LARGE SCALE PREDICTION AND REVERSE GENETICS ANALYSIS OF

PROGRAMMED CELL DEATH GENES IN *CHLAMYDOMONAS REINHARDTII*

_____

A Thesis

Presented to

The Faculty of the Department of Biological Sciences

Sam Houston State University

_____

In Partial Fulfillment

of the Requirements for the Degree of

Master of Science

_____

by

Matthew Robert Breuer

December, 2018

LARGE SCALE PREDICTION AND REVERSE GENETICS ANALYSIS OF

PROGRAMMED CELL DEATH GENES IN *CHLAMYDOMONAS REINHARDTII*


by

Matthew Robert Breuer


_____


APPROVED:


Anne Gaillard, PhD
Committee Director


Madhusudan Choudhary, PhD
Committee Member


James Harper, PhD
Committee Member


John Pascarella, PhD
Dean, College of Science and Engineering
Technology

# ABSTRACT

Breuer, Matthew R., *Large scale prediction and reverse genetics analysis of programmed cell death genes in Chlamydomonas reinhardtii.* Master of Science (Biology), December, 2018, Sam Houston State University, Huntsville, Texas.

Programmed cell death (PCD) refers to any form of cell death that is coordinated by the genome. PCD consists of complex molecular pathways which directly cause the death of a cell. The most well-known form of PCD is an animal-specific process known as apoptosis, of which the underlying molecular pathways are well-characterized. However, despite the observation that PCD occurs ubiquitously throughout the tree of life, little is known regarding the molecular mechanisms of PCD in non-animal systems. In response to a number of different environmental stressors, *Chlamydomonas reinhardtii* undergoes a form of PCD which exhibits characteristics of apoptosis, including DNA laddering, accumulation of reactive oxygen species, and externalization of phosphatidylserine. The presence of these shared features between *C. reinhardtii* PCD and apoptosis suggests that similar molecular pathways may underlie the two processes. Despite this, many of the genes required for apoptosis in animals appear to be absent in *C. reinhardtii.* In the present study, we first employed a large-scale, homology-based, bioinformatics approach to predict the gene products that contribute to *C. reinhardtii* PCD. From the list of sequences that were obtained by these methods, we selected several entries to study in further detail using a reverse genetic approach. We obtained *C. reinhardtii* mutant strains, each with an insertional mutation in one of the selected genes, from the *Chlamydomonas* Library Project (CLiP) and validated that the insert had been mapped accurately in each of the mutant strains. To determine the effects of losing any of these selected genes, we subjected the mutant and parental background strains to a PCD-

inducing heat stress. First, we sought to determine if the loss of a single selected gene

would affect the ability of cells to undergo PCD in response to stress. Second, we wanted

to determine if the loss of one of the selected genes would alter either the timing or

intensity of phenotypes characteristic of *C. reinhardtii* PCD. Our results suggest a role

for several of the selected genes in PCD, and future studies will be aimed at further

characterizing these roles in more detail.

KEY WORDS:  Programmed cell death, *Chlamydomonas reinhardtii*, Apoptosis,
necrosis, *bi-1*, *ire-1*, *e2f*, *tat-d*, *lsd-1*, *pig3.*

# ACKNOWLEDGEMENTS

**TABLE OF CONTENTS**

# LIST OF TABLES

# LIST OF FIGURES

**CHAPTER I**

**Introduction to Programmed Cell Death**

**A Brief History and Overview of Programmed Cell Death**

Programmed cell death (PCD) can be defined as any form of cell death that is initiated and mediated by the cell's own genome. During PCD, a series of tightly controlled and energy-dependent molecular events directly results in the death of the cell.[1] PCD is distinct from accidental cell death (ACD), which is not a biological process, and hence is not coded for by the genome.[2] Rather, ACD nearly always comes about as a result of an intense and rapid-onset physical or environmental stressor. The primary distinction between these two broad categories of cell death is, while ACD occurs passively, PCD is an active process which requires the input and expenditure of energy to carry out.[2]

