10.18632/oncotarget.9415

Locher, K. P. (n.d.). *Mechanistic diversity in ATP-binding cassette (ABC) transporters*. https://doi.org/10.3929/ethz-b-000117629

Mandal, S. K., Adhikari, R., Sharma, A., Chandravanshi, M., Gogoi, P., & Kanaujia, S. P. (2019). Designating ligand specificities to metal uptake ABC transporters in: Thermus thermophilus HB8. *Metallomics*, *11*(3), 597–612. https://doi.org/10.1039/C8MT00374B

Mangal, V., Donaldson, M. E., Lewis, A., Saville, B. J., & Guéguen, C. (2022). Identifying Euglena Gracilis Metabolic and Transcriptomic Adaptations in Response to Mercury Stress. *Frontiers in Environmental Science*, *10*. https://doi.org/10.3389/fenvs.2022.836732

Mangal, V., & Guéguen, C. (2015). Examining concentrations and molecular weights of thiols in microorganism cultures and in Churchill River (Manitoba) using a fluorescent-labeling method coupled to asymmetrical flow field-flow fractionation. *Analytical and Bioanalytical Chemistry*, *407*(15). https://doi.org/10.1007/s00216-015-8599-0

Mastalerz, M., & Drobniak, A. (2007). Arsenic, cadmium, lead, and zinc in the Danville and Springfield coal members (Pennsylvanian) from Indiana. *International Journal of Coal Geology*, *71*(1), 37–53. https://doi.org/10.1016/J.COAL.2006.05.005

Matović, V., Buha, A., Bulat, Z., & Dukić-ĆOsić, D. (2011). Cadmium toxicity revisited: Focus on oxidative stress induction and interactions with zinc and magnesium. In *Arhiv za Higijenu Rada i Toksikologiju* (Vol. 62, Issue 1, pp. 65–76). https://doi.org/10.2478/10004-1254-62-2011-2075

Matzov, D., Taoka, M., Nobe, Y., Yamauchi, Y., Halfon, Y., Asis, N., Zimermann, E., Rozenberg, H., Bashan, A., Bhushan, S., Isobe, T., Gray, M. W., Yonath, A., & Shalev-Benami, M. (2020). Cryo-EM structure of the highly atypical cytoplasmic ribosome of Euglena gracilis. *Nucleic Acids Research*, *48*(20), 11750–11761. https://doi.org/10.1093/nar/gkaa893

Mendoza-Cozatl, D., Devars, S., Loza-Tavera, H., & Moreno-Sánchez, R. (2002). *Cadmium accumulation in the chloroplast of Euglena gracilis*.

Mendoza-Cózatl, D. G., & Moreno-Sánchez, R. (2005). Cd2+ transport and storage in the chloroplast of Euglena gracilis. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1706*(1–2), 88–97. https://doi.org/10.1016/J.BBABIO.2004.09.010

Mendoza-Cózatl, D., Loza-Tavera, H., Hernández-Navarro, A., & Moreno-Sánchez, R. (n.d.). *Sulfur assimilation and glutathione metabolism under cadmium stress in yeast, protists and plants*. https://doi.org/10.1016/j.femsre.2004.09.004

MERAKLI, N., BULDUK, İ., & MEMON, A. (2021). CU-INDUCED GENES EXPRESSION IN THE ACCUMULATOR PLANT BRASSICA NIGRA L. *Trakya University Journal of Natural Sciences*. https://doi.org/10.23902/trkjnat.978842

Mobin, M., & Khan, N. A. (2007). Photosynthetic activity, pigment composition and antioxidative response of two mustard (Brassica juncea) cultivars differing in photosynthetic capacity subjected to cadmium stress. *Journal of Plant Physiology*, *164*(5), 601–610. https://doi.org/10.1016/J.JPLPH.2006.03.003

Muhammad, N., Nafees, M., Ge, L., Khan, M. H., Bilal, M., Chan, W. P., & Lisak, G. (2021). Assessment of industrial wastewater for potentially toxic elements, human health (dermal) risks, and pollution sources: A case study of Gadoon Amazai industrial estate, Swabi, Pakistan. *Journal of Hazardous Materials*, *419*, 126450. https://doi.org/10.1016/J.JHAZMAT.2021.126450

Muhammad, T., Zhang, J., Ma, Y., Li, Y., Zhang, F., Zhang, Y., & Liang, Y. (n.d.). *molecules Overexpression of a Mitogen-Activated Protein Kinase SlMAPK3 Positively Regulates Tomato Tolerance to Cadmium and Drought Stress*. https://doi.org/10.3390/molecules24030556

Nikolaidis, C., Zafiriadis, I., Mathioudakis, V., & Constantinidis, T. (2010). Heavy metal pollution associated with an abandoned lead-zinc mine in the Kirki Region, NE Greece.

Bulletin of Environmental Contamination and Toxicology, *85*(3), 307–312. https://doi.org/10.1007/s00128-010-0079-9

Norici, A., Hell, R., & Giordano, M. (2005). Sulfur and primary production in aquatic environments: An ecological perspective. In *Photosynthesis Research* (Vol. 86, Issue 3, pp. 409–417). https://doi.org/10.1007/s11120-005-3250-0

O'Neill, E. C., Trick, M., Hill, L., Rejzek, M., Dusi, R. G., Hamilton, C. J., Zimba, P. v., Henrissat, B., & Field, R. A. (2015). The transcriptome of Euglena gracilis reveals unexpected metabolic capabilities for carbohydrate and natural product biochemistry. *Molecular BioSystems*, *11*(10), 2808–2820. https://doi.org/10.1039/c5mb00319a

Owa, F. D. (2013). Water pollution: Sources, effects, control and management. *Mediterranean Journal of Social Sciences*, *4*(8), 65–68. https://doi.org/10.5901/mjss.2013.v4n8p65

Paul, K., Brunstetter, D., Titen, S., & Blair, D. F. (2011). A molecular mechanism of direction switching in the flagellar motor of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(41), 17171–17176. https://doi.org/10.1073/pnas.1110111108

Pinto, A. P., Mota, A. M., de Varennes, A., & Pinto, F. C. (2004). Influence of organic matter on the uptake of cadmium, zinc, copper and iron by sorghum plants. *Science of The Total Environment*, *326*(1–3), 239–247. https://doi.org/10.1016/J.SCITOTENV.2004.01.004

Purchase, D., Scholes, L. N. L., Revitt, D. M., & Shutes, R. B. E. (2009). Effects of temperature of metal tolerance and the accumulation of Zn and Pb by metal-tolerant fungi isolated from urban runoff treatment wetlands. *Journal of Applied Microbiology* , *106*(4), 1163–1174.

Rakhshaee, R., Giahi, M., & Pourahmad, A. (2009). Studying effect of cell wall's carboxyl–carboxylate ratio change of Lemna minor to remove heavy metals from aqueous solution. *Journal of Hazardous Materials*, *163*(1), 165–173. https://doi.org/10.1016/J.JHAZMAT.2008.06.074

Ramaiah, A., & Shanmugasundaram, E. R. B. (1962). Effect of sulphur compounds on the uptake of molybdenum by Neurospora crassa. *Biochimica et Biophysica Acta*, *60*(2), 386–392. https://doi.org/10.1016/0006-3002(62)90414-6

Richter, P. R., Liu, Y., An, Y., Li, X., Nasir, A., Strauch, S. M., Becker, I., Krüger, J., Schuster, M., Ntefidou, M., Daiker, V., Haag, F. W. M., Aiach, A., & Lebert, M. (2015). Amino acids as possible alternative nitrogen source for growth of Euglena gracilis Z in life support systems. *Life Sciences in Space Research*, *4*, 1–5. https://doi.org/10.1016/J.LSSR.2014.11.001

Riechers, D. E., Romero-Puertas, M. C., Moenne, A., González, A., & Laporte, D. (2021). *Cadmium Accumulation Involves Synthesis of Glutathione and Phytochelatins, and Activation of CDPK, CaMK, CBLPK, and MAPK Signaling Pathways in Ulva compressa*. https://doi.org/10.3389/fpls.2021.669096

Rodriguez-Couto, S., Pienkos, P. T., Sunna, A., Gissibl, A., Sun, A., Care, A., & Nevalainen, H. (2019). Bioproducts From Euglena gracilis: Synthesis and Applications. *Frontiers in Bioengineering and Biotechnology | Www.Frontiersin.Org*, *1*, 108. https://doi.org/10.3389/fbioe.2019.00108

Rodríguez-Zavala, J. S., García-García, J. D., Ortiz-Cruz, M. A., Moreno-Sánchez, R., Rodríguez-Zavala, S., Moreno, R., Jos´, J., Rodr´iguez, J. S., Rodr´iguez-Zavala, R., Garc´ia, J. D., Garc´ia-Garc´ia, G., Garc´ia, G., Moreno-S, R., & Anchez, ´. (2007). Journal of Environmental Science and Health, Part A Toxic/Hazardous Substances and Environmental Engineering Molecular mechanisms of resistance to heavy metals in the protist Euglena gracilis Molecular mechanisms of resistance to heavy metals in the protist Euglena gracilis. *Journal of Environmental Science and Health, Part A*, *42*(10), 1365–1378. https://doi.org/10.1080/10934520701480326

Saidha, T., Na, S.-Q., Li, J., & Schiff, J. A. (1988). A sulphate metabolizing centre in Euglena mitochondria. In *Biochem. J* (Vol. 253).

Sambrook, J., & Russell, D. W. (2001). *Molecular Cloning: A Laboratory Manual* (Third, Vol. 1). Cold Spring Harbor Laboratory Press .

Sanità Di Toppi, L., & Gabbrielli, R. (1999). Response to cadmium in higher plants. In *Environmental and Experimental Botany* (Vol. 41).

Sarkar, M. K., Paul, K., & Blair, D. (2010). Chemotaxis signaling protein CheY binds to the rotor protein FliN to control the direction of flagellar rotation in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(20), 9370–9375. https://doi.org/10.1073/pnas.1000935107

Schantz, M.-L., & Schantz, R. (1989). Sequence of a cDNA clone encoding /} tubnlin from Eugiena gracUis. In *Nucleic Acids Research* (Vol. 17). https://academic.oup.com/nar/article/17/16/6727/1074728

Shabbir, S., Wang, W., Chen, M., Wu, B., Nawaz, M., Liu, P., Dai, Y., Gou, Q., Liu, R., & Hu, G. (2021). *Molecular Mechanism of Engineered Zymomonas Mobilis Response to Furfural and Acetic acid Stress*. https://doi.org/10.21203/rs.3.rs-629707/v1

Sharma, D., Garg, A., Kumar, M., Rashid, F., & Khan, A. U. (2019). Down-Regulation of Flagellar, Fimbriae, and Pili Proteins in Carbapenem-Resistant Klebsiella pneumoniae (NDM-4) Clinical Isolates: A Novel Linkage to Drug Resistance. *Frontiers in Microbiology*, *10*. https://doi.org/10.3389/fmicb.2019.02865

Song, W.-Y., Park, J., Eisenach, C., Maeshima, M., Lee, Y., & Martinoia, E. (2014). *ABC Transporters and Heavy Metals* (pp. 1–17). https://doi.org/10.1007/978-3-319-06511-3_1

Stout, L., & Nüsslein, K. (2010). Biotechnological potential of aquatic plant–microbe interactions. *Current Opinion in Biotechnology*, *21*(3), 339–345. https://doi.org/10.1016/J.COPBIO.2010.04.004

Tessier, L. H., Keller, M., Chan, R. L., Fournier, R., Weil, J. H., & Imbault, P. (1991). Short leader sequences may be transferred from small RNAs to pre-mature mRNAs by trans-splicing in Euglena. *EMBO Journal*, *10*(9), 2621–2625. https://doi.org/10.1002/j.1460-2075.1991.tb07804.x

Volesky, B. (2001). Detoxification of metal-bearing effluents: biosorption for the next century. *Hydrometallurgy*, *59*(2–3), 203–216. https://doi.org/10.1016/S0304-386X(00)00160-2

Wang, B., Li, Y., Shao, C., Tan, Y., & Cai, L. (2012). Cadmium and Its Epigenetic Effects. *Current Medicinal Chemistry*, *19*(16), 2611–2620. https://doi.org/10.2174/092986712800492913

Wang, R., Wu, J., Deng, X., Liu, D., & Yan, Y. (2018). Drought-responsive protein identification in developing grains of a wheat-Haynaldia villosa 6VS/6AL translocation line. *Crop and Pasture Science*, *69*(12), 1182–1196. https://doi.org/10.1071/CP18303

Waterhouse, R. M., Seppey, M., Sim~, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. v, Zdobnov, E. M., & Rosenberg, M. (n.d.). *BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics*. https://doi.org/10.1093/molbev/msx319

Weyens, N., van der Lelie, D., Taghavi, S., & Vangronsveld, J. (2009). Phytoremediation: plant–endophyte partnerships take the challenge. *Current Opinion in Biotechnology*, *20*(2), 248–254. https://doi.org/10.1016/J.COPBIO.2009.02.012

Wheeler, A. E., Zingaro, R. A., Irgolic, K., & Bottino, N. R. (1982). The effect of selenate, selenite, and sulfate on the growth of six unicellular marine algae. *Journal of Experimental Marine Biology and Ecology*, *57*(2–3), 181–194. https://doi.org/10.1016/0022-0981(82)90191-5

Winters, C., Guéguen, C., & Noble, A. (2017). Equilibrium and kinetic studies of Cu(II) and Ni(II) sorption on living Euglena gracilis. *Journal of Applied Phycology*, *29*(3), 1391–1398. https://doi.org/10.1007/s10811-016-1040-z

Winters, C., Mangal, V., Lewis, A., Gueguen, C., Noble, A., & Long, A. (2021). *Methods and uses of encapsulated exudates and dried euglena biomass for minding metal* (Patent No. US20210292192A1).

World Health Organization. (2022, March 21). *Drinking Water* .

Wright, I. A., McCarthy, B., Belmer, N., & Price, P. (2015). Subsidence from an Underground Coal Mine and Mine Wastewater Discharge Causing Water Pollution and Degradation of Aquatic Ecosystems. *Water, Air, and Soil Pollution*, *226*(10). https://doi.org/10.1007/s11270-015-2598-9

Yang, H., Xiong, Z., Xu, Z., & Liu, R. (2021). Interactive Effects of Lanthanum and Calcium on Cadmium Accumulation in Wheat with Special Reference to TaNramp5 Expression Regulated by Calmodulin. *Cite This: J. Agric. Food Chem*, *69*, 6878. https://doi.org/10.1021/acs.jafc.1c00365

Yoshida, Y., Tomiyama, T., Maruta, T., Tomita, M., Ishikawa, T., & Arakawa, K. (2016). *De novo assembly and comparative transcriptome analysis of Euglena gracilis in response to anaerobic conditions*. https://doi.org/10.1186/s12864-016-2540-6

Yotsova, E., Dobrikova, A., Stefanov, M., Misheva, S., Bardáčová, M., Matušíková, I., Žideková, L., Blehová, A., & Apostolova, E. (2020). Effects of cadmium on two wheat cultivars depending on different nitrogen supply. *Plant Physiology and Biochemistry*, *155*, 789–799. https://doi.org/10.1016/J.PLAPHY.2020.06.042

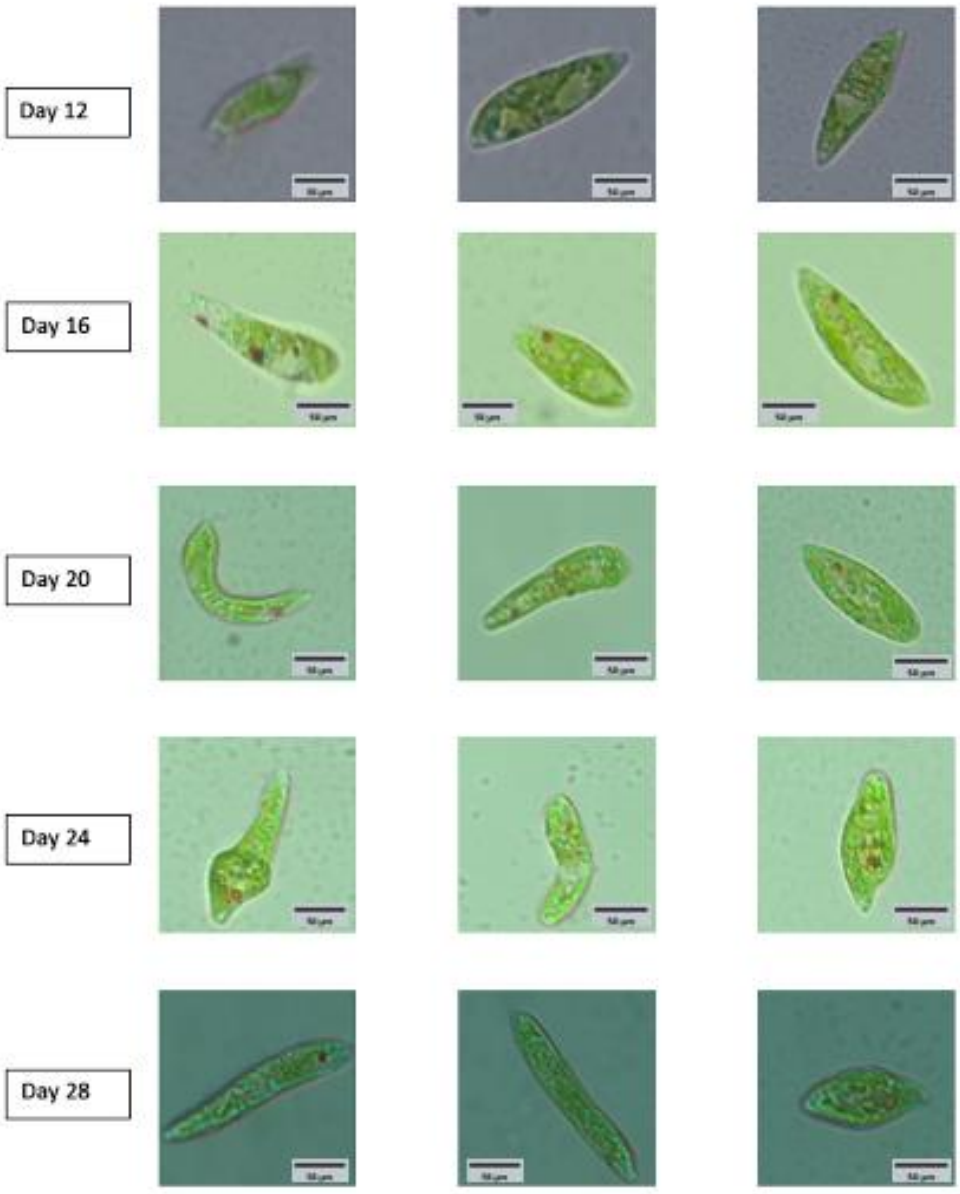
Yung, M. C., Ma, J., Salemi, M. R., Phinney, B. S., Bowman, G. R., & Jiao, Y. (2014). Shotgun proteomic analysis unveils survival and detoxification strategies by caulobacter crescentus during exposure to uranium, chromium, and cadmium. *Journal of Proteome Research*, *13*(4), 1833–1847. https://doi.org/10.1021/pr400880s

Zhang, H., Forman, H. J., & Choi, J. (2005). γ-Glutamyl Transpeptidase in Glutathione Biosynthesis. *Methods in Enzymology*, *401*, 468–483. https://doi.org/10.1016/S0076-6879(05)01028-1

Zhang, L., Thomas, A. E., Ae, H., & Melis, A. (n.d.). *Biochemical and morphological characterization of sulfur-deprived and H 2-producing Chlamydomonas reinhardtii (green alga)*. https://doi.org/10.1007/s004250100660

Zhao, X., Oll B, W. H. H., & Yun, G. (2002). Elimination of cadmium trace contaminations from drinking water. In *Water Research* (Vol. 36).

SUPPLEMENTARY MATERIALS



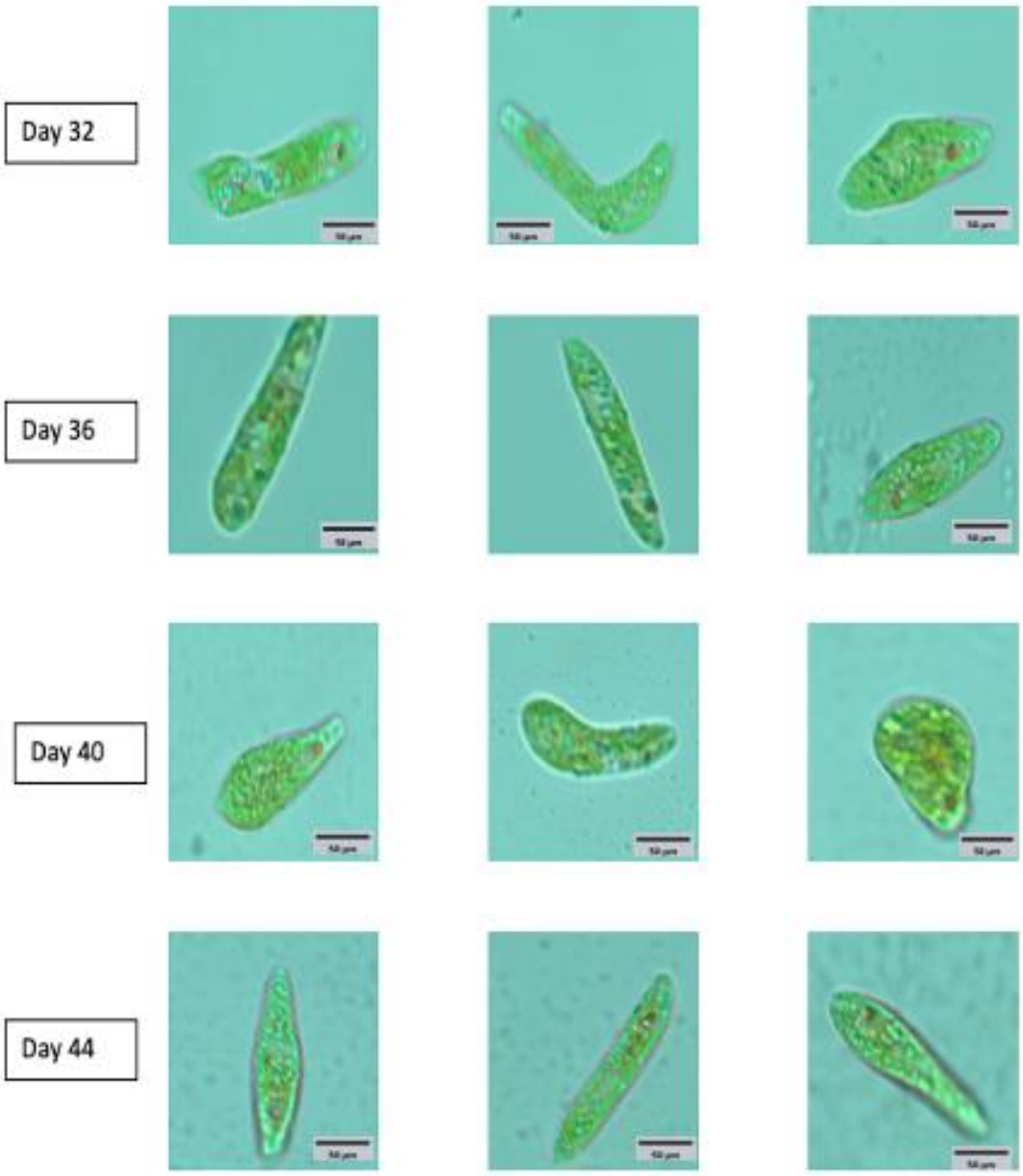
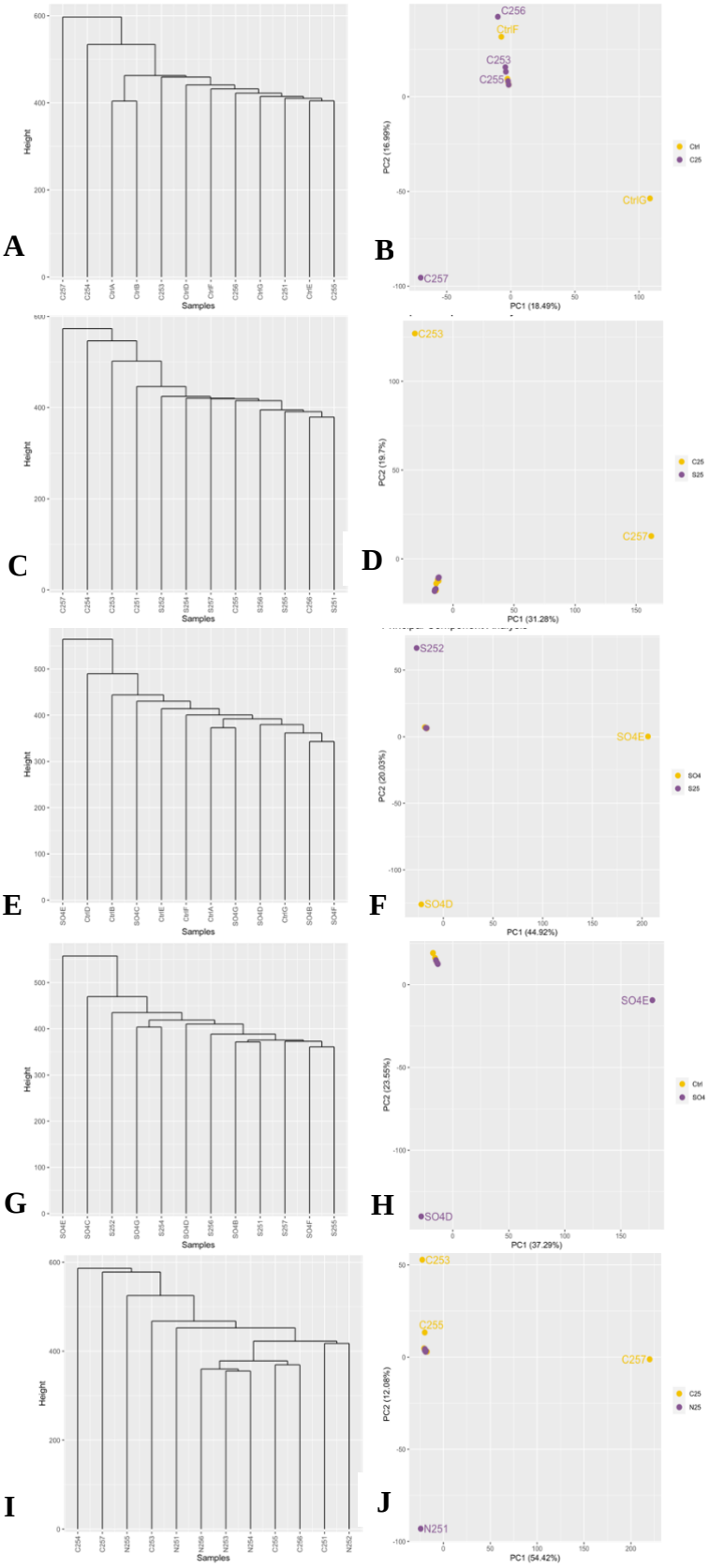


Figure S1: *Euglena gracilis* cultures grown in MAM, MAM+ 60 mM $MgSO_4*7H_2O$ and MAM + 76 mM NH_4NO_3. Cultures were grown for a period of 44 days and were centrifuged, and the entire volume was transferred into fresh media every 4 days. Photos were taken at 40x magnification. Scale bar represents 50 µm.


A
B
C
D
E
F
G
H
I
J
Height
Samples
PC1 (18.49%)
PC2 (16.69%)
PC1 (31.28%)
PC2 (19.7%)
PC1 (44.92%)
PC2 (20.03%)
PC1 (37.29%)
PC2 (23.55%)
PC1 (54.42%)
PC2 (12.08%)
C256
CtrlF
C253
C255
CtrlG
C257
S252
SO4E
SO4D
N251
Ctrl
C25
S25
SO4
N25

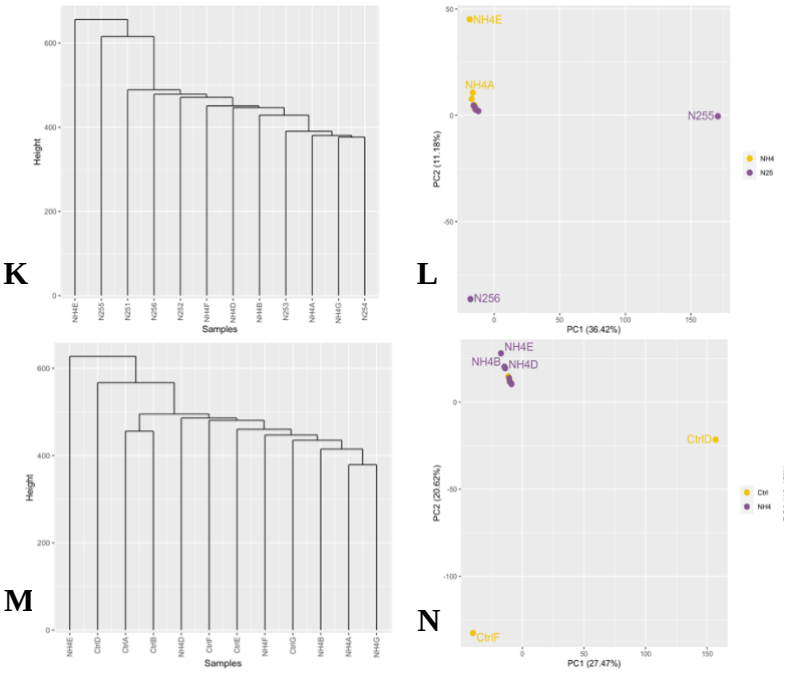


Figure S2: Clustering analysis showed no separation between either non-pretreated, sulfur pretreated and nitrogen pretreated *Euglena gracilis* cultures with or without $CdCl_2$.A-B) Clustering analysis done using a dendrogram (A) and principal component analysis (B), (Ctrl = yellow circles, Ctrl + 25 µM $CdCl_2$ = purple circles). **C-D)** dendrogram (C) and principal component analysis (D), (Ctrl + 25 µM $CdCl_2$= yellow circles, Sulfur + 25 µM $CdCl_2$ = purple circles). **E-F)** dendrogram (E) amd principal component analysis (F), (Sulfur = yellow circles, Sulfur + 25 µM $CdCl_2$ = purple circles). **G-H)** dendrogam (G) and principal component analysis (H), (Ctrl = yellow circles, Sulfur = purple circles). **I-J)** dendrogram (I) and principal component analysis (J), (Ctrl + 25 µM $CdCl_2$ = yellow circles, Nitrogen + 25 µM $CdCl_2$ = purple circles). **K-L)** dendrogram (K) and principal component analysis (L), (Nitrogen= yellow circles, Nitrogen+ 25 µM $CdCl_2$ = purple circles). **M-N)** dendrogram (M) and principal component analysis (N), (Ctrl= yellow circles, Nitrogen = purple circles).

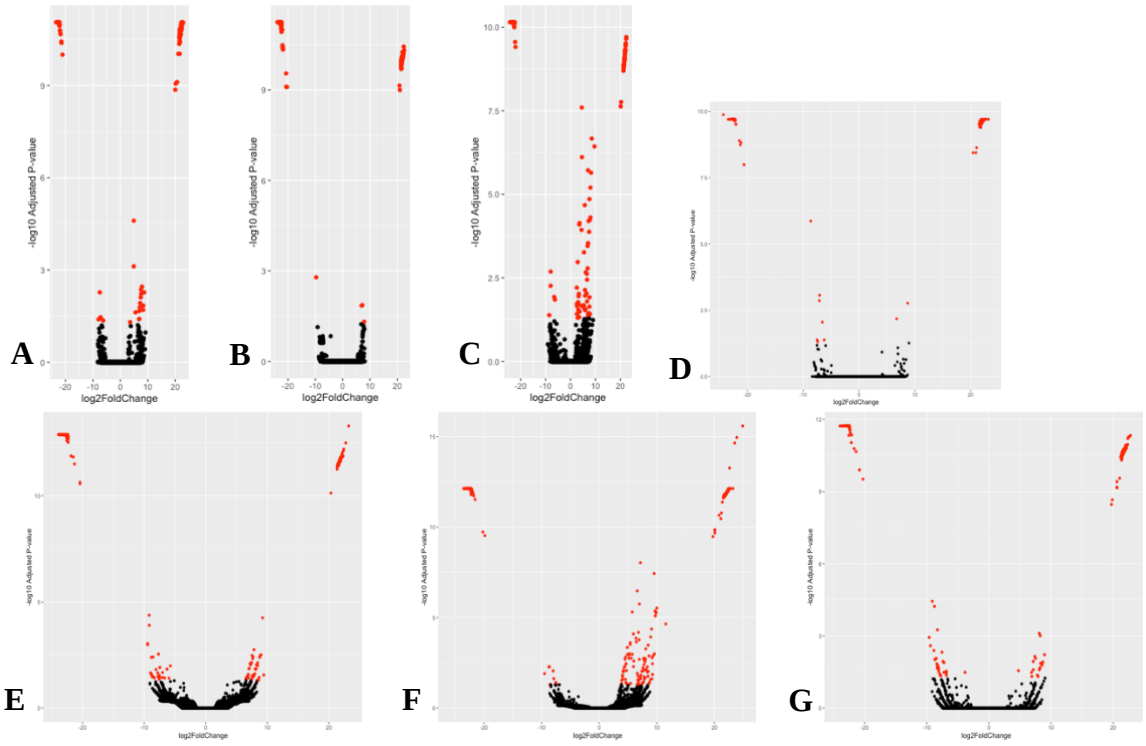


Figure S3: DGE analysis of non-pretreated, sulfur pretreated and nitrogen pretreated *Euglena gracilis* cultures grown in the presence and absence of cadmium shows a difference in transcript levels. DESEq2 was used at a significance level of 0.05 to test for differential transcript expression (>2 or <-2) across non-pretreated, sulfur pretreated and nitrogen pretreated *E. gracilis* cultures with and without$CdCl_2$. A-G) volcano plots comparing the -log10 adjusted p-value (y-axis) to the log2 fold change (x- axis) between A) Non-pretreated cultures with and without $CdCl_2$, B) Sulfur pretreated cultures vs non-pretreated cultures with $CdCl_2$, C) Sulfur pretreated cultures with and without $CdCl_2$, D) Sulfur pretreated cultures vs non-pretreated cultures, E) Nitrogen pretreated cultures vs Non-pretreated cultures both with $CdCl_2$, F) Nitrogen pretreated cultures in the presence and absence of $CdCl_2$, and G) Nitrogen pretreated cultures vs non-pretreated cultures. Red dots indicate genes that are differentially expressed and are statistically significant, while black dots indicate genes whose expression was not significantly different.

Table S1: List of sample abbreviations sent to the Centre for Applied Genomics with the corresponding pretreatment and $CdCl_2$ concentration.

Sample Name	Pretreatment	Cd Concentration
Ctrl A-G	MAM	0 µM
C25 1-7	MAM	25 µM
S A-G	MAM + $MgSO_4*7H_2O$	0 µM
S25 1-7	MAM +$MgSO_4*7H_2O$	25 µM
N A-G	MAM + NH_4NO_3	0 µM
N25 1-7	MAM +NH_4NO_3	25 µM

Table S2: List of abbreviations used during differential gene expression analysis with DESeq2.

Comparison	Abbreviation
E. gracilis in MAM in the presence and absence of $CdCl_2$	C25 vs Ctrl
Sulfur pretreated *E. gracilis* compared to *E. gracilis*, both in the presence of $CdCl_2$	S25 vs C25
Sulfur pretreated *E. gracilis* in the presence and absence of $CdCl_2$	S25 vs SO_4

Sulfur pretreated cultures vs *E. gracilis* in MAM	SO_4 vs Ctrl
Nitrogen pretreated compared to *E. gracilis* cells in MAM, both exposed to $CdCl_2$	N25 vs C25
Nitrogen pretreated *E. gracilis* in the presence and absence of $CdCl_2$	N25 vs NH_4
Nitrogen pretreated cultures vs *E. gracilis* in MAM	NH_4 vs Ctrl

Table S3: blastx results of differentially expressed transcripts defined by DESeq2 for non-pretreated *Euglena gracilis* cultures in the presence and absence of $CdCl_2$ using the NCBI non-redundant protein database.