The term programmed cell death was introduced in 1964 and was first used to describe a phenomenon observed during the development of the silkmoth musculature.[3] The process was described as a "programmed" death, due to the ability of the researchers to predict which cells would die, and when, during the developmental process. Prior to the coinage of this term, researchers had observed that individual cells of multicellular organisms often die during normal physiological processes.[4] However, it wasn't until 1972 that a clear case was made for the existence of PCD as a form of death distinct from necrosis, which was, at the time, defined as the unintended death of the cell that is brought about as the result of injury.[5,6] Subsequent genetic-based studies in the nematode *Caenorhabditis elegans* and the fruit fly *Drosophila melanogaster* revealed that specific

genes, coded for by the cell's own genome, were responsible for the death of certain cells at specific times during development.[7–9]

In recent years, the scope of organisms which have been utilized for PCD studies has increased dramatically. Studies in PCD have been, and are currently being, conducted in a wide variety of organisms, including mammals, fruit flies, plants, yeast, and even bacteria. The understanding of the molecular and cellular basis for PCD processes has had an extraordinary impact on mankind's understanding of development, homeostasis, cellular communication, and human disease.

**Functional Aspects of PCD Under Normal Conditions**

Though the concept of a genetically-coded "suicide" program superficially seems paradoxical, PCD is an essential and regularly-occurring component of life.

**Multicellular Organisms** Currently, PCD is best understood in the context of multicellular organisms. In such systems, PCD is an essential and regularly-occurring process which plays critical roles in the normal development and homeostasis of the organism. Dysregulation of the ability to undergo PCD is a major contributor to a number of human diseases, including cancer[10], Alzheimer's disease[11], and Parkinson's disease[12].

During development, PCD occurs both in gametes and the developing embryo[13]. In germline cells, PCD is induced when deleterious mutations accumulate or when DNA damage is induced. In this way, PCD functions to protect the integrity of the genome by preventing a genetically-compromised gamete from passing on unfavorable traits to the organism's offspring.[13,14] During animal embryonic development, certain cells are "predetermined" to die at specific times. In this context, PCD facilitates the formation and differentiation of physical assemblies, and also plays a key role in the elimination of

unnecessary physical features, such as vestigial interdigital webbing and sex-specific structures.[15,16] Similarly to animals, PCD also plays a vital role in the proper development of plant-specific structures, such as the root cap and xylem tracheary elements.[17]

For maturing and post-developmental multicellular organisms, PCD is one of the primary means by which tissue homeostasis is maintained.[18] In rapidly-renewing tissues, such as the skin, hematopoietic system, and lining of the intestine, the deletion and replacement of specific differentiated cell types is a continually-ongoing process. This controlled cell turnover serves several purposes, including the maintenance of tissue size and shape, the elimination of aged or damaged cells, and the removal of cell types which frequently come into contact with the outside environment, as such cells have an inherently higher basal mutation rate relative to other cell types.[18]

**Unicellular Organisms** Though best understood in the context of multicellular organisms, PCD is also known to be a regularly-occurring process in unicellular organisms. For many years, the role of PCD in unicellular organisms has been unclear. In the context of multicellular organisms, the purpose of PCD is somewhat intuitive: the death of specific cells, under certain conditions, benefits the organism as a whole. In other words, cellular PCD programs were likely selected for their ability to increase the fitness potential of the organism. The observation that unicellular organisms can undergo PCD is inconsistent with this notion. In the case of unicellular organisms, the cell comprises the organism. Consequently, the death of the cell is also the death of the organism and thus could not increase the organism's reproductive fitness. Why then, would such a seemingly detrimental program be maintained in unicellular organisms? One of the most prominent hypotheses presented to explain the existence and selection

for PCD in unicellular organisms is that, in such systems, PCD acts at the population, rather than the organismal, level.[19]