NR_Isoform Id	NR_sseqid	NR_evalue	NR_description
TRINITY_DN200360_c0_g1_i1	WP_162809807.1	7.69E-38	chemotaxis protein, partial [Vibrio cholerae]
TRINITY_DN207373_c0_g1_i1	WP_008450649.1	4.85E-33	acetate kinase [Prevotella amnii]
TRINITY_DN270125_c0_g1_i1	EPI54498.1	6.16E-45	dTMP kinase domain protein, partial [Gardnerella vaginalis JCP7276]
TRINITY_DN238436_c0_g1_i1	WP_164799065.1	1.85E-16	peptidase [Lactobacillus iners]
TRINITY_DN281312_c0_g1_i1	DAM60650.1	2.38E-39	TPA: MAG TPA: Integrase [Herelleviridae sp.]
TRINITY_DN231511_c0_g1_i1	WP_201775009.1	1.20E-49	Holliday junction branch migration DNA helicase RuvB [Fannyhessea vaginae]
TRINITY_DN295572_c0_g1_i1	RIY23104.1	4.99E-17	peptidase, partial [Bifidobacteriaceae bacterium NR021]
TRINITY_DN236134_c0_g1_i1	EPI52247.1	1.92E-06	hypothetical protein HMPREF1575_00740 [Gardnerella vaginalis JCP7672]
TRINITY_DN192008_c0_g1_i1	MBS7185617.1	2.95E-35	phosphoglycerate kinase [Clostridium sp.]
TRINITY_DN180686_c0_g1_i1	EPI55698.1	1.71E-39	hypothetical protein HMPREF1573_01091, partial [Gardnerella vaginalis JCP7276]
TRINITY_DN158784_c2_g1_i1	CEG80874.1	1.31E-06	Putative Recombinational repair protein [Rhizopus 88lpine88ores]
TRINITY_DN109450_c0_g1_i1	PSC67172.1	4.37E-19	cell division cycle 20 [Micractinium conductrix]
TRINITY_DN157530_c0_g1_i1	WP_234805216.1	3.46E-31	hypothetical protein, partial [Salmonella enterica]
TRINITY_DN182881_c0_g1_i1	EFO70065.1	8.15E-46	proline—tRNA ligase [Lactobacillus iners LactinV 01V1-a]
TRINITY_DN129540_c0_g1_i1	KAF0135724.1	1.12E-06	hypothetical protein FD153_2052, partial [Rhodospirillaceae bacterium]

TRINITY_DN129540_c0_g1_i10	KAF0135724.1	9.99E-08	hypothetical protein FD153_2052, partial [Rhodospirillaceae bacterium]
TRINITY_DN129540_c0_g1_i11	MBC7635469.1	4.24E-35	integrase arm-type DNA-binding domain-containing protein [Acetobacteraceae bacterium]
TRINITY_DN129540_c0_g1_i12	MBC7635469.1	1.96E-25	integrase arm-type DNA-binding domain-containing protein [Acetobacteraceae bacterium]
TRINITY_DN129540_c0_g1_i13	OYV50480.1	4.87E-118	hypothetical protein B7Z77_05760 [Acidocella sp. 20-58-15]
TRINITY_DN129540_c0_g1_i15	KAF0135724.1	3.70E-08	hypothetical protein FD153_2052, partial [Rhodospirillaceae bacterium]
TRINITY_DN129540_c0_g1_i2	KAF0135724.1	1.07E-07	hypothetical protein FD153_2052, partial [Rhodospirillaceae bacterium]
TRINITY_DN129540_c0_g1_i3	KAF0135724.1	2.66E-08	hypothetical protein FD153_2052, partial [Rhodospirillaceae bacterium]
TRINITY_DN129540_c0_g1_i4	MBC7635469.1	3.35E-15	integrase arm-type DNA-binding domain-containing protein [Acetobacteraceae bacterium]
TRINITY_DN129540_c0_g1_i5	KAF0135724.1	2.69E-08	hypothetical protein FD153_2052, partial [Rhodospirillaceae bacterium]
TRINITY_DN129540_c0_g1_i6	KAF0135724.1	6.30E-08	hypothetical protein FD153_2052, partial [Rhodospirillaceae bacterium]
TRINITY_DN129540_c0_g1_i7	KAF0135724.1	3.07E-08	hypothetical protein FD153_2052, partial [Rhodospirillaceae bacterium]
TRINITY_DN129540_c0_g1_i8	KAF0135724.1	2.30E-08	hypothetical protein FD153_2052, partial [Rhodospirillaceae bacterium]
TRINITY_DN129540_c0_g1_i9	KAF0135724.1	3.80E-08	hypothetical protein FD153_2052, partial [Rhodospirillaceae bacterium]
TRINITY_DN115854_c0_g1_i1	AAY34139.1	1.49E-09	snu13p-like [Euglena gracilis]
TRINITY_DN161247_c0_g1_i1	EPI56452.1	1.57E-27	hypothetical protein HMPREF1573_00869 [Gardnerella vaginalis JCP7276]
TRINITY_DN161247_c0_g1_i2	OKY55971.1	1.43E-27	hypothetical protein BHS10_00250 [Gardnerella vaginalis]
TRINITY_DN318734_c0_g1_i1	BAJ81851.1	6.46E-39	30S ribosomal protein S9 [Acidiphilium multivorum AIU301]
TRINITY_DN305840_c0_g1_i1	WP_004121707.1	2.72E-39	pyruvate formate lyase-activating protein [Gardnerella vaginalis]
TRINITY_DN321149_c0_g1_i1	TXH99032.1	6.06E-33	ATP-binding cassette domain-containing protein [Pseudomonas sp.]
TRINITY_DN320484_c0_g1_i1	OEU22037.1	2.48E-17	hypothetical protein FRACYDRAFT_232192 [Fragilariopsis cylindrus CCMP1102]
TRINITY_DN343342_c0_g1_i1	MBP6614511.1	9.16E-38	class I polyI-hydroxyalkanoic acid synthase [Aquabacterium sp.]
TRINITY_DN347737_c0_g1_i1	WP_102694911.1	2.27E-40	low molecular weight phosphotyrosine protein phosphatase [Gardnerella vaginalis]
TRINITY_DN91468_c0_g1_i1	WP_007421477.1	1.02E-72	MULTISPECIES: ferredoxin family protein [Acidiphilium]
TRINITY_DN91468_c0_g1_i2	WP_007421477.1	5.38E-40	MULTISPECIES: ferredoxin family protein [Acidiphilium]
TRINITY_DN91468_c0_g1_i3	WP_007421477.1	3.10E-74	MULTISPECIES: ferredoxin family protein [Acidiphilium]
TRINITY_DN91468_c0_g1_i4	WP_007421477.1	4.26E-74	MULTISPECIES: ferredoxin family protein [Acidiphilium]
TRINITY_DN21430_c17_g1_i1	KAF8091070.1	4.69E-18	hypothetical protein N665_0454s0011 [Sinapis alba]
TRINITY_DN75107_c3_g1_i1	EHH51985.1	4.26E-33	hypothetical protein EGM_12341, partial [Macaca fascicularis]
TRINITY_DN75107_c3_g1_i2	EHH18371.1	4.05E-38	hypothetical protein EGK_14950, partial [Macaca mulatta]
TRINITY_DN41657_c1_g1_i1	BCS94975.1	8.37E-08	radical SAM protein [Desulfoluna sp. ASN36]
TRINITY_DN17767_c0_g2_i1	YP_001315110.1	2.42E-06	putative reverse transcriptase and intron maturase [Chlorokybus atmophyticus]
TRINITY_DN17767_c0_g2_i1	YP_001315110.1	3.06E-06	putative reverse transcriptase and intron maturase [Chlorokybus atmophyticus]
TRINITY_DN17767_c0_g2_i1	YP_001315110.1	3.06E-06	putative reverse transcriptase and intron maturase [Chlorokybus atmophyticus]
TRINITY_DN47524_c0_g1_i1	GFH12567.1	9.42E-07	eIF3 p110, partial [Haematococcus lacustris]
TRINITY_DN64336_c2_g1_i1	RHY92403.1	8.24E-07	hypothetical protein DYB35_001108 [Aphanomyces astaci]
TRINITY_DN64336_c2_g1_i1	RHY92403.1	8.24E-07	hypothetical protein DYB35_001108 [Aphanomyces astaci]
TRINITY_DN70997_c8_g1_i1	PBH82200.1	1.03E-12	hypothetical protein BGU59_19325 [Clostridioides difficile]
TRINITY_DN20156_c2_g1_i1	MBT3636022.1	8.94E-12	30S ribosomal protein S13 [Opitutae bacterium]
TRINITY_DN80163_c0_g4_i1	AAR04017.2	1.43E-47	chloroplast trisophosphate isomerase [Euglena gracilis]

TRINITY_DN16652_c0_g1_i1	KAB0794669.1	7.33E-07	hypothetical protein PPYR_11508 [Photinus pyralis]
TRINITY_DN51336_c0_g1_i1	NUO71914.1	1.96E-19	nitronate monooxygenase [Frateuria sp.]
TRINITY_DN42857_c0_g1_i1	KAA6384641.1	1.38E-06	putative Serine/Threonine kinase domain protein [Streblomastix strix]
TRINITY_DN57958_c0_g1_i2	KAF9270907.1	4.85E-06	hypothetical protein BGZ68_004636, partial [Mortierel90lpineina]
TRINITY_DN57958_c0_g1_i3	KAF9270907.1	9.30E-06	hypothetical protein BGZ68_004636, partial [Mortierel90lpineina]
TRINITY_DN57409_c0_g1_i1	HJE71029.1	5.09E-34	NADP-dependent isocitrate dehydrogenase [Pseudomonas oryzihabitans]
TRINITY_DN57409_c0_g1_i2	WP_019019862.1	8.84E-23	NADP-dependent isocitrate dehydrogenase [Halomonas lutea]
TRINITY_DN57409_c0_g1_i2	WP_019019862.1	2.31E-08	NADP-dependent isocitrate dehydrogenase [Halomonas lutea]
TRINITY_DN74275_c1_g1_i1	6TDU_M	8.37E-09	Chain M, ATPEG1 [Euglena gracilis]
TRINITY_DN55596_c0_g1_i1	WP_081432878.1	7.35E-94	MULTISPECIES: flagellar motor switch protein FliM [unclassified Acidiphilium]
TRINITY_DN55596_c0_g1_i10	WP_081432878.1	1.44E-53	MULTISPECIES: flagellar motor switch protein FliM [unclassified Acidiphilium]
TRINITY_DN55596_c0_g1_i2	WP_081432878.1	9.08E-15	MULTISPECIES: flagellar motor switch protein FliM [unclassified Acidiphilium]
TRINITY_DN55596_c0_g1_i3	WP_081432878.1	2.28E-58	MULTISPECIES: flagellar motor switch protein FliM [unclassified Acidiphilium]
TRINITY_DN55596_c0_g1_i4	WP_081432878.1	1.54E-41	MULTISPECIES: flagellar motor switch protein FliM [unclassified Acidiphilium]
TRINITY_DN55596_c0_g1_i5	WP_081432878.1	8.22E-61	MULTISPECIES: flagellar motor switch protein FliM [unclassified Acidiphilium]
TRINITY_DN55596_c0_g1_i6	WP_081432878.1	2.32E-153	MULTISPECIES: flagellar motor switch protein FliM [unclassified Acidiphilium]
TRINITY_DN55596_c0_g1_i7	WP_081432878.1	5.43E-41	MULTISPECIES: flagellar motor switch protein FliM [unclassified Acidiphilium]
TRINITY_DN55596_c0_g1_i8	WP_081432878.1	2.07E-29	MULTISPECIES: flagellar motor switch protein FliM [unclassified Acidiphilium]
TRINITY_DN55596_c0_g1_i9	WP_081432878.1	2.00E-58	MULTISPECIES: flagellar motor switch protein FliM [unclassified Acidiphilium]

Table S4: blastx results of differentially expressed transcripts identified by DESeq2 for non-pretreated *Euglena gracilis* cultures in the presence and absence of $CdCl_2$ using the NCBI SWISS-PROT protein database.

SWISS_Isoform Id	SWISS_sseqid	SWISS_evalue	SWISS_description
TRINITY_DN207373_c0_g1_i1	A6KXS0.1	1.04E-23	RecName: Full=Acetate kinase; AltName: Full=Acetokinase [Phocaeicola vulgatus ATCC 8482]
TRINITY_DN270125_c0_g1_i1	Q5SHX3.1	4.93E-09	RecName: Full=Thymidylate kinase; AltName: Full=dTMP kinase [Thermus thermophilus HB8]
TRINITY_DN231511_c0_g1_i1	A4J537.1	4.69E-31	RecName: Full=Holliday junction ATP-dependent DNA helicase RuvB [Desulfotomaculum reducens MI-1]
TRINITY_DN192008_c0_g1_i1	A0PYP1.1	2.49E-37	RecName: Full=Phosphoglycerate kinase [Clostridium novyi NT]
TRINITY_DN158784_c2_g1_i1	Q9LFE3.1	7.60E-08	RecName: Full=Amino acid transporter AVT1H; Short=AtAvt1H [Arabidopsis thaliana]
TRINITY_DN109450_c0_g1_i1	Q9SZA4.1	2.99E-19	RecName: Full=Cell division cycle 20.1, cofactor of APC complex; Short=AtCDC20.1 [Arabidopsis thaliana]
TRINITY_DN157530_c0_g1_i1	O00370.1	4.42E-31	RecName: Full=LINE-1 retrotransposable element ORF2 protein; Short=ORF2p; Includes: RecName: Full=Reverse transcriptase; Includes: RecName: Full=Endonuclease [Homo sapiens]
TRINITY_DN182881_c0_g1_i1	Q5FJN0.1	2.27E-42	RecName: Full=Proline—tRNA ligase; AltName: Full=Prolyl-tRNA synthetase; Short=ProRS [Lactobacillus acidophilus NCFM]
TRINITY_DN129540_c0_g1_i13	P37317.1	3.88E-11	RecName: Full=Integrase [Shigella virus Sf6]
TRINITY_DN318734_c0_g1_i1	B5ZXZ5.1	2.40E-22	RecName: Full=30S ribosomal protein S9 [Rhizobium leguminosarum bv. Trifolii WSM2304]
TRINITY_DN305840_c0_g1_i1	Q46267.1	1.27E-06	RecName: Full=Pyruvate formate-lyase-activating enzyme; Short=PFL-activating enzyme; AltName:

			Full=Formate-C-acetyltransferase-activating enzyme [Clostridium pasteurianum]
TRINITY_DN321149_c0_g1_i1	Q9FNU2.1	2.93E-08	RecName: Full=ABC transporter B family member 25; Short=ABC transporter ABCB.25; Short=OsABCB25; AltName: Full=Protein ALS1 homolog; AltName: Full=Protein ALUMINUM SENSITIVE 1; Short=OsALS1 [Oryza sativa Japonica Group]
TRINITY_DN343342_c0_g1_i1	P23608.1	2.78E-25	RecName: Full=Poly(3-hydroxyalkanoate) polymerase subunit PhaC; Short=PHA polymerase; AltName: Full=PHB synthase subunit PhaC; AltName: Full=Poly(3-hydroxybutyrate) polymerase subunit PhaC; Short=PHB polymerase; Short=Poly-beta-hydroxybutyrate polymerase; AltName: Full=Polyhydroxyalkanoate synthase subunit PhaC; Short=PHA synthase [Cupriavidus necator H16]
TRINITY_DN91468_c0_g1_i1	P80448.1	8.18E-45	RecName: Full=Ferredoxin-2; AltName: Full=Ferredoxin II; Short=FdII [Rhodospirillum rubrum]
TRINITY_DN91468_c0_g1_i2	Q44037.1	1.43E-36	RecName: Full=Ferredoxin-1 [Afipia felis]
TRINITY_DN91468_c0_g1_i3	P80448.1	3.45E-46	RecName: Full=Ferredoxin-2; AltName: Full=Ferredoxin II; Short=FdII [Rhodospirillum rubrum]
TRINITY_DN91468_c0_g1_i4	P80448.1	4.64E-46	RecName: Full=Ferredoxin-2; AltName: Full=Ferredoxin II; Short=FdII [Rhodospirillum rubrum]
TRINITY_DN75107_c3_g1_i1	Q96MD7.1	4.34E-30	RecName: Full=Uncharacterized protein C9orf85 [Homo sapiens]
TRINITY_DN75107_c3_g1_i2	Q8N7I0.1	7.52E-29	RecName: Full=Protein GVQW1; AltName: Full=GVQW motif-containing protein 1; AltName: Full=Tigger transposable element-derived 1-like protein 2; Flags: Precursor [Homo sapiens]
TRINITY_DN47524_c0_g1_i1	B4P5F7.1	3.87E-06	RecName: Full=Eukaryotic translation initiation factor 3 subunit B; Short=eIF3b; AltName: Full=Eukaryotic translation initiation factor 3 subunit 9 [Drosophila yakuba]
TRINITY_DN64336_c2_g1_i1	Q55BV5.1	1.09E-07	RecName: Full=26S proteasome regulatory subunit 4 homolog; AltName: Full=Tat-binding protein alpha; Short=DdTBPalpha [Dictyostelium discoideum]
TRINITY_DN64336_c2_g1_i1	Q55BV5.1	1.09E-07	RecName: Full=26S proteasome regulatory subunit 4 homolog; AltName: Full=Tat-binding protein alpha; Short=DdTBPalpha [Dictyostelium discoideum]
TRINITY_DN82768_c1_g1_i1	Q00372.2	4.93E-06	RecName: Full=Carbon catabolite-derepressing protein kinase [[Candida] glabrata CBS 138]
TRINITY_DN20156_c2_g1_i1	Q8FS35.2	2.49E-13	RecName: Full=30S ribosomal protein S13 [Corynebacterium efficiens YS-314]
TRINITY_DN16652_c0_g1_i1	Q9LK64.1	9.31E-07	RecName: Full=ABC transporter C family member 3; Short=ABC transporter ABCC.3; Short=AtABCC3; AltName: Full=ATP-energized glutathione S-conjugate pump 3; AltName: Full=Glutathione S-conjugate-transporting ATPase 3; AltName: Full=Multidrug resistance-associated protein 3 [Arabidopsis thaliana]
TRINITY_DN42857_c0_g1_i1	Q63184.1	9.02E-07	RecName: Full=Interferon-induced, double-stranded RNA-activated protein kinase; AltName: Full=Eukaryotic translation initiation factor 2-alpha kinase 2; Short=eIF-2A protein kinase 2; AltName: Full=Interferon-inducible RNA-dependent protein kinase; AltName: Full=Protein kinase RNA-activated; Short=PKR; Short=Protein kinase R; AltName: Full=Tyrosine-protein kinase EIF2AK2 [Rattus norvegicus]
TRINITY_DN57409_c0_g1_i1	P16100.5	2.10E-31	RecName: Full=Isocitrate dehydrogenase [NADP]; Short=IDH; AltName: Full=Oxalosuccinate decarboxylase [Azotobacter vinelandii]
TRINITY_DN57409_c0_g1_i2	P16100.5	1.38E-20	RecName: Full=Isocitrate dehydrogenase [NADP]; Short=IDH; AltName: Full=Oxalosuccinate decarboxylase [Azotobacter vinelandii]
TRINITY_DN57409_c0_g1_i2	P16100.5	2.32E-11	RecName: Full=Isocitrate dehydrogenase [NADP]; Short=IDH; AltName: Full=Oxalosuccinate decarboxylase [Azotobacter vinelandii]
TRINITY_DN55596_c0_g1_i1	B8GXB5.1	5.05E-20	RecName: Full=Flagellar motor switch protein FliM [Caulobacter vibrioides NA1000]

TRINITY_DN55596_c0_g1_i10	B8GXB5.1	1.34E-07	RecName: Full=Flagellar motor switch protein FliM [Caulobacter vibrioides NA1000]
TRINITY_DN55596_c0_g1_i3	B8GXB5.1	4.90E-08	RecName: Full=Flagellar motor switch protein FliM [Caulobacter vibrioides NA1000]
TRINITY_DN55596_c0_g1_i5	B8GXB5.1	1.74E-08	RecName: Full=Flagellar motor switch protein FliM [Caulobacter vibrioides NA1000]
TRINITY_DN55596_c0_g1_i6	B8GXB5.1	6.18E-39	RecName: Full=Flagellar motor switch protein FliM [Caulobacter vibrioides NA1000]
TRINITY_DN55596_c0_g1_i9	B8GXB5.1	3.25E-08	RecName: Full=Flagellar motor switch protein FliM [Caulobacter vibrioides NA1000]

Table S5: blastx results of differentially expressed transcripts identified by DESeq2 for non-pretreated *Euglena gracilis* cultures in the presence and absence of $CdCl_2$ using the *Arabidopsis thaiana* Ensembl database.

Athaliana_Isoform Id	Athaliana_sseqid	Athaliana_evalue	Athaliana_description
TRINITY_DN203263_c0_g1_i1	AT1G76010.2	5.55E-05	Alba DNA/RNA-binding protein
TRINITY_DN203263_c0_g1_i1	AT1G76010.2	5.55E-05	Alba DNA/RNA-binding protein
TRINITY_DN192008_c0_g1_i1	AT1G79550.1	1.20E-20	Phosphoglycerate kinase
TRINITY_DN158784_c2_g1_i1	AT5G16740.1	8.73E-09	Amino acid transporter AVT1H
TRINITY_DN109450_c0_g1_i1	AT4G33270.1	3.44E-20	Cell division cycle 20.1, cofactor of APC complex
TRINITY_DN318734_c0_g1_i1	AT3G49080.1	1.46E-08	30S ribosomal protein S9, mitochondrial]
TRINITY_DN321149_c0_g1_i1	AT5G39040.1	1.20E-08	ABC transporter B family member 27
TRINITY_DN47524_c0_g1_i1	AT5G25780.1	1.88E-06	Eukaryotic translation initiation factor 3 subunit B
TRINITY_DN64336_c2_g1_i1	AT2G20140.1	4.09E-06	RPT2b
TRINITY_DN82768_c1_g1_i1	AT2G26980.5	6.36E-05	Non-specific serine/threonine protein kinase
TRINITY_DN20156_c2_g1_i1	AT5G14320.2	4.57E-08	30S ribosomal protein S13, chloroplastic
TRINITY_DN16652_c0_g1_i1	AT3G13080.1	1.07E-07	MRP3
TRINITY_DN42857_c0_g1_i1	AT5G04510.3	5.01E-08	3-phosphoinositide-dependent protein kinase 1

Table S6: blastx results of differentially expressed transcripts identified by DESeq2 for non-pretreated *Euglena gracilis* cultures in the presence and absence of $CdCl_2$ using the *Chlamydomonas reinhardtti* Ensembl database.

Creinhardtii_Isoform Id	Creinhardtii_sseqid	Creinhardtii_evalue	Creinhardtii_description
TRINITY_DN207373_c0_g1_i1	PNW70194	4.26E-14	hypothetical protein
TRINITY_DN192008_c0_g1_i1	PNW76650	4.51E-20	hypothetical protein
TRINITY_DN109450_c0_g1_i1	PNW78378	3.56E-19	hypothetical protein
TRINITY_DN321149_c0_g1_i1	PNW86357	6.58E-07	hypothetical protein
TRINITY_DN320484_c0_g1_i1	PNW84031	2.53E-08	hypothetical protein
TRINITY_DN64336_c2_g1_i1	PNW80789	7.48E-09	hypothetical protein
TRINITY_DN64336_c2_g1_i1	PNW80789	7.48E-09	hypothetical protein
TRINITY_DN20156_c2_g1_i1	PNW74498	1.25E-07	hypothetical protein
TRINITY_DN42857_c0_g1_i1	PNW87550	2.03E-09	hypothetical protein
TRINITY_DN57958_c0_g1_i2	PNW71910	2.31E-04	hypothetical protein
TRINITY_DN57958_c0_g1_i3	PNW71910	2.18E-04	hypothetical protein

Table S7: blastx results of differentially expressed transcripts identified by DESeq2 for non-pretreated *Euglena gracilis* cultures in the presence and absence of $CdCl_2$ using the *Synechocystis sp.* Ensembl database.

Synechocystis_Isoform Id	Synechocystis_sseqid	Synechocystis_evalue	Synechocystis_description
TRINITY_DN207373_c0_g1_i1	AIE74091	1.14E-15	description:Acetate kinase
TRINITY_DN231511_c0_g1_i1	AIE74274	3.22E-28	Holliday junction DNA helicase RuvB
TRINITY_DN192008_c0_g1_i1	AIE75237	2.94E-17	Phosphoglycerate kinase
TRINITY_DN182881_c0_g1_i1	AIE73902	3.19E-21	Prolyl-tRNA synthetase, bacterial type
TRINITY_DN318734_c0_g1_i1	AIE75030	5.12E-08	SSU ribosomal protein S9p (S16e)
TRINITY_DN321149_c0_g1_i1	AIE75212	1.29E-05	ATP-binding protein of ABC transporter
TRINITY_DN91468_c0_g1_i1	AIE75376	4.68E-06	:4Fe-4S ferredoxin, iron-sulfur binding
TRINITY_DN91468_c0_g1_i2	AIE75376	8.09E-07	4Fe-4S ferredoxin, iron-sulfur binding
TRINITY_DN91468_c0_g1_i3	AIE75376	1.85E-06	4Fe-4S ferredoxin, iron-sulfur binding
TRINITY_DN91468_c0_g1_i4	AIE75376	2.10E-06	4Fe-4S ferredoxin, iron-sulfur binding
TRINITY_DN42857_c0_g1_i1	AIE74546	3.65E-04	eukaryotic protein kinase

Table S8: blastx results of differentially expressed transcripts identified by DESeq2 for non-pretreated *Euglena. gracilis* cultures in the presence and absence of CdCl2 using the *Homo sapiens* Ensembl database.

Hsapiens_Isoform Id	Hsapiens_sseqid	Hsapiens_evalue	Hsapiens_description
TRINITY_DN192008_c0_g1_i1	ENSP00000305995.3	2.23E-11	phosphoglycerate kinase 2
TRINITY_DN109450_c0_g1_i1	ENSP00000361540.1	3.29E-17	cell division cycle 20
TRINITY_DN182881_c0_g1_i1	ENSP00000360327.3	2.25E-16	prolyl-tRNA synthetase 2, mitochondrial
TRINITY_DN321149_c0_g1_i1	ENSP00000440138.1	6.97E-09	ATP binding cassette subfamily B member 9
TRINITY_DN75107_c3_g1_i1	ENSP00000505475.1	3.69E-35	VPS53 subunit of GARP complex
TRINITY_DN75107_c3_g1_i2	ENSP00000352210.4	3.34E-35	PDZ and LIM domain 5
TRINITY_DN64336_c2_g1_i1	ENSP00000261303.8	5.77E-07	proteasome 26S subunit, ATPase 1
TRINITY_DN64336_c2_g1_i1	ENSP00000261303.8	5.77E-07	proteasome 26S subunit, ATPase 1
TRINITY_DN82768_c1_g1_i1	ENSP00000380317.2	1.01E-05	protein kinase AMP-activated catalytic subunit alpha 1
TRINITY_DN16652_c0_g1_i1	ENSP00000507301.1	2.66E-04	ATP binding cassette subfamily C member 6
TRINITY_DN42857_c0_g1_i1	ENSP00000486558.1	5.78E-06	maternal embryonic leucine zipper kinase
TRINITY_DN57958_c0_g1_i2	ENSP00000222823.4	8.34E-05	nucleotide binding oligomerization domain containing 1
TRINITY_DN57958_c0_g1_i3	ENSP00000222823.4	1.00E-04	nucleotide binding oligomerization domain containing 1

Table S9: blastx results of differentially expressed transcripts identified by DESeq2 for non-pretreated *Euglena gracilis* cultures in the presence and absence of $CdCl_2$ using the *Trypanosoma brucei* Ensembl database.

Tbrucei_Isoforms Id	Tbrucei_sseqid	Tbrucei_evalue	Tbrucei_description
TRINITY_DN192008_c0_g1_i1	CAJ15978	2.99E-13	phosphoglycerate kinase
TRINITY_DN109450_c0_g1_i1	AAZ13414	3.13E-13	cell division cycle protein, putative
TRINITY_DN115854_c0_g1_i1	EAN76582	2.04E-04	ribosomal protein S6, putative
TRINITY_DN64336_c2_g1_i1	EAN79343	4.47E-11	proteasome regulatory ATPase subunit 2
TRINITY_DN64336_c2_g1_i1	EAN79343	4.47E-11	proteasome regulatory ATPase subunit 2
TRINITY_DN82768_c1_g1_i1	EAN77478	2.73E-04	protein kinase, putative
TRINITY_DN42857_c0_g1_i1	EAN79784	1.15E-06	protein kinase, putative
TRINITY_DN57958_c0_g1_i2	EAN78352	4.45E-05	hypothetical protein, conserved
TRINITY_DN57958_c0_g1_i3	EAN78352	5.33E-05	hypothetical protein, conserved

Table S10: blastx results of differentially expressed transcripts identified by DESeq2 for sulfur pretreated *Euglena gracilis* cultures compared to non-pretreated cultures in the presence of $CdCl_2$ using the NCBI non-redundant protein database.

NR_Isoform Id	NR_sseqid	NR_evalue	NR_description
TRINITY_DN200360_c0_g1_i1	WP_162809807.1	7.69E-38	chemotaxis protein, partial [Vibrio cholerae]
TRINITY_DN277332_c0_g1_i1	WP_114911319.1	8.64E-41	ABC transporter substrate-binding protein [Acidibrevibacterium fodinaquatile]
TRINITY_DN207373_c0_g1_i1	WP_008450649.1	4.85E-33	acetate kinase [Prevotella amnii]
TRINITY_DN270125_c0_g1_i1	EPI54498.1	6.16E-45	dTMP kinase domain protein, partial [Gardnerella vaginalis JCP7276]
TRINITY_DN204984_c0_g1_i1	GER28773.1	1.49E-26	protein kinase superfamily protein [Striga asiatica]
TRINITY_DN238436_c0_g1_i1	WP_164799065.1	1.85E-16	peptidase [Lactobacillus iners]
TRINITY_DN231511_c0_g1_i1	WP_201775009.1	1.20E-49	Holliday junction branch migration DNA helicase RuvB [Fannyhessea vaginae]
TRINITY_DN295572_c0_g1_i1	RIY23104.1	4.99E-17	peptidase, partial [Bifidobacteriaceae bacterium NR021]
TRINITY_DN236134_c0_g1_i1	EPI52247.1	1.92E-06	hypothetical protein HMPREF1575_00740 [Gardnerella vaginalis JCP7672]
TRINITY_DN239594_c0_g1_i1	WP_090177237.1	5.66E-07	hypothetical protein [Luteibacter sp. UNC138MFCol5.1]
TRINITY_DN196392_c0_g1_i1	WP_114911399.1	3.94E-34	hypothetical protein [Acidibrevibacterium fodinaquatile]
TRINITY_DN158535_c10_g1_i1	AAQ24863.1	3.41E-20	heat shock protein 70, partial [Euglena gracilis]
TRINITY_DN180686_c0_g1_i1	EPI55698.1	1.71E-39	hypothetical protein HMPREF1573_01091, partial [Gardnerella vaginalis JCP7276]
TRINITY_DN157530_c0_g1_i1	WP_234805216.1	3.46E-31	hypothetical protein, partial [Salmonella enterica]
TRINITY_DN182881_c0_g1_i1	EFO70065.1	8.15E-46	proline--tRNA ligase [Lactobacillus iners LactinV 01V1-a]
TRINITY_DN101872_c0_g1_i1	WP_068467488.1	1.13E-08	RMD1 family protein [Parachlamydia sp. C2]
TRINITY_DN160293_c1_g1_i1	QLA09623.1	8.85E-12	60S large subunit ribosomal protein eL19 [Euglena gracilis]
TRINITY_DN306459_c0_g1_i1	MBL8788430.1	4.44E-27	VOC family protein [Deltaproteobacteria bacterium]
TRINITY_DN321149_c0_g1_i1	TXH99032.1	6.06E-33	ATP-binding cassette domain-containing protein [Pseudomonas sp.]
TRINITY_DN343792_c0_g1_i1	WP_019036358.1	5.22E-32	UDP-N-acetylmuramate--L-alanine ligase [Prevotella amnii]
TRINITY_DN301869_c0_g1_i1	WP_004128804.1	2.76E-38	BspA family leucine-rich repeat surface protein [Gardnerella vaginalis]
TRINITY_DN343342_c0_g1_i1	MBP6614511.1	9.16E-38	class I poly(R)-hydroxyalkanoic acid synthase [Aquabacterium sp.]
TRINITY_DN347737_c0_g1_i1	WP_102694911.1	2.27E-40	low molecular weight phosphotyrosine protein phosphatase [Gardnerella vaginalis]
TRINITY_DN10914_c0_g2_i1	XP_005759141.1	7.72E-29	hypothetical protein EMIHUDRAFT_460075 [Emiliania huxleyi CCMP1516]
TRINITY_DN15843_c0_g3_i1	XP_020527713.1	1.81E-09	uncharacterized protein LOC18442033 [Amborella trichopoda]

TRINITY_DN21430_c17_g1_i1	KAF8091070.1	4.69E-18	hypothetical protein N665_0454s0011 [Sinapis alba]
TRINITY_DN21061_c12_g1_i1	APF46182.1	2.54E-51	plastid uroporphyrinogen decarboxylase isoform 1, partial [Euglena gracilis]
TRINITY_DN14547_c4_g2_i1	CAD1825229.1	3.53E-06	unnamed protein product [Ananas comosus var. bracteatus]
TRINITY_DN31000_c0_g2_i1	PSC75257.1	2.44E-09	aldo keto reductase [Micractinium conductrix]
TRINITY_DN48075_c3_g1_i1	EPY29463.1	2.63E-07	centromere protein J [Angomonas deanei]
TRINITY_DN70997_c8_g1_i1	PBH82200.1	1.03E-12	hypothetical protein BGU59_19325 [Clostridioides difficile]
TRINITY_DN16652_c0_g1_i1	KAB0794669.1	7.33E-07	hypothetical protein PPYR_11508 [Photinus pyralis]
TRINITY_DN51336_c0_g1_i1	NUO71914.1	1.96E-19	nitronate monooxygenase [Frateuria sp.]
TRINITY_DN2850_c12_g1_i1	BAD20741.1	3.09E-20	putative membrane-bound adenylyl cyclase [Euglena gracilis]
TRINITY_DN28590_c2_g1_i1	NBC85862.1	3.84E-12	hydroxyacid dehydrogenase [Bacteroidetes bacterium]
TRINITY_DN79385_c1_g1_i1	PNX55782.1	9.50E-20	long chain acyl-CoA synthetase 8-like protein, partial [Trifolium pratense]
TRINITY_DN74275_c1_g1_i1	6TDU_M	8.37E-09	Chain M, ATPEG1 [Euglena gracilis]

Table S11: blastx results of differentially expressed transcripts identified by DESeq2 for sulfur pretreated *Euglena gracilis* cultures compared to non-pretreated cultures both in the presence of $CdCl_2$ using the NCBI SWISS-PROT protein database.

SWISS_Isoform Id	SWISS_sseqid	SWISS_evalue	SWISS_description
TRINITY_DN207373_c0_g1_i1	A6KXS0.1	1.04E-23	RecName: Full=Acetate kinase; AltName: Full=Acetokinase [Phocaeicola vulgatus ATCC 8482]
TRINITY_DN270125_c0_g1_i1	Q5SHX3.1	4.93E-09	RecName: Full=Thymidylate kinase; AltName: Full=dTMP kinase [Thermus thermophilus HB8]
TRINITY_DN204984_c0_g1_i1	Q9LIC2.1	8.91E-28	RecName: Full=Transmembrane 9 superfamily member 7; AltName: Full=Endomembrane protein 5; AltName: Full=Transmembrane nine protein 7; Short=AtTMN7; Flags: Precursor [Arabidopsis thaliana]
TRINITY_DN231511_c0_g1_i1	A4J537.1	4.69E-31	RecName: Full=Holliday junction ATP-dependent DNA helicase RuvB [Desulfotomaculum reducens MI-1]
TRINITY_DN158535_c10_g1_i1	Q92260.1	6.52E-20	RecName: Full=Heat shock 70 kDa protein; AltName: Allergen=Pen c 19 [Penicillium citrinum]
TRINITY_DN157530_c0_g1_i1	O00370.1	4.42E-31	RecName: Full=LINE-1 retrotransposable element ORF2 protein; Short=ORF2p; Includes: RecName: Full=Reverse transcriptase; Includes: RecName: Full=Endonuclease [Homo sapiens]
TRINITY_DN182881_c0_g1_i1	Q5FJN0.1	2.27E-42	RecName: Full=Proline--tRNA ligase; AltName: Full=Prolyl-tRNA synthetase; Short=ProRS [Lactobacillus acidophilus NCFM]
TRINITY_DN160293_c1_g1_i1	P36241.2	6.03E-12	RecName: Full=60S ribosomal protein L19 [Drosophila melanogaster]
TRINITY_DN321149_c0_g1_i1	Q9FNU2.1	2.93E-08	RecName: Full=ABC transporter B family member 25; Short=ABC transporter ABCB.25; Short=OsABCB25; AltName: Full=Protein ALS1 homolog; AltName: Full=Protein ALUMINUM SENSITIVE 1; Short=OsALS1 [Oryza sativa Japonica Group]
TRINITY_DN343792_c0_g1_i1	A6L070.1	3.13E-13	RecName: Full=UDP-N-acetylmuramate--L-alanine ligase; AltName: Full=UDP-N-acetylmuramoyl-L-alanine synthetase [Phocaeicola vulgatus ATCC 8482]
TRINITY_DN343342_c0_g1_i1	P23608.1	2.78E-25	RecName: Full=Poly(3-hydroxyalkanoate) polymerase subunit PhaC; Short=PHA polymerase; AltName: Full=PHB synthase subunit PhaC; AltName: Full=Poly(3-hydroxybutyrate) polymerase subunit PhaC; Short=PHB polymerase; Short=Poly-beta-hydroxybutyrate polymerase; AltName: Full=Polyhydroxyalkanoate synthase subunit PhaC; Short=PHA synthase [Cupriavidus necator H16]
TRINITY_DN10914_c0_g2_i1	Q73KL7.1	7.98E-16	RecName: Full=L-methionine gamma-lyase; Short=MGL; AltName: Full=Homocysteine desulfhydrase; AltName: Full=L-methionine-alpha-deamino-gamma-mercaptomethane-lyase; Short=METase [Treponema denticola ATCC 35405]

TRINITY_DN15843_c0_g3_i1	F4I6V0.1	1.54E-10	RecName: Full=Beta-arabinofuranosyltransferase RAY1; AltName: Full=Protein REDUCED ARABINOSE YARIV 1 [Arabidopsis thaliana]
TRINITY_DN21061_c12_g1_i1	Q9V595.1	6.59E-18	RecName: Full=Uroporphyrinogen decarboxylase; Short=UPD; Short=URO-D [Drosophila melanogaster]
TRINITY_DN16652_c0_g1_i1	Q9LK64.1	9.31E-07	RecName: Full=ABC transporter C family member 3; Short=ABC transporter ABCC.3; Short=AtABCC3; AltName: Full=ATP-energized glutathione S-conjugate pump 3; AltName: Full=Glutathione S-conjugate-transporting ATPase 3; AltName: Full=Multidrug resistance-associated protein 3 [Arabidopsis thaliana]
TRINITY_DN2850_c12_g1_i1	Q27675.1	3.33E-15	RecName: Full=Receptor-type adenylate cyclase A; AltName: Full=ATP pyrophosphate-lyase; AltName: Full=Adenylyl cyclase [Leishmania donovani]
TRINITY_DN79385_c1_g1_i1	Q9CAP8.1	3.92E-20	RecName: Full=Long chain acyl-CoA synthetase 9, chloroplastic [Arabidopsis thaliana]

Table S12: blastx results of differentially expressed transcripts identified by DESeq2 for sulfur pretreated *Euglena gracilis* cultures compared to non-pretreated cultures in the presence of $CdCl_2$ using the *Arabidopsis thaliana* Ensembl database.

Athaliana_Isoform Id	Athaliana_sseqid	Athaliana_evalue	Athaliana_description
TRINITY_DN204984_c0_g1_i1	AT3G13772.1	1.02E-28	Transmembrane 9 superfamily member
TRINITY_DN203263_c0_g1_i1	AT1G76010.2	5.55E-05	Alba DNA/RNA-binding protein
TRINITY_DN203263_c0_g1_i1	AT1G76010.2	5.55E-05	Alba DNA/RNA-binding protein
TRINITY_DN158535_c10_g1_i1	AT3G12580.1	5.62E-20	Probable mediator of RNA polymerase II transcription subunit 37c
TRINITY_DN160293_c1_g1_i1	AT4G02230.1	1.36E-11	60S ribosomal protein L19-3
TRINITY_DN321149_c0_g1_i1	AT5G39040.1	1.20E-08	description:ABC transporter B family member 27
TRINITY_DN10914_c0_g2_i1	AT3G01120.1	5.65E-12	MTO1
TRINITY_DN15843_c0_g3_i1	AT1G70630.5	1.50E-11	Nucleotide-diphospho-sugar transferase family protein
TRINITY_DN21061_c12_g1_i1	AT2G40490.1	7.40E-08	Uroporphyrinogen decarboxylase 2, chloroplastic
TRINITY_DN14547_c4_g2_i1	AT2G36000.1	1.03E-05	At2g36000/F11F19.9
TRINITY_DN16652_c0_g1_i1	AT3G13080.1	1.07E-07	MRP3
TRINITY_DN79385_c1_g1_i1	AT1G77590.2	4.40E-21	Long chain acyl-CoA synthetase 9, chloroplastic

Table S13: blastx results of differentially expressed transcripts identified by DESeq2 for sulfur pretreated *Euglena gracilis* cultures compared to non-pretreated cultures in the presence of $CdCl_2$ using the *Chlamydomonas reinhardtti* Ensembl database.

Creinhardtii_Isoform Id	Creinhardtii_sseqid	Creinhardtii_evalue	Creinhardtii_description
TRINITY_DN207373_c0_g1_i1	PNW70194	4.26E-14	hypothetical protein
TRINITY_DN204984_c0_g1_i1	PNW77512	3.45E-29	hypothetical protein
TRINITY_DN196392_c0_g1_i1	PNW79088	3.39E-10	hypothetical protein
TRINITY_DN158535_c10_g1_i1	PNW79920	2.39E-17	hypothetical protein
TRINITY_DN101872_c0_g1_i1	PNW78931	1.79E-05	hypothetical protein
TRINITY_DN160293_c1_g1_i1	PNW86172	3.26E-11	hypothetical protein
TRINITY_DN321149_c0_g1_i1	PNW86357	6.58E-07	hypothetical protein
TRINITY_DN10914_c0_g2_i1	PNW84424	1.50E-07	hypothetical protein
TRINITY_DN15843_c0_g3_i1	PNW83849	1.02E-05	hypothetical protein
TRINITY_DN21061_c12_g1_i1	PNW76686	1.37E-09	hypothetical protein
TRINITY_DN31000_c0_g2_i1	PNW69962	5.58E-09	hypothetical protein
TRINITY_DN2850_c12_g1_i1	PNW85731	3.47E-06	hypothetical protein
TRINITY_DN79385_c1_g1_i1	PNW85346	3.69E-10	hypothetical protein

Table S14: blastx results of differentially expressed transcripts identified by DESeq2 for sulfur pretreated *Euglena gracilis* cultures compared to non-pretreated cultures both in the presence of $CdCl_2$ using the *Synechocystis sp.* Ensembl database.

Synechocytis_Isoform Id	Synechocytis_sseqid	Synechocytis_evalue	Synechocytis_description
TRINITY_DN207373_c0_g1_i1	AIE74091	1.14E-15	description:Acetate kinase
TRINITY_DN231511_c0_g1_i1	AIE74274	3.22E-28	Holliday junction DNA helicase RuvB
TRINITY_DN182881_c0_g1_i1	AIE73902	3.19E-21	Prolyl-tRNA synthetase, bacterial type
TRINITY_DN321149_c0_g1_i1	AIE75212	1.29E-05	ATP-binding protein of ABC transporter
TRINITY_DN21061_c12_g1_i1	AIE75456	4.26E-06	Uroporphyrinogen III decarboxylase

Table S15: blastx results of differentially expressed transcripts identified by DESeq2 for sulfur pretreated *Euglena gracilis* cultures compared to non-pretreated cultures both in the presence of $CdCl_2$ using the *Homo. sapiens* Ensembl database.

Hsapiens_Isoform Id	Hsapiens_sseqid	Hsapiens_evalue	Hsapiens_description
TRINITY_DN204984_c0_g1_i1	ENSP00000217315.5	1.21E-23	transmembrane 9 superfamily member 4
TRINITY_DN158535_c10_g1_i1	ENSP00000433584.1	6.91E-18	heat shock protein family A (Hsp70) member 8
TRINITY_DN182881_c0_g1_i1	ENSP00000360327.3	2.25E-16	prolyl-tRNA synthetase 2, mitochondrial
TRINITY_DN160293_c1_g1_i1	ENSP00000225430.4	2.70E-10	ribosomal protein L19
TRINITY_DN321149_c0_g1_i1	ENSP00000440138.1	6.97E-09	ATP binding cassette subfamily B member 9
TRINITY_DN10914_c0_g2_i1	ENSP00000413407.2	2.33E-08	cystathionine gamma-lyase
TRINITY_DN21061_c12_g1_i1	ENSP00000498668.1	1.15E-11	uroporphyrinogen decarboxylase
TRINITY_DN16652_c0_g1_i1	ENSP00000507301.1	2.66E-04	ATP binding cassette subfamily C member 6
TRINITY_DN79385_c1_g1_i1	ENSP00000506881.1	3.29E-19	acyl-CoA synthetase long chain family member 4

Table S16: blastx results of differentially expressed transcripts identified by DESeq2 for sulfur pretreated *Euglena gracilis* cultures compared to non-pretreated cultures both in the presence of $CdCl_2$ using the *Trypanosoma brucei* Ensembl database.

Tbrucei_Isoforms Id	Tbrucei_sseqid	Tbrucei_evalue	Tbrucei_description
TRINITY_DN204984_c0_g1_i1	AAZ10452	1.60E-15	endosomal integral membrane protein, putative
TRINITY_DN158535_c10_g1_i1	EAN80086	1.03E-15	heat shock protein 70
TRINITY_DN101872_c0_g1_i1	EAN79339	2.12E-07	hypothetical protein, conserved
TRINITY_DN160293_c1_g1_i1	AAZ12564	4.10E-09	60S ribosomal protein L19, putative
TRINITY_DN10914_c0_g2_i1	EAN77147	3.69E-27	cystathione gamma lyase, putative
TRINITY_DN4765_c0_g2_i1	EAN78212	1.28E-05	hypothetical protein, conserved
TRINITY_DN48075_c3_g1_i1	EAN79300	5.41E-10	hypothetical protein, conserved
TRINITY_DN2850_c12_g1_i1	EAN79127	4.94E-17	receptor-type adenylate cyclase GRESAG 4, putative
TRINITY_DN45223_c0_g1_i1	EAN79716	2.49E-6	cystathionine beta-synthase, putative

Table S17: blastx results of differentially expressed transcripts identified by DESeq2 for sulfur pretreated *Euglena gracilis* cultures in the presence and absence of $CdCl_2$ using the NCBI non-redundant protein database.

NR_Isoform Id	NR_sseqid	NR_evalue	NR_description
TRINITY_DN211561_c0_g1_i1	MBF1508352.1	1.01E-41	hypothetical protein [Prevotella pallens]
TRINITY_DN158535_c10_g1_i1	AAQ24863.1	3.41E-20	heat shock protein 70, partial [Euglena gracilis]
TRINITY_DN166865_c1_g1_i1	WP_142894646.1	3.77E-33	electron transfer flavoprotein subunit beta/FixA family protein [Denitrobaculum tricleocarpae]
TRINITY_DN168885_c0_g1_i1	VXC31200.1	1.49E-16	conserved hypothetical protein [Burkholderia sp. 8Y]
TRINITY_DN101872_c0_g1_i1	WP_068467488.1	1.13E-08	RMD1 family protein [Parachlamydia sp. C2]
TRINITY_DN101613_c0_g1_i1	MBR2285283.1	6.99E-06	anion permease [Paludibacteraceae bacterium]
TRINITY_DN110614_c0_g1_i1	WP_070151516.1	3.37E-22	phosphate regulon transcriptional regulator PhoB [Sphingobium phenoxybenzoativorans]
TRINITY_DN110614_c0_g1_i2	WP_164078977.1	1.67E-29	response regulator, partial [Stenotrophomonas maltophilia]
TRINITY_DN103825_c0_g1_i2	XP_041465966.1	1.37E-20	ankyrin repeat domain-containing protein 29-like [Lytechinus variegatus]
TRINITY_DN103825_c0_g1_i2	XP_041465966.1	9.04E-08	ankyrin repeat domain-containing protein 29-like [Lytechinus variegatus]
TRINITY_DN103825_c0_g1_i2	XP_041465966.1	1.57E-07	ankyrin repeat domain-containing protein 29-like [Lytechinus variegatus]
TRINITY_DN103825_c0_g1_i6	XP_011673447.2	6.17E-19	ankyrin repeat domain-containing protein 50-like [Strongylocentrotus purpuratus]
TRINITY_DN103825_c0_g1_i6	XP_011673447.2	5.75E-10	ankyrin repeat domain-containing protein 50-like [Strongylocentrotus purpuratus]
TRINITY_DN306459_c0_g1_i1	MBL8788430.1	4.44E-27	VOC family protein [Deltaproteobacteria bacterium]
TRINITY_DN341964_c0_g1_i1	KAB1265972.1	1.47E-38	Carboxypeptidase D [Camelus dromedarius]
TRINITY_DN320484_c0_g1_i1	OEU22037.1	2.48E-17	hypothetical protein FRACYDRAFT_232192 [Fragilariopsis cylindrus CCMP1102]
TRINITY_DN53929_c1_g1_i1	CCA20060.1	8.52E-08	conserved unknown protein putative [Albugo laibachii Nc14]
TRINITY_DN4416_c1_g1_i1	KAF9328496.1	5.05E-20	hypothetical protein BGZ91_000971, partial [Linnemannia 98aveolat]
TRINITY_DN4416_c1_g1_i1	KAF9328496.1	5.87E-16	hypothetical protein BGZ91_000971, partial [Linnemannia 98aveolat]
TRINITY_DN4416_c1_g1_i1	KAF9328496.1	6.89E-15	hypothetical protein BGZ91_000971, partial [Linnemannia 98aveolat]
TRINITY_DN41880_c0_g1_i1	VWX60559.1	1.95E-40	conserved hypothetical protein [Burkholderiales bacterium 8X]
TRINITY_DN41880_c0_g1_i2	VWX60559.1	1.87E-42	conserved hypothetical protein [Burkholderiales bacterium 8X]
TRINITY_DN15234_c0_g1_i1	MBA3417570.1	1.97E-07	serine/threonine-protein kinase [Geodermatophilaceae bacterium]
TRINITY_DN1240_c1_g1_i10	TRY97330.1	1.69E-21	hypothetical protein DNTS_020614 [Danionella translucida]
TRINITY_DN1240_c1_g1_i3	TRY97330.1	4.11E-21	hypothetical protein DNTS_020614 [Danionella translucida]
TRINITY_DN1240_c1_g1_i8	TRY97330.1	4.12E-21	hypothetical protein DNTS_020614 [Danionella translucida]
TRINITY_DN1240_c1_g1_i9	TRY97330.1	6.41E-21	hypothetical protein DNTS_020614 [Danionella translucida]
TRINITY_DN21430_c1_g3_i1	KAF1858519.1	5.73E-69	hypothetical protein Lal_00015036 [Lupinus albus]
TRINITY_DN21430_c1_g3_i1	KAF1858519.1	2.35E-12	hypothetical protein Lal_00015036 [Lupinus albus]

TRINITY_DN80964_c0_g1_i1	SIT47157.1	4.77E-27	hypothetical protein BN2475_710089 [Paraburkholderia ribeironis]
TRINITY_DN80964_c0_g1_i2	SIT47157.1	2.88E-26	hypothetical protein BN2475_710089 [Paraburkholderia ribeironis]
TRINITY_DN22260_c0_g1_i12	XP_012756544.1	3.45E-82	hypothetical protein SAMD00019534_029570 [Acytostelium subglobosum LB1]
TRINITY_DN22260_c0_g1_i14	XP_012756544.1	2.60E-82	hypothetical protein SAMD00019534_029570 [Acytostelium subglobosum LB1]
TRINITY_DN22260_c0_g1_i19	XP_012756544.1	3.63E-82	hypothetical protein SAMD00019534_029570 [Acytostelium subglobosum LB1]
TRINITY_DN22260_c0_g1_i20	XP_012756544.1	4.26E-82	hypothetical protein SAMD00019534_029570 [Acytostelium subglobosum LB1]
TRINITY_DN22260_c0_g1_i22	XP_012756544.1	1.02E-81	hypothetical protein SAMD00019534_029570 [Acytostelium subglobosum LB1]
TRINITY_DN22260_c0_g1_i24	XP_012756544.1	7.92E-83	hypothetical protein SAMD00019534_029570 [Acytostelium subglobosum LB1]
TRINITY_DN22260_c0_g1_i7	XP_012756544.1	1.80E-81	hypothetical protein SAMD00019534_029570 [Acytostelium subglobosum LB1]
TRINITY_DN37329_c3_g1_i1	MCA9453434.1	7.70E-06	hypothetical protein [Nitrospira sp.]
TRINITY_DN32900_c0_g1_i10	CAB5278204.1	6.13E-43	hypothetical protein IST495A_06018 [Burkholderia multivorans]
TRINITY_DN32900_c0_g1_i2	CBA31922.1	3.30E-28	hypothetical protein Csp_D29540 [Curvibacter putative symbiont of Hydra magnipapillata]
TRINITY_DN32900_c0_g1_i3	BAJ11784.1	8.97E-30	dehydration responsive protein, partial [Corchorus olitorius]
TRINITY_DN32900_c0_g1_i4	BAJ11784.1	1.13E-26	dehydration responsive protein, partial [Corchorus olitorius]
TRINITY_DN32900_c0_g1_i5	CBA31922.1	9.95E-29	hypothetical protein Csp_D29540 [Curvibacter putative symbiont of Hydra magnipapillata]
TRINITY_DN32900_c0_g1_i6	CBA31922.1	1.18E-30	hypothetical protein Csp_D29540 [Curvibacter putative symbiont of Hydra magnipapillata]
TRINITY_DN32900_c0_g1_i6	CBA31922.1	1.18E-30	hypothetical protein Csp_D29540 [Curvibacter putative symbiont of Hydra magnipapillata]
TRINITY_DN32900_c0_g1_i7	CBA31922.1	2.99E-28	hypothetical protein Csp_D29540 [Curvibacter putative symbiont of Hydra magnipapillata]
TRINITY_DN84431_c1_g1_i1	CAE7192839.1	1.07E-12	anks1b [Symbiodinium natans]
TRINITY_DN8007_c0_g1_i1	CAJ30045.1	5.23E-45	conserved hypothetical protein [Magnetospirillum gryphiswaldense MSR-1]
TRINITY_DN8007_c0_g1_i3	EAZ75437.1	1.96E-89	conserved hypothetical protein [Vibrio cholerae B33]
TRINITY_DN8007_c0_g1_i4	KUK40615.1	3.47E-43	Uncharacterized protein XD69_1335 [Clostridia bacterium 62_21]
TRINITY_DN8007_c0_g1_i5	KMS65245.1	1.89E-58	hypothetical protein BVRB_037930, partial [Beta vulgaris subsp. Vulgaris]
TRINITY_DN8007_c0_g1_i5	KMS65245.1	1.89E-58	hypothetical protein BVRB_037930, partial [Beta vulgaris subsp. Vulgaris]
TRINITY_DN11676_c2_g4_i1	OEU22037.1	1.32E-08	hypothetical protein FRACYDRAFT_232192 [Fragilariopsis cylindrus CCMP1102]
TRINITY_DN69807_c0_g1_i1	MBQ5160593.1	1.89E-42	hypothetical protein [Neisseria meningitidis]
TRINITY_DN69807_c0_g1_i10	EEV67789.1	7.17E-48	conserved hypothetical protein [Staphylococcus aureus A9719]
TRINITY_DN69807_c0_g1_i11	MBQ5160593.1	5.82E-42	hypothetical protein [Neisseria meningitidis]
TRINITY_DN69807_c0_g1_i13	KAG9455510.1	2.98E-42	hypothetical protein H6P81_000018 [Aristolochia fimbriata]
TRINITY_DN69807_c0_g1_i14	EFQ45725.1	2.66E-33	hypothetical protein LBGG_01984 [Lactobacillus gasseri MV-22]

TRINITY_DN69807_c0_g1_i2	MBQ5160593.1	4.89E-44	hypothetical protein [Neisseria meningitidis]
TRINITY_DN69807_c0_g1_i3	EFQ45725.1	1.74E-29	hypothetical protein LBGG_01984 [Lactobacillus gasseri MV-22]
TRINITY_DN69807_c0_g1_i6	AYM55783.1	1.28E-36	hypothetical protein [uncultured bacterium]
TRINITY_DN69807_c0_g1_i7	MBQ5160593.1	4.16E-47	hypothetical protein [Neisseria meningitidis]
TRINITY_DN69807_c0_g1_i9	KAG9455510.1	5.57E-49	hypothetical protein H6P81_000018 [Aristolochia fimbriata]
TRINITY_DN40078_c0_g1_i3	CUG92858.1	8.99E-35	casein kinase, putative [Bodo saltans]
TRINITY_DN40078_c0_g1_i4	CUG92858.1	1.56E-35	casein kinase, putative [Bodo saltans]
TRINITY_DN9316_c12_g1_i1	APF46140.1	1.13E-43	plastid coproporphyrinogen oxidase isoform 4, partial [Euglena gracilis]
TRINITY_DN304_c0_g2_i1	VWX60559.1	1.06E-45	conserved hypothetical protein [Burkholderiales bacterium 8X]
TRINITY_DN7086_c0_g1_i2	KAF1857789.1	8.89E-25	hypothetical protein Lal_00041211 [Lupinus albus]
TRINITY_DN7086_c0_g1_i3	KAF1857789.1	1.21E-22	hypothetical protein Lal_00041211 [Lupinus albus]
TRINITY_DN7086_c0_g1_i4	KAF1858446.1	2.73E-30	hypothetical protein Lal_00014961 [Lupinus albus]
TRINITY_DN7086_c3_g2_i1	CCG06586.1	8.59E-61	unnamed protein product, partial [Pararhodospirillum photometricum DSM 122]
TRINITY_DN7086_c3_g2_i3	EDP24804.1	4.80E-42	hypothetical protein PEPMIC_00056 [Parvimonas micra ATCC 33270]
TRINITY_DN2267_c4_g1_i1	KAF8950195.1	1.87E-15	hypothetical protein BGZ46_004675, partial [Entomortierella lignicola]
TRINITY_DN2267_c4_g1_i1	KAF8950195.1	2.09E-09	hypothetical protein BGZ46_004675, partial [Entomortierella lignicola]
TRINITY_DN2267_c4_g1_i1	KAF8950195.1	1.90E-08	hypothetical protein BGZ46_004675, partial [Entomortierella lignicola]
TRINITY_DN2267_c4_g1_i1	KAF8950195.1	2.13E-08	hypothetical protein BGZ46_004675, partial [Entomortierella lignicola]
TRINITY_DN2267_c4_g1_i1	KAF8950195.1	5.70E-08	hypothetical protein BGZ46_004675, partial [Entomortierella lignicola]
TRINITY_DN2267_c4_g1_i1	KAF8950195.1	4.10E-07	hypothetical protein BGZ46_004675, partial [Entomortierella lignicola]
TRINITY_DN2267_c4_g1_i2	EDO43196.1	1.25E-08	predicted protein, partial [Nematostella vectensis]
TRINITY_DN2267_c4_g1_i2	EDO43196.1	1.25E-08	predicted protein, partial [Nematostella vectensis]
TRINITY_DN2267_c4_g1_i2	EDO43196.1	1.62E-08	predicted protein, partial [Nematostella vectensis]
TRINITY_DN2267_c4_g1_i2	EDO43196.1	2.11E-08	predicted protein, partial [Nematostella vectensis]
TRINITY_DN2267_c4_g1_i2	EDO43196.1	2.11E-08	predicted protein, partial [Nematostella vectensis]
TRINITY_DN2267_c4_g1_i2	EDO43196.1	2.11E-08	predicted protein, partial [Nematostella vectensis]
TRINITY_DN2267_c4_g1_i2	EDO43196.1	2.11E-08	predicted protein, partial [Nematostella vectensis]
TRINITY_DN2267_c4_g1_i2	EDO43196.1	2.11E-08	predicted protein, partial [Nematostella vectensis]
TRINITY_DN2267_c4_g1_i2	EDO43196.1	2.11E-08	predicted protein, partial [Nematostella vectensis]
TRINITY_DN2267_c4_g1_i2	EDO43196.1	3.38E-08	predicted protein, partial [Nematostella vectensis]
TRINITY_DN2267_c4_g1_i2	EDO43196.1	7.19E-07	predicted protein, partial [Nematostella vectensis]
TRINITY_DN2267_c4_g1_i2	EDO43196.1	1.96E-06	predicted protein, partial [Nematostella vectensis]
TRINITY_DN2267_c4_g1_i2	EDO43196.1	4.25E-06	predicted protein, partial [Nematostella vectensis]
TRINITY_DN71410_c3_g1_i1	KAH3765117.1	4.18E-09	putative E3 ubiquitin-protein ligase 100aveola-1 [Pelomyxa schiedti]
TRINITY_DN79385_c1_g1_i1	PNX55782.1	9.50E-20	long chain acyl-CoA synthetase 8-like protein, partial [Trifolium pratense]
TRINITY_DN10642_c0_g1_i1	XP_020900175.1	5.89E-22	E3 ubiquitin-protein ligase TRIM9 [Exaiptasia diaphana]

TRINITY_DN30550_c0_g1_i2	MBS7312439.1	5.80E-10	aldehyde dehydrogenase family protein [Bacteroidales bacterium]
TRINITY_DN76342_c0_g2_i1	XP_020631299.1	3.44E-10	E3 ubiquitin-protein ligase TRIM71-like [Orbicella 101aveolate]
TRINITY_DN76342_c0_g2_i2	XP_020631354.1	3.46E-09	tripartite motif-containing protein 45-like [Orbicella 101aveolate]
TRINITY_DN76342_c0_g2_i8	XP_020631299.1	3.30E-10	E3 ubiquitin-protein ligase TRIM71-like [Orbicella 101aveolate]

Table S18: blastx results of differentially expressed transcripts identified by DESeq2 for sulfur pretreated *Euglena gracilis* cultures in the presence and absence of $CdCl_2$ using the NCBI SWISS-PROT protein database.

SWISS_Isoform Id	SWISS_sseqid	SWISS_evalue	SWISS_description
TRINITY_DN158535_c10_g1_i1	Q92260.1	6.52E-20	RecName: Full=Heat shock 70 kDa protein; AltName: Allergen=Pen c 19 [Penicillium citrinum]
TRINITY_DN166865_c1_g1_i1	Q9HZP6.1	4.45E-27	RecName: Full=Electron transfer flavoprotein subunit beta; Short=Beta-ETF; AltName: Full=Electron transfer flavoprotein small subunit; Short=ETFSS [Pseudomonas aeruginosa PAO1]
TRINITY_DN110614_c0_g1_i1	Q93CB8.2	2.98E-08	RecName: Full=DNA-binding response regulator MtrA [Mycobacterium avium subsp. paratuberculosis K-10]
TRINITY_DN110614_c0_g1_i2	P0AFJ5.1	2.08E-15	RecName: Full=Phosphate regulon transcriptional regulatory protein PhoB [Escherichia coli K-12]
TRINITY_DN103825_c0_g1_i2	Q9ULJ7.4	6.11E-13	RecName: Full=Ankyrin repeat domain-containing protein 50 [Homo sapiens]
TRINITY_DN103825_c0_g1_i2	Q9ULJ7.4	9.57E-13	RecName: Full=Ankyrin repeat domain-containing protein 50 [Homo sapiens]
TRINITY_DN103825_c0_g1_i2	Q9ULJ7.4	4.03E-07	RecName: Full=Ankyrin repeat domain-containing protein 50 [Homo sapiens]
TRINITY_DN103825_c0_g1_i2	Q9ULJ7.4	4.24E-07	RecName: Full=Ankyrin repeat domain-containing protein 50 [Homo sapiens]
TRINITY_DN103825_c0_g1_i6	Q6JAN1.1	1.11E-12	RecName: Full=Inversin; AltName: Full=Inversion of embryo turning protein; AltName: Full=Nephrocystin-2 [Canis lupus familiaris]
TRINITY_DN341964_c0_g1_i1	O75976.2	1.31E-40	RecName: Full=Carboxypeptidase D; AltName: Full=Metallocarboxypeptidase D; AltName: Full=gp180; Flags: Precursor [Homo sapiens]
TRINITY_DN341964_c0_g1_i1	O75976.2	1.43E-09	RecName: Full=Carboxypeptidase D; AltName: Full=Metallocarboxypeptidase D; AltName: Full=gp180; Flags: Precursor [Homo sapiens]
TRINITY_DN341964_c0_g1_i1	O75976.2	6.98E-08	RecName: Full=Carboxypeptidase D; AltName: Full=Metallocarboxypeptidase D; AltName: Full=gp180; Flags: Precursor [Homo sapiens]
TRINITY_DN4416_c1_g1_i1	Q5DU56.2	8.68E-13	RecName: Full=Protein NLRC3 [Mus musculus]
TRINITY_DN4416_c1_g1_i1	Q5DU56.2	1.47E-11	RecName: Full=Protein NLRC3 [Mus musculus]
TRINITY_DN4416_c1_g1_i1	Q5DU56.2	3.91E-11	RecName: Full=Protein NLRC3 [Mus musculus]
TRINITY_DN4416_c1_g1_i1	Q5DU56.2	8.65E-10	RecName: Full=Protein NLRC3 [Mus musculus]
TRINITY_DN4416_c1_g1_i1	Q5DU56.2	1.07E-09	RecName: Full=Protein NLRC3 [Mus musculus]
TRINITY_DN4416_c1_g1_i1	Q5DU56.2	2.06E-09	RecName: Full=Protein NLRC3 [Mus musculus]
TRINITY_DN4416_c1_g1_i1	Q5DU56.2	1.61E-08	RecName: Full=Protein NLRC3 [Mus musculus]
TRINITY_DN4416_c1_g1_i1	Q5DU56.2	1.10E-07	RecName: Full=Protein NLRC3 [Mus musculus]
TRINITY_DN15234_c0_g1_i1	O22938.1	9.74E-06	RecName: Full=Leucine-rich repeat receptor-like tyrosine-protein kinase PXC3; AltName: Full=Protein PXY/TDR-CORRELATED 3; Flags: Precursor [Arabidopsis thaliana]
TRINITY_DN1240_c1_g1_i10	Q80VI1.1	8.11E-17	RecName: Full=E3 ubiquitin-protein ligase TRIM56; AltName: Full=RING-type E3 ubiquitin transferase TRIM56; AltName: Full=Tripartite motif-containing protein 56 [Mus musculus]

TRINITY_DN1240_c1_g1_i3	Q80VI1.1	1.70E-16	RecName: Full=E3 ubiquitin-protein ligase TRIM56; AltName: Full=RING-type E3 ubiquitin transferase TRIM56; AltName: Full=Tripartite motif-containing protein 56 [Mus musculus]
TRINITY_DN1240_c1_g1_i8	Q80VI1.1	1.70E-16	RecName: Full=E3 ubiquitin-protein ligase TRIM56; AltName: Full=RING-type E3 ubiquitin transferase TRIM56; AltName: Full=Tripartite motif-containing protein 56 [Mus musculus]
TRINITY_DN1240_c1_g1_i9	Q80VI1.1	1.36E-16	RecName: Full=E3 ubiquitin-protein ligase TRIM56; AltName: Full=RING-type E3 ubiquitin transferase TRIM56; AltName: Full=Tripartite motif-containing protein 56 [Mus musculus]
TRINITY_DN22260_c0_g1_i12	Q6L8G0.1	6.90E-27	RecName: Full=Zinc transporter 5; AltName: Full=ZRT/IRT-like protein 5; Short=OsZIP5; Flags: Precursor [Oryza sativa Japonica Group]
TRINITY_DN22260_c0_g1_i14	Q6L8G0.1	6.97E-27	RecName: Full=Zinc transporter 5; AltName: Full=ZRT/IRT-like protein 5; Short=OsZIP5; Flags: Precursor [Oryza sativa Japonica Group]
TRINITY_DN22260_c0_g1_i19	Q6L8G0.1	6.98E-27	RecName: Full=Zinc transporter 5; AltName: Full=ZRT/IRT-like protein 5; Short=OsZIP5; Flags: Precursor [Oryza sativa Japonica Group]
TRINITY_DN22260_c0_g1_i20	Q6L8G0.1	7.75E-27	RecName: Full=Zinc transporter 5; AltName: Full=ZRT/IRT-like protein 5; Short=OsZIP5; Flags: Precursor [Oryza sativa Japonica Group]
TRINITY_DN22260_c0_g1_i22	Q6L8G0.1	1.73E-26	RecName: Full=Zinc transporter 5; AltName: Full=ZRT/IRT-like protein 5; Short=OsZIP5; Flags: Precursor [Oryza sativa Japonica Group]
TRINITY_DN22260_c0_g1_i24	Q6L8G0.1	5.16E-27	RecName: Full=Zinc transporter 5; AltName: Full=ZRT/IRT-like protein 5; Short=OsZIP5; Flags: Precursor [Oryza sativa Japonica Group]
TRINITY_DN22260_c0_g1_i7	Q6L8G0.1	1.95E-26	RecName: Full=Zinc transporter 5; AltName: Full=ZRT/IRT-like protein 5; Short=OsZIP5; Flags: Precursor [Oryza sativa Japonica Group]
TRINITY_DN663_c3_g1_i2	Q9NUK0.2	7.01E-06	RecName: Full=Muscleblind-like protein 3; AltName: Full=Cys3His CCG1-required protein; AltName: Full=Muscleblind-like X-linked protein; AltName: Full=Protein HCHCR [Homo sapiens]
TRINITY_DN40078_c0_g1_i3	B9VVJ6.1	2.28E-31	RecName: Full=Casein kinase I isoform 2 [Trypanosoma brucei brucei]
TRINITY_DN40078_c0_g1_i4	B9VVJ6.1	7.98E-31	RecName: Full=Casein kinase I isoform 2 [Trypanosoma brucei brucei]
TRINITY_DN9316_c12_g1_i1	Q93Z96.1	9.72E-24	RecName: Full=Coproporphyrinogen-III oxidase 2, chloroplastic; Short=AtCPO-II; Short=Coprogen oxidase; Short=Coproporphyrinogenase; Flags: Precursor [Arabidopsis thaliana]
TRINITY_DN2267_c4_g1_i1	Q7RTR2.2	7.40E-11	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN2267_c4_g1_i1	Q7RTR2.2	4.33E-10	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN2267_c4_g1_i1	Q7RTR2.2	1.82E-09	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]

TRINITY_DN2267_c4_g1_i1	Q7RTR2.2	1.95E-09	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN2267_c4_g1_i1	Q7RTR2.2	1.21E-08	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN2267_c4_g1_i1	Q7RTR2.2	1.09E-06	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN2267_c4_g1_i2	Q7RTR2.2	2.79E-06	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN2267_c4_g1_i2	Q7RTR2.2	4.42E-06	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN71410_c3_g1_i1	Q9LVX0.1	3.55E-07	RecName: Full=Probable E3 ubiquitin-protein ligase ARI3; AltName: Full=ARIADNE-like protein ARI3; AltName: Full=Protein ariadne homolog 3; AltName: Full=RING-type E3 ubiquitin transferase ARI3 [Arabidopsis thaliana]
TRINITY_DN79385_c1_g1_i1	Q9CAP8.1	3.92E-20	RecName: Full=Long chain acyl-CoA synthetase 9, chloroplastic [Arabidopsis thaliana]
TRINITY_DN10642_c0_g1_i1	O15344.1	2.14E-18	RecName: Full=E3 ubiquitin-protein ligase Midline-1; AltName: Full=Midin; AltName: Full=Putative transcription factor XPRF; AltName: Full=RING finger protein 59; AltName: Full=RING finger protein Midline-1; AltName: Full=RING-type E3 ubiquitin transferase Midline-1; AltName: Full=Tripartite motif-containing protein 18 [Homo sapiens]
TRINITY_DN30550_c0_g1_i2	Q70DU8.2	3.63E-07	RecName: Full=Aldehyde dehydrogenase family 3 member H1; Short=AtALDH4; Short=Ath-ALDH4 [Arabidopsis thaliana]

Table S19: blastx results of differentially expressed transcripts identified by DESeq2 for sulfur pretreated *E. gracilis* cultures in the presence and absence of $CdCl_2$ using the Ensembl *Arabidopsis thaliana* database.

Athaliana_Isoform Id	Athaliana_sseqid	Athaliana_evalue	Athaliana_description
TRINITY_DN158535_c10_g1_i1	AT3G12580.1	5.62E-20	Probable mediator of RNA polymerase II transcription subunit 37c
TRINITY_DN166865_c1_g1_i1	AT5G43430.3	2.22E-23	electron transfer flavoprotein beta
TRINITY_DN103825_c0_g1_i2	AT2G03430.1	6.04E-10	:Ankyrin repeat family protein
TRINITY_DN103825_c0_g1_i6	AT4G19150.1	3.27E-09	Ankyrin repeat family protein
TRINITY_DN341964_c0_g1_i1	AT1G71696.1	6.98E-07	Carboxypeptidase SOL1
TRINITY_DN4416_c1_g1_i1	AT1G10510.1	6.94E-10	RNI-like superfamily protein
TRINITY_DN4416_c1_g1_i1	AT1G10510.1	1.35E-09	RNI-like superfamily protein
TRINITY_DN4416_c1_g1_i1	AT1G10510.1	3.42E-09	RNI-like superfamily protein
TRINITY_DN4416_c1_g1_i1	AT1G10510.1	1.65E-08	RNI-like superfamily protein
TRINITY_DN4416_c1_g1_i1	AT1G10510.1	2.54E-08	RNI-like superfamily protein
TRINITY_DN4416_c1_g1_i1	AT1G10510.1	3.68E-07	RNI-like superfamily protein
TRINITY_DN4416_c1_g1_i1	AT1G10510.1	1.58E-05	RNI-like superfamily protein
TRINITY_DN15234_c0_g1_i1	AT5G24080.2	3.04E-07	G-type lectin S-receptor-like serine/threonine-protein kinase At5g24080
TRINITY_DN1240_c1_g1_i10	AT5G13530.1	8.87E-06	E3 ubiquitin-protein ligase KEG
TRINITY_DN1240_c1_g1_i3	AT5G13530.1	1.50E-05	E3 ubiquitin-protein ligase KEG
TRINITY_DN1240_c1_g1_i8	AT5G13530.1	1.50E-05	E3 ubiquitin-protein ligase KEG
TRINITY_DN1240_c1_g1_i9	AT5G13530.1	1.40E-05	E3 ubiquitin-protein ligase KEG
TRINITY_DN22260_c0_g1_i12	AT4G33020.2	1.81E-25	ZIP9
TRINITY_DN22260_c0_g1_i14	AT4G33020.2	1.90E-25	ZIP9
TRINITY_DN22260_c0_g1_i19	AT4G33020.2	1.82E-25	ZIP9
TRINITY_DN22260_c0_g1_i20	AT4G33020.2	2.08E-25	ZIP9
TRINITY_DN22260_c0_g1_i22	AT4G33020.2	2.42E-25	ZIP9
TRINITY_DN22260_c0_g1_i24	AT4G33020.2	1.46E-25	ZIP9
TRINITY_DN22260_c0_g1_i7	AT4G33020.2	2.69E-25	ZIP9
TRINITY_DN45428_c0_g2_i2	AT2G44020.1	5.50E-05	Expressed protein
TRINITY_DN40078_c0_g1_i3	AT4G14340.1	4.75E-29	Casein kinase 1-like protein 11
TRINITY_DN40078_c0_g1_i4	AT1G04440.1	5.47E-29	Casein kinase 1-like protein 13
TRINITY_DN9316_c12_g1_i1	AT4G03205.1	1.12E-24	Coproporphyrinogen III oxidase
TRINITY_DN2267_c4_g1_i1	AT1G10510.1	7.95E-10	RNI-like superfamily protein
TRINITY_DN2267_c4_g1_i1	AT1G10510.1	1.79E-09	RNI-like superfamily protein
TRINITY_DN2267_c4_g1_i1	AT1G10510.1	1.44E-07	RNI-like superfamily protein
TRINITY_DN2267_c4_g1_i1	AT1G10510.1	2.38E-07	RNI-like superfamily protein
TRINITY_DN2267_c4_g1_i1	AT1G10510.1	2.60E-07	RNI-like superfamily protein
TRINITY_DN2267_c4_g1_i1	AT1G10510.1	3.37E-05	RNI-like superfamily protein
TRINITY_DN2267_c4_g1_i1	AT1G10510.1	3.09E-04	RNI-like superfamily protein
TRINITY_DN2267_c4_g1_i1	AT1G10510.1	8.91E-04	RNI-like superfamily protein
TRINITY_DN2267_c4_g1_i2	AT1G10510.1	5.18E-06	RNI-like superfamily protein
TRINITY_DN2267_c4_g1_i2	AT1G10510.1	1.15E-04	RNI-like superfamily protein
TRINITY_DN2267_c4_g1_i2	AT1G10510.1	2.81E-04	RNI-like superfamily protein
TRINITY_DN2267_c4_g1_i2	AT1G10510.1	4.57E-04	RNI-like superfamily protein
TRINITY_DN71410_c3_g1_i1	AT3G27710.1	4.08E-08	Probable E3 ubiquitin-protein ligase ARI3
TRINITY_DN79385_c1_g1_i1	AT1G77590.2	4.40E-21	Long chain acyl-CoA synthetase 9, chloroplastic
TRINITY_DN10642_c0_g1_i1	AT5G22750.1	4.53E-04	DNA repair protein RAD5A
TRINITY_DN30550_c0_g1_i2	AT1G44170.3	3.84E-08	Aldehyde dehydrogenase family 3 member H1
TRINITY_DN76342_c0_g2_i1	AT5G05130.1	6.37E-04	Putative SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 3-like 1
TRINITY_DN76342_c0_g2_i2	AT5G05130.1	5.32E-04	Putative SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 3-like 1
TRINITY_DN76342_c0_g2_i8	AT5G05130.1	6.19E-04	Putative SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 3-like 1

Table S20: blastx results of differentially expressed transcripts identified by DESeq2 for sulfur pretreated *Euglena gracilis* cultures in the presence and absence of $CdCl_2$ using the Ensembl *Chlamydomonas reinhardtti* database.