To understand the phenomenon of unicellular PCD and how it relates to the community as a whole, it is first necessary to briefly review intercellular communication in unicellular organisms. The processes by which unicellular organisms sense and respond to changes in cell density are referred to as quorum sensing (QS). Although QS mechanisms have been characterized primarily in bacteria, similar density-sensing mechanisms have been observed in unicellular eukaryotes as well.[20] Unicellular organisms utilize these signaling systems to synchronize gene expression and, at least to some degree, act as a multicellular entity.[21] The existence of these social behaviors in unicellular organisms, in conjunction with the observation that the death of some cells benefits surrounding kin, has led to the notion that unicellular PCD acts at a level above the organismal level.[22]

Indeed, a compelling justification has been presented to support the hypothesis that PCD in unicellular organisms functions at the population level. In sporulating bacteria, such as *Bacillus* and *Clostridium*, a subgroup of the population forms endospores in response to nutrient depletion.[23,24] During sporulation, cell division produces both a prespore, which will eventually develop into the mature endospore, and a mother cell, within which the prespore resides during the maturation process. Following the maturation of the prespore into the endospore, the mother cell undergoes autolysis, and the mature spore is released from the corpse of the mother cell.[19]

In response to low nutrient abundance, the gram-negative soil bacterium *Myxococcus xanthus* aggregates to form a three-dimensional structure known as a fruiting

body[25]. Within the fruiting body, *M. xanthus* cells differentiate into one of three cell types: 10% of the population form spores, 30% of the population differentiate into rods, and the remaining cells, localize to space between the spore-forming cells and the peripheral rod cells. Each cell type has a defined role within the fruiting body. The cells which sporulate will, at the completion of sporogenesis, become spores that are highly resistant to a number of extracellular conditions. The spores are located at the top of the fruiting body and are dispersed from the into the surrounding area, in search of a favorable growth environment. Cells which have differentiated into rods localize to the periphery of the assembly and act as physical supports for the growing structure. The undifferentiated cells exist between the peripheral rod cells and the innermost sporulating cells, and undergo autolytic PCD, supplying the adjacent cells with a source of nutrients.[26,27]

The life cycle of the eukaryote *Dictyostelium discoideum* bears many similarities to that of *M. xanthus*, in that individual *D. discoideum* cells amass to form a multicellular fruiting body in response to nutrient starvation.[28] Moreover, the formation of *D. discoideum* fruiting bodies is dependent on a subpopulation of cells undergoing PCD.[29] During the initial stages of fruiting body formation, individual *D. discoideum* cells aggregate to form a multi-cellular mound, referred to as a slug.[29] Cells within the slug differentiate into two broad cell types: prestalk cells and prespore cells.[28] Once integrated into the slug, differentiated cells localize to specific locations of the structure in preparation of fruiting body formation. Pre-stalk cells migrate to the front of the structure, while pre-spore cells assemble the rear.[28] During the transitionary stages between the slug and fruiting body stages, the cells comprising the stalk enlarge through

the formation of intracellular vacuoles. This enlargement serves to elevate the prespore cells at the top of the fruiting body, allowing for an increased area of spore dispersal. In addition, cells comprising the stalk contribute to the overall stability of the structure by secreting a cellulose coat which is integrated into both the extracellular matrices of individual cells and to the periphery of the stalk.[30] Approximately 20% of the cells of the fruiting body die as during this process.

Controlled cell death and lysis also play crucial roles in maintaining the integrity and homeostasis of biofilms.[19] Biofilms are complex biological communities which form on a solid surface, and are comprised of microorganisms and a matrix of biomolecules, called extracellular polymeric substances (EPS).[31] The EPS is produced by the cells comprising this structure, and consists of proteins, polysaccharides, and extracellular DNA (eDNA). The incorporation of eDNA into the biofilm is facilitated by QS signals, through which the death and lysis of some cells is achieved.[32,33] eDNA contributes to the overall stability of the biofilm by stably binding to other EPS polymers in the matrix.[34] Furthermore, eDNA aids in the adhesion of cells to one another, as well as to the solid surface.[32,33,35] Following the primary attachment of cells to a solid surface, the affixed microorganisms initialize the production of EPS components, allowing the integration of other cells into the growing biofilm. In addition to contributing the release of eDNA to growing biofilms, PCD may also play roles in established biofilms by providing a source of nutrients under starvation.[31] Moreover, the death of some members of the mature biofilm allows for the detachment of other members into the area. This mechanism of severance facilitates the dispersal of cells into the surrounding environment, providing these cells with the opportunity to establish a new biofilm at a new location.