Creinhardtii_Isoform Id	Creinhardtii_sseqid	Creinhardtii_evalue	Creinhardtii_description
TRINITY_DN158535_c10_g1_i1	PNW79920	2.39E-17	hypothetical protein
TRINITY_DN166865_c1_g1_i1	PNW76352	1.11E-21	hypothetical protein
TRINITY_DN101872_c0_g1_i1	PNW78931	1.79E-05	hypothetical protein
TRINITY_DN110614_c0_g1_i2	PNW76815	3.94E-04	hypothetical protein
TRINITY_DN103825_c0_g1_i2	PNW83376	9.25E-17	hypothetical protein
TRINITY_DN103825_c0_g1_i2	PNW83376	7.55E-13	hypothetical protein
TRINITY_DN103825_c0_g1_i2	PNW83376	1.44E-12	hypothetical protein
TRINITY_DN103825_c0_g1_i2	PNW83376	6.11E-11	hypothetical protein
TRINITY_DN103825_c0_g1_i2	PNW83376	2.80E-10	hypothetical protein
TRINITY_DN103825_c0_g1_i2	PNW83376	1.18E-09	hypothetical protein
TRINITY_DN103825_c0_g1_i2	PNW83376	2.28E-05	hypothetical protein
TRINITY_DN103825_c0_g1_i6	PNW83376	5.70E-17	hypothetical protein
TRINITY_DN103825_c0_g1_i6	PNW83376	8.65E-13	hypothetical protein
TRINITY_DN103825_c0_g1_i6	PNW83376	1.45E-11	hypothetical protein
TRINITY_DN103825_c0_g1_i6	PNW83376	7.66E-09	hypothetical protein
TRINITY_DN103825_c0_g1_i6	PNW83376	8.87E-09	hypothetical protein
TRINITY_DN103825_c0_g1_i6	PNW83376	2.08E-05	hypothetical protein
TRINITY_DN341964_c0_g1_i1	PNW80934	4.03E-04	hypothetical protein
TRINITY_DN320484_c0_g1_i1	PNW84031	2.53E-08	hypothetical protein
TRINITY_DN4416_c1_g1_i1	PNW88806	5.01E-12	hypothetical protein
TRINITY_DN4416_c1_g1_i1	PNW88806	8.45E-10	hypothetical protein
TRINITY_DN15234_c0_g1_i1	PNW77419	2.20E-08	hypothetical protein
TRINITY_DN26392_c1_g3_i1	PNW75014	1.31E-04	hypothetical protein
TRINITY_DN22260_c0_g1_i12	PNW73817	8.57E-15	hypothetical protein
TRINITY_DN22260_c0_g1_i14	PNW73817	1.11E-14	hypothetical protein
TRINITY_DN22260_c0_g1_i19	PNW73817	8.62E-15	hypothetical protein
TRINITY_DN22260_c0_g1_i20	PNW73817	1.18E-14	hypothetical protein
TRINITY_DN22260_c0_g1_i22	PNW73817	3.28E-14	hypothetical protein
TRINITY_DN22260_c0_g1_i24	PNW73817	9.57E-15	hypothetical protein
TRINITY_DN22260_c0_g1_i7	PNW73817	3.51E-14	hypothetical protein
TRINITY_DN37329_c3_g1_i1	PNW84142	5.62E-07	hypothetical protein
TRINITY_DN40078_c0_g1_i3	PNW76048	1.12E-28	hypothetical protein
TRINITY_DN40078_c0_g1_i4	PNW76048	4.94E-28	hypothetical protein
TRINITY_DN9316_c12_g1_i1	PNW86414	6.33E-23	hypothetical protein
TRINITY_DN2267_c4_g1_i1	PNW77156	3.45E-08	hypothetical protein
TRINITY_DN2267_c4_g1_i1	PNW77156	3.18E-04	hypothetical protein
TRINITY_DN2267_c4_g1_i2	PNW76530	2.33E-06	hypothetical protein
TRINITY_DN71410_c3_g1_i1	PNW72103	5.04E-10	hypothetical protein
TRINITY_DN79385_c1_g1_i1	PNW85346	3.69E-10	hypothetical protein
TRINITY_DN10642_c0_g1_i1	PNW83336	9.89E-05	hypothetical protein

Table S21: blastx results of differentially expressed transcripts identified by DESeq2 for sulfur pretreated *Euglena gracilis* cultures in the presence and absence of $CdCl_2$ using the Ensembl *Synechcytis sp.* database.

Synechocytis_Isoform Id	Synechocytis_sseqid	Synechocytis_evalue	Synechocytis_description
TRINITY_DN110614_c0_g1_i1	AIE72797	1.68E-04	Nitrogen-responsive response regulator NrrA
TRINITY_DN110614_c0_g1_i2	AIE72797	9.40E-10	Nitrogen-responsive response regulator NrrA
TRINITY_DN103825_c0_g1_i2	AIE75326	2.19E-07	Ankyrin
TRINITY_DN103825_c0_g1_i6	AIE75326	1.12E-05	Ankyrin
TRINITY_DN15234_c0_g1_i1	AIE74158	1.82E-04	serine/threonine protein kinase
TRINITY_DN9316_c12_g1_i1	AIE74482	1.54E-12	:Coproporphyrinogen III oxidase, aerobic
TRINITY_DN30550_c0_g1_i2	AIE74710	3.07E-08	Aldehyde dehydrogenase
TRINITY_DN30550_c0_g1_i3	AIE74710	2.30E-05	Aldehyde dehydrogenase

Table S22: blastx results of differentially expressed transcripts identified by DESeq2 for sulfur pretreated *Euglena gracilis* cultures in the presence and absence of $CdCl_2$ using the Ensembl *Homo sapiens* database.

Hsapiens_Isoform Id	Hsapiens_sseqid	Hsapiens_evalue	Hsapiens_description
TRINITY_DN158535_c10_g1_i1	ENSP00000433584.1	6.91E-18	heat shock protein family A (Hsp70) member 8
TRINITY_DN166865_c1_g1_i1	ENSP00000311930.3	7.34E-24	electron transfer flavoprotein subunit beta
TRINITY_DN103825_c0_g1_i2	ENSP00000425658.1	1.54E-13	ankyrin repeat domain 50
TRINITY_DN103825_c0_g1_i2	ENSP00000425658.1	2.42E-13	ankyrin repeat domain 50
TRINITY_DN103825_c0_g1_i2	ENSP00000425658.1	1.91E-12	ankyrin repeat domain 50
TRINITY_DN103825_c0_g1_i2	ENSP00000425658.1	1.77E-08	ankyrin repeat domain 50
TRINITY_DN103825_c0_g1_i2	ENSP00000425658.1	1.02E-07	ankyrin repeat domain 50
TRINITY_DN103825_c0_g1_i2	ENSP00000425658.1	1.07E-07	ankyrin repeat domain 50

Table S23: blastx results of differentially expressed transcripts identified by DESeq2 for sulfur pretreated *Euglena gracilis* cultures in the presence and absence of $CdCl_2$ using the Ensembl *Trypanosoma brucei* database.

Tbrucei_Isoforms Id	Tbrucei_sseqid	Tbrucei_evalue	Tbrucei_description
TRINITY_DN158535_c10_g1_i1	EAN80086	1.03E-15	heat shock protein 70
TRINITY_DN166865_c1_g1_i1	EAN78998	7.80E-27	electron transfer flavoprotein, putative
TRINITY_DN101872_c0_g1_i1	EAN79339	2.12E-07	hypothetical protein, conserved
TRINITY_DN103825_c0_g1_i2	EAN77209	6.41E-07	ankyrin repeat protein, putative
TRINITY_DN103825_c0_g1_i6	EAN77209	5.33E-07	ankyrin repeat protein, putative
TRINITY_DN4416_c1_g1_i1	AAZ10329	5.21E-10	hypothetical protein, conserved
TRINITY_DN4416_c1_g1_i1	AAZ10329	7.81E-08	hypothetical protein, conserved
TRINITY_DN4416_c1_g1_i1	AAZ10329	3.05E-07	hypothetical protein, conserved
TRINITY_DN4416_c1_g1_i1	AAZ10329	4.13E-06	hypothetical protein, conserved
TRINITY_DN4416_c1_g1_i1	AAZ10329	1.20E-05	:hypothetical protein, conserved
TRINITY_DN4416_c1_g1_i1	AAZ10329	1.49E-04	hypothetical protein, conserved
TRINITY_DN15234_c0_g1_i1	EAN78798	3.27E-07	protein kinase, putative
TRINITY_DN1240_c1_g1_i10	EAN78012	5.02E-04	hypothetical protein, conserved
TRINITY_DN1240_c1_g1_i3	EAN78012	8.15E-04	hypothetical protein, conserved
TRINITY_DN1240_c1_g1_i8	EAN78012	8.16E-04	hypothetical protein, conserved
TRINITY_DN1240_c1_g1_i9	EAN78012	6.58E-04	:hypothetical protein, conserved
TRINITY_DN22260_c0_g1_i12	EAN79853	2.58E-51	cation transporter, putative
TRINITY_DN22260_c0_g1_i14	EAN79853	7.44E-52	cation transporter, putative
TRINITY_DN22260_c0_g1_i19	EAN79853	2.67E-51	cation transporter, putative
TRINITY_DN22260_c0_g1_i20	EAN79853	1.03E-51	cation transporter, putative
TRINITY_DN22260_c0_g1_i22	EAN79853	1.40E-50	cation transporter, putative
TRINITY_DN22260_c0_g1_i24	EAN79853	3.32E-52	cation transporter, putative
TRINITY_DN22260_c0_g1_i7	EAN79853	2.00E-50	cation transporter, putative
TRINITY_DN40078_c0_g1_i3	AAZ11194	6.09E-37	casein kinase I, epsilon isoform, putative
TRINITY_DN40078_c0_g1_i4	AAZ11194	7.49E-37	casein kinase I, epsilon isoform, putative

TRINITY_DN2267_c4_g1_i1	CAJ16624	1.44E-08	hypothetical protein, conserved; leucine-rich repeat protein (LRRP), putative
TRINITY_DN2267_c4_g1_i2	EAN78352	4.54E-06	hypothetical protein, conserved
TRINITY_DN71410_c3_g1_i1	AAZ11937	8.72E-07	hypothetical protein, conserved
TRINITY_DN10642_c0_g1_i1	AAZ11923	4.85E-07	hypothetical protein, conserved
TRINITY_DN30550_c0_g1_i2	AAZ11867	3.90E-06	aldehyde dehydrogenase family, putative
TRINITY_DN76342_c0_g2_i1	EAN78012	2.27E-04	hypothetical protein, conserved
TRINITY_DN76342_c0_g2_i2	EAN78012	3.36E-04	hypothetical protein, conserved
TRINITY_DN76342_c0_g2_i8	EAN78012	2.21E-04	hypothetical protein, conserved

Table S24: blastx results of differentially expressed transcripts identified by DESeq2 for sulfur pretreated *Euglena gracilis* cultures compared to the non-pretreated cultures using the NCBI non-redundant protein databases.

NR_Isoform Id	NR_sseqid	NR_evalue	NR_description
TRINITY_DN211561_c0_g1_i1	MBF1508352.1	1.01E-41	hypothetical protein [Prevotella pallens]
TRINITY_DN281312_c0_g1_i1	DAM60650.1	2.38E-39	TPA: MAG TPA: Integrase [Herelleviridae sp.]
TRINITY_DN192008_c0_g1_i1	MBS7185617.1	2.95E-35	phosphoglycerate kinase [Clostridium sp.]
TRINITY_DN110614_c0_g1_i1	WP_070151516.1	3.37E-22	phosphate regulon transcriptional regulator PhoB [Sphingobium phenoxybenzoativorans]
TRINITY_DN110614_c0_g1_i2	WP_164078977.1	1.67E-29	response regulator, partial [Stenotrophomonas maltophilia]
TRINITY_DN318734_c0_g1_i1	BAJ81851.1	6.46E-39	30S ribosomal protein S9 [Acidiphilium multivorum AIU301]
TRINITY_DN305840_c0_g1_i1	WP_004121707.1	2.72E-39	pyruvate formate lyase-activating protein [Gardnerella vaginalis]
TRINITY_DN47161_c0_g1_i1	CEO94715.1	2.46E-13	hypothetical protein PBRA_000501 [Plasmodiophora brassicae]
TRINITY_DN47161_c0_g1_i1	CEO94715.1	2.73E-11	hypothetical protein PBRA_000501 [Plasmodiophora brassicae]
TRINITY_DN47161_c0_g1_i1	CEO94715.1	1.39E-10	hypothetical protein PBRA_000501 [Plasmodiophora brassicae]
TRINITY_DN47161_c0_g1_i1	CEO94715.1	4.91E-10	hypothetical protein PBRA_000501 [Plasmodiophora brassicae]
TRINITY_DN47161_c0_g1_i1	CEO94715.1	5.16E-10	hypothetical protein PBRA_000501 [Plasmodiophora brassicae]
TRINITY_DN47161_c0_g1_i1	CEO94715.1	2.35E-08	hypothetical protein PBRA_000501 [Plasmodiophora brassicae]
TRINITY_DN47161_c0_g1_i1	CEO94715.1	1.63E-07	hypothetical protein PBRA_000501 [Plasmodiophora brassicae]
TRINITY_DN47161_c0_g1_i1	CEO94715.1	2.57E-07	hypothetical protein PBRA_000501 [Plasmodiophora brassicae]
TRINITY_DN47161_c0_g1_i1	CEO94715.1	2.44E-06	hypothetical protein PBRA_000501 [Plasmodiophora brassicae]
TRINITY_DN26365_c4_g2_i1	KAH3725539.1	2.20E-10	hypothetical protein DPMN_051384 [Dreissena polymorpha]
TRINITY_DN26365_c4_g2_i1	KAH3725539.1	1.03E-06	hypothetical protein DPMN_051384 [Dreissena polymorpha]
TRINITY_DN17767_c0_g2_i1	YP_001315110.1	2.42E-06	putative reverse transcriptase and intron maturase [Chlorokybus atmophyticus]
TRINITY_DN17767_c0_g2_i1	YP_001315110.1	3.06E-06	putative reverse transcriptase and intron maturase [Chlorokybus atmophyticus]
TRINITY_DN17767_c0_g2_i1	YP_001315110.1	3.06E-06	putative reverse transcriptase and intron maturase [Chlorokybus atmophyticus]
TRINITY_DN47524_c0_g1_i1	GFH12567.1	9.42E-07	eIF3 p110, partial [Haematococcus lacustris]
TRINITY_DN48256_c4_g1_i1	KAH7656872.1	3.33E-13	2-acylglycerol O-acyltransferase protein [Dioscorea alata]
TRINITY_DN1151_c0_g3_i1	KAG7672561.1	6.78E-10	hypothetical protein KSW81_001523 [Chlorella desiccata (nom. nud.)]
TRINITY_DN21823_c0_g1_i1	WP_122717726.1	2.65E-06	PDR/VanB family oxidoreductase [Pseudomonas viridiflava]
TRINITY_DN18691_c0_g1_i1	NDC83818.1	5.48E-10	hypothetical protein [bacterium]
TRINITY_DN1094_c1_g1_i1	CAD8192067.1	2.91E-09	unnamed protein product [Paramecium pentaurelia]
TRINITY_DN57409_c0_g1_i1	HJE71029.1	5.09E-34	NADP-dependent isocitrate dehydrogenase [Pseudomonas oryzihabitans]

TRINITY_DN57409_c0_g1_i2	WP_019019862.1	8.84E-23	NADP-dependent isocitrate dehydrogenase [Halomonas lutea]
TRINITY_DN57409_c0_g1_i2	WP_019019862.1	2.31E-08	NADP-dependent isocitrate dehydrogenase [Halomonas lutea]
TRINITY_DN49095_c2_g1_i1	MBA4421078.1	6.85E-07	NADH dehydrogenase (quinone) subunit G [Anaerolinea sp.]
TRINITY_DN5273_c0_g4_i1	GET92788.1	3.53E-17	syntaxin, putative [Leishmania tarentolae]

Table S25: blastx results of differentially expressed transcripts identified by DESeq2 for sulfur pretreated *Euglena gracilis* cultures compared to the non-pretreated cultures using the NCBI SWISS-PROT protein databases.

SWISS_Isoform Id	SWISS_sseqid	SWISS_evalue	SWISS_description
TRINITY_DN192008_c0_g1_i1	A0PYP1.1	2.49E-37	RecName: Full=Phosphoglycerate kinase [Clostridium novyi NT]
TRINITY_DN110614_c0_g1_i1	Q93CB8.2	2.98E-08	RecName: Full=DNA-binding response regulator MtrA [Mycobacterium avium subsp. paratuberculosis K-10]
TRINITY_DN110614_c0_g1_i2	P0AFJ5.1	2.08E-15	RecName: Full=Phosphate regulon transcriptional regulatory protein PhoB [Escherichia coli K-12]
TRINITY_DN318734_c0_g1_i1	B5ZXZ5.1	2.40E-22	RecName: Full=30S ribosomal protein S9 [Rhizobium leguminosarum bv. trifolii WSM2304]
TRINITY_DN305840_c0_g1_i1	Q46267.1	1.27E-06	RecName: Full=Pyruvate formate-lyase-activating enzyme; Short=PFL-activating enzyme; AltName: Full=Formate-C-acetyltransferase-activating enzyme [Clostridium pasteurianum]
TRINITY_DN26365_c4_g2_i1	P43974.2	1.73E-06	RecName: Full=Putative glycosyltransferase HI_0258 [Haemophilus influenzae Rd KW20]
TRINITY_DN47524_c0_g1_i1	B4P5F7.1	3.87E-06	RecName: Full=Eukaryotic translation initiation factor 3 subunit B; Short=eIF3b; AltName: Full=Eukaryotic translation initiation factor 3 subunit 9 [Drosophila yakuba]
TRINITY_DN48256_c4_g1_i1	Q9C942.1	4.44E-11	RecName: Full=Caffeoylshikimate esterase; AltName: Full=Lysophospholipase 2; Short=LysoPL2 [Arabidopsis thaliana]
TRINITY_DN1151_c0_g3_i1	Q55G18.1	9.89E-06	RecName: Full=Proteasome maturation protein homolog [Dictyostelium discoideum]
TRINITY_DN1094_c1_g1_i1	O22040.2	8.37E-10	RecName: Full=Mitogen-activated protein kinase kinase kinase ANP1; AltName: Full=Arabidopsis NPK1-related kinase 1 [Arabidopsis thaliana]
TRINITY_DN57409_c0_g1_i1	P16100.5	2.10E-31	RecName: Full=Isocitrate dehydrogenase [NADP]; Short=IDH; AltName: Full=Oxalosuccinate decarboxylase [Azotobacter vinelandii]
TRINITY_DN57409_c0_g1_i2	P16100.5	1.38E-20	RecName: Full=Isocitrate dehydrogenase [NADP]; Short=IDH; AltName: Full=Oxalosuccinate decarboxylase [Azotobacter vinelandii]
TRINITY_DN57409_c0_g1_i2	P16100.5	2.32E-11	RecName: Full=Isocitrate dehydrogenase [NADP]; Short=IDH; AltName: Full=Oxalosuccinate decarboxylase [Azotobacter vinelandii]
TRINITY_DN49095_c2_g1_i1	P24918.2	5.51E-07	RecName: Full=NADH-ubiquinone oxidoreductase 78 kDa subunit, mitochondrial; AltName: Full=Complex I-78kD; Short=CI-78kD; Flags: Precursor [Neurospora crassa OR74A]
TRINITY_DN5273_c0_g4_i1	O14662.3	5.08E-14	RecName: Full=Syntaxin-16; Short=Syn16 [Homo sapiens]

Table S26: blastx results of differentially expressed transcripts identified by DESeq2 for sulfur pretreated *Euglena gracilis* cultures compared to the non-pretreated cultures using the Ensembl *Arabidopsis thaliana* database.

Athaliana_Isoform Id	Athaliana_sseqid	Athaliana_evalue	Athaliana_description
TRINITY_DN192008_c0_g1_i1	AT1G79550.1	1.20E-20	Phosphoglycerate kinase
TRINITY_DN127354_c0_g1_i1	AT5G16740.1	2.34E-05	Amino acid transporter AVT1H
TRINITY_DN318734_c0_g1_i1	AT3G49080.1	1.46E-08	30S ribosomal protein S9, mitochondrial
TRINITY_DN47161_c0_g1_i1	AT2G31890.1	1.11E-04	RAP domain-containing protein, chloroplastic
TRINITY_DN47524_c0_g1_i1	AT5G25780.1	1.88E-06	Eukaryotic translation initiation factor 3 subunit B
TRINITY_DN48256_c4_g1_i1	AT1G11090.1	3.87E-15	Alpha/beta-Hydrolases superfamily protein
TRINITY_DN1151_c0_g3_i1	AT1G67250.1	8.14E-09	At1g67250
TRINITY_DN21823_c0_g1_i1	AT2G27510.1	6.99E-05	Ferredoxin
TRINITY_DN1094_c1_g1_i1	AT1G09000.1	9.62E-11	Mitogen-activated protein kinase kinase kinase ANP1
TRINITY_DN49095_c2_g1_i1	AT5G37510.1	1.10E-05	NADH dehydrogenase [ubiquinone] iron-sulfur protein 1, mitochondrial
TRINITY_DN5273_c0_g4_i1	AT4G02195.1	3.09E-13	Syntaxin-42

Table S27: blastx results of differentially expressed transcripts identified by DESeq2 for sulfur pretreated *Euglena gracilis* cultures compared to the non-pretreated cultures using the Ensembl *Chlamydomonas reinhardtti* database.

Creinhardtii_Isoform Id	Creinhardtii_sseqid	Creinhardtii_evalue	Creinhardtii_description
TRINITY_DN158535_c10_g1_i1	PNW79920	2.39E-17	hypothetical protein
TRINITY_DN166865_c1_g1_i1	PNW76352	1.11E-21	hypothetical protein
TRINITY_DN101872_c0_g1_i1	PNW78931	1.79E-05	hypothetical protein
TRINITY_DN110614_c0_g1_i2	PNW76815	3.94E-04	hypothetical protein
TRINITY_DN103825_c0_g1_i2	PNW83376	9.25E-17	hypothetical protein
TRINITY_DN103825_c0_g1_i2	PNW83376	7.55E-13	hypothetical protein
TRINITY_DN103825_c0_g1_i2	PNW83376	1.44E-12	hypothetical protein
TRINITY_DN103825_c0_g1_i2	PNW83376	6.11E-11	hypothetical protein
TRINITY_DN103825_c0_g1_i2	PNW83376	2.80E-10	hypothetical protein
TRINITY_DN103825_c0_g1_i2	PNW83376	1.18E-09	hypothetical protein
TRINITY_DN103825_c0_g1_i2	PNW83376	2.28E-05	hypothetical protein
TRINITY_DN103825_c0_g1_i6	PNW83376	5.70E-17	hypothetical protein
TRINITY_DN103825_c0_g1_i6	PNW83376	8.65E-13	hypothetical protein
TRINITY_DN103825_c0_g1_i6	PNW83376	1.45E-11	hypothetical protein
TRINITY_DN103825_c0_g1_i6	PNW83376	7.66E-09	hypothetical protein
TRINITY_DN103825_c0_g1_i6	PNW83376	8.87E-09	hypothetical protein
TRINITY_DN103825_c0_g1_i6	PNW83376	2.08E-05	hypothetical protein
TRINITY_DN341964_c0_g1_i1	PNW80934	4.03E-04	hypothetical protein
TRINITY_DN320484_c0_g1_i1	PNW84031	2.53E-08	hypothetical protein
TRINITY_DN4416_c1_g1_i1	PNW88806	5.01E-12	hypothetical protein
TRINITY_DN4416_c1_g1_i1	PNW88806	8.45E-10	hypothetical protein
TRINITY_DN15234_c0_g1_i1	PNW77419	2.20E-08	hypothetical protein
TRINITY_DN26392_c1_g3_i1	PNW75014	1.31E-04	hypothetical protein
TRINITY_DN22260_c0_g1_i12	PNW73817	8.57E-15	hypothetical protein
TRINITY_DN22260_c0_g1_i14	PNW73817	1.11E-14	hypothetical protein
TRINITY_DN22260_c0_g1_i19	PNW73817	8.62E-15	hypothetical protein
TRINITY_DN22260_c0_g1_i20	PNW73817	1.18E-14	hypothetical protein
TRINITY_DN22260_c0_g1_i22	PNW73817	3.28E-14	hypothetical protein
TRINITY_DN22260_c0_g1_i24	PNW73817	9.57E-15	hypothetical protein
TRINITY_DN22260_c0_g1_i7	PNW73817	3.51E-14	hypothetical protein
TRINITY_DN37329_c3_g1_i1	PNW84142	5.62E-07	hypothetical protein
TRINITY_DN40078_c0_g1_i3	PNW76048	1.12E-28	hypothetical protein
TRINITY_DN40078_c0_g1_i4	PNW76048	4.94E-28	hypothetical protein
TRINITY_DN9316_c12_g1_i1	PNW86414	6.33E-23	hypothetical protein
TRINITY_DN2267_c4_g1_i1	PNW77156	3.45E-08	hypothetical protein
TRINITY_DN2267_c4_g1_i1	PNW77156	3.18E-04	hypothetical protein
TRINITY_DN2267_c4_g1_i2	PNW76530	2.33E-06	hypothetical protein
TRINITY_DN71410_c3_g1_i1	PNW72103	5.04E-10	hypothetical protein
TRINITY_DN79385_c1_g1_i1	PNW85346	3.69E-10	hypothetical protein
TRINITY_DN10642_c0_g1_i1	PNW83336	9.89E-05	hypothetical protein

Table S28: blastx results of differentially expressed transcripts identified by DESeq2 for sulfur pretreated *Euglena gracilis* cultures compared to the non-pretreated cultures using the Ensembl *Synechocystis sp.* database.

Synechocytis_Isoform Id	Synechocytis_sseqid	Synechocytis_evalue	Synechocytis_description
TRINITY_DN192008_c0_g1_i1	AIE75237	2.94E-17	Phosphoglycerate kinase
TRINITY_DN110614_c0_g1_i1	AIE72797	1.68E-04	Nitrogen-responsive response regulator NrrA
TRINITY_DN110614_c0_g1_i2	AIE72797	9.40E-10	Nitrogen-responsive response regulator NrrA
TRINITY_DN318734_c0_g1_i1	AIE75030	5.12E-08	SSU ribosomal protein S9p (S16e)
TRINITY_DN21823_c0_g1_i1	AIE72625	4.07E-05	soluble [2Fe-2S] ferredoxin
TRINITY_DN1094_c1_g1_i1	AIE74158	1.93E-05	serine/threonine protein kinase
TRINITY_DN49095_c2_g1_i1	AIE75156	2.15E-06	NAD-reducing hydrogenase subunit HoxU

Table S29: blastx results of differentially expressed transcripts identified by DESeq2 for sulfur pretreated *Euglena gracilis* cultures compared to the non-pretreated cultures using the Ensembl *Homo sapiens* database.

Hsapiens_Isoform Id	Hsapiens_sseqid	Hsapiens_evalue	Hsapiens_description
TRINITY_DN192008_c0_g1_i1	ENSP00000305995.3	2.23E-11	phosphoglycerate kinase 2
TRINITY_DN48256_c4_g1_i1	ENSP00000417060.1	9.42E-08	monoglyceride lipase
TRINITY_DN1094_c1_g1_i1	ENSP00000462086.1	2.73E-08	mitogen-activated protein kinase kinase kinase 3
TRINITY_DN49095_c2_g1_i1	ENSP00000409766.1	3.57E-06	NADH:ubiquinone oxidoreductase core subunit S1
TRINITY_DN5273_c0_g4_i1	ENSP00000352634.4	9.93E-15	syntaxin 16

Table S30: blastx results of differentially expressed transcripts identified by DESeq2 for sulfur pretreated *Euglena gracilis* cultures compared to the non-pretreated cultures using the Ensembl *Trypanosoma brucei* database.

Tbrucei_Isoforms Id	Tbrucei_sseqid	Tbrucei_evalue	Tbrucei_description
TRINITY_DN192008_c0_g1_i1	CAJ15978	2.99E-13	phosphoglycerate kinase
TRINITY_DN47161_c0_g1_i1	AAZ11414	1.59E-06	hypothetical protein, conserved
TRINITY_DN47161_c0_g1_i1	AAZ11414	5.42E-05	hypothetical protein, conserved
TRINITY_DN1151_c0_g3_i1	EAN78875	1.54E-04	hypothetical protein, conserved
TRINITY_DN1094_c1_g1_i1	EAN79777	5.50E-10	protein kinase, putative
TRINITY_DN49095_c2_g1_i1	EAN78631	2.39E-04	electron transfer protein, putative
TRINITY_DN5273_c0_g4_i1	EAN77203	8.05E-17	syntaxin, putative

Table S31: blastx results for differentially expressed transcripts identified by DESeq2 with $CdCl_2$ exposed nitrogen pretreated *Euglena gracilis* cultures compared to non-pretreated cultures using the NCBI non-redundant protein database.

NR_Isoform Id	NR_sseqid	NR_evalue	NR_description
TRINITY_DN200360_c0_g1_i1	WP_162809807.1	7.69E-38	chemotaxis protein, partial [Vibrio cholerae]
TRINITY_DN277332_c0_g1_i1	WP_114911319.1	8.64E-41	ABC transporter substrate-binding protein [Acidibrevibacterium fodinaquatile]
TRINITY_DN228713_c0_g1_i1	RYR63465.1	5.77E-16	hypothetical protein Ahy_A04g021264 isoform A [Arachis hypogaea]
TRINITY_DN228713_c0_g1_i1	RYR63465.1	5.77E-16	hypothetical protein Ahy_A04g021264 isoform A [Arachis hypogaea]
TRINITY_DN207373_c0_g1_i1	WP_008450649.1	4.85E-33	acetate kinase [Prevotella amnii]
TRINITY_DN270125_c0_g1_i1	EPI54498.1	6.16E-45	dTMP kinase domain protein, partial [Gardnerella vaginalis JCP7276]
TRINITY_DN238436_c0_g1_i1	WP_164799065.1	1.85E-16	peptidase [Lactobacillus iners]
TRINITY_DN279331_c0_g1_i1	WP_025496050.1	9.19E-25	MULTISPECIES: hypothetical protein [Paraburkholderia]
TRINITY_DN206388_c0_g1_i1	MBW8833736.1	2.63E-27	acyl-CoA dehydrogenase family protein [Burkholderia sp.]
TRINITY_DN287751_c0_g1_i1	WP_106355748.1	5.51E-42	SDR family oxidoreductase [Paraburkholderia insulsa]
TRINITY_DN231511_c0_g1_i1	WP_201775009.1	1.20E-49	Holliday junction branch migration DNA helicase RuvB [Fannyhessea vaginae]
TRINITY_DN282077_c0_g1_i1	TMB11299.1	1.94E-07	hypothetical protein E6J71_25560 [Deltaproteobacteria bacterium]
TRINITY_DN217547_c0_g1_i1	KAH0443421.1	7.93E-08	hypothetical protein KCU90_g1340, partial [Aureobasidium melanogenum]
TRINITY_DN236134_c0_g1_i1	EPI52247.1	1.92E-06	hypothetical protein HMPREF1575_00740 [Gardnerella vaginalis JCP7672]
TRINITY_DN239594_c0_g1_i1	WP_090177237.1	5.66E-07	hypothetical protein [Luteibacter sp. UNC138MFCol5.1]
TRINITY_DN196392_c0_g1_i1	WP_114911399.1	3.94E-34	hypothetical protein [Acidibrevibacterium fodinaquatile]
TRINITY_DN135460_c0_g1_i1	AAU11484.1	1.22E-10	cytosolic phosphoglycerate kinase [Euglena gracilis]
TRINITY_DN135460_c0_g1_i1	AAU11484.1	1.22E-10	cytosolic phosphoglycerate kinase [Euglena gracilis]
TRINITY_DN180686_c0_g1_i1	EPI55698.1	1.71E-39	hypothetical protein HMPREF1573_01091, partial [Gardnerella vaginalis JCP7276]
TRINITY_DN146132_c0_g1_i1	WP_007423908.1	2.43E-47	MULTISPECIES: ATP-binding cassette domain-containing protein [Acidiphilium]
TRINITY_DN146132_c0_g1_i2	WP_007423908.1	2.41E-28	MULTISPECIES: ATP-binding cassette domain-containing protein [Acidiphilium]
TRINITY_DN146132_c0_g1_i3	WP_007423908.1	9.23E-20	MULTISPECIES: ATP-binding cassette domain-containing protein [Acidiphilium]
TRINITY_DN146132_c0_g1_i4	WP_007423908.1	3.28E-19	MULTISPECIES: ATP-binding cassette domain-containing protein [Acidiphilium]
TRINITY_DN146132_c0_g1_i5	WP_007423908.1	4.51E-10	MULTISPECIES: ATP-binding cassette domain-containing protein [Acidiphilium]
TRINITY_DN146132_c0_g1_i6	WP_007423908.1	2.45E-19	MULTISPECIES: ATP-binding cassette domain-containing protein [Acidiphilium]
TRINITY_DN146132_c0_g1_i7	WP_007423908.1	7.98E-19	MULTISPECIES: ATP-binding cassette domain-containing protein [Acidiphilium]
TRINITY_DN109432_c5_g1_i1	KAF7744766.1	3.33E-13	hypothetical protein DSO57_002184 [Entomophthora muscae]
TRINITY_DN157530_c0_g1_i1	WP_234805216.1	3.46E-31	hypothetical protein, partial [Salmonella enterica]
TRINITY_DN182881_c0_g1_i1	EFO70065.1	8.15E-46	proline--tRNA ligase [Lactobacillus iners LactinV 01V1-a]
TRINITY_DN118511_c0_g1_i1	GAN74957.1	8.92E-35	hypothetical protein Apmu_0247_01 [Acidiphilium multivorum AIU301]
TRINITY_DN107490_c0_g1_i1	CAB3731036.1	7.95E-63	hypothetical protein LMG24238_05738 [Paraburkholderia sediminicola]
TRINITY_DN107490_c0_g1_i2	WP_218623398.1	5.41E-39	hypothetical protein, partial [Paraburkholderia fungorum]
TRINITY_DN109647_c0_g1_i1	CUG58006.1	1.42E-27	Hypothetical protein, putative [Bodo saltans]
TRINITY_DN118365_c0_g2_i1	GIL95107.1	1.41E-10	hypothetical protein Vretimale_1165 [Volvox reticuliferus]
TRINITY_DN161247_c0_g1_i1	EPI56452.1	1.57E-27	hypothetical protein HMPREF1573_00869 [Gardnerella vaginalis JCP7276]

TRINITY_DN161247_c0_g1_i2	OKY55971.1	1.43E-27	hypothetical protein BHS10_00250 [Gardnerella vaginalis]
TRINITY_DN305510_c0_g1_i1	WP_183800302.1	2.19E-40	LysR substrate-binding domain-containing protein [Paraburkholderia fungorum]
TRINITY_DN311292_c0_g1_i1	WP_042273652.1	1.29E-39	LysR family transcriptional regulator [Paraburkholderia fungorum]
TRINITY_DN321149_c0_g1_i1	TXH99032.1	6.06E-33	ATP-binding cassette domain-containing protein [Pseudomonas sp.]
TRINITY_DN343792_c0_g1_i1	WP_019036358.1	5.22E-32	UDP-N-acetylmuramate--L-alanine ligase [Prevotella amnii]
TRINITY_DN301869_c0_g1_i1	WP_004128804.1	2.76E-38	BspA family leucine-rich repeat surface protein [Gardnerella vaginalis]
TRINITY_DN307446_c0_g1_i1	NYH18689.1	6.55E-44	arginine/lysine/ornithine decarboxylase [Paraburkholderia bryophila]
TRINITY_DN320314_c1_g1_i1	QLA09599.1	1.13E-36	60S large subunit ribosomal protein uL2 [Euglena gracilis]
TRINITY_DN343342_c0_g1_i1	MBP6614511.1	9.16E-38	class I poly(R)-hydroxyalkanoic acid synthase [Aquabacterium sp.]
TRINITY_DN347737_c0_g1_i1	WP_102694911.1	2.27E-40	low molecular weight phosphotyrosine protein phosphatase [Gardnerella vaginalis]
TRINITY_DN15843_c0_g3_i1	XP_020527713.1	1.81E-09	uncharacterized protein LOC18442033 [Amborella trichopoda]
TRINITY_DN21430_c5_g1_i1	CAJ30046.1	3.93E-16	hypothetical protein mgI388 [Magnetospirillum gryphiswaldense MSR-1]
TRINITY_DN21430_c5_g1_i2	EOB10733.1	2.23E-38	hypothetical protein H376_690 [Rickettsia prowazekii str. GvF12]
TRINITY_DN21430_c17_g1_i1	KAF8091070.1	4.69E-18	hypothetical protein N665_0454s0011 [Sinapis alba]
TRINITY_DN1848_c17_g1_i1	BAD20741.1	2.95E-37	putative membrane-bound adenylyl cyclase [Euglena gracilis]
TRINITY_DN75107_c3_g1_i1	EHH51985.1	4.26E-33	hypothetical protein EGM_12341, partial [Macaca fascicularis]
TRINITY_DN75107_c3_g1_i2	EHH18371.1	4.05E-38	hypothetical protein EGK_14950, partial [Macaca mulatta]
TRINITY_DN80950_c0_g1_i1	MBW9255936.1	1.83E-20	hypothetical protein [Acidithiobacillus ferriphilus]
TRINITY_DN80950_c0_g1_i2	MBW9255936.1	5.91E-10	hypothetical protein [Acidithiobacillus ferriphilus]
TRINITY_DN80950_c0_g1_i3	KDM65718.1	1.09E-23	hypothetical protein ACIDI_92c00270 [Acidiphilium sp. JA12-A1]
TRINITY_DN80950_c0_g1_i4	KDM65718.1	3.37E-14	hypothetical protein ACIDI_92c00270 [Acidiphilium sp. JA12-A1]
TRINITY_DN80950_c0_g1_i5	MBW9255936.1	3.91E-09	hypothetical protein [Acidithiobacillus ferriphilus]
TRINITY_DN16485_c1_g1_i1	CUW39324.1	1.56E-24	conserved protein of unknown function [Magnetospirillum sp. XM-1]
TRINITY_DN22815_c1_g2_i1	GDX61073.1	2.03E-35	hypothetical protein LBMAG32_06070 [Nitrosomonadaceae bacterium]
TRINITY_DN60139_c0_g1_i1	CDW84025.1	1.65E-06	UNKNOWN [Stylonychia lemnae]
TRINITY_DN57515_c0_g1_i11	WP_035187580.1	########	FAD-dependent oxidoreductase [Acidiphilium sp. JA12-A1]
TRINITY_DN14547_c4_g2_i1	CAD1825229.1	3.53E-06	unnamed protein product [Ananas comosus var. bracteatus]
TRINITY_DN85282_c0_g1_i13	EDM97731.1	3.03E-43	hypothetical protein BACCAP_04474 [Pseudoflavonifractor capillosus ATCC 29799]
TRINITY_DN85282_c0_g1_i14	GAD10793.1	1.50E-41	hypothetical protein GFGA_1d0358 [Gluconobacter frateurii NBRC 103465]
TRINITY_DN85282_c0_g1_i18	CDN41090.1	1.14E-25	hypothetical protein BN871_AB_00880 [Paenibacillus sp. P22]
TRINITY_DN85282_c0_g1_i3	AOE12499.1	1.69E-37	hypothetical protein [uncultured bacterium]
TRINITY_DN85282_c0_g1_i5	CDB39267.1	6.63E-54	putative uncharacterized protein [Azospirillum sp. CAG:260]
TRINITY_DN85282_c0_g1_i9	EDM97731.1	9.54E-30	hypothetical protein BACCAP_04474 [Pseudoflavonifractor capillosus ATCC 29799]
TRINITY_DN4780_c6_g1_i1	P14963.1	7.70E-27	RecName: Full=Elongation factor 1-alpha; Short=EF-1-alpha [Euglena gracilis]
TRINITY_DN43046_c0_g1_i1	KXL46362.1	5.64E-26	hypothetical protein FE78DRAFT_79182 [Acidomyces sp. 'richmondensis']
TRINITY_DN43046_c0_g1_i2	KXL46362.1	9.17E-28	hypothetical protein FE78DRAFT_79182 [Acidomyces sp. 'richmondensis']
TRINITY_DN43046_c0_g1_i3	CAG8451736.1	3.75E-45	11792_t:CDS:10 [Diversispora eburnea]
TRINITY_DN43046_c0_g1_i3	CAG8451736.1	1.16E-21	11792_t:CDS:10 [Diversispora eburnea]
TRINITY_DN43046_c0_g1_i4	XP_040746828.1	4.76E-29	P-loop containing nucleoside triphosphate hydrolase protein [Linderina pennispora]

TRINITY_DN43046_c0_g1_i4	XP_040746828.1	1.66E-08	P-loop containing nucleoside triphosphate hydrolase protein [Linderina pennispora]
TRINITY_DN31000_c0_g2_i1	PSC75257.1	2.44E-09	aldo keto reductase [Micractinium conductrix]
TRINITY_DN39055_c2_g1_i1	XP_045033802.1	1.31E-16	heat shock 70 kDa protein 4L-like [Daphnia magna]
TRINITY_DN66263_c0_g1_i1	GAN73150.1	7.53E-08	hypothetical protein Apmu_0056_42 [Acidiphilium multivorum AIU301]
TRINITY_DN66263_c0_g1_i10	GAN73150.1	3.24E-07	hypothetical protein Apmu_0056_42 [Acidiphilium multivorum AIU301]
TRINITY_DN66263_c0_g1_i11	GAN73150.1	3.77E-07	hypothetical protein Apmu_0056_42 [Acidiphilium multivorum AIU301]
TRINITY_DN66263_c0_g1_i12	GAN73150.1	2.77E-07	hypothetical protein Apmu_0056_42 [Acidiphilium multivorum AIU301]
TRINITY_DN66263_c0_g1_i13	GAN73150.1	8.26E-35	hypothetical protein Apmu_0056_42 [Acidiphilium multivorum AIU301]
TRINITY_DN66263_c0_g1_i14	GAN73150.1	4.65E-07	hypothetical protein Apmu_0056_42 [Acidiphilium multivorum AIU301]
TRINITY_DN66263_c0_g1_i16	GAN73150.1	2.13E-07	hypothetical protein Apmu_0056_42 [Acidiphilium multivorum AIU301]
TRINITY_DN66263_c0_g1_i2	GAN73150.1	8.68E-50	hypothetical protein Apmu_0056_42 [Acidiphilium multivorum AIU301]
TRINITY_DN66263_c0_g1_i3	GAN73150.1	2.82E-19	hypothetical protein Apmu_0056_42 [Acidiphilium multivorum AIU301]
TRINITY_DN66263_c0_g1_i4	GAN73150.1	1.19E-10	hypothetical protein Apmu_0056_42 [Acidiphilium multivorum AIU301]
TRINITY_DN66263_c0_g1_i5	GAN73150.1	1.13E-48	hypothetical protein Apmu_0056_42 [Acidiphilium multivorum AIU301]
TRINITY_DN66263_c0_g1_i6	GAN73150.1	3.42E-26	hypothetical protein Apmu_0056_42 [Acidiphilium multivorum AIU301]
TRINITY_DN66263_c0_g1_i7	GAN73150.1	2.61E-07	hypothetical protein Apmu_0056_42 [Acidiphilium multivorum AIU301]
TRINITY_DN66263_c0_g1_i8	GAN73150.1	5.76E-23	hypothetical protein Apmu_0056_42 [Acidiphilium multivorum AIU301]
TRINITY_DN66263_c0_g1_i9	GAN73150.1	5.78E-08	hypothetical protein Apmu_0056_42 [Acidiphilium multivorum AIU301]
TRINITY_DN71275_c3_g1_i1	XP_045591142.1	7.97E-07	mesocentin-like [Procambarus clarkii]
TRINITY_DN31808_c5_g1_i1	RHZ05580.1	2.98E-17	hypothetical protein DYB26_010314, partial [Aphanomyces astaci]
TRINITY_DN28694_c4_g1_i1	XP_013897766.1	2.01E-16	Inorganic phosphate transporter 1-6 [Monoraphidium neglectum]
TRINITY_DN87196_c0_g1_i1	WP_050807168.1	4.25E-22	DEAD/DEAH box helicase, partial [Acidiphilium sp. PM]
TRINITY_DN87196_c0_g1_i13	WP_050807168.1	7.93E-63	DEAD/DEAH box helicase, partial [Acidiphilium sp. PM]
TRINITY_DN87196_c0_g1_i2	WP_050807168.1	1.17E-57	DEAD/DEAH box helicase, partial [Acidiphilium sp. PM]
TRINITY_DN87196_c0_g1_i6	WP_211476745.1	3.59E-06	DEAD/DEAH box helicase [Acidiphilium multivorum]
TRINITY_DN70425_c2_g1_i1	XP_022531639.1	6.61E-07	eukaryotic translation initiation factor 2-alpha kinase 3-like [Astyanax mexicanus]
TRINITY_DN70997_c8_g1_i1	PBH82200.1	1.03E-12	hypothetical protein BGU59_19325 [Clostridioides difficile]
TRINITY_DN29805_c3_g1_i1	KAG0045693.1	5.22E-11	hypothetical protein BGZ90_008386, partial [Linnemannia elongata]
TRINITY_DN40718_c0_g2_i1	XP_039028193.1	6.28E-07	calcium-dependent protein kinase 10-like [Hibiscus syriacus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]

TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	6.09E-09	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN4284_c20_g1_i1	PNF28120.1	2.94E-06	hypothetical protein B7P43_G11869, partial [Cryptotermes secundus]
TRINITY_DN16652_c0_g1_i1	KAB0794669.1	7.33E-07	hypothetical protein PPYR_11508 [Photinus pyralis]
TRINITY_DN51336_c0_g1_i1	NUO71914.1	1.96E-19	nitronate monooxygenase [Frateuria sp.]
TRINITY_DN66798_c1_g1_i2	MBN1921813.1	4.91E-12	SDR family oxidoreductase [Anaerolineae bacterium]
TRINITY_DN42857_c0_g1_i1	KAA6384641.1	1.38E-06	putative Serine/Threonine kinase domain protein [Streblomastix strix]
TRINITY_DN18667_c2_g1_i3	KAF1854556.1	4.49E-28	hypothetical protein Lal_00008066 [Lupinus albus]
TRINITY_DN18667_c2_g1_i4	KAF1854556.1	9.97E-29	hypothetical protein Lal_00008066 [Lupinus albus]
TRINITY_DN18667_c2_g1_i5	KAF1854518.1	1.63E-28	hypothetical protein Lal_00013077 [Lupinus albus]
TRINITY_DN18667_c2_g1_i8	KAH9292424.1	2.23E-34	hypothetical protein KI387_042380, partial [Taxus chinensis]
TRINITY_DN13465_c0_g2_i1	NP_041924.1	8.60E-16	ribosomal protein S2 [Euglena gracilis]
TRINITY_DN13465_c0_g2_i1	NP_041924.1	8.60E-16	ribosomal protein S2 [Euglena gracilis]
TRINITY_DN13465_c0_g2_i2	NP_041924.1	6.55E-11	ribosomal protein S2 [Euglena gracilis]
TRINITY_DN96040_c0_g1_i1	OYV81687.1	1.26E-18	DNA mismatch repair protein MutL, partial [Acidiphilium sp. 21-68-69]
TRINITY_DN96040_c0_g1_i2	OYV81687.1	8.61E-20	DNA mismatch repair protein MutL, partial [Acidiphilium sp. 21-68-69]
TRINITY_DN96040_c0_g1_i3	OYV81687.1	6.26E-18	DNA mismatch repair protein MutL, partial [Acidiphilium sp. 21-68-69]
TRINITY_DN96040_c0_g1_i4	WP_048858130.1	########	DNA mismatch repair endonuclease MutL, partial [Acidiphilium multivorum]
TRINITY_DN96040_c0_g1_i5	OYV81687.1	1.52E-17	DNA mismatch repair protein MutL, partial [Acidiphilium sp. 21-68-69]

TRINITY_DN96040_c0_g1_i6	OYV81687.1	1.48E-18	DNA mismatch repair protein MutL, partial [Acidiphilium sp. 21-68-69]
TRINITY_DN96040_c0_g1_i7	OYV81687.1	4.96E-19	DNA mismatch repair protein MutL, partial [Acidiphilium sp. 21-68-69]
TRINITY_DN36144_c2_g1_i1	KAF4521887.1	1.95E-09	hypothetical protein B566_EDAN008388 [Ephemera danica]

Table S32: blastx results for differentially expressed transcripts identified by DESeq2 with $CdCl_2$ exposed nitrogen pretreated *E. gracilis* cultures compared to non-pretreated cultures using the NCBI SWISS-PROT protein database.

SWISS_Isoform Id	SWISS_sseqid	SWISS_evalue	SWISS_description
TRINITY_DN228713_c0_g1_i1	P31584.1	1.42E-16	RecName: Full=GTP-binding protein yptV1 [Volvox carteri]
TRINITY_DN228713_c0_g1_i1	P31584.1	1.42E-16	RecName: Full=GTP-binding protein yptV1 [Volvox carteri]
TRINITY_DN207373_c0_g1_i1	A6KXS0.1	1.04E-23	RecName: Full=Acetate kinase; AltName: Full=Acetokinase [Phocaeicola vulgatus ATCC 8482]
TRINITY_DN270125_c0_g1_i1	Q5SHX3.1	4.93E-09	RecName: Full=Thymidylate kinase; AltName: Full=dTMP kinase [Thermus thermophilus HB8]
TRINITY_DN206388_c0_g1_i1	Q2KHZ9.1	1.18E-09	RecName: Full=Glutaryl-CoA dehydrogenase, mitochondrial; Short=GCD; Flags: Precursor [Bos taurus]
TRINITY_DN287751_c0_g1_i1	P39483.1	1.46E-10	RecName: Full=Glucose 1-dehydrogenase 2; AltName: Full=GLCDH-II [Priestia megaterium]
TRINITY_DN231511_c0_g1_i1	A4J537.1	4.69E-31	RecName: Full=Holliday junction ATP-dependent DNA helicase RuvB [Desulfotomaculum reducens MI-1]
TRINITY_DN135460_c0_g1_i1	P50312.1	1.04E-06	RecName: Full=Phosphoglycerate kinase, glycosomal; AltName: Full=Phosphoglycerate kinase C; AltName: Full=gPGK [Leishmania major]
TRINITY_DN135460_c0_g1_i1	P50312.1	1.04E-06	RecName: Full=Phosphoglycerate kinase, glycosomal; AltName: Full=Phosphoglycerate kinase C; AltName: Full=gPGK [Leishmania major]
TRINITY_DN146132_c0_g1_i1	Q6BEX0.1	1.79E-11	RecName: Full=Galactofuranose transporter ATP-binding protein YtfR [Escherichia coli K-12]
TRINITY_DN109432_c5_g1_i1	Q5RC80.1	3.26E-09	RecName: Full=RNA-binding protein 39; AltName: Full=RNA-binding motif protein 39 [Pongo abelii]
TRINITY_DN157530_c0_g1_i1	O00370.1	4.42E-31	RecName: Full=LINE-1 retrotransposable element ORF2 protein; Short=ORF2p; Includes: RecName: Full=Reverse transcriptase; Includes: RecName: Full=Endonuclease [Homo sapiens]
TRINITY_DN182881_c0_g1_i1	Q5FJN0.1	2.27E-42	RecName: Full=Proline--tRNA ligase; AltName: Full=Prolyl-tRNA synthetase; Short=ProRS [Lactobacillus acidophilus NCFM]
TRINITY_DN109647_c0_g1_i1	A8JB22.2	1.06E-19	RecName: Full=Dynein regulatory complex subunit 2; AltName: Full=Coiled-coil domain-containing protein 65 homolog; AltName: Full=Flagellar-associated protein 250 [Chlamydomonas reinhardtii]
TRINITY_DN118365_c0_g2_i1	Q9SZ96.1	6.82E-10	RecName: Full=UDP-rhamnose/UDP-galactose transporter 5; Short=UDP-Rha/UDP-Gal transporter 5; AltName: Full=Bi-functional UDP-rhamnose/UDP-galactose transporter [Arabidopsis thaliana]
TRINITY_DN305510_c0_g1_i1	P16931.3	2.22E-13	RecName: Full=HTH-type transcriptional regulator DgdR; AltName: Full=2,2-dialkylglycine decarboxylase repressor [Burkholderia cepacia]
TRINITY_DN321149_c0_g1_i1	Q9FNU2.1	2.93E-08	RecName: Full=ABC transporter B family member 25; Short=ABC transporter ABCB.25; Short=OsABCB25; AltName: Full=Protein ALS1 homolog; AltName: Full=Protein ALUMINUM SENSITIVE 1; Short=OsALS1 [Oryza sativa Japonica Group]
TRINITY_DN343792_c0_g1_i1	A6L070.1	3.13E-13	RecName: Full=UDP-N-acetylmuramate--L-alanine ligase; AltName: Full=UDP-N-acetylmuramoyl-L-alanine synthetase [Phocaeicola vulgatus ATCC 8482]
TRINITY_DN307446_c0_g1_i1	Q9I2S7.1	3.56E-08	RecName: Full=Lysine decarboxylase LdcA [Pseudomonas aeruginosa PAO1]

TRINITY_DN320314_c1_g1_i1	P29766.1	1.65E-24	RecName: Full=60S ribosomal protein L8; AltName: Full=L2; AltName: Full=Ribosomal protein TL2 [Solanum lycopersicum]
TRINITY_DN343342_c0_g1_i1	P23608.1	2.78E-25	RecName: Full=Poly(3-hydroxyalkanoate) polymerase subunit PhaC; Short=PHA polymerase; AltName: Full=PHB synthase subunit PhaC; AltName: Full=Poly(3-hydroxybutyrate) polymerase subunit PhaC; Short=PHB polymerase; Short=Poly-beta-hydroxybutyrate polymerase; AltName: Full=Polyhydroxyalkanoate synthase subunit PhaC; Short=PHA synthase [Cupriavidus necator H16]
TRINITY_DN15843_c0_g3_i1	F4I6V0.1	1.54E-10	RecName: Full=Beta-arabinofuranosyltransferase RAY1; AltName: Full=Protein REDUCED ARABINOSE YARIV 1 [Arabidopsis thaliana]
TRINITY_DN1848_c17_g1_i1	Q99280.1	1.43E-09	RecName: Full=Receptor-type adenylate cyclase GRESAG 4.3; AltName: Full=ATP pyrophosphate-lyase; AltName: Full=Adenylyl cyclase [Trypanosoma brucei brucei]
TRINITY_DN75107_c3_g1_i1	Q96MD7.1	4.34E-30	RecName: Full=Uncharacterized protein C9orf85 [Homo sapiens]
TRINITY_DN75107_c3_g1_i2	Q8N7I0.1	7.52E-29	RecName: Full=Protein GVQW1; AltName: Full=GVQW motif-containing protein 1; AltName: Full=Tigger transposable element-derived 1-like protein 2; Flags: Precursor [Homo sapiens]
TRINITY_DN57515_c0_g1_i11	B7JBP8.1	5.86E-50	RecName: Full=Sulfide-quinone reductase; Short=SQR; AltName: Full=Sulfide:quinone oxidoreductase [Acidithiobacillus ferrooxidans ATCC 23270]
TRINITY_DN4780_c6_g1_i1	P14963.1	8.06E-30	RecName: Full=Elongation factor 1-alpha; Short=EF-1-alpha [Euglena gracilis]
TRINITY_DN43046_c0_g1_i1	Q9FWX8.2	1.55E-24	RecName: Full=ABC transporter B family member 12; Short=ABC transporter ABCB.12; Short=AtABCB12; AltName: Full=Multidrug resistance protein 16; AltName: Full=P-glycoprotein 12 [Arabidopsis thaliana]
TRINITY_DN43046_c0_g1_i1	Q9FWX8.2	1.51E-18	RecName: Full=ABC transporter B family member 12; Short=ABC transporter ABCB.12; Short=AtABCB12; AltName: Full=Multidrug resistance protein 16; AltName: Full=P-glycoprotein 12 [Arabidopsis thaliana]
TRINITY_DN43046_c0_g1_i2	O80725.1	5.06E-27	RecName: Full=ABC transporter B family member 4; Short=ABC transporter ABCB.4; Short=AtABCB4; AltName: Full=Multidrug resistance protein 4; AltName: Full=P-glycoprotein 4 [Arabidopsis thaliana]
TRINITY_DN43046_c0_g1_i2	O80725.1	2.40E-24	RecName: Full=ABC transporter B family member 4; Short=ABC transporter ABCB.4; Short=AtABCB4; AltName: Full=Multidrug resistance protein 4; AltName: Full=P-glycoprotein 4 [Arabidopsis thaliana]
TRINITY_DN43046_c0_g1_i3	Q54BT3.1	3.31E-42	RecName: Full=ABC transporter B family member 2; AltName: Full=ABC transporter ABCB.2 [Dictyostelium discoideum]
TRINITY_DN43046_c0_g1_i3	Q54BT3.1	1.07E-27	RecName: Full=ABC transporter B family member 2; AltName: Full=ABC transporter ABCB.2 [Dictyostelium discoideum]
TRINITY_DN43046_c0_g1_i4	A0A059JJ46.1	1.04E-26	RecName: Full=ABC multidrug transporter MDR2; AltName: Full=Multidrug resistance protein 2 [Trichophyton interdigitale MR816]
TRINITY_DN43046_c0_g1_i4	A0A059JJ46.1	3.68E-19	RecName: Full=ABC multidrug transporter MDR2; AltName: Full=Multidrug resistance protein 2 [Trichophyton interdigitale MR816]
TRINITY_DN39055_c2_g1_i1	Q0IIM3.1	1.47E-15	RecName: Full=Heat shock protein 105 kDa; AltName: Full=Heat shock 110 kDa protein [Bos taurus]
TRINITY_DN31808_c5_g1_i1	Q06572.2	2.53E-09	RecName: Full=Pyrophosphate-energized vacuolar membrane proton pump; AltName: Full=Pyrophosphate-energized inorganic pyrophosphatase; Short=H(+)-PPase [Hordeum vulgare]
TRINITY_DN28694_c4_g1_i1	G7KDA1.1	6.13E-17	RecName: Full=Low affinity inorganic phosphate transporter 8; Short=MtPT8; Short=MtPht1;8; AltName: Full=Arbuscular mycorrhiza-induced

			phosphate transporter PT8; Short=AM-induced phosphate transporter PT8; AltName: Full=H(+)/Pi cotransporter PT8 [Medicago truncatula]
TRINITY_DN87196_c0_g1_i13	Q4I830.1	4.71E-18	RecName: Full=ATP-dependent RNA helicase DRS1 [Fusarium graminearum PH-1]
TRINITY_DN87196_c0_g1_i2	Q09719.1	1.89E-15	RecName: Full=ATP-dependent RNA helicase dbp10 [Schizosaccharomyces pombe 972h-]
TRINITY_DN70425_c2_g1_i1	Q19192.2	3.92E-08	RecName: Full=Eukaryotic translation initiation factor 2-alpha kinase pek-1; AltName: Full=CePEK; Short=PEK; AltName: Full=PRKR-like endoplasmic reticulum kinase; Short=PERK; Flags: Precursor [Caenorhabditis elegans]
TRINITY_DN16652_c0_g1_i1	Q9LK64.1	9.31E-07	RecName: Full=ABC transporter C family member 3; Short=ABC transporter ABCC.3; Short=AtABCC3; AltName: Full=ATP-energized glutathione S-conjugate pump 3; AltName: Full=Glutathione S-conjugate-transporting ATPase 3; AltName: Full=Multidrug resistance-associated protein 3 [Arabidopsis thaliana]
TRINITY_DN66798_c1_g1_i2	P38004.3	8.06E-07	RecName: Full=3-oxoacyl-[acyl-carrier-protein] reductase FabG; AltName: Full=3-ketoacyl-acyl carrier protein reductase; AltName: Full=Beta-Ketoacyl-acyl carrier protein reductase; AltName: Full=Beta-ketoacyl-ACP reductase [Chlamydia trachomatis D/UW-3/CX]
TRINITY_DN42857_c0_g1_i1	Q63184.1	9.02E-07	RecName: Full=Interferon-induced, double-stranded RNA-activated protein kinase; AltName: Full=Eukaryotic translation initiation factor 2-alpha kinase 2; Short=eIF-2A protein kinase 2; AltName: Full=Interferon-inducible RNA-dependent protein kinase; AltName: Full=Protein kinase RNA-activated; Short=PKR; Short=Protein kinase R; AltName: Full=Tyrosine-protein kinase EIF2AK2 [Rattus norvegicus]
TRINITY_DN13465_c0_g2_i1	P30389.1	9.00E-19	RecName: Full=30S ribosomal protein S2, chloroplastic [Euglena gracilis]
TRINITY_DN13465_c0_g2_i1	P30389.1	9.00E-19	RecName: Full=30S ribosomal protein S2, chloroplastic [Euglena gracilis]
TRINITY_DN13465_c0_g2_i2	P30389.1	6.86E-14	RecName: Full=30S ribosomal protein S2, chloroplastic [Euglena gracilis]
TRINITY_DN96040_c0_g1_i1	A5FW68.1	3.08E-21	RecName: Full=DNA mismatch repair protein MutL [Acidiphilium cryptum JF-5]
TRINITY_DN96040_c0_g1_i2	A5FW68.1	2.51E-22	RecName: Full=DNA mismatch repair protein MutL [Acidiphilium cryptum JF-5]
TRINITY_DN96040_c0_g1_i3	A5FW68.1	1.48E-20	RecName: Full=DNA mismatch repair protein MutL [Acidiphilium cryptum JF-5]
TRINITY_DN96040_c0_g1_i4	A5FW68.1	3.30E-105	RecName: Full=DNA mismatch repair protein MutL [Acidiphilium cryptum JF-5]
TRINITY_DN96040_c0_g1_i5	A5FW68.1	3.25E-20	RecName: Full=DNA mismatch repair protein MutL [Acidiphilium cryptum JF-5]
TRINITY_DN96040_c0_g1_i6	A5FW68.1	3.64E-21	RecName: Full=DNA mismatch repair protein MutL [Acidiphilium cryptum JF-5]
TRINITY_DN96040_c0_g1_i7	A5FW68.1	1.23E-21	RecName: Full=DNA mismatch repair protein MutL [Acidiphilium cryptum JF-5]
TRINITY_DN36144_c2_g1_i1	Q6B6N4.1	4.38E-11	RecName: Full=Proliferating cell nuclear antigen; Short=PCNA [Haplochromis burtoni]

Table S33: blastx results for differentially expressed transcripts identified by DESeq2 with $CdCl_2$ exposed nitrogen pretreated *Euglena gracilis* cultures compared to non-pretreated cultures using the Ensembl *Arabidopsis thaliana* database.

Athaliana_Isoform Id	Athaliana_sseqid	Athaliana_evalue	Athaliana_description
TRINITY_DN228713_c0_g1_i1	AT3G11730.1	2.29E-16	Ras-related protein RABD1
TRINITY_DN228713_c0_g1_i1	AT3G11730.1	2.29E-16	Ras-related protein RABD1
TRINITY_DN287751_c0_g1_i1	AT1G63380.2	5.23E-09	F2K11.24
TRINITY_DN146132_c0_g1_i1	AT1G65410.1	1.31E-05	Protein TRIGALACTOSYLDIACYLGLYCEROL 3, chloroplastic
TRINITY_DN109432_c5_g1_i1	AT5G09880.1	1.01E-06	Splicing factor, CC1-like protein
TRINITY_DN118365_c0_g2_i1	AT4G09810.2	6.83E-11	UDP-rhamnose/UDP-galactose transporter 5
TRINITY_DN321149_c0_g1_i1	AT5G39040.1	1.20E-08	ABC transporter B family member 27
TRINITY_DN320314_c1_g1_i1	AT4G36130.1	7.69E-25	60S ribosomal protein L8-3
TRINITY_DN15843_c0_g3_i1	AT1G70630.5	1.50E-11	Nucleotide-diphospho-sugar transferase family protein
TRINITY_DN14547_c4_g2_i1	AT2G36000.1	1.03E-05	At2g36000/F11F19.9
TRINITY_DN4780_c6_g1_i1	AT1G07930.2	1.79E-18	Elongation factor 1-alpha 4
TRINITY_DN43046_c0_g1_i1	AT1G02530.1	1.78E-25	PGP12
TRINITY_DN43046_c0_g1_i1	AT1G02530.1	1.74E-19	PGP12
TRINITY_DN43046_c0_g1_i2	AT2G47000.6	5.81E-28	ABC transporter B family member 4
TRINITY_DN43046_c0_g1_i2	AT2G47000.6	2.76E-25	ABC transporter B family member 4
TRINITY_DN43046_c0_g1_i3	AT5G46540.1	1.37E-40	ABC transporter B family member 7
TRINITY_DN43046_c0_g1_i3	AT5G46540.1	9.41E-33	ABC transporter B family member 7 [
TRINITY_DN43046_c0_g1_i4	AT2G47000.6	9.19E-27	ABC transporter B family member 4
TRINITY_DN43046_c0_g1_i4	AT2G47000.6	3.84E-25	ABC transporter B family member 4
TRINITY_DN39055_c2_g1_i1	AT1G79920.3	1.72E-15	Heat shock protein 70 (Hsp 70) family protein
TRINITY_DN31808_c5_g1_i1	AT1G15690.1	1.16E-09	VHP1
TRINITY_DN28694_c4_g1_i1	AT1G76430.1	7.29E-18	Putative phosphate transporter
TRINITY_DN87196_c0_g1_i13	AT1G16280.1	3.39E-16	DEAD-box ATP-dependent RNA helicase 36
TRINITY_DN87196_c0_g1_i2	AT2G47330.1	7.19E-13	DEAD-box ATP-dependent RNA helicase 24
TRINITY_DN70425_c2_g1_i1	AT2G38620.2	3.38E-06	Cyclin-dependent kinase B1-2
TRINITY_DN29805_c3_g1_i1	AT1G10510.1	4.51E-07	RNI-like superfamily protein
TRINITY_DN29805_c3_g1_i1	AT1G10510.1	2.55E-04	RNI-like superfamily protein
TRINITY_DN29805_c3_g1_i1	AT1G10510.1	6.30E-04	RNI-like superfamily protein
TRINITY_DN16652_c0_g1_i1	AT3G13080.1	1.07E-07	MRP3
TRINITY_DN66798_c1_g1_i2	AT1G07440.1	3.94E-05	Tropinone reductase homolog At1g07440
TRINITY_DN42857_c0_g1_i1	AT5G04510.3	5.01E-08	3-phosphoinositide-dependent protein kinase 1
TRINITY_DN13465_c0_g2_i1	ATCG00160.1	1.57E-06	30S ribosomal protein S2, chloroplastic
TRINITY_DN13465_c0_g2_i1	ATCG00160.1	1.57E-06	30S ribosomal protein S2, chloroplastic
TRINITY_DN13465_c0_g2_i2	ATCG00160.1	2.80E-04	:30S ribosomal protein S2, chloroplastic
TRINITY_DN96040_c0_g1_i1	AT4G09140.1	9.36E-12	DNA mismatch repair protein MLH1
TRINITY_DN96040_c0_g1_i2	AT4G09140.1	7.43E-12	DNA mismatch repair protein MLH1
TRINITY_DN96040_c0_g1_i3	AT4G09140.1	1.63E-11	DNA mismatch repair protein MLH1
TRINITY_DN96040_c0_g1_i4	AT4G09140.1	2.94E-39	NA mismatch repair protein MLH1
TRINITY_DN96040_c0_g1_i5	AT4G09140.1	2.68E-11	DNA mismatch repair protein MLH1
TRINITY_DN96040_c0_g1_i6	AT4G09140.1	1.54E-11	DNA mismatch repair protein MLH1
TRINITY_DN96040_c0_g1_i7	AT4G09140.1	8.14E-12	DNA mismatch repair protein MLH1
TRINITY_DN36144_c2_g1_i1	AT1G07370.1	4.32E-09	Proliferating cellular nuclear antigen 1
TRINITY_DN36144_c2_g1_i2	AT1G07370.1	5.07E-04	Proliferating cellular nuclear antigen 1
TRINITY_DN36144_c2_g1_i2	AT1G07370.1	5.07E-04	Proliferating cellular nuclear antigen 1
TRINITY_DN8486_c7_g1_i1	AT3G17900.1	0.001	AT3g17900/MEB5_12

Table S34: blastx results for differentially expressed transcripts identified by DESeq2 with $CdCl_2$ exposed nitrogen pretreated *Euglena gracilis* cultures compared to non-pretreated cultures using the Ensembl *Chlamydomonas reinhardtti* database.

Creinhardtii_Isoform Id	Creinhardtii_sseqid	Creinhardtii_evalue	Creinhardtii_description
TRINITY_DN228713_c0_g1_i1	PNW75798	1.95E-17	hypothetical protein
TRINITY_DN228713_c0_g1_i1	PNW75798	1.95E-17	hypothetical protein
TRINITY_DN207373_c0_g1_i1	PNW70194	4.26E-14	hypothetical protein
TRINITY_DN287751_c0_g1_i1	PNW87873	4.15E-04	hypothetical protein
TRINITY_DN196392_c0_g1_i1	PNW79088	3.39E-10	hypothetical protein
TRINITY_DN146132_c0_g1_i1	PNW84615	1.19E-07	hypothetical protein
TRINITY_DN109432_c5_g1_i1	PNW76201	7.36E-06	hypothetical protein

TRINITY_DN109647_c0_g1_i1	PNW74416	8.44E-21	hypothetical protein
TRINITY_DN118365_c0_g2_i1	PNW87474	2.41E-13	hypothetical protein
TRINITY_DN321149_c0_g1_i1	PNW86357	6.58E-07	hypothetical protein
TRINITY_DN320314_c1_g1_i1	PNW75662	7.66E-24	hypothetical protein
TRINITY_DN15843_c0_g3_i1	PNW83849	1.02E-05	hypothetical protein
TRINITY_DN1848_c17_g1_i1	PNW77695	1.82E-04	hypothetical protein
TRINITY_DN4780_c6_g1_i1	PNW82579	8.47E-05	hypothetical protein
TRINITY_DN43046_c0_g1_i1	PNW70564	5.87E-21	hypothetical protein
TRINITY_DN43046_c0_g1_i1	PNW70564	1.87E-17	hypothetical protein
TRINITY_DN43046_c0_g1_i2	PNW70564	1.43E-24	hypothetical protein
TRINITY_DN43046_c0_g1_i2	PNW70564	2.01E-21	hypothetical protein
TRINITY_DN43046_c0_g1_i3	PNW70565	6.30E-34	hypothetical protein
TRINITY_DN43046_c0_g1_i3	PNW70565	2.40E-27	hypothetical protein
TRINITY_DN43046_c0_g1_i4	PNW70564	2.42E-24	hypothetical protein
TRINITY_DN43046_c0_g1_i4	PNW70564	2.36E-20	hypothetical protein
TRINITY_DN31000_c0_g2_i1	PNW69962	5.58E-09	hypothetical protein
TRINITY_DN39055_c2_g1_i1	PNW72223	4.36E-13	hypothetical protein
TRINITY_DN31808_c5_g1_i1	PNW78917	3.40E-10	hypothetical protein
TRINITY_DN28694_c4_g1_i1	PNW71995	1.55E-08	hypothetical protein
TRINITY_DN87196_c0_g1_i13	PNW75127	1.05E-17	hypothetical protein
TRINITY_DN87196_c0_g1_i2	PNW75127	2.50E-14	hypothetical protein
TRINITY_DN70425_c2_g1_i1	PNW77588	1.84E-09	hypothetical protein
TRINITY_DN29805_c3_g1_i1	PNW76530	4.00E-09	hypothetical protein
TRINITY_DN66798_c1_g1_i2	PNW88420	6.89E-06	hypothetical protein
TRINITY_DN42857_c0_g1_i1	PNW87550	2.03E-09	hypothetical protein
TRINITY_DN96040_c0_g1_i1	PNW88590	7.01E-13	hypothetical protein
TRINITY_DN96040_c0_g1_i2	PNW88590	5.59E-13	hypothetical protein
TRINITY_DN96040_c0_g1_i3	PNW88590	1.07E-11	hypothetical protein
TRINITY_DN96040_c0_g1_i4	PNW88590	6.04E-38	hypothetical protein
TRINITY_DN96040_c0_g1_i5	PNW88590	1.06E-11	hypothetical protein
TRINITY_DN96040_c0_g1_i6	PNW88590	5.65E-13	hypothetical protein
TRINITY_DN96040_c0_g1_i7	PNW88590	7.30E-13	hypothetical protein
TRINITY_DN36144_c2_g1_i1	PNW75173	4.86E-11	hypothetical protein
TRINITY_DN36144_c2_g1_i2	PNW75173	4.55E-06	hypothetical protein
TRINITY_DN36144_c2_g1_i2	PNW75173	4.55E-06	hypothetical protein

Table S35: blastx results for differentially expressed transcripts identified by DESeq2 with $CdCl_2$ exposed nitrogen pretreated *Euglena gracilis* cultures compared to non-pretreated cultures using the Ensembl *Synechocystis sp.* database.