Cumulatively, these examples provide evidence supporting the hypothesis that PCD in unicellular organisms acts at the population, rather than the organismal, level. That the death of individual cells can, in some instances, provide a benefit to surrounding cells remains uncontested. However, whether the death of some cells in a unicellular community is truly an altruistic death is a matter of debate.[27] For example, in the case of fruiting body formation in *M. xanthus*, it has been proposed that the cells which undergo autolysis are induced to do so by neighboring cells.[27] Indeed, a precedence for such behaviors has been established, as other examples of inter-cellular "self-killing" have been documented in bacteria. For example, during the early stages of sporulation in *B. subtilis*, the sporulating cells produce extracellular substances which induce PCD in neighboring non-sporulating cells.[36] Cannibalization of these dead cells provides a source of nutrients to the mother cell, which may be sufficient to reverse the process of sporulation.[36] Another example of self-killing is a phenomenon observed in *Streptococcus pneumoniae* known as fratricide, in which competent cells induce the death of nearby non-competent sister cells in order to take up the released DNA and integrate it into their own genome via recombination.[37] Thus, at least in some in instances of unicellular PCD, the death of the cell is induced by neighboring cells. As such, in the population sense, it is not entirely clear whether unicellular PCD came to be as the result of intraspecies competition or as an endogenous system of altruism.

**PCD as a Response to Stress**

In addition to occurring under normal conditions, PCD is also a cellular response to "emergency" situations, such as stress. Under such conditions, the cell detects the stress event and initiates the PCD process as a response if unable to adequately cope

(Figure 1). Common external sources of PCD-inducing stress include osmotic stress, high or low temperatures, UV light, anoxia, pathogens, and various chemicals/compounds.[38–42] Internal sources of PCD-inducing stresses include oxidative stress, DNA damage, and the unfolded protein response (UPR).[43–45]
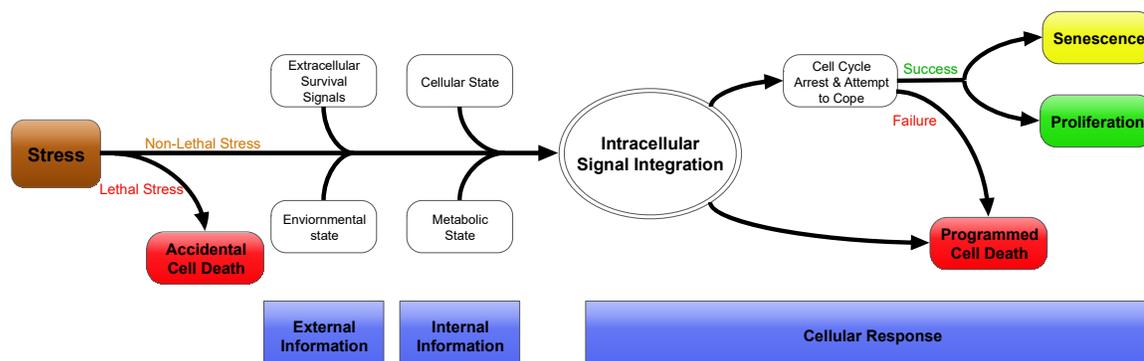


Figure 1. Cellular integration of stress and other signals determines cell fate. Figure was, in part, from Reape and McCabe[46] and Galluzzi et. al[47].

As mentioned previously, the primary characteristic that distinguished PCD from ACD is that PCD is a genetically-encoded process. At the onset of a mild to moderate stress event, both pro-death and pro-survival genes are upregulated and/or activated. This activation of confounding pathways primes the cell to quickly initiate pathways which lead to either survival or cell death.[48] Thus, stress-induced PCD is brought about as the result of pro-death signaling pathways prevailing over pro-survival pathways.[48] Here, it should be reiterated that the activation of these pathways is demonstrative of PCD as an active and energy-dependent process. It should also be noted that while extreme stresses do induce cell death, the death is passive and occurs as a direct result of the stressor itself rather than as a consequence of genetically-encoded molecular pathways.[48] It is by this distinguishing feature that the death of a cell can be determined as PCD or ACD in instances of stress-induced cell death.[48]

**Molecular Mechanisms of PCD**

In the following paragraphs, a brief overview of the PCD field as it stands today will be presented. In particular, the primary molecular mechanisms by which stress-induced PCD is initiated and executed within the cell will be summarized, and an emphasis will be placed on highlighting both the shared and unique features of PCD programs in phylogenetically-diverse organisms.

**Animals** By far, PCD has been most well-characterized in mammals, primarily due to the immediate applicability of mammalian PCD to medicine. The most well-studied form of PCD in any system is a process termed apoptosis.[15] Apoptosis can proceed along either an extrinsic or intrinsic pathway.[49] The extrinsic apoptotic pathway is activated by the binding of extracellular ligands to the ectodomains of cell surface death receptors, which then pass the signal to the cell's interior, activating the intracellular apoptotic pathways. The intrinsic pathway of apoptosis is activated by intracellular signaling which occurs in response to an internal stress signal, such as hypoxia, ER stress, DNA damage, or the accumulation of reactive oxygen species (ROS).

Members of the B cell lymphoma 2 (BCL-2) family play a core role in the regulation of apoptosis. BCL-2 family proteins are divided into three classes, depending on the combination of BCL-2 homology (BH) domains that each possesses.[50] The first group of BCL-2 proteins is the pro-apoptotic effectors BAX, BAK, and BOK, which contain BH1, BH2, and BH3 domains. The second group, the pro-apoptotic BH3-only proteins, is comprised of BID, BIM, BAD, BCL-2-interacting killer (BIK), BCL-2 modifying factor (BMF), BCL-2– adenovirus E1B 19 kDa-interacting protein 3 (BNIP3), activator of apoptosis harakiri (HRK), NOXA and PUMA. The final group is the anti-

apoptotic proteins, which contain BH1, BH2, BH3, and BH4 domains. These include BCL-2, BCL-X$_L$, BCL-W, and MCL1.[50]

While the molecular events involving the BCL-2 family members have yet to be fully elucidated, the activities of these proteins appear to centralize around the oligermization of the pro-apoptotic BCL-2 effector proteins within the outer mitochondrial membrane.[50] The BH3-only proteins appear to be the primary regulators of this process and act by either blocking the inhibitory actions of the anti-apoptotic BCL-2 proteins, or by enhancing the functions of the BCL-2 effector proteins.[50]

Regardless of whether the apoptotic process is initiated along the extrinsic or intrinsic pathway, the end result is the same in both cases: the recruitment of BH3-only proteins results in the activation of BCL-2 effector proteins, which translocate to the mitochondrion and form a pore in the outer membrane called the mitochondrial permeability transition pore (MPTP). The formation of the MPTP results in the loss of mitochondrial outer membrane potential (MOMP) and the release of contents from the mitochondrial intermembrane space (IMS).[51]

The formation of the MPTP represents a crucial step in the apoptotic process which is colloquially referred to as the "point of no return". Prior to the loss of MMP, the death process is considered to be reversible, either by endogenous signaling pathways or pharmacological intervention.[48] Following the formation of the MPTP, however, the cell is irrevocably committed to carrying out the PCD program. The MPTP brings about the death of the cell by two primary mechanisms.[51] First, the MPTP directly disrupts energy-producing processes carried out by the mitochondrion, resulting in the cessation of ATP production.[51] Second, the release of pro-apoptotic molecules from the mitochondrion

results in the activation of pathways which lead to the disassembly and degradation of intracellular components.[48]