Synechocytis_Isoform Id	Synechocytis_sseqid	Synechocytis_evalue	Synechocytis_description
TRINITY_DN207373_c0_g1_i1	AIE74091	1.14E-15	Acetate kinase
TRINITY_DN287751_c0_g1_i1	AIE73343	3.75E-10	3-oxoacyl-[acyl-carrier protein] reductase
TRINITY_DN231511_c0_g1_i1	AIE74274	3.22E-28	Holliday junction DNA helicase RuvB
TRINITY_DN146132_c0_g1_i1	AIE75560	3.55E-11	Branched-chain amino acid transport ATP-binding protein LivG
TRINITY_DN182881_c0_g1_i1	AIE73902	3.19E-21	Prolyl-tRNA synthetase, bacterial type
TRINITY_DN305510_c0_g1_i1	AIE74291	6.51E-06	Transcriptional regulator, LysR family
TRINITY_DN321149_c0_g1_i1	AIE75212	1.29E-05	ATP-binding protein of ABC transporter
TRINITY_DN12141_c0_g2_i1	AIE75111	7.85E-04	photosystem I subunit II (PsaD)
TRINITY_DN57515_c0_g1_i1 1	AIE74868	3.12E-07	NADH dehydrogenase
TRINITY_DN59165_c2_g1_i1	AIE73471	2.43E-04	Excinuclease ABC subunit B
TRINITY_DN43046_c0_g1_i1	AIE75638	1.13E-14	Lipid A export ATP-binding/permease protein MsbA
TRINITY_DN43046_c0_g1_i2	AIE75638	1.03E-19	Lipid A export ATP-binding/permease protein MsbA
TRINITY_DN43046_c0_g1_i3	AIE75638	2.94E-20	Lipid A export ATP-binding/permease protein MsbA
TRINITY_DN43046_c0_g1_i4	AIE75638	6.86E-19	Lipid A export ATP-binding/permease protein MsbA

TRINITY_DN87196_c0_g1_i13	AIE75077	1.56E-14	Cold-shock DEAD-box protein A
TRINITY_DN87196_c0_g1_i2	AIE75077	1.07E-11	Cold-shock DEAD-box protein A
TRINITY_DN70425_c2_g1_i1	AIE74546	3.65E-05	eukaryotic protein kinase
TRINITY_DN66798_c1_g1_i2	AIE72830	4.04E-04	3-oxoacyl-[acyl-carrier protein] reductase
TRINITY_DN42857_c0_g1_i1	AIE74546	3.65E-04	eukaryotic protein kinase
TRINITY_DN13465_c0_g2_i1	AIE74756	7.40E-09	SSU ribosomal protein S2p (SAe)
TRINITY_DN13465_c0_g2_i1	AIE74756	7.40E-09	SSU ribosomal protein S2p (SAe)
TRINITY_DN13465_c0_g2_i2	AIE74756	6.41E-06	SSU ribosomal protein S2p (SAe)
TRINITY_DN96040_c0_g1_i1	AIE73395	3.18E-10	DNA mismatch repair protein MutL
TRINITY_DN96040_c0_g1_i2	AIE73395	4.38E-10	DNA mismatch repair protein MutL
TRINITY_DN96040_c0_g1_i3	AIE73395	5.96E-10	DNA mismatch repair protein MutL
TRINITY_DN96040_c0_g1_i4	AIE73395	5.44E-34	DNA mismatch repair protein MutL
TRINITY_DN96040_c0_g1_i5	AIE73395	4.05E-10	DNA mismatch repair protein MutL
TRINITY_DN96040_c0_g1_i6	AIE73395	2.10E-10	DNA mismatch repair protein MutL
TRINITY_DN96040_c0_g1_i7	AIE73395	1.23E-10	DNA mismatch repair protein MutL

Table S36: blastx results for differentially expressed transcripts identified by DESeq2 with $CdCl_2$ exposed nitrogen pretreated *Euglena gracilis* cultures compared to non-pretreated cultures using the Ensembl *Homo sapiens* database.

Hsapiens_Isoform Id	Hsapiens_sseqid	Hsapiens_evalue	Hsapiens_description
TRINITY_DN228713_c0_g1_i1	ENSP00000387286.3	1.01E-16	RAB1A, member RAS oncogene family
TRINITY_DN228713_c0_g1_i1	ENSP00000387286.3	1.01E-16	RAB1A, member RAS oncogene family
TRINITY_DN206388_c0_g1_i1	ENSP00000468584.1	2.91E-10	glutaryl-CoA dehydrogenase
TRINITY_DN287751_c0_g1_i1	ENSP00000395512.1	3.99E-06	peroxisomal trans-2-enoyl-CoA reductase
TRINITY_DN146132_c0_g1_i1	ENSP00000428032.1	1.29E-06	ATP binding cassette subfamily A member 10
TRINITY_DN146132_c0_g1_i1	ENSP00000428032.1	7.86E-04	ATP binding cassette subfamily A member 10
TRINITY_DN109432_c5_g1_i1	ENSP00000489196.1	1.31E-10	RNA binding motif protein 39
TRINITY_DN182881_c0_g1_i1	ENSP00000360327.3	2.25E-16	prolyl-tRNA synthetase 2, mitochondrial
TRINITY_DN109647_c0_g1_i1	ENSP00000312706.4	1.31E-15	coiled-coil domain containing 65
TRINITY_DN118365_c0_g2_i1	ENSP00000501065.1	1.37E-06	solute carrier family 35 member E3
TRINITY_DN321149_c0_g1_i1	ENSP00000440138.1	6.97E-09	ATP binding cassette subfamily B member 9
TRINITY_DN320314_c1_g1_i1	ENSP00000436460.1	1.29E-24	ribosomal protein L8
TRINITY_DN75107_c3_g1_i1	ENSP00000505475.1	3.69E-35	VPS53 subunit of GARP complex
TRINITY_DN75107_c3_g1_i2	ENSP00000352210.4	3.34E-35	PDZ and LIM domain 5
TRINITY_DN4780_c6_g1_i1	ENSP00000298049.8	3.77E-15	eukaryotic translation elongation factor 1 alpha 2
TRINITY_DN43046_c0_g1_i1	ENSP00000384881.2	2.96E-21	ATP binding cassette subfamily B member 5
TRINITY_DN43046_c0_g1_i2	ENSP00000384881.2	8.22E-21	ATP binding cassette subfamily B member 5
TRINITY_DN43046_c0_g1_i3	ENSP00000444095.1	4.38E-41	ATP binding cassette subfamily B member 1 [
TRINITY_DN43046_c0_g1_i3	ENSP00000444095.1	1.22E-31	ATP binding cassette subfamily B member 1
TRINITY_DN43046_c0_g1_i4	ENSP00000384881.2	8.75E-20	ATP binding cassette subfamily B member 5
TRINITY_DN39055_c2_g1_i1	ENSP00000369768.4	3.18E-16	heat shock protein family H (Hsp110) member 1
TRINITY_DN87196_c0_g1_i13	ENSP00000448477.1	2.10E-16	DEAD-box helicase 54
TRINITY_DN87196_c0_g1_i2	ENSP00000448477.1	5.40E-14	DEAD-box helicase 54
TRINITY_DN70425_c2_g1_i1	ENSP00000378914.4	2.96E-07	unc-51 like autophagy activating kinase 2
TRINITY_DN371_c7_g1_i1	ENSP00000419347.1	8.13E-06	muscleblind like splicing regulator 1
TRINITY_DN29805_c3_g1_i1	ENSP00000482302.1	1.27E-06	NLR family CARD domain containing 3
TRINITY_DN29805_c3_g1_i1	ENSP00000482302.1	6.72E-06	NLR family CARD domain containing 3
TRINITY_DN29805_c3_g1_i1	ENSP00000482302.1	7.69E-05	NLR family CARD domain containing 3
TRINITY_DN29805_c3_g1_i1	ENSP00000482302.1	1.51E-04	NLR family CARD domain containing 3
TRINITY_DN16652_c0_g1_i1	ENSP00000507301.1	2.66E-04	ATP binding cassette subfamily C member 6
TRINITY_DN42857_c0_g1_i1	ENSP00000486558.1	5.78E-06	maternal embryonic leucine zipper kinase
TRINITY_DN96040_c0_g1_i1	ENSP00000501030.1	3.24E-12	mutL homolog 1
TRINITY_DN96040_c0_g1_i2	ENSP00000231790.3	3.30E-12	mutL homolog 1
TRINITY_DN96040_c0_g1_i3	ENSP00000501030.1	6.36E-12	mutL homolog 1
TRINITY_DN96040_c0_g1_i4	ENSP00000501030.1	3.29E-38	mutL homolog 1
TRINITY_DN96040_c0_g1_i5	ENSP00000231790.3	8.53E-12	mutL homolog 1
TRINITY_DN96040_c0_g1_i6	ENSP00000501030.1	5.47E-12	mutL homolog 1
TRINITY_DN96040_c0_g1_i7	ENSP00000231790.3	3.44E-12	mutL homolog 1
TRINITY_DN36144_c2_g1_i1	ENSP00000368458.3	8.64E-10	proliferating cell nuclear antigen

Table S37: blastx results for differentially expressed transcripts identified by DESeq2 with $CdCl_2$ exposed nitrogen pretreated *Euglena gracilis* cultures compared to non-pretreated cultures using the Ensembl *Trypanosoma brucei* database.

Tbrucei_Isoforms Id	Tbrucei_sseqid	Tbrucei_evalue	Tbrucei_description
TRINITY_DN228713_c0_g1_i1	AAZ12860	1.83E-17	small GTP-binding protein Rab1, putative
TRINITY_DN228713_c0_g1_i1	AAZ12860	1.83E-17	small GTP-binding protein Rab1, putative
TRINITY_DN135460_c0_g1_i1	CAJ15976	4.22E-08	phosphoglycerate kinase
TRINITY_DN135460_c0_g1_i1	CAJ15976	4.22E-08	phosphoglycerate kinase
TRINITY_DN146132_c0_g1_i1	EAN78912	5.85E-08	ABC transporter, putative
TRINITY_DN109432_c5_g1_i1	EAN79031	1.20E-06	RNA-binding protein, putative
TRINITY_DN109647_c0_g1_i1	EAN79690	2.06E-19	hypothetical protein, conserved
TRINITY_DN118365_c0_g2_i1	AAZ11955	6.25E-05	hypothetical protein, conserved
TRINITY_DN320314_c1_g1_i1	AAZ11226	4.68E-26	60S ribosomal protein L2, putative
TRINITY_DN1848_c17_g1_i1	EAN80648	1.05E-12	expression site-associated gene (ESAG) protein, putative
TRINITY_DN4765_c0_g2_i1	EAN78212	1.28E-05	hypothetical protein, conserved
TRINITY_DN4780_c6_g1_i1	EAN77638	1.46E-20	elongation factor 1-alpha
TRINITY_DN43046_c0_g1_i1	EAN80637	2.73E-12	ABC transporter, putative
TRINITY_DN43046_c0_g1_i2	EAN80637	8.40E-18	ABC transporter, putative
TRINITY_DN43046_c0_g1_i3	EAN80637	3.53E-16	ABC transporter, putative
TRINITY_DN43046_c0_g1_i4	EAN80637	2.13E-16	ABC transporter, putative
TRINITY_DN39055_c2_g1_i1	EAN78648	3.43E-11	heat shock protein, putative
TRINITY_DN31808_c5_g1_i1	AAZ13557	8.25E-10	vacuolar-type proton translocating pyrophosphatase 1
TRINITY_DN87196_c0_g1_i13	AAZ10854	1.68E-17	ATP-dependent DEAD/H RNA helicase, putative
TRINITY_DN87196_c0_g1_i2	AAZ10854	3.61E-14	ATP-dependent DEAD/H RNA helicase, putative
TRINITY_DN70425_c2_g1_i1	AAZ10425	2.20E-07	serine/threonine-protein kinase, putative
TRINITY_DN29805_c3_g1_i1	EAN80633	3.03E-04	hypothetical protein, conserved
TRINITY_DN29805_c3_g1_i1	EAN80633	5.99E-04	hypothetical protein, conserved
TRINITY_DN42857_c0_g1_i1	EAN79784	1.15E-06	protein kinase, putative
TRINITY_DN96040_c0_g1_i1	AAZ13448	1.96E-10	mismatch repair protein MLH1
TRINITY_DN96040_c0_g1_i2	AAZ13448	5.85E-11	mismatch repair protein MLH1
TRINITY_DN96040_c0_g1_i3	AAZ13448	1.00E-10	mismatch repair protein MLH1
TRINITY_DN96040_c0_g1_i4	AAZ13448	6.64E-39	mismatch repair protein MLH1
TRINITY_DN96040_c0_g1_i5	AAZ13448	4.01E-10	mismatch repair protein MLH1
TRINITY_DN96040_c0_g1_i6	AAZ13448	2.48E-10	mismatch repair protein MLH1
TRINITY_DN96040_c0_g1_i7	AAZ13448	1.20E-10	mismatch repair protein MLH1
TRINITY_DN36144_c2_g1_i1	EAN76586	2.95E-08	proliferative cell nuclear antigen (PCNA), putative

Table S38: blastx results for differentially expressed transcripts identified by DESeq2 in nitrogen pretreated *Euglena gracilis* cultures grown in the presence and absence of $CdCl_2$ using the NCBI non-redundant protein database.

NR_Isoform Id	NR_sseqid	NR_evalue	NR_description
TRINITY_DN284159_c0_g1_i1	MCA3930948.1	2.42E-36	30S ribosomal protein S12 methylthiotransferase RimO [Burkholderia sp.]
TRINITY_DN228713_c0_g1_i1	RYR63465.1	5.77E-16	hypothetical protein Ahy_A04g021264 isoform A [Arachis hypogaea]
TRINITY_DN228713_c0_g1_i1	RYR63465.1	5.77E-16	hypothetical protein Ahy_A04g021264 isoform A [Arachis hypogaea]
TRINITY_DN214222_c0_g1_i1	WP_007422099.1	2.07E-43	MULTISPECIES: tetratricopeptide repeat protein [Acidiphilium]
TRINITY_DN230405_c0_g1_i1	WP_211476008.1	2.15E-38	NADH-quinone oxidoreductase subunit D [Acidiphilium multivorum]
TRINITY_DN279331_c0_g1_i1	WP_025496050.1	9.19E-25	MULTISPECIES: hypothetical protein [Paraburkholderia]
TRINITY_DN206388_c0_g1_i1	MBW8833736.1	2.63E-27	acyl-CoA dehydrogenase family protein [Burkholderia sp.]
TRINITY_DN236785_c0_g1_i1	WP_007180614.1	1.54E-35	MULTISPECIES: cold-shock protein [Burkholderiaceae]
TRINITY_DN214757_c0_g1_i1	WP_089446250.1	3.61E-30	hypothetical protein [Burkholderia sp. AU33423]
TRINITY_DN287751_c0_g1_i1	WP_106355748.1	5.51E-42	SDR family oxidoreductase [Paraburkholderia insulsa]
TRINITY_DN298051_c0_g1_i1	KAH0445222.1	1.31E-39	hypothetical protein KCU90_g520, partial [Aureobasidium melanogenum]

TRINITY_DN221978_c0_g1_i1	WP_151723175.1	6.64E-40	polynucleotide adenylyltransferase PcnB [Acinetobacter junii]
TRINITY_DN282077_c0_g1_i1	TMB11299.1	1.94E-07	hypothetical protein E6J71_25560 [Deltaproteobacteria bacterium]
TRINITY_DN217547_c0_g1_i1	KAH0443421.1	7.93E-08	hypothetical protein KCU90_g1340, partial [Aureobasidium melanogenum]
TRINITY_DN225489_c1_g1_i1	HBF97028.1	3.18E-17	ATP-dependent chaperone ClpB [Alphaproteobacteria bacterium]
TRINITY_DN230573_c0_g1_i1	TIB19257.1	7.37E-06	hypothetical protein E3P89_04130, partial [Wallemia ichthyophaga]
TRINITY_DN230573_c0_g1_i2	EPS71058.1	4.30E-14	hypothetical protein M569_03701 [Genlisea aurea]
TRINITY_DN225293_c0_g1_i1	WP_035185626.1	1.15E-38	hypothetical protein [Acidiphilium sp. JA12-A1]
TRINITY_DN283226_c0_g1_i1	MBV8626797.1	1.62E-37	circularly permuted type 2 ATP-grasp protein [Paraburkholderia sp.]
TRINITY_DN146132_c0_g1_i1	WP_007423908.1	2.43E-47	MULTISPECIES: ATP-binding cassette domain-containing protein [Acidiphilium]
TRINITY_DN146132_c0_g1_i2	WP_007423908.1	2.41E-28	MULTISPECIES: ATP-binding cassette domain-containing protein [Acidiphilium]
TRINITY_DN146132_c0_g1_i3	WP_007423908.1	9.23E-20	MULTISPECIES: ATP-binding cassette domain-containing protein [Acidiphilium]
TRINITY_DN146132_c0_g1_i4	WP_007423908.1	3.28E-19	MULTISPECIES: ATP-binding cassette domain-containing protein [Acidiphilium]
TRINITY_DN146132_c0_g1_i5	WP_007423908.1	4.51E-10	MULTISPECIES: ATP-binding cassette domain-containing protein [Acidiphilium]
TRINITY_DN146132_c0_g1_i6	WP_007423908.1	2.45E-19	MULTISPECIES: ATP-binding cassette domain-containing protein [Acidiphilium]
TRINITY_DN146132_c0_g1_i7	WP_007423908.1	7.98E-19	MULTISPECIES: ATP-binding cassette domain-containing protein [Acidiphilium]
TRINITY_DN168885_c0_g1_i1	VXC31200.1	1.49E-16	conserved hypothetical protein [Burkholderia sp. 8Y]
TRINITY_DN107488_c0_g4_i1	KAE9161490.1	1.11E-36	hypothetical protein PF005_g31226, partial [Phytophthora fragariae]
TRINITY_DN187589_c0_g1_i1	CAE6808465.1	3.73E-40	hypothetical protein R69658_05293 [Paraburkholderia aspalathi]
TRINITY_DN178315_c0_g1_i1	WP_231295419.1	8.01E-22	DHA2 family efflux MFS transporter permease subunit [Acidiphilium cryptum]
TRINITY_DN181129_c0_g1_i1	WP_028199273.1	6.04E-28	MULTISPECIES: polyamine ABC transporter ATP-binding protein [Paraburkholderia]
TRINITY_DN140645_c0_g1_i1	WP_042775870.1	9.07E-17	hypothetical protein [Sinorhizobium fredii]
TRINITY_DN127322_c0_g1_i1	ABQ31020.1	1.15E-31	hypothetical protein Acry_1818 [Acidiphilium cryptum JF-5]
TRINITY_DN190444_c0_g1_i1	KFX62623.1	3.33E-26	alpha-amylase [Burkholderia sp. K24]
TRINITY_DN172861_c0_g1_i1	MBN3757506.1	2.57E-39	error-prone DNA polymerase [Paraburkholderia sp. Tr-20389]
TRINITY_DN100609_c0_g1_i1	GEZ67244.1	2.01E-22	hypothetical protein [Tanacetum cinerariifolium]
TRINITY_DN100609_c0_g1_i2	KZR99951.1	5.45E-25	Uncharacterized protein APZ42_003968 [Daphnia magna]
TRINITY_DN305510_c0_g1_i1	WP_183800302.1	2.19E-40	LysR substrate-binding domain-containing protein [Paraburkholderia fungorum]
TRINITY_DN311292_c0_g1_i1	WP_042273652.1	1.29E-39	LysR family transcriptional regulator [Paraburkholderia fungorum]
TRINITY_DN330320_c0_g1_i1	XP_029242822.1	9.56E-20	phospholipid-transporting ATPase 1-like protein [Trypanosoma rangeli]
TRINITY_DN307446_c0_g1_i1	NYH18689.1	6.55E-44	arginine/lysine/ornithine decarboxylase [Paraburkholderia bryophila]
TRINITY_DN320484_c0_g1_i1	OEU22037.1	2.48E-17	hypothetical protein FRACYDRAFT_232192 [Fragilariopsis cylindrus CCMP1102]
TRINITY_DN12296_c2_g3_i1	GFR48109.1	2.74E-19	hypothetical protein Agub_g9933, partial [Astrephomene gubernaculifera]
TRINITY_DN86541_c0_g1_i1	TFJ87788.1	5.06E-11	hypothetical protein NSK_001135 [Nannochloropsis salina CCMP1776]
TRINITY_DN86541_c0_g1_i2	EWM30255.1	1.29E-23	fumarate hydratase [Nannochloropsis gaditana]
TRINITY_DN39413_c1_g1_i1	PWU89065.1	1.01E-42	putative casein kinase [Trypanosoma cruzi]
TRINITY_DN60853_c1_g1_i1	BAJ81420.1	6.51E-30	4-hydroxy-3-methylbut-2-enyl diphosphate reductase [Acidiphilium multivorum AIU301]
TRINITY_DN60853_c1_g1_i2	WP_035188442.1	2.76E-51	MULTISPECIES: 4-hydroxy-3-methylbut-2-enyl diphosphate reductase [unclassified Acidiphilium]

TRINITY_DN4361_c1_g2_i1	OJW26686.1	3.03E-10	hypothetical protein BGO49_00415 [Planctomycetales bacterium 71-10]
TRINITY_DN4361_c1_g2_i10	XP_006812781.1	5.58E-11	PREDICTED: F-box only protein 21-like [Saccoglossus kowalevskii]
TRINITY_DN4361_c1_g2_i12	XP_006812781.1	2.54E-10	PREDICTED: F-box only protein 21-like [Saccoglossus kowalevskii]
TRINITY_DN4361_c1_g2_i14	XP_006812781.1	1.10E-10	PREDICTED: F-box only protein 21-like [Saccoglossus kowalevskii]
TRINITY_DN4361_c1_g2_i4	OJW26686.1	2.84E-09	hypothetical protein BGO49_00415 [Planctomycetales bacterium 71-10]
TRINITY_DN4361_c1_g2_i6	MBV8761887.1	1.48E-09	transglutaminase family protein [Deltaproteobacteria bacterium]
TRINITY_DN4361_c1_g2_i8	XP_006812781.1	3.58E-11	PREDICTED: F-box only protein 21-like [Saccoglossus kowalevskii]
TRINITY_DN20707_c0_g1_i1	WP_093634842.1	0	porin [Paraburkholderia aspalathi]
TRINITY_DN20707_c0_g1_i2	WP_169485870.1	1.94E-38	porin [Paraburkholderia polaris]
TRINITY_DN20707_c0_g1_i3	WP_169485870.1	2.73E-25	porin [Paraburkholderia polaris]
TRINITY_DN20707_c0_g1_i4	WP_169485870.1	1.95E-71	porin [Paraburkholderia polaris]
TRINITY_DN20707_c0_g1_i5	WP_093634842.1	0	porin [Paraburkholderia aspalathi]
TRINITY_DN6500_c4_g1_i1	RLU75026.1	5.00E-19	hypothetical protein CTZ27_38775, partial [Streptomyces griseocarneus]
TRINITY_DN6500_c4_g1_i2	OKY54895.1	2.37E-34	hypothetical protein BHS10_01364 [Gardnerella vaginalis]
TRINITY_DN21430_c1_g3_i1	KAF1858519.1	5.73E-69	hypothetical protein Lal_00015036 [Lupinus albus]
TRINITY_DN21430_c1_g3_i1	KAF1858519.1	2.35E-12	hypothetical protein Lal_00015036 [Lupinus albus]
TRINITY_DN8527_c0_g2_i3	KAG1928647.1	3.50E-14	interferon-induced very large GTPase 1-like [Pimephales promelas]
TRINITY_DN80950_c0_g1_i1	MBW9255936.1	1.83E-20	hypothetical protein [Acidithiobacillus ferriphilus]
TRINITY_DN80950_c0_g1_i2	MBW9255936.1	5.91E-10	hypothetical protein [Acidithiobacillus ferriphilus]
TRINITY_DN80950_c0_g1_i3	KDM65718.1	1.09E-23	hypothetical protein ACIDI_92c00270 [Acidiphilium sp. JA12-A1]
TRINITY_DN80950_c0_g1_i4	KDM65718.1	3.37E-14	hypothetical protein ACIDI_92c00270 [Acidiphilium sp. JA12-A1]
TRINITY_DN80950_c0_g1_i5	MBW9255936.1	3.91E-09	hypothetical protein [Acidithiobacillus ferriphilus]
TRINITY_DN2484_c0_g2_i1	CUG92219.1	5.86E-21	acetyl CoA synthetase, putative [Bodo saltans]
TRINITY_DN22815_c1_g2_i1	GDX61073.1	2.03E-35	hypothetical protein LBMAG32_06070 [Nitrosomonadaceae bacterium]
TRINITY_DN29024_c1_g1_i1	KAF7288696.1	6.85E-08	Threonine dehydratase [Mycena chlorophos]
TRINITY_DN3722_c0_g1_i1	KAH3744493.1	6.94E-11	vesicular integral-membrane protein VIP36 [Pelomyxa schiedti]
TRINITY_DN83701_c0_g1_i4	WP_007423481.1	4.25E-06	MULTISPECIES: EexN family lipoprotein [Acidiphilium]
TRINITY_DN83701_c0_g1_i5	WP_007423481.1	1.43E-06	MULTISPECIES: EexN family lipoprotein [Acidiphilium]
TRINITY_DN32900_c0_g1_i10	CAB5278204.1	6.13E-43	hypothetical protein IST495A_06018 [Burkholderia multivorans]
TRINITY_DN32900_c0_g1_i2	CBA31922.1	3.30E-28	hypothetical protein Csp_D29540 [Curvibacter putative symbiont of Hydra magnipapillata]
TRINITY_DN32900_c0_g1_i3	BAJ11784.1	8.97E-30	dehydration responsive protein, partial [Corchorus olitorius]
TRINITY_DN32900_c0_g1_i4	BAJ11784.1	1.13E-26	dehydration responsive protein, partial [Corchorus olitorius]
TRINITY_DN32900_c0_g1_i5	CBA31922.1	9.95E-29	hypothetical protein Csp_D29540 [Curvibacter putative symbiont of Hydra magnipapillata]
TRINITY_DN32900_c0_g1_i6	CBA31922.1	1.18E-30	hypothetical protein Csp_D29540 [Curvibacter putative symbiont of Hydra magnipapillata]
TRINITY_DN32900_c0_g1_i6	CBA31922.1	1.18E-30	hypothetical protein Csp_D29540 [Curvibacter putative symbiont of Hydra magnipapillata]
TRINITY_DN32900_c0_g1_i7	CBA31922.1	2.99E-28	hypothetical protein Csp_D29540 [Curvibacter putative symbiont of Hydra magnipapillata]
TRINITY_DN71275_c3_g1_i1	XP_045591142.1	7.97E-07	mesocentin-like [Procambarus clarkii]
TRINITY_DN1146_c1_g1_i1	XP_033123426.1	1.79E-11	ATP-binding cassette sub-family C member 9-like isoform X3 [Anneissia japonica]
TRINITY_DN1146_c1_g1_i2	XP_012538805.1	5.86E-12	probable multidrug resistance-associated protein lethal(2)03659 [Monomorium pharaonis]
TRINITY_DN8007_c0_g1_i1	CAJ30045.1	5.23E-45	conserved hypothetical protein [Magnetospirillum gryphiswaldense MSR-1]
TRINITY_DN8007_c0_g1_i3	EAZ75437.1	1.96E-89	conserved hypothetical protein [Vibrio cholerae B33]
TRINITY_DN8007_c0_g1_i4	KUK40615.1	3.47E-43	Uncharacterized protein XD69_1335 [Clostridia bacterium 62_21]
TRINITY_DN8007_c0_g1_i5	KMS65245.1	1.89E-58	hypothetical protein BVRB_037930, partial [Beta vulgaris subsp. vulgaris]

TRINITY_DN8007_c0_g1_i5	KMS65245.1	1.89E-58	hypothetical protein BVRB_037930, partial [Beta vulgaris subsp. vulgaris]
TRINITY_DN38135_c0_g1_i1	MBR9985138.1	1.15E-17	FAD-binding oxidoreductase [Desulfosarcina sp.]
TRINITY_DN21585_c0_g2_i1	QLD47788.1	1.10E-38	hypothetical protein C9419_01235 [Paraburkholderia fungorum]
TRINITY_DN21585_c0_g2_i1	QLD47788.1	2.43E-23	hypothetical protein C9419_01235 [Paraburkholderia fungorum]
TRINITY_DN21585_c0_g2_i11	WP_028200626.1	0	MULTISPECIES: 4-aminobutyrate--2-oxoglutarate transaminase [Paraburkholderia]
TRINITY_DN21585_c0_g2_i12	QLD47788.1	5.15E-39	hypothetical protein C9419_01235 [Paraburkholderia fungorum]
TRINITY_DN21585_c0_g2_i12	QLD47788.1	4.21E-14	hypothetical protein C9419_01235 [Paraburkholderia fungorum]
TRINITY_DN21585_c0_g2_i12	QLD47788.1	4.21E-14	hypothetical protein C9419_01235 [Paraburkholderia fungorum]
TRINITY_DN21585_c0_g2_i13	MBB4519561.1	4.21E-21	hypothetical protein [Paraburkholderia fungorum]
TRINITY_DN21585_c0_g2_i14	WP_028198416.1	2.84E-48	MULTISPECIES: hypothetical protein [Paraburkholderia]
TRINITY_DN21585_c0_g2_i17	KAH0441021.1	0	2-isopropylmalate synthase, partial [Aureobasidium melanogenum]
TRINITY_DN21585_c0_g2_i18	WP_146153228.1	7.78E-14	hypothetical protein [Paraburkholderia insulsa]
TRINITY_DN21585_c0_g2_i19	WP_028196708.1	0	MULTISPECIES: SpoVR family protein [Paraburkholderia]
TRINITY_DN21585_c0_g2_i21	WP_147329533.1	1.14E-17	hypothetical protein [Paraburkholderia sp. DHOC27]
TRINITY_DN21585_c0_g2_i22	WP_028198416.1	8.39E-50	MULTISPECIES: hypothetical protein [Paraburkholderia]
TRINITY_DN21585_c0_g2_i23	PRZ47698.1	1.44E-14	general secretion pathway protein K [Paraburkholderia insulsa]
TRINITY_DN21585_c0_g2_i24	WP_028198416.1	1.07E-49	MULTISPECIES: hypothetical protein [Paraburkholderia]
TRINITY_DN21585_c0_g2_i28	CAE6814243.1	4.07E-14	hypothetical protein R69888_05799 [Paraburkholderia haematera]
TRINITY_DN21585_c0_g2_i29	QLD47788.1	5.00E-39	hypothetical protein C9419_01235 [Paraburkholderia fungorum]
TRINITY_DN21585_c0_g2_i29	QLD47788.1	5.20E-11	hypothetical protein C9419_01235 [Paraburkholderia fungorum]
TRINITY_DN21585_c0_g2_i30	WP_028359528.1	0	MULTISPECIES: sigma-54 dependent transcriptional regulator [Paraburkholderia]
TRINITY_DN21585_c0_g2_i32	MBB4519561.1	4.20E-18	hypothetical protein [Paraburkholderia fungorum]
TRINITY_DN21585_c0_g2_i4	WP_028359528.1	0	MULTISPECIES: sigma-54 dependent transcriptional regulator [Paraburkholderia]
TRINITY_DN21585_c0_g2_i5	KAH0441021.1	########	2-isopropylmalate synthase, partial [Aureobasidium melanogenum]
TRINITY_DN21585_c0_g2_i6	WP_028196708.1	0	MULTISPECIES: SpoVR family protein [Paraburkholderia]
TRINITY_DN21585_c0_g2_i7	WP_028200626.1	0	MULTISPECIES: 4-aminobutyrate--2-oxoglutarate transaminase [Paraburkholderia]
TRINITY_DN21585_c0_g2_i8	WP_028359528.1	0	MULTISPECIES: sigma-54 dependent transcriptional regulator [Paraburkholderia]
TRINITY_DN1341_c0_g1_i1	XP_002681209.1	########	serine/threonine protein kinase [Naegleria gruberi]
TRINITY_DN75565_c0_g1_i1	WP_007423928.1	1.57E-25	glycosyltransferase family 4 protein, partial [Acidiphilium sp. PM]
TRINITY_DN90231_c0_g1_i1	SIT43366.1	5.47E-06	hypothetical protein BN2476_350065 [Paraburkholderia piptadeniae]
TRINITY_DN90231_c0_g1_i2	SIT43366.1	1.68E-10	hypothetical protein BN2476_350065 [Paraburkholderia piptadeniae]
TRINITY_DN90231_c0_g1_i3	WP_081833058.1	8.70E-56	MULTISPECIES: hypothetical protein [Burkholderiaceae]
TRINITY_DN90231_c0_g1_i4	SIT43366.1	3.72E-10	hypothetical protein BN2476_350065 [Paraburkholderia piptadeniae]
TRINITY_DN11676_c2_g4_i1	OEU22037.1	1.32E-08	hypothetical protein FRACYDRAFT_232192 [Fragilariopsis cylindrus CCMP1102]
TRINITY_DN304_c0_g2_i1	VWX60559.1	1.06E-45	conserved hypothetical protein [Burkholderiales bacterium 8X]
TRINITY_DN40718_c0_g2_i1	XP_039028193.1	6.28E-07	calcium-dependent protein kinase 10-like [Hibiscus syriacus]
TRINITY_DN23974_c0_g1_i1	BAC05484.2	4.64E-44	ascorbate peroxidase [Euglena gracilis]
TRINITY_DN23974_c0_g1_i1	BAC05484.2	2.64E-26	ascorbate peroxidase [Euglena gracilis]
TRINITY_DN36224_c0_g4_i1	KAH7054824.1	4.16E-17	leucine rich repeat-containing proteins-like protein [Linnemannia elongata]

TRINITY_DN36224_c0_g4_i1	KAH7054824.1	1.66E-16	leucine rich repeat-containing proteins-like protein [Linnemannia elongata]
TRINITY_DN36224_c0_g4_i1	KAH7054824.1	1.95E-15	leucine rich repeat-containing proteins-like protein [Linnemannia elongata]
TRINITY_DN36224_c0_g4_i1	KAH7054824.1	5.32E-15	leucine rich repeat-containing proteins-like protein [Linnemannia elongata]
TRINITY_DN36224_c0_g4_i1	KAH7054824.1	2.00E-14	leucine rich repeat-containing proteins-like protein [Linnemannia elongata]
TRINITY_DN36224_c0_g4_i1	KAH7054824.1	3.60E-14	leucine rich repeat-containing proteins-like protein [Linnemannia elongata]
TRINITY_DN36224_c0_g4_i1	KAH7054824.1	3.74E-14	leucine rich repeat-containing proteins-like protein [Linnemannia elongata]
TRINITY_DN36224_c0_g4_i1	KAH7054824.1	8.73E-14	leucine rich repeat-containing proteins-like protein [Linnemannia elongata]
TRINITY_DN36224_c0_g4_i1	KAH7054824.1	1.42E-12	leucine rich repeat-containing proteins-like protein [Linnemannia elongata]
TRINITY_DN36224_c0_g4_i1	KAH7054824.1	8.89E-12	leucine rich repeat-containing proteins-like protein [Linnemannia elongata]
TRINITY_DN36224_c0_g4_i1	KAH7054824.1	1.61E-11	leucine rich repeat-containing proteins-like protein [Linnemannia elongata]
TRINITY_DN36224_c0_g4_i1	KAH7054824.1	1.85E-11	leucine rich repeat-containing proteins-like protein [Linnemannia elongata]
TRINITY_DN36224_c0_g4_i1	KAH7054824.1	4.36E-11	leucine rich repeat-containing proteins-like protein [Linnemannia elongata]
TRINITY_DN36224_c0_g4_i1	KAH7054824.1	6.57E-11	leucine rich repeat-containing proteins-like protein [Linnemannia elongata]
TRINITY_DN36224_c0_g4_i1	KAH7054824.1	1.31E-10	leucine rich repeat-containing proteins-like protein [Linnemannia elongata]
TRINITY_DN36224_c0_g4_i1	KAH7054824.1	1.90E-10	leucine rich repeat-containing proteins-like protein [Linnemannia elongata]
TRINITY_DN36224_c0_g4_i1	KAH7054824.1	3.96E-10	leucine rich repeat-containing proteins-like protein [Linnemannia elongata]
TRINITY_DN36224_c0_g4_i1	KAH7054824.1	1.40E-09	leucine rich repeat-containing proteins-like protein [Linnemannia elongata]
TRINITY_DN36224_c0_g4_i1	KAH7054824.1	2.69E-09	leucine rich repeat-containing proteins-like protein [Linnemannia elongata]
TRINITY_DN36224_c0_g4_i1	KAH7054824.1	4.34E-09	leucine rich repeat-containing proteins-like protein [Linnemannia elongata]
TRINITY_DN36224_c0_g4_i1	KAH7054824.1	1.19E-08	leucine rich repeat-containing proteins-like protein [Linnemannia elongata]
TRINITY_DN36224_c0_g4_i1	KAH7054824.1	8.05E-08	leucine rich repeat-containing proteins-like protein [Linnemannia elongata]
TRINITY_DN36224_c0_g4_i1	KAH7054824.1	1.03E-07	leucine rich repeat-containing proteins-like protein [Linnemannia elongata]
TRINITY_DN36224_c0_g4_i1	KAH7054824.1	1.04E-07	leucine rich repeat-containing proteins-like protein [Linnemannia elongata]
TRINITY_DN36224_c0_g4_i1	KAH7054824.1	1.21E-07	leucine rich repeat-containing proteins-like protein [Linnemannia elongata]
TRINITY_DN36224_c0_g4_i1	KAH7054824.1	1.33E-07	leucine rich repeat-containing proteins-like protein [Linnemannia elongata]
TRINITY_DN36224_c0_g4_i1	KAH7054824.1	1.61E-07	leucine rich repeat-containing proteins-like protein [Linnemannia elongata]
TRINITY_DN36224_c0_g4_i1	KAH7054824.1	2.83E-06	leucine rich repeat-containing proteins-like protein [Linnemannia elongata]
TRINITY_DN15094_c2_g1_i1	KAF8363536.1	2.54E-22	hypothetical protein PRIPAC_90459, partial [Pristionchus pacificus]
TRINITY_DN6789_c6_g1_i1	KAH3772145.1	2.58E-07	hypothetical protein DPMN_173481 [Dreissena polymorpha]
TRINITY_DN6789_c6_g1_i1	KAH3772145.1	3.85E-07	hypothetical protein DPMN_173481 [Dreissena polymorpha]
TRINITY_DN6789_c6_g1_i1	KAH3772145.1	3.85E-07	hypothetical protein DPMN_173481 [Dreissena polymorpha]
TRINITY_DN6789_c6_g1_i1	KAH3772145.1	6.53E-07	hypothetical protein DPMN_173481 [Dreissena polymorpha]
TRINITY_DN6789_c6_g1_i1	KAH3772145.1	9.18E-06	hypothetical protein DPMN_173481 [Dreissena polymorpha]
TRINITY_DN8390_c2_g1_i1	AGG11438.1	3.38E-06	hypothetical protein, partial [Paratrypanosoma confusum]
TRINITY_DN10642_c0_g1_i1	XP_020900175.1	5.89E-22	E3 ubiquitin-protein ligase TRIM9 [Exaiptasia diaphana]
TRINITY_DN30823_c0_g2_i1	KZM92177.1	1.55E-13	hypothetical protein DCAR_020458 [Daucus carota subsp. sativus]
TRINITY_DN29663_c1_g1_i1	CUE72018.1	1.88E-34	kinesin, putative [Bodo saltans]
TRINITY_DN17647_c1_g1_i12	WP_039360812.1	1.84E-14	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i12	WP_039360812.1	4.57E-12	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i12	WP_039360812.1	1.95E-11	hypothetical protein [Candidatus Protochlamydia amoebophila]

TRINITY_DN17647_c1_g1_i12	WP_039360812.1	2.36E-11	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i12	WP_039360812.1	4.08E-10	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i12	WP_039360812.1	3.25E-08	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i13	WP_039360812.1	7.16E-12	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i13	WP_039360812.1	1.89E-10	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i13	WP_039360812.1	8.60E-10	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i13	WP_039360812.1	2.58E-09	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i13	WP_039360812.1	3.73E-09	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i13	WP_039360812.1	2.92E-08	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i13	WP_039360812.1	9.03E-07	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i19	WP_039360812.1	3.52E-15	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i19	WP_039360812.1	1.10E-12	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i19	WP_039360812.1	4.96E-12	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i19	WP_039360812.1	6.07E-12	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i19	WP_039360812.1	1.18E-10	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i19	WP_039360812.1	1.11E-08	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i19	WP_039360812.1	6.79E-06	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i19	WP_039360812.1	7.49E-06	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i28	WP_039360812.1	1.84E-14	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i28	WP_039360812.1	4.57E-12	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i28	WP_039360812.1	1.95E-11	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i28	WP_039360812.1	2.36E-11	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i28	WP_039360812.1	4.08E-10	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i28	WP_039360812.1	3.25E-08	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i29	WP_039360812.1	1.14E-14	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i29	WP_039360812.1	3.05E-12	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i29	WP_039360812.1	1.33E-11	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i29	WP_039360812.1	1.61E-11	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i29	WP_039360812.1	2.88E-10	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i29	WP_039360812.1	2.41E-08	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i30	WP_039360812.1	1.04E-11	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i30	WP_039360812.1	2.74E-10	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i30	WP_039360812.1	1.24E-09	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i30	WP_039360812.1	4.05E-09	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i30	WP_039360812.1	6.78E-09	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i30	WP_039360812.1	4.22E-08	hypothetical protein [Candidatus Protochlamydia amoebophila]

TRINITY_DN17647_c1_g1_i30	WP_039360812.1	1.30E-06	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i33	WP_039360812.1	5.69E-11	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i33	WP_039360812.1	1.30E-09	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i33	WP_039360812.1	5.54E-09	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i33	WP_039360812.1	1.71E-08	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i33	WP_039360812.1	2.80E-08	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i33	WP_039360812.1	1.61E-07	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN17647_c1_g1_i33	WP_039360812.1	4.29E-06	hypothetical protein [Candidatus Protochlamydia amoebophila]
TRINITY_DN60075_c0_g1_i1	MBU6357870.1	7.93E-28	NAD(P)-dependent oxidoreductase [Rhodospirillales bacterium]
TRINITY_DN60075_c0_g1_i2	WP_012040244.1	1.20E-19	NAD(P)-dependent oxidoreductase [Acidiphilium cryptum]
TRINITY_DN47942_c0_g1_i1	EGO96075.1	6.58E-21	Imidazolonepropionase, partial [Acidiphilium sp. PM]
TRINITY_DN47942_c0_g1_i2	EGO96075.1	6.72E-10	Imidazolonepropionase, partial [Acidiphilium sp. PM]
TRINITY_DN52904_c0_g1_i9	WP_211476062.1	2.88E-43	LrgB family protein [Acidiphilium multivorum]
TRINITY_DN76342_c0_g2_i1	XP_020631299.1	3.44E-10	E3 ubiquitin-protein ligase TRIM71-like [Orbicella faveolata]
TRINITY_DN76342_c0_g2_i2	XP_020631354.1	3.46E-09	tripartite motif-containing protein 45-like [Orbicella faveolata]
TRINITY_DN76342_c0_g2_i8	XP_020631299.1	3.30E-10	E3 ubiquitin-protein ligase TRIM71-like [Orbicella faveolata]

Table S39: blastx results for differentially expressed transcripts identified by DESeq2 in nitrogen pretreated *Euglena gracilis* cultures grown in the presence and absence of $CdCl_2$ using the NCBI SWISS-PROT protein database.

SWISS_Isoform Id	SWISS_sseqid	SWISS_evalue	SWISS_description
TRINITY_DN284159_c0_g1_i1	B2JG80.1	1.84E-37	RecName: Full=Ribosomal protein S12 methylthiotransferase RimO; Short=S12 MTTase; Short=S12 methylthiotransferase; AltName: Full=Ribosomal protein S12 (aspartate-C(3))-methylthiotransferase; AltName: Full=Ribosome maturation factor RimO [Paraburkholderia phymatum STM815]
TRINITY_DN228713_c0_g1_i1	P31584.1	1.42E-16	RecName: Full=GTP-binding protein yptV1 [Volvox carteri]
TRINITY_DN228713_c0_g1_i1	P31584.1	1.42E-16	RecName: Full=GTP-binding protein yptV1 [Volvox carteri]
TRINITY_DN230405_c0_g1_i1	A5FX11.1	2.30E-41	RecName: Full=NADH-quinone oxidoreductase subunit D; AltName: Full=NADH dehydrogenase I subunit D; AltName: Full=NDH-1 subunit D [Acidiphilium cryptum JF-5]
TRINITY_DN206388_c0_g1_i1	Q2KHZ9.1	1.18E-09	RecName: Full=Glutaryl-CoA dehydrogenase, mitochondrial; Short=GCD; Flags: Precursor [Bos taurus]
TRINITY_DN236785_c0_g1_i1	P72366.1	1.54E-24	RecName: Full=Cold shock-like protein CspA [Stigmatella aurantiaca DW4/3-1]
TRINITY_DN287751_c0_g1_i1	P39483.1	1.46E-10	RecName: Full=Glucose 1-dehydrogenase 2; AltName: Full=GLCDH-II [Priestia megaterium]
TRINITY_DN225489_c1_g1_i1	Q9A9T4.1	6.31E-17	RecName: Full=Chaperone protein ClpB [Caulobacter vibrioides CB15]
TRINITY_DN283226_c0_g1_i1	Q55587.1	1.70E-13	RecName: Full=Uncharacterized protein sll0335 [Synechocystis sp. PCC 6803 substr. Kazusa]
TRINITY_DN146132_c0_g1_i1	Q6BEX0.1	1.79E-11	RecName: Full=Galactofuranose transporter ATP-binding protein YtfR [Escherichia coli K-12]
TRINITY_DN107488_c0_g4_i1	P33625.1	1.38E-37	RecName: Full=Tubulin alpha chain [Euglena gracilis]
TRINITY_DN190444_c0_g1_i1	E6TMM3.1	4.87E-09	RecName: Full=Maltokinase; Short=MaK; AltName: Full=Maltose-1-phosphate synthase [Mycolicibacterium gilvum Spyr1]
TRINITY_DN172861_c0_g1_i1	Q47EP3.1	1.15E-20	RecName: Full=Error-prone DNA polymerase [Dechloromonas aromatica RCB]

TRINITY_DN305510_c0_g1_i1	P16931.3	2.22E-13	RecName: Full=HTH-type transcriptional regulator DgdR; AltName: Full=2,2-dialkylglycine decarboxylase repressor [Burkholderia cepacia]
TRINITY_DN330320_c0_g1_i1	P39524.2	5.70E-18	RecName: Full=Probable phospholipid-transporting ATPase DRS2 [Saccharomyces cerevisiae S288C]
TRINITY_DN307446_c0_g1_i1	Q9I2S7.1	3.56E-08	RecName: Full=Lysine decarboxylase LdcA [Pseudomonas aeruginosa PAO1]
TRINITY_DN12296_c2_g3_i1	Q9VP61.1	7.30E-16	RecName: Full=Acetyl-coenzyme A synthetase; AltName: Full=Acetate--CoA ligase; AltName: Full=Acetyl-CoA synthetase; Short=ACS; Short=AceCS; AltName: Full=Acyl-activating enzyme [Drosophila melanogaster]
TRINITY_DN86541_c0_g1_i2	P14407.2	8.32E-11	RecName: Full=Fumarate hydratase class I, anaerobic; AltName: Full=D-tartrate dehydratase; AltName: Full=Fumarase B [Escherichia coli K-12]
TRINITY_DN39413_c1_g1_i1	B9VVJ6.1	3.52E-42	RecName: Full=Casein kinase I isoform 2 [Trypanosoma brucei brucei]
TRINITY_DN60853_c1_g1_i1	Q7VYS2.1	1.00E-17	RecName: Full=4-hydroxy-3-methylbut-2-enyl diphosphate reductase; Short=HMBPP reductase [Bordetella pertussis Tohama I]
TRINITY_DN60853_c1_g1_i2	Q5NP61.1	1.60E-30	RecName: Full=4-hydroxy-3-methylbut-2-enyl diphosphate reductase; Short=HMBPP reductase [Zymomonas mobilis subsp. mobilis ZM4 = ATCC 31821]
TRINITY_DN4361_c1_g2_i1	Q87DH6.1	1.82E-06	RecName: Full=UPF0162 protein PD_0709 [Xylella fastidiosa Temecula1]
TRINITY_DN4361_c1_g2_i10	Q87DH6.1	2.24E-06	RecName: Full=UPF0162 protein PD_0709 [Xylella fastidiosa Temecula1]
TRINITY_DN4361_c1_g2_i8	Q87DH6.1	1.38E-06	RecName: Full=UPF0162 protein PD_0709 [Xylella fastidiosa Temecula1]
TRINITY_DN20707_c0_g1_i1	Q04064.1	8.05E-21	RecName: Full=Outer membrane porin protein BP0840; Flags: Precursor [Bordetella pertussis Tohama I]
TRINITY_DN20707_c0_g1_i2	P24305.1	6.88E-10	RecName: Full=Outer membrane porin protein 32; Short=OMP32; Flags: Precursor [Delftia acidovorans]
TRINITY_DN20707_c0_g1_i4	Q04064.1	1.06E-19	RecName: Full=Outer membrane porin protein BP0840; Flags: Precursor [Bordetella pertussis Tohama I]
TRINITY_DN20707_c0_g1_i5	Q04064.1	4.52E-21	RecName: Full=Outer membrane porin protein BP0840; Flags: Precursor [Bordetella pertussis Tohama I]
TRINITY_DN8527_c0_g2_i3	E1BD59.1	4.68E-12	RecName: Full=E3 ubiquitin-protein ligase TRIM56; AltName: Full=RING-type E3 ubiquitin transferase TRIM56; AltName: Full=Tripartite motif-containing protein 56 [Bos taurus]
TRINITY_DN2484_c0_g2_i1	Q9VP61.1	1.10E-19	RecName: Full=Acetyl-coenzyme A synthetase; AltName: Full=Acetate--CoA ligase; AltName: Full=Acetyl-CoA synthetase; Short=ACS; Short=AceCS; AltName: Full=Acyl-activating enzyme [Drosophila melanogaster]
TRINITY_DN29024_c1_g1_i1	O42615.1	5.96E-08	RecName: Full=Threonine dehydratase, mitochondrial; AltName: Full=Threonine deaminase; Flags: Precursor [Blastobotrys adeninivorans]
TRINITY_DN3722_c0_g1_i1	Q62902.1	6.81E-08	RecName: Full=Protein ERGIC-53; AltName: Full=ER-Golgi intermediate compartment 53 kDa protein; AltName: Full=Lectin mannose-binding 1; AltName: Full=p58; Flags: Precursor [Rattus norvegicus]
TRINITY_DN1146_c1_g1_i1	Q96J66.1	3.31E-09	RecName: Full=ATP-binding cassette sub-family C member 11; AltName: Full=Multidrug resistance-associated protein 8 [Homo sapiens]
TRINITY_DN1146_c1_g1_i2	Q96J66.1	1.18E-09	RecName: Full=ATP-binding cassette sub-family C member 11; AltName: Full=Multidrug resistance-associated protein 8 [Homo sapiens]
TRINITY_DN21585_c0_g2_i11	P50457.1	0	RecName: Full=4-aminobutyrate aminotransferase PuuE; AltName: Full=GABA aminotransferase; Short=GABA-AT; AltName: Full=Gamma-amino-N-butyrate transaminase; Short=GABA transaminase; AltName: Full=Glutamate:succinic semialdehyde transaminase [Escherichia coli K-12]
TRINITY_DN21585_c0_g2_i17	Q13WM0.1	0	RecName: Full=2-isopropylmalate synthase; AltName: Full=Alpha-IPM synthase; AltName: Full=Alpha-isopropylmalate synthase [Paraburkholderia xenovorans LB400]
TRINITY_DN21585_c0_g2_i19	P29013.2	0	RecName: Full=Uncharacterized protein YcgB [Escherichia coli K-12]

TRINITY_DN21585_c0_g2_i30	O85057.1	1.14E-73	RecName: Full=Limonene hydroxylase; AltName: Full=(S)-limonene 6-monooxygenase; AltName: Full=(S)-limonene 7-monooxygenase; AltName: Full=Carveol dehydrogenase; AltName: Full=Perillyl-alcohol dehydrogenase [Geobacillus stearothermophilus]
TRINITY_DN21585_c0_g2_i4	O85057.1	4.11E-73	RecName: Full=Limonene hydroxylase; AltName: Full=(S)-limonene 6-monooxygenase; AltName: Full=(S)-limonene 7-monooxygenase; AltName: Full=Carveol dehydrogenase; AltName: Full=Perillyl-alcohol dehydrogenase [Geobacillus stearothermophilus]
TRINITY_DN21585_c0_g2_i5	Q13WM0.1	4.34E-151	RecName: Full=2-isopropylmalate synthase; AltName: Full=Alpha-IPM synthase; AltName: Full=Alpha-isopropylmalate synthase [Paraburkholderia xenovorans LB400]
TRINITY_DN21585_c0_g2_i6	P29013.2	0	RecName: Full=Uncharacterized protein YcgB [Escherichia coli K-12]
TRINITY_DN21585_c0_g2_i7	P50457.1	0	RecName: Full=4-aminobutyrate aminotransferase PuuE; AltName: Full=GABA aminotransferase; Short=GABA-AT; AltName: Full=Gamma-amino-N-butyrate transaminase; Short=GABA transaminase; AltName: Full=Glutamate:succinic semialdehyde transaminase [Escherichia coli K-12]
TRINITY_DN21585_c0_g2_i8	O85057.1	1.80E-72	RecName: Full=Limonene hydroxylase; AltName: Full=(S)-limonene 6-monooxygenase; AltName: Full=(S)-limonene 7-monooxygenase; AltName: Full=Carveol dehydrogenase; AltName: Full=Perillyl-alcohol dehydrogenase [Geobacillus stearothermophilus]
TRINITY_DN1341_c0_g1_i1	P54644.1	7.05E-112	RecName: Full=RAC family serine/threonine-protein kinase homolog [Dictyostelium discoideum]
TRINITY_DN23974_c0_g1_i1	Q42564.1	7.48E-25	RecName: Full=L-ascorbate peroxidase 3; Short=AtAPx03 [Arabidopsis thaliana]
TRINITY_DN36224_c0_g4_i1	Q7RTR2.2	3.00E-15	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN36224_c0_g4_i1	Q7RTR2.2	4.15E-13	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN36224_c0_g4_i1	Q7RTR2.2	4.91E-12	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN36224_c0_g4_i1	Q7RTR2.2	6.64E-12	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN36224_c0_g4_i1	Q7RTR2.2	1.46E-10	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN36224_c0_g4_i1	Q7RTR2.2	1.72E-10	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName:

			Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN36224_c0_g4_i1	Q7RTR2.2	6.84E-10	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN36224_c0_g4_i1	Q7RTR2.2	1.00E-09	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN36224_c0_g4_i1	Q7RTR2.2	1.95E-09	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN36224_c0_g4_i1	Q7RTR2.2	7.66E-09	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN36224_c0_g4_i1	Q7RTR2.2	1.38E-06	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN15094_c2_g1_i1	M4IRL6.1	4.28E-21	RecName: Full=Probable CoA ligase CCL7; Short=HlCCL7 [Humulus lupulus]
TRINITY_DN8390_c2_g1_i1	B9VVJ6.1	2.92E-06	RecName: Full=Casein kinase I isoform 2 [Trypanosoma brucei brucei]
TRINITY_DN10642_c0_g1_i1	O15344.1	2.14E-18	RecName: Full=E3 ubiquitin-protein ligase Midline-1; AltName: Full=Midin; AltName: Full=Putative transcription factor XPRF; AltName: Full=RING finger protein 59; AltName: Full=RING finger protein Midline-1; AltName: Full=RING-type E3 ubiquitin transferase Midline-1; AltName: Full=Tripartite motif-containing protein 18 [Homo sapiens]
TRINITY_DN29663_c1_g1_i1	P46869.1	4.78E-07	RecName: Full=Kinesin-like protein FLA10; AltName: Full=Protein KHP1 [Chlamydomonas reinhardtii]
TRINITY_DN17647_c1_g1_i12	Q7RTR2.2	8.41E-13	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN17647_c1_g1_i12	Q7RTR2.2	7.45E-09	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN17647_c1_g1_i12	Q7RTR2.2	5.62E-08	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN17647_c1_g1_i12	Q7RTR2.2	2.00E-07	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2;

			Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN17647_c1_g1_i13	Q7RTR2.2	8.07E-11	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN17647_c1_g1_i13	Q7RTR2.2	1.16E-09	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN17647_c1_g1_i13	Q7RTR2.2	3.54E-07	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN17647_c1_g1_i13	Q7RTR2.2	5.54E-06	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN17647_c1_g1_i13	Q7RTR2.2	7.19E-06	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN17647_c1_g1_i19	Q7RTR2.2	2.58E-13	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN17647_c1_g1_i19	Q7RTR2.2	3.15E-09	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN17647_c1_g1_i19	Q7RTR2.2	2.52E-08	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN17647_c1_g1_i19	Q7RTR2.2	9.30E-08	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN17647_c1_g1_i28	Q7RTR2.2	8.41E-13	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]

TRINITY_DN17647_c1_g1_i28	Q7RTR2.2	7.45E-09	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN17647_c1_g1_i28	Q7RTR2.2	5.62E-08	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN17647_c1_g1_i28	Q7RTR2.2	2.00E-07	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN17647_c1_g1_i29	Q7RTR2.2	6.05E-13	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN17647_c1_g1_i29	Q7RTR2.2	5.85E-09	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN17647_c1_g1_i29	Q7RTR2.2	4.47E-08	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN17647_c1_g1_i29	Q7RTR2.2	1.61E-07	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN17647_c1_g1_i30	Q7RTR2.2	4.36E-11	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN17647_c1_g1_i30	Q7RTR2.2	1.39E-09	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN17647_c1_g1_i30	Q7RTR2.2	1.92E-07	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN17647_c1_g1_i30	Q7RTR2.2	4.85E-06	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and

			CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN17647_c1_g1_i33	Q7RTR2.2	1.61E-10	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN17647_c1_g1_i33	Q7RTR2.2	4.44E-09	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN17647_c1_g1_i33	Q7RTR2.2	5.08E-07	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN60075_c0_g1_i1	O33730.2	5.27E-17	RecName: Full=Uncharacterized oxidoreductase Sfri_1503 [Shewanella frigidimarina NCIMB 400]
TRINITY_DN60075_c0_g1_i2	O33730.2	3.74E-11	RecName: Full=Uncharacterized oxidoreductase Sfri_1503 [Shewanella frigidimarina NCIMB 400]
TRINITY_DN47942_c0_g1_i1	A5FZB8.1	9.15E-23	RecName: Full=Imidazolonepropionase; AltName: Full=Imidazolone-5-propionate hydrolase [Acidiphilium cryptum JF-5]
TRINITY_DN47942_c0_g1_i2	A5FZB8.1	2.14E-12	RecName: Full=Imidazolonepropionase; AltName: Full=Imidazolone-5-propionate hydrolase [Acidiphilium cryptum JF-5]
TRINITY_DN52904_c0_g1_i9	P45146.1	7.04E-07	RecName: Full=Uncharacterized protein HI_1298 [Haemophilus influenzae Rd KW20]

Table S40: blastx results for differentially expressed transcripts identified by DESeq2 in nitrogen pretreated *Euglena gracilis* cultures grown in the presence and absence of $CdCl_2$ using the Ensembl *Arabidopsis thaliana* database.

Athaliana_Isoform Id	Athaliana_sseqid	Athaliana_evalue	Athaliana_description
TRINITY_DN228713_c0_g1_i1	AT3G11730.1	2.29E-16	Ras-related protein RABD1
TRINITY_DN228713_c0_g1_i1	AT3G11730.1	2.29E-16	Ras-related protein RABD1
TRINITY_DN230405_c0_g1_i1	ATMG00510.1	1.65E-12	NADH dehydrogenase subunit 7
TRINITY_DN236785_c0_g1_i1	AT2G17870.1	4.74E-15	Cold shock domain-containing protein 3
TRINITY_DN287751_c0_g1_i1	AT1G63380.2	5.23E-09	F2K11.24
TRINITY_DN225489_c1_g1_i1	AT1G74310.1	1.57E-17	Chaperone protein ClpB1
TRINITY_DN146132_c0_g1_i1	AT1G65410.1	1.31E-05	Protein TRIGALACTOSYLDIACYLGLYCEROL 3, chloroplastic
TRINITY_DN107488_c0_g4_i1	AT4G14960.1	1.42E-36	Tubulin alpha chain
TRINITY_DN330320_c0_g1_i1	AT1G59820.1	1.89E-15	Phospholipid-transporting ATPase
TRINITY_DN12296_c2_g3_i1	AT5G36880.2	1.60E-15	Acetyl-coenzyme A synthetase, chloroplastic/glyoxysomal
TRINITY_DN39413_c1_g1_i1	AT5G43320.2	4.33E-39	Casein kinase 1-like protein 8
TRINITY_DN51592_c1_g1_i1	AT3G01085.3	1.55E-05	Protein kinase superfamily protein
TRINITY_DN2484_c0_g2_i1	AT5G36880.4	1.31E-20	Acetyl-coenzyme A synthetase, chloroplastic/glyoxysomal
TRINITY_DN29024_c1_g1_i1	AT3G10050.1	1.93E-06	Threonine dehydratase
TRINITY_DN1146_c1_g1_i1	AT3G13100.2	3.38E-09	ABC transporter C family member 7
TRINITY_DN1146_c1_g1_i2	AT3G13100.2	7.41E-10	ABC transporter C family member 7
TRINITY_DN21585_c0_g2_i11	AT5G46180.1	2.62E-58	Ornithine aminotransferase, mitochondrial
TRINITY_DN21585_c0_g2_i17	AT1G74040.2	1.85E-28	2-isopropylmalate synthase 2, chloroplastic
TRINITY_DN21585_c0_g2_i5	AT1G74040.3	1.48E-10	2-isopropylmalate synthase 2, chloroplastic
TRINITY_DN21585_c0_g2_i7	AT5G46180.1	1.79E-57	Ornithine aminotransferase, mitochondrial
TRINITY_DN1341_c0_g1_i1	AT3G08730.1	1.14E-85	Non-specific serine/threonine protein kinase

TRINITY_DN23974_c0_g1_i1	AT4G35000.1	8.60E-26	L-ascorbate peroxidase 3
TRINITY_DN36224_c0_g4_i1	AT1G10510.1	2.93E-09	RNI-like superfamily protein
TRINITY_DN36224_c0_g4_i1	AT1G10510.1	5.42E-09	RNI-like superfamily protein
TRINITY_DN36224_c0_g4_i1	AT1G10510.1	3.70E-08	RNI-like superfamily protein
TRINITY_DN36224_c0_g4_i1	AT1G10510.1	5.06E-08	RNI-like superfamily protein
TRINITY_DN36224_c0_g4_i1	AT1G10510.1	2.40E-07	RNI-like superfamily protein
TRINITY_DN36224_c0_g4_i1	AT1G10510.1	2.73E-07	RNI-like superfamily protein
TRINITY_DN36224_c0_g4_i1	AT1G10510.1	1.68E-06	RNI-like superfamily protein
TRINITY_DN36224_c0_g4_i1	AT1G10510.1	6.50E-06	RNI-like superfamily protein
TRINITY_DN15094_c2_g1_i1	AT4G05160.1	2.09E-21	4-coumarate——CoA ligase-like 7
TRINITY_DN8390_c2_g1_i1	AT4G14340.1	1.35E-05	Casein kinase 1-like protein 11
TRINITY_DN98943_c0_g1_i1	AT1G66410.1	4.98E-04	Calmodulin 4
TRINITY_DN10642_c0_g1_i1	AT5G22750.1	4.53E-04	DNA repair protein RAD5A
TRINITY_DN30823_c0_g2_i1	AT5G01710.1	1.60E-15	Methyltransferase
TRINITY_DN29663_c1_g1_i1	AT3G44730.1	6.34E-06	Kinesin-like protein 1
TRINITY_DN17647_c1_g1_i12	AT1G10510.1	9.60E-10	RNI-like superfamily protein
TRINITY_DN17647_c1_g1_i12	AT1G10510.1	1.75E-07	RNI-like superfamily protein
TRINITY_DN17647_c1_g1_i13	AT1G10510.1	3.28E-09	RNI-like superfamily protein
TRINITY_DN17647_c1_g1_i13	AT1G10510.1	5.09E-07	RNI-like superfamily protein
TRINITY_DN17647_c1_g1_i13	AT1G10510.1	1.04E-04	RNI-like superfamily protein
TRINITY_DN17647_c1_g1_i19	AT1G10510.1	4.26E-10	RNI-like superfamily protein
TRINITY_DN17647_c1_g1_i19	AT1G10510.1	8.81E-08	RNI-like superfamily protein
TRINITY_DN17647_c1_g1_i28	AT1G10510.1	9.60E-10	RNI-like superfamily protein
TRINITY_DN17647_c1_g1_i28	AT1G10510.1	1.75E-07	RNI-like superfamily protein
TRINITY_DN17647_c1_g1_i29	AT1G10510.1	7.62E-10	RNI-like superfamily protein
TRINITY_DN17647_c1_g1_i29	AT1G10510.1	1.43E-07	RNI-like superfamily protein
TRINITY_DN17647_c1_g1_i30	AT1G10510.1	2.30E-09	RNI-like superfamily protein
TRINITY_DN17647_c1_g1_i30	AT1G10510.1	3.62E-07	RNI-like superfamily protein
TRINITY_DN17647_c1_g1_i30	AT1G10510.1	8.22E-05	RNI-like superfamily protein
TRINITY_DN17647_c1_g1_i33	AT1G10510.1	6.29E-09	RNI-like superfamily protein
TRINITY_DN17647_c1_g1_i33	AT1G10510.1	8.41E-07	RNI-like superfamily protein
TRINITY_DN17647_c1_g1_i33	AT1G10510.1	1.65E-04	RNI-like superfamily protein
TRINITY_DN60075_c0_g1_i1	AT4G29120.1	1.05E-04	Probable 3-hydroxyisobutyrate dehydrogenase-like 1, mitochondrial
TRINITY_DN8486_c7_g1_i1	AT3G17900.1	0.001	AT3g17900/MEB5_12
TRINITY_DN76342_c0_g2_i1	AT5G05130.1	6.37E-04	Putative SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 3-like 1
TRINITY_DN76342_c0_g2_i2	AT5G05130.1	5.32E-04	Putative SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 3-like 1
TRINITY_DN76342_c0_g2_i8	AT5G05130.1	6.19E-04	Putative SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 3-like 1

Table S41: blastx results for differentially expressed transcripts identified by DESeq2 in nitrogen pretreated *Euglena gracilis* cultures grown in the presence and absence of $CdCl_2$ using the Ensembl *Chlamydomonas reinhardtti* database.

Creinhardtii_Isoform Id	Creinhardtii_sseqid	Creinhardtii_evalue	Creinhardtii_description
TRINITY_DN284159_c0_g1_i1	PNW82252	1.14E-07	hypothetical protein
TRINITY_DN228713_c0_g1_i1	PNW75798	1.95E-17	hypothetical protein
TRINITY_DN228713_c0_g1_i1	PNW75798	1.95E-17	hypothetical protein
TRINITY_DN230405_c0_g1_i1	PNW79199	9.83E-16	hypothetical protein
TRINITY_DN236785_c0_g1_i1	PNW81973	7.85E-11	hypothetical protein
TRINITY_DN287751_c0_g1_i1	PNW87873	4.15E-04	hypothetical protein
TRINITY_DN225489_c1_g1_i1	PNW76502	5.70E-13	hypothetical protein
TRINITY_DN146132_c0_g1_i1	PNW84615	1.19E-07	hypothetical protein
TRINITY_DN107488_c0_g4_i1	PNW83770	1.85E-38	hypothetical protein
TRINITY_DN330320_c0_g1_i1	PNW75666	1.50E-15	hypothetical protein
TRINITY_DN320484_c0_g1_i1	PNW84031	2.53E-08	hypothetical protein
TRINITY_DN12296_c2_g3_i1	PNW89031	2.47E-21	hypothetical protein
TRINITY_DN86541_c0_g1_i2	PNW81648	1.15E-14	hypothetical protein
TRINITY_DN39413_c1_g1_i1	PNW76048	7.18E-40	hypothetical protein
TRINITY_DN51592_c1_g1_i1	PNW83964	1.41E-04	hypothetical protein
TRINITY_DN2484_c0_g2_i1	PNW89031	1.73E-22	hypothetical protein

TRINITY_DN1146_c1_g1_i1	PNW85154	2.19E-11	hypothetical protein
TRINITY_DN1146_c1_g1_i2	PNW85154	1.76E-13	hypothetical protein
TRINITY_DN38135_c0_g1_i1	PNW82667	4.95E-09	hypothetical protein
TRINITY_DN21585_c0_g2_i11	PNW79407	3.96E-75	hypothetical protein
TRINITY_DN21585_c0_g2_i17	PNW81754	4.56E-11	hypothetical protein
TRINITY_DN21585_c0_g2_i30	PNW74733	1.32E-05	hypothetical protein
TRINITY_DN21585_c0_g2_i4	PNW74733	1.54E-05	hypothetical protein
TRINITY_DN21585_c0_g2_i7	PNW79407	4.36E-74	hypothetical protein
TRINITY_DN21585_c0_g2_i8	PNW74733	1.81E-05	hypothetical protein
TRINITY_DN1341_c0_g1_i1	PNW71866	3.78E-61	hypothetical protein
TRINITY_DN23974_c0_g1_i1	PNW79135	5.15E-24	hypothetical protein
TRINITY_DN36224_c0_g4_i1	PNW88806	1.59E-11	hypothetical protein
TRINITY_DN36224_c0_g4_i1	PNW88806	5.99E-09	hypothetical protein
TRINITY_DN36224_c0_g4_i1	PNW88806	1.42E-07	hypothetical protein
TRINITY_DN36224_c0_g4_i1	PNW88806	3.91E-06	hypothetical protein
TRINITY_DN15094_c2_g1_i1	PNW74853	8.95E-12	hypothetical protein
TRINITY_DN8390_c2_g1_i1	PNW76048	6.31E-04	hypothetical protein
TRINITY_DN98943_c0_g1_i1	PNW85257	4.92E-05	hypothetical protein
TRINITY_DN10642_c0_g1_i1	PNW83336	9.89E-05	hypothetical protein
TRINITY_DN29663_c1_g1_i1	PNW70698	3.79E-08	hypothetical protein
TRINITY_DN17647_c1_g1_i12	PNW84437	3.53E-09	hypothetical protein
TRINITY_DN17647_c1_g1_i13	PNW84437	4.12E-08	hypothetical protein
TRINITY_DN17647_c1_g1_i19	PNW84437	1.53E-09	hypothetical protein
TRINITY_DN17647_c1_g1_i28	PNW84437	3.53E-09	hypothetical protein
TRINITY_DN17647_c1_g1_i29	PNW84437	2.78E-09	hypothetical protein
TRINITY_DN17647_c1_g1_i30	PNW84437	4.12E-08	hypothetical protein
TRINITY_DN17647_c1_g1_i33	PNW84437	1.08E-07	hypothetical protein

Table S42: blastx results for differentially expressed transcripts identified by DESeq2 in nitrogen pretreated *Euglena gracilis* cultures grown in the presence and absence of $CdCl_2$ using the Ensembl *Synechocystis sp.* database.

Synechocytis_Isoform Id	Synechocytis_sseqid	Synechocytis_evalue	Synechocytis_description
TRINITY_DN284159_c0_g1_i1	AIE75078	4.66E-08	Ribosomal protein S12p Asp88 methylthiotransferase
TRINITY_DN230405_c0_g1_i1	AIE74458	3.38E-10	NAD(P)H-quinone oxidoreductase chain H
TRINITY_DN287751_c0_g1_i1	AIE73343	3.75E-10	3-oxoacyl-[acyl-carrier protein] reductase
TRINITY_DN225489_c1_g1_i1	AIE72935	7.00E-15	ClpB protein
TRINITY_DN283226_c0_g1_i1	AIE75085	1.48E-15	Protein containing domains DUF404, DUF407
TRINITY_DN146132_c0_g1_i1	AIE75560	3.55E-11	Branched-chain amino acid transport ATP-binding protein LivG
TRINITY_DN305510_c0_g1_i1	AIE74291	6.51E-06	Transcriptional regulator, LysR family
TRINITY_DN12296_c2_g3_i1	AIE74180	8.84E-11	description:Acetyl-coenzyme A synthetase
TRINITY_DN60853_c1_g1_i1	AIE73929	2.40E-04	Glycosyltransferase
TRINITY_DN60853_c1_g1_i2	AIE73929	8.41E-05	Glycosyltransferase
TRINITY_DN2484_c0_g2_i1	AIE74180	7.41E-17	Acetyl-coenzyme A synthetase
TRINITY_DN29024_c1_g1_i1	AIE72618	2.95E-09	Threonine dehydratase biosynthetic
TRINITY_DN59165_c2_g1_i1	AIE73471	2.43E-04	Excinuclease ABC subunit B
TRINITY_DN21585_c0_g2_i11	AIE73711	1.97E-57	Acetylornithine aminotransferase
TRINITY_DN21585_c0_g2_i17	AIE73388	4.82E-20	2-isopropylmalate synthase
TRINITY_DN21585_c0_g2_i30	AIE74026	9.76E-10	Transcriptional regulatory protein tyrR
TRINITY_DN21585_c0_g2_i4	AIE74026	1.17E-09	Transcriptional regulatory protein tyrR
TRINITY_DN21585_c0_g2_i5	AIE73388	1.67E-04	2-isopropylmalate synthase
TRINITY_DN21585_c0_g2_i7	AIE73711	1.23E-56	:Acetylornithine aminotransferase
TRINITY_DN21585_c0_g2_i8	AIE74026	1.41E-09	Transcriptional regulatory protein tyrR
TRINITY_DN1341_c0_g1_i1	AIE75313	1.22E-13	eukaryotic protein kinase
TRINITY_DN23974_c0_g1_i1	AIE75792	9.94E-10	Catalase / Peroxidase
TRINITY_DN15094_c2_g1_i1	AIE74353	6.30E-11	Long-chain-fatty-acid--CoA ligase

Table S43: blastx results for differentially expressed transcripts identified by DESeq2 in nitrogen pretreated *Euglena gracilis* cultures grown in the presence and absence of $CdCl_2$ using the Ensembl *Homo sapiens* database.

Hsapiens_Isoform Id	Hsapiens_sseqid	Hsapiens_evalue	Hsapiens_description
TRINITY_DN228713_c0_g1_i1	ENSP00000387286.3	1.01E-16	RAB1A, member RAS oncogene family
TRINITY_DN228713_c0_g1_i1	ENSP00000387286.3	1.01E-16	RAB1A, member RAS oncogene family
TRINITY_DN230405_c0_g1_i1	ENSP00000504215.1	6.15E-13	NADH:ubiquinone oxidoreductase core subunit S2
TRINITY_DN206388_c0_g1_i1	ENSP00000468584.1	2.91E-10	glutaryl-CoA dehydrogenase
TRINITY_DN236785_c0_g1_i1	ENSP00000007699.5	3.85E-09	:Y-box binding protein 2
TRINITY_DN287751_c0_g1_i1	ENSP00000395512.1	3.99E-06	peroxisomal trans-2-enoyl-CoA reductase
TRINITY_DN225489_c1_g1_i1	ENSP00000340385.5	0.001	caseinolytic mitochondrial matrix peptidase chaperone subunit B
TRINITY_DN146132_c0_g1_i1	ENSP00000428032.1	1.29E-06	ATP binding cassette subfamily A member 10
TRINITY_DN146132_c0_g1_i1	ENSP00000428032.1	7.86E-04	ATP binding cassette subfamily A member 10
TRINITY_DN107488_c0_g4_i1	ENSP00000396212.1	8.98E-37	tubulin alpha 4a
TRINITY_DN330320_c0_g1_i1	ENSP00000371084.5	8.98E-18	ATPase phospholipid transporting 8A1
TRINITY_DN12296_c2_g3_i1	ENSP00000417783.1	8.60E-15	acyl-CoA synthetase short chain family member 2
TRINITY_DN39413_c1_g1_i1	ENSP00000463757.1	4.74E-40	casein kinase 1 delta
TRINITY_DN4361_c1_g2_i4	ENSP00000483508.1	7.08E-05	F-box protein 21
TRINITY_DN4361_c1_g2_i6	ENSP00000483508.1	7.29E-05	FBXO21 description:F-box protein 21
TRINITY_DN51592_c1_g1_i1	ENSP00000476533.1	4.28E-06	mitogen-activated protein kinase kinase kinase kinase 3
TRINITY_DN8527_c0_g2_i3	ENSP00000305161.6	4.49E-12	:tripartite motif containing 56
TRINITY_DN2484_c0_g2_i1	ENSP00000388793.2	1.08E-15	acyl-CoA synthetase short chain family member 1
TRINITY_DN3722_c0_g1_i1	ENSP00000251047.4	2.54E-07	lectin, mannose binding 1
TRINITY_DN1146_c1_g1_i1	ENSP00000311326.6	8.36E-10	ATP binding cassette subfamily C member 11
TRINITY_DN1146_c1_g1_i2	ENSP00000311326.6	2.98E-10	ATP binding cassette subfamily C member 11
TRINITY_DN2114_c0_g2_i1	ENSP00000509123.1	1.27E-04	protein O-linked mannose N-acetylglucosaminyltransferase 2 (beta 1,4-)
TRINITY_DN21585_c0_g2_i11	ENSP00000357838.5	4.26E-51	ornithine aminotransferase
TRINITY_DN21585_c0_g2_i7	ENSP00000357838.5	2.23E-50	aminotransferase
TRINITY_DN1341_c0_g1_i1	ENSP00000375892.2	1.00E-86	AKT serine/threonine kinase 2
TRINITY_DN36224_c0_g4_i1	ENSP00000352039.6	7.58E-16	NLR family CARD domain containing 3
TRINITY_DN15094_c2_g1_i1	ENSP00000426150.1	1.38E-12	acyl-CoA synthetase long chain family member 1
TRINITY_DN98943_c0_g1_i1	ENSP00000406112.1	3.29E-05	calmodulin 2
TRINITY_DN10642_c0_g1_i1	ENSP00000391154.1	5.00E-19	midline 1
TRINITY_DN29663_c1_g1_i1	ENSP00000395443.2	4.03E-05	kinesin family member 1A
TRINITY_DN17647_c1_g1_i12	ENSP00000352039.6	2.12E-13	NLR family CARD domain containing 3
TRINITY_DN17647_c1_g1_i13	ENSP00000323897.9	1.70E-11	NLR family CARD domain containing 3
TRINITY_DN17647_c1_g1_i19	ENSP00000352039.6	6.51E-14	NLR family CARD domain containing 3

TRINITY_DN17647_c1_g1_i28	ENSP00000352039.6	2.12E-13	NLR family CARD domain containing 3
TRINITY_DN17647_c1_g1_i29	ENSP00000352039.6	1.53E-13	NLR family CARD domain containing 3
TRINITY_DN17647_c1_g1_i30	ENSP00000323897.9	1.04E-11	NLR family CARD domain containing 3
TRINITY_DN17647_c1_g1_i33	ENSP00000323897.9	3.88E-11	NLR family CARD domain containing 3
TRINITY_DN76342_c0_g2_i1	ENSP00000508734.1	3.17E-06	midline 1
TRINITY_DN76342_c0_g2_i2	ENSP00000508734.1	1.15E-05	midline 1
TRINITY_DN76342_c0_g2_i8	ENSP00000508734.1	3.07E-06	midline 1

Table S44: blastx results for differentially expressed transcripts identified by DESeq2 in nitrogen pretreated *Euglena gracilis* cultures grown in the presence and absence of $CdCl_2$ using the Ensembl *Trypanosoma brucei* database.

Tbrucei_Isoforms Id	Tbrucei_sseqid	Tbrucei_evalue	Tbrucei_description
TRINITY_DN228713_c0_g1_i1	AAZ12860	1.83E-17	small GTP-binding protein Rab1, putative
TRINITY_DN228713_c0_g1_i1	AAZ12860	1.83E-17	:small GTP-binding protein Rab1, putative
TRINITY_DN236785_c0_g1_i1	EAN79753	1.65E-06	:mitochondrial RNA binding protein
TRINITY_DN225489_c1_g1_i1	AAQ16055	2.27E-15	ATP-dependent Clp protease subunit, heat shock protein 100 (HSP100), putative
TRINITY_DN146132_c0_g1_i1	EAN78912	5.85E-08	ABC transporter, putative
TRINITY_DN107488_c0_g4_i1	CAJ16360	1.96E-37	alpha tubulin
TRINITY_DN330320_c0_g1_i1	EAN79305	4.20E-21	phospholipid-transporting ATPase 1-like protein, putative
TRINITY_DN21967_c0_g1_i1	AAZ11811	1.59E-04	hypothetical protein, conserved
TRINITY_DN12296_c2_g3_i1	AAZ13019	1.30E-14	acetyl-CoA synthetase, putative
TRINITY_DN86541_c0_g1_i2	AAZ10483	6.49E-09	fumarate hydratase, putative
TRINITY_DN39413_c1_g1_i1	AAZ11195	8.51E-44	casein kinase, putative
TRINITY_DN51592_c1_g1_i1	AAZ12797	2.08E-05	cell division control protein 2 homolog 2
TRINITY_DN8527_c0_g2_i3	AAZ11923	3.59E-15	hypothetical protein, conserved
TRINITY_DN2484_c0_g2_i1	AAZ13019	9.06E-20	acetyl-CoA synthetase, putative
TRINITY_DN3722_c0_g1_i1	EAN79384	5.92E-07	lectin, putative
TRINITY_DN1146_c1_g1_i1	AAZ12983	4.64E-04	multidrug resistance protein A
TRINITY_DN1146_c1_g1_i2	AAZ12983	2.57E-05	multidrug resistance protein A
TRINITY_DN78986_c0_g1_i1	AAZ11551	4.69E-04	hypothetical protein, conserved
TRINITY_DN1341_c0_g1_i1	AAZ11787	2.64E-101	rac serine-threonine kinase, putative
TRINITY_DN36224_c0_g4_i1	EAN78352	3.31E-13	hypothetical protein, conserved
TRINITY_DN36224_c0_g4_i1	EAN78352	1.03E-08	hypothetical protein, conserved
TRINITY_DN36224_c0_g4_i1	EAN78352	1.99E-07	hypothetical protein, conserved
TRINITY_DN36224_c0_g4_i1	EAN78352	3.36E-05	hypothetical protein, conserved
TRINITY_DN36224_c0_g4_i1	EAN78352	4.64E-05	hypothetical protein, conserved
TRINITY_DN36224_c0_g4_i1	EAN78352	5.14E-04	hypothetical protein, conserved
TRINITY_DN15094_c2_g1_i1	EAN77749	5.68E-12	fatty acyl CoA synthetase, putative
TRINITY_DN8390_c2_g1_i1	AAZ11194	2.83E-09	casein kinase I, epsilon isoform, putative
TRINITY_DN98943_c0_g1_i1	EAN80250	1.33E-04	calmodulin
TRINITY_DN10642_c0_g1_i1	AAZ11923	4.85E-07	hypothetical protein, conserved
TRINITY_DN29663_c1_g1_i1	AAZ13259	7.18E-07	kinesin, putative
TRINITY_DN17647_c1_g1_i12	EAN79168	1.65E-08	hypothetical protein, conserved
TRINITY_DN17647_c1_g1_i13	EAN79168	2.67E-09	hypothetical protein, conserved
TRINITY_DN17647_c1_g1_i19	EAN79168	8.55E-09	hypothetical protein, conserved
TRINITY_DN17647_c1_g1_i28	EAN79168	1.65E-08	hypothetical protein, conserved
TRINITY_DN17647_c1_g1_i29	EAN79168	1.36E-08	hypothetical protein, conserved
TRINITY_DN17647_c1_g1_i30	EAN79168	3.18E-09	:hypothetical protein, conserved
TRINITY_DN17647_c1_g1_i33	EAN79168	7.62E-09	hypothetical protein, conserved
TRINITY_DN76342_c0_g2_i1	EAN78012	2.27E-04	hypothetical protein, conserved
TRINITY_DN76342_c0_g2_i2	EAN78012	3.36E-04	hypothetical protein, conserved
TRINITY_DN76342_c0_g2_i8	EAN78012	2.21E-04	hypothetical protein, conserved

Table S45: blastx results for differentially expressed transcripts identified by DESeq2 in nitrogen pretreated *Euglena gracilis* cultures compared to non-pretreated cultures using the NCBI non-redundant protein database.

NR_Isoform Id	NR_sseqid	NR_evalue	NR_description
TRINITY_DN225669_c4_g1_i1	CAG8815642.1	6.67E-19	11253_t:CDS:2, partial [Dentiscutata erythropus]
TRINITY_DN230405_c0_g1_i1	WP_211476008.1	2.15E-38	NADH-quinone oxidoreductase subunit D [Acidiphilium multivorum]
TRINITY_DN281312_c0_g1_i1	DAM60650.1	2.38E-39	TPA: MAG TPA: Integrase [Herelleviridae sp.]
TRINITY_DN192008_c0_g1_i1	MBS7185617.1	2.95E-35	phosphoglycerate kinase [Clostridium sp.]
TRINITY_DN127853_c1_g1_i1	AMN16513.1	1.60E-24	heat shock protein 90, partial [Euglena agilis]
TRINITY_DN118511_c0_g1_i1	GAN74957.1	8.92E-35	hypothetical protein Apmu_0247_01 [Acidiphilium multivorum AIU301]
TRINITY_DN199816_c0_g1_i1	MCA7118743.1	9.26E-38	MFS transporter [Acidibrevibacterium sp.]
TRINITY_DN178315_c0_g1_i1	WP_231295419.1	8.01E-22	DHA2 family efflux MFS transporter permease subunit [Acidiphilium cryptum]
TRINITY_DN140645_c0_g1_i1	WP_042775870.1	9.07E-17	hypothetical protein [Sinorhizobium fredii]
TRINITY_DN132906_c0_g2_i1	WP_199905840.1	2.07E-34	DUF1725 domain-containing protein [Mycobacterium tuberculosis]
TRINITY_DN132906_c0_g2_i2	MRB78795.1	1.01E-21	DUF1725 domain-containing protein [Bacillus thuringiensis]
TRINITY_DN170684_c0_g1_i1	WP_114913969.1	4.35E-09	CHRD domain-containing protein [Acidibrevibacterium fodinaquatile]
TRINITY_DN170684_c0_g1_i2	WP_114913969.1	2.43E-20	CHRD domain-containing protein [Acidibrevibacterium fodinaquatile]
TRINITY_DN100609_c0_g1_i1	GEZ67244.1	2.01E-22	hypothetical protein [Tanacetum cinerariifolium]
TRINITY_DN100609_c0_g1_i2	KZR99951.1	5.45E-25	Uncharacterized protein APZ42_003968 [Daphnia magna]
TRINITY_DN318734_c0_g1_i1	BAJ81851.1	6.46E-39	30S ribosomal protein S9 [Acidiphilium multivorum AIU301]
TRINITY_DN305840_c0_g1_i1	WP_004121707.1	2.72E-39	pyruvate formate lyase-activating protein [Gardnerella vaginalis]
TRINITY_DN330320_c0_g1_i1	XP_029242822.1	9.56E-20	phospholipid-transporting ATPase 1-like protein [Trypanosoma rangeli]
TRINITY_DN4481_c1_g2_i1	XP_029231101.1	1.04E-16	putative proteasome regulatory non-ATPase subunit 3 [Trypanosoma conorhini]
TRINITY_DN14102_c0_g2_i1	BAD20740.1	8.22E-11	putative membrane-bound adenylyl cyclase [Euglena gracilis]
TRINITY_DN809_c27_g1_i1	MBO8631107.1	4.18E-12	aminotransferase class I/II-fold pyridoxal phosphate-dependent enzyme [Staphylococcus aureus]
TRINITY_DN75002_c0_g1_i1	KAG0049188.1	8.00E-23	hypothetical protein BGZ83_005991 [Gryganskiella cystojenkinii]
TRINITY_DN75002_c0_g1_i1	KAG0049188.1	4.11E-14	hypothetical protein BGZ83_005991 [Gryganskiella cystojenkinii]
TRINITY_DN75002_c0_g1_i1	KAG0049188.1	1.78E-13	hypothetical protein BGZ83_005991 [Gryganskiella cystojenkinii]
TRINITY_DN17767_c0_g2_i1	YP_001315110.1	2.42E-06	putative reverse transcriptase and intron maturase [Chlorokybus atmophyticus]
TRINITY_DN17767_c0_g2_i1	YP_001315110.1	3.06E-06	putative reverse transcriptase and intron maturase [Chlorokybus atmophyticus]
TRINITY_DN17767_c0_g2_i1	YP_001315110.1	3.06E-06	putative reverse transcriptase and intron maturase [Chlorokybus atmophyticus]
TRINITY_DN3722_c0_g1_i1	KAH3744493.1	6.94E-11	vesicular integral-membrane protein VIP36 [Pelomyxa schiedti]
TRINITY_DN17032_c0_g2_i1	QLA09621.1	1.16E-23	60S large subunit ribosomal protein eL15 [Euglena gracilis]
TRINITY_DN43053_c2_g1_i1	KAG4917006.1	2.44E-22	hypothetical protein JHK87_054563 [Glycine soja]
TRINITY_DN31051_c0_g1_i1	CAF1577965.1	1.20E-18	unnamed protein product [Didymodactylos carnosus]
TRINITY_DN38135_c0_g1_i1	MBR9985138.1	1.15E-17	FAD-binding oxidoreductase [Desulfosarcina sp.]
TRINITY_DN75565_c0_g1_i1	WP_007423928.1	1.57E-25	glycosyltransferase family 4 protein, partial [Acidiphilium sp. PM]
TRINITY_DN23974_c0_g1_i1	BAC05484.2	4.64E-44	ascorbate peroxidase [Euglena gracilis]
TRINITY_DN23974_c0_g1_i1	BAC05484.2	2.64E-26	ascorbate peroxidase [Euglena gracilis]
TRINITY_DN27215_c2_g1_i1	PKN77467.1	7.25E-15	hypothetical protein CVU51_16700 [Deltaproteobacteria bacterium HGW-Deltaproteobacteria-1]
TRINITY_DN27215_c2_g1_i2	PKN77467.1	3.36E-13	hypothetical protein CVU51_16700 [Deltaproteobacteria bacterium HGW-Deltaproteobacteria-1]

TRINITY_DN27215_c2_g1_i2	PKN77467.1	2.09E-08	hypothetical protein CVU51_16700 [Deltaproteobacteria bacterium HGW-Deltaproteobacteria-1]
TRINITY_DN7321_c1_g1_i1	KAF0714762.1	1.01E-12	hypothetical protein As57867_003693 [Aphanomyces stellatus]
TRINITY_DN60075_c0_g1_i1	MBU6357870.1	7.93E-28	NAD(P)-dependent oxidoreductase [Rhodospirillales bacterium]
TRINITY_DN60075_c0_g1_i2	WP_012040244.1	1.20E-19	NAD(P)-dependent oxidoreductase [Acidiphilium cryptum]
TRINITY_DN3638_c5_g2_i1	OAQ25058.1	6.32E-13	RNI-like protein [Linnemannia elongata AG-77]
TRINITY_DN3638_c5_g2_i1	OAQ25058.1	2.70E-11	RNI-like protein [Linnemannia elongata AG-77]
TRINITY_DN3638_c5_g2_i1	OAQ25058.1	1.08E-09	RNI-like protein [Linnemannia elongata AG-77]
TRINITY_DN3638_c5_g2_i1	OAQ25058.1	2.06E-09	RNI-like protein [Linnemannia elongata AG-77]
TRINITY_DN3638_c5_g2_i1	OAQ25058.1	3.74E-09	RNI-like protein [Linnemannia elongata AG-77]
TRINITY_DN3638_c5_g2_i1	OAQ25058.1	3.97E-09	RNI-like protein [Linnemannia elongata AG-77]
TRINITY_DN3638_c5_g2_i1	OAQ25058.1	5.42E-09	RNI-like protein [Linnemannia elongata AG-77]
TRINITY_DN3638_c5_g2_i1	OAQ25058.1	1.56E-08	RNI-like protein [Linnemannia elongata AG-77]
TRINITY_DN3638_c5_g2_i1	OAQ25058.1	4.75E-08	RNI-like protein [Linnemannia elongata AG-77]
TRINITY_DN3638_c5_g2_i1	OAQ25058.1	5.03E-08	RNI-like protein [Linnemannia elongata AG-77]
TRINITY_DN3638_c5_g2_i1	OAQ25058.1	2.81E-07	RNI-like protein [Linnemannia elongata AG-77]
TRINITY_DN3638_c5_g2_i1	OAQ25058.1	6.33E-07	RNI-like protein [Linnemannia elongata AG-77]
TRINITY_DN3638_c5_g2_i1	OAQ25058.1	1.13E-06	RNI-like protein [Linnemannia elongata AG-77]

Table S46: blastx results for differentially expressed transcripts identified by DESeq2 in nitrogen pretreated *Euglena gracilis* cultures compared to non-pretreated cultures using the NCBI SWISS-PROT protein database.

SWISS_Isoform Id	SWISS_sseqid	SWISS_evalue	SWISS_description
TRINITY_DN225669_c4_g1_i1	P17804.1	3.26E-19	RecName: Full=Heat shock 70 kDa protein [Leishmania donovani]
TRINITY_DN230405_c0_g1_i1	A5FX11.1	2.30E-41	RecName: Full=NADH-quinone oxidoreductase subunit D; AltName: Full=NADH dehydrogenase I subunit D; AltName: Full=NDH-1 subunit D [Acidiphilium cryptum JF-5]
TRINITY_DN192008_c0_g1_i1	A0PYP1.1	2.49E-37	RecName: Full=Phosphoglycerate kinase [Clostridium novyi NT]
TRINITY_DN127853_c1_g1_i1	P27323.3	1.81E-22	RecName: Full=Heat shock protein 90-1; Short=AtHSP90.1; Short=AtHsp90-1; AltName: Full=Heat shock protein 81-1; Short=Hsp81-1; AltName: Full=Heat shock protein 83 [Arabidopsis thaliana]
TRINITY_DN132906_c0_g2_i1	O00370.1	2.73E-24	RecName: Full=LINE-1 retrotransposable element ORF2 protein; Short=ORF2p; Includes: RecName: Full=Reverse transcriptase; Includes: RecName: Full=Endonuclease [Homo sapiens]
TRINITY_DN132906_c0_g2_i2	O00370.1	1.06E-13	RecName: Full=LINE-1 retrotransposable element ORF2 protein; Short=ORF2p; Includes: RecName: Full=Reverse transcriptase; Includes: RecName: Full=Endonuclease [Homo sapiens]
TRINITY_DN318734_c0_g1_i1	B5ZXZ5.1	2.40E-22	RecName: Full=30S ribosomal protein S9 [Rhizobium leguminosarum bv. trifolii WSM2304]
TRINITY_DN305840_c0_g1_i1	Q46267.1	1.27E-06	RecName: Full=Pyruvate formate-lyase-activating enzyme; Short=PFL-activating enzyme; AltName: Full=Formate-C-acetyltransferase-activating enzyme [Clostridium pasteurianum]
TRINITY_DN330320_c0_g1_i1	P39524.2	5.70E-18	RecName: Full=Probable phospholipid-transporting ATPase DRS2 [Saccharomyces cerevisiae S288C]
TRINITY_DN4481_c1_g2_i1	P14685.3	6.32E-16	RecName: Full=26S proteasome non-ATPase regulatory subunit 3; AltName: Full=26S proteasome regulatory subunit RPN3; AltName: Full=26S proteasome regulatory subunit S3; AltName:

			Full=Proteasome subunit p58; AltName: Full=Transplantation antigen P91A; AltName: Full=Tum-P91A antigen [Mus musculus]
TRINITY_DN809_c27_g1_i1	P28734.1	2.09E-13	RecName: Full=Aspartate aminotransferase, cytoplasmic; AltName: Full=Transaminase A [Daucus carota]
TRINITY_DN1225_c3_g1_i1	Q504A5.1	3.35E-06	RecName: Full=Probable thiopurine S-methyltransferase; Short=Thiopurine methyltransferase [Danio rerio]
TRINITY_DN75002_c0_g1_i1	Q7RTR2.2	5.41E-15	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN75002_c0_g1_i1	Q7RTR2.2	6.96E-15	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN75002_c0_g1_i1	Q7RTR2.2	2.70E-14	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN75002_c0_g1_i1	Q7RTR2.2	5.82E-14	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN75002_c0_g1_i1	Q7RTR2.2	8.19E-14	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN75002_c0_g1_i1	Q7RTR2.2	1.41E-13	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN75002_c0_g1_i1	Q7RTR2.2	4.16E-12	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]

TRINITY_DN75002_c0_g1_i1	Q7RTR2.2	1.08E-11	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN75002_c0_g1_i1	Q7RTR2.2	1.30E-10	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN75002_c0_g1_i1	Q7RTR2.2	8.53E-09	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN75002_c0_g1_i1	Q7RTR2.2	1.49E-07	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN3722_c0_g1_i1	Q62902.1	6.81E-08	RecName: Full=Protein ERGIC-53; AltName: Full=ER-Golgi intermediate compartment 53 kDa protein; AltName: Full=Lectin mannose-binding 1; AltName: Full=p58; Flags: Precursor [Rattus norvegicus]
TRINITY_DN17032_c0_g2_i1	Q86K01.1	1.10E-11	RecName: Full=60S ribosomal protein L15 [Dictyostelium discoideum]
TRINITY_DN43053_c2_g1_i1	Q39249.1	7.03E-23	RecName: Full=Violaxanthin de-epoxidase, chloroplastic; Short=AtVxDE; AltName: Full=Protein NON-PHOTOCHEMICAL QUENCHING 1; Flags: Precursor [Arabidopsis thaliana]
TRINITY_DN23974_c0_g1_i1	Q42564.1	7.48E-25	RecName: Full=L-ascorbate peroxidase 3; Short=AtAPx03 [Arabidopsis thaliana]
TRINITY_DN27215_c2_g1_i1	Q04520.1	9.56E-13	RecName: Full=Diacetyl reductase [(S)-acetoin forming]; AltName: Full=Acetoin(diacetyl) reductase; Short=AR; AltName: Full=Meso-2,3-butanediol dehydrogenase [Raoultella terrigena]
TRINITY_DN27215_c2_g1_i2	Q48436.2	2.24E-11	RecName: Full=Diacetyl reductase [(S)-acetoin forming]; AltName: Full=Acetoin(diacetyl) reductase; Short=AR; AltName: Full=Meso-2,3-butanediol dehydrogenase [Klebsiella pneumoniae]
TRINITY_DN7321_c1_g1_i1	Q94A18.2	5.15E-09	RecName: Full=ABC transporter G family member 29; Short=ABC transporter ABCG.29; Short=AtABCG29; AltName: Full=Pleiotropic drug resistance protein 1 [Arabidopsis thaliana]
TRINITY_DN60075_c0_g1_i1	O33730.2	5.27E-17	RecName: Full=Uncharacterized oxidoreductase Sfri_1503 [Shewanella frigidimarina NCIMB 400]
TRINITY_DN60075_c0_g1_i2	O33730.2	3.74E-11	RecName: Full=Uncharacterized oxidoreductase Sfri_1503 [Shewanella frigidimarina NCIMB 400]
TRINITY_DN3638_c5_g2_i1	Q7RTR2.2	2.14E-10	RecName: Full=NLR family CARD domain-containing protein 3; AltName:

			Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN3638_c5_g2_i1	Q7RTR2.2	1.47E-09	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN3638_c5_g2_i1	Q7RTR2.2	2.52E-09	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN3638_c5_g2_i1	Q7RTR2.2	2.83E-09	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN3638_c5_g2_i1	Q7RTR2.2	7.10E-09	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN3638_c5_g2_i1	Q7RTR2.2	3.64E-08	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN3638_c5_g2_i1	Q7RTR2.2	5.03E-08	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]
TRINITY_DN3638_c5_g2_i1	Q7RTR2.2	7.11E-06	RecName: Full=NLR family CARD domain-containing protein 3; AltName: Full=CARD15-like protein; AltName: Full=Caterpiller protein 16.2; Short=CLR16.2; AltName: Full=NACHT, LRR and CARD domains-containing protein 3; AltName: Full=Nucleotide-binding oligomerization domain protein 3 [Homo sapiens]

Table S47: blastx results for differentially expressed transcripts identified by DESeq2 in nitrogen pretreated *Euglena gracilis* cultures compared to non-pretreated cultures using the Ensembl *Arabidopsis thaliana* database.

Athaliana_Isoform Id	Athaliana_sseqid	Athaliana_evalue	Athaliana_description
TRINITY_DN225669_c4_g1_i1	AT3G12580.1	8.39E-17	Probable mediator of RNA polymerase II transcription subunit 37c
TRINITY_DN230405_c0_g1_i1	ATMG00510.1	1.65E-12	NADH dehydrogenase subunit 7
TRINITY_DN192008_c0_g1_i1	AT1G79550.1	1.20E-20	Phosphoglycerate kinase
TRINITY_DN127853_c1_g1_i1	AT5G52640.1	2.08E-23	Heat shock protein 90-1
TRINITY_DN318734_c0_g1_i1	AT3G49080.1	1.46E-08	30S ribosomal protein S9, mitochondrial
TRINITY_DN330320_c0_g1_i1	AT1G59820.1	1.89E-15	Phospholipid-transporting ATPase
TRINITY_DN4481_c1_g2_i1	AT1G75990.1	6.42E-12	26S proteasome non-ATPase regulatory subunit 3 homolog B
TRINITY_DN809_c27_g1_i1	AT1G62800.1	1.42E-13	Aspartate aminotransferase
TRINITY_DN75002_c0_g1_i1	AT1G10510.1	3.53E-13	RNI-like superfamily protein
TRINITY_DN17032_c0_g2_i1	AT4G17390.1	1.70E-10	60S ribosomal protein L15-2
TRINITY_DN43053_c2_g1_i1	AT1G08550.2	8.08E-24	NPQ1
TRINITY_DN23974_c0_g1_i1	AT4G35000.1	8.60E-26	L-ascorbate peroxidase 3
TRINITY_DN27215_c2_g1_i1	AT3G05260.1	1.67E-07	Glucose and ribitol dehydrogenase homolog 2
TRINITY_DN27215_c2_g1_i2	AT3G05260.1	2.95E-06	Glucose and ribitol dehydrogenase homolog 2
TRINITY_DN7321_c1_g1_i1	AT3G16340.1	5.91E-10	ABC transporter G family member 29
TRINITY_DN60075_c0_g1_i1	AT4G29120.1	1.05E-04	Probable 3-hydroxyisobutyrate dehydrogenase-like 1, mitochondrial
TRINITY_DN3638_c5_g2_i1	AT1G10510.1	8.11E-09	RNI-like superfamily protein

Table S48: blastx results for differentially expressed transcripts identified by DESeq2 in nitrogen pretreated *Euglena gracilis* cultures compared to non-pretreated cultures using the Ensembl *Chlamydomonas reinhardtti* database.

Creinhardtii_Isoform Id	Creinhardtii_sseqid	Creinhardtii_evalue	Creinhardtii_description
TRINITY_DN225669_c4_g1_i1	PNW79920	4.87E-13	hypothetical protein
TRINITY_DN230405_c0_g1_i1	PNW79199	9.83E-16	hypothetical protein
TRINITY_DN192008_c0_g1_i1	PNW76650	4.51E-20	hypothetical protein
TRINITY_DN127853_c1_g1_i1	PNW78238	7.01E-22	hypothetical protein
TRINITY_DN330320_c0_g1_i1	PNW75666	1.50E-15	hypothetical protein
TRINITY_DN4481_c1_g2_i1	PNW82137	2.43E-11	hypothetical protein
TRINITY_DN809_c27_g1_i1	PNW78716	7.95E-07	hypothetical protein
TRINITY_DN75002_c0_g1_i1	PNW88806	5.27E-11	hypothetical protein
TRINITY_DN75002_c0_g1_i1	PNW88806	5.34E-10	hypothetical protein
TRINITY_DN75002_c0_g1_i1	PNW88806	1.77E-05	hypothetical protein
TRINITY_DN75002_c0_g1_i1	PNW88806	3.85E-04	hypothetical protein
TRINITY_DN17032_c0_g2_i1	PNW86545	1.11E-12	hypothetical protein
TRINITY_DN38135_c0_g1_i1	PNW82667	4.95E-09	hypothetical protein
TRINITY_DN23974_c0_g1_i1	PNW79135	5.15E-24	hypothetical protein
TRINITY_DN27215_c2_g1_i1	PNW80395	2.91E-06	hypothetical protein
TRINITY_DN27215_c2_g1_i2	PNW80395	2.26E-05	hypothetical protein
TRINITY_DN7321_c1_g1_i1	PNW76450	2.92E-12	hypothetical protein
TRINITY_DN3638_c5_g2_i1	PNW76530	5.58E-09	hypothetical protein
TRINITY_DN3638_c5_g2_i1	PNW76530	1.09E-04	hypothetical protein
TRINITY_DN80917_c2_g1_i1	PNW70022	9.70E-06	hypothetical protein

Table S49: blastx results for differentially expressed transcripts identified by DESeq2 in nitrogen pretreated *Euglena gracilis* cultures compared to non-pretreated cultures using the Ensembl *Synechocystis sp.* database.

Synechocytis_Isoform Id	Synechocytis_sseqid	Synechocytis_evalue	Synechocytis_description
TRINITY_DN230405_c0_g1_i1	AIE74458	3.38E-10	NAD(P)H-quinone oxidoreductase chain H
TRINITY_DN192008_c0_g1_i1	AIE75237	2.94E-17	Phosphoglycerate kinase
TRINITY_DN318734_c0_g1_i1	AIE75030	5.12E-08	SSU ribosomal protein S9p (S16e)
TRINITY_DN23974_c0_g1_i1	AIE75792	9.94E-10	Catalase / Peroxidase
TRINITY_DN27215_c2_g1_i1	AIE72830	2.48E-11	3-oxoacyl-[acyl-carrier protein] reductase
TRINITY_DN27215_c2_g1_i2	AIE72830	1.83E-11	3-oxoacyl-[acyl-carrier protein] reductase
TRINITY_DN80917_c2_g1_i1	AIE75325	4.49E-04	hypothetical protein

Table S50: blastx results for differentially expressed transcripts identified by DESeq2 in nitrogen pretreated *Euglena gracilis* cultures compared to non-pretreated cultures using the Ensembl *H. sapiens* database.

Hsapiens_Isoform Id	Hsapiens_sseqid	Hsapiens_evalue	Hsapiens_description
TRINITY_DN225669_c4_g1_i1	ENSP00000324173.6	3.17E-15	heat shock protein family A (Hsp70) member 5
TRINITY_DN230405_c0_g1_i1	ENSP00000504215.1	6.15E-13	NADH:ubiquinone oxidoreductase core subunit S2
TRINITY_DN192008_c0_g1_i1	ENSP00000305995.3	2.23E-11	phosphoglycerate kinase 2
TRINITY_DN127853_c1_g1_i1	ENSP00000481908.1	2.29E-17	heat shock protein 90 alpha family class B member 1
TRINITY_DN132906_c0_g2_i1	ENSP00000482892.1	1.32E-26	REV3 like, DNA directed polymerase zeta catalytic subunit
TRINITY_DN132906_c0_g2_i2	ENSP00000482892.1	1.85E-13	REV3 like, DNA directed polymerase zeta catalytic subunit
TRINITY_DN330320_c0_g1_i1	ENSP00000371084.5	8.98E-18	ATPase phospholipid transporting 8A1
TRINITY_DN4481_c1_g2_i1	ENSP00000264639.4	2.78E-16	proteasome 26S subunit, non-ATPase 3
TRINITY_DN809_c27_g1_i1	ENSP00000359539.5	2.81E-07	glutamic-oxaloacetic transaminase 1
TRINITY_DN75002_c0_g1_i1	ENSP00000482302.1	1.98E-16	NLR family CARD domain containing 3
TRINITY_DN3722_c0_g1_i1	ENSP00000251047.4	2.54E-07	lectin, mannose binding 1
TRINITY_DN17032_c0_g2_i1	ENSP00000483260.1	5.20E-07	ribosomal protein L15
TRINITY_DN27215_c2_g1_i1	ENSP00000263278.3	2.63E-11	hydroxysteroid 17-beta dehydrogenase 14
TRINITY_DN27215_c2_g1_i2	ENSP00000472746.1	7.80E-10	hydroxysteroid 17-beta dehydrogenase 14
TRINITY_DN3638_c5_g2_i1	ENSP00000352039.6	5.41E-11	NLR family CARD domain containing 3

Table S51: blastx results for differentially expressed transcripts identified by DESeq2 in nitrogen pretreated *Euglena gracilis* cultures compared to non-pretreated cultures using the Ensembl *Trypanosoma brucei* database.

Tbrucei_Isoforms Id	Tbrucei_sseqid	Tbrucei_evalue	Tbrucei_description
TRINITY_DN225669_c4_g1_i1	EAN80086	2.42E-20	heat shock protein 70
TRINITY_DN192008_c0_g1_i1	CAJ15978	2.99E-13	phosphoglycerate kinase
TRINITY_DN127853_c1_g1_i1	EAN78479	7.97E-21	heat shock protein 83
TRINITY_DN330320_c0_g1_i1	EAN79305	4.20E-21	phospholipid-transporting ATPase 1-like protein, putative
TRINITY_DN21967_c0_g1_i1	AAZ11811	1.59E-04	hypothetical protein, conserved
TRINITY_DN4481_c1_g2_i1	EAN79185	6.69E-20	proteasome regulatory non-ATP-ase subunit 3
TRINITY_DN809_c27_g1_i1	EAN77787	1.37E-07	aspartate aminotransferase
TRINITY_DN75002_c0_g1_i1	EAN78352	2.93E-11	hypothetical protein, conserved
TRINITY_DN75002_c0_g1_i1	EAN78352	1.14E-08	hypothetical protein, conserved
TRINITY_DN75002_c0_g1_i1	EAN78352	6.17E-08	hypothetical protein, conserved
TRINITY_DN75002_c0_g1_i1	EAN78352	3.35E-06	hypothetical protein, conserved
TRINITY_DN75002_c0_g1_i1	EAN78352	4.98E-04	hypothetical protein, conserved
TRINITY_DN3722_c0_g1_i1	EAN79384	5.92E-07	lectin, putative
TRINITY_DN17032_c0_g2_i1	AAZ12062	1.01E-09	ribosomal protein L15, putative
TRINITY_DN27215_c2_g1_i1	EAN78571	3.44E-10	NAD or NADP dependent oxidoreductase, putative
TRINITY_DN27215_c2_g1_i2	EAN78571	4.14E-09	NAD or NADP dependent oxidoreductase, putative
TRINITY_DN3638_c5_g2_i1	EAN78352	1.78E-09	hypothetical protein, conserved