The most established apoptotic factor released from the mitochondrion is cytochrome c, which binds and activates cytosolic APAF-1.[51] Activated APAF-1 monomers homopolymerize to form a complex called the apoptosome. Formation of the apoptosome leads to the activation of caspase 9, which then cleaves and activates procaspases 3 and 7, which then proceed to activate other executioners of the apoptotic program.[49]

The formation of the mitochondrial pore also results in the release of other proteins that are crucial to the execution phase of the apoptotic process: apoptosis-inducing factor (AIF), endonuclease G (endoG), high-temperature requirement 2 (HtrA2/Omi), and second mitochondria-derived activator of caspase/direct IAP binding protein with a low pI (Smac/DIABLO). Following its release from the mitochondrial intermembrane space into the cytosol, AIF translocates to the nucleus, where it is believed to mediate chromatin condensation and genomic fragmentation.[52,53] Upon the loss of mitochondrial membrane potential, endoG translocates to the nucleus, where it carries out its endonuclease activity and creates nucleosomal fragments by cleaving chromatin DNA.[54] Finally, HtrA2/Omi and Smac/DIABLO obstruct the anti-apoptotic process by binding and inhibiting the inhibitor of apoptosis (IAP) family members, which would normally function to impede the apoptotic functions of caspase-3 and caspase-9.[51]

The transcription factor p53, perhaps the most renowned participant in the apoptotic process, is highly active during the initiation and execution stages of both the intrinsic and extrinsic pathways.[55] Under normal conditions, p53 is continuously

expressed at constant levels. The inactive form is present in the cytosol, bound to its negative regulator, MDM2. At the onset of stress, MDM2 is ubiquinated, tagging it for destruction by the proteasome. With its inhibitor destroyed, p53 rapidly accumulates within the cytoplasm. Following activation by post-translational modifications, p53 is translocated to the nucleus, where it stimulates the transcription of a number of different genes. First, p53 activates stress-related genes in an attempt to cope with the stress.[56] If the cell succumbs to the stress, p53 activates pro-apoptotic genes to execute apoptosis.[51]

Cells which have undergone apoptosis exhibit a number of characteristics which distinguish them from non-apoptotic cells. As such, these phenotypes are useful in the detection of cells that have undergone apoptosis. These phenotypic changes include, but are not limited to, cell shrinkage[57], nuclear condensation[58], DNA fragmentation[59], a change in the mitochondrial membrane potential (MMP)[60], the rapid accumulation of reactive oxygen species (ROS)[61], and the externalization of the phospholipid phosphatidylserine (PS) to the outer leaflet of the cytoplasmic membrane[62].

Though the current understanding of PCD has come primarily from studying the apoptotic process, multiple forms of PCD have been documented in animals.[22] In recent years, numerous examples of these so-called "alternative" PCD types have been identified. These include necroptosis[63,64], pyroptosis[65], ferroptosis[66], and autophagic death[67], to name a few. Each variant of PCD appears to rely on a different set of molecular pathways, resulting in features that can sometimes be used to distinguish between PCD phenotypes.

**Plants** During PCD, plant cells exhibit many of the of the characteristics of apoptosis observed in animals.[68] These features include cell shrinkage, chromatin

condensation, phosphatidylserine externalization, DNA laddering, TUNEL-positive nuclei, mitochondrial permeabilization and depolarization, and delocalization of cytochrome c from the mitochondrion.[68] Importantly, however, while apoptotic cells display all of these phenotypes and others, whether a plant cell displays any or all of these features during PCD is dependent on the cell type, as well as the nature, intensity, and duration of the stressor.[68] Furthermore, as is the case with animal cells, plant cells similarly appear to have the capacity to undergo multiple and distinct forms of PCD, each of which results in a different combination of detectable phenotypes. Several schemes have been proposed for classifying the various subroutines of plant cell death, but a single standardized method of categorizing plant PCD has yet to be adopted by the plant PCD community.[69–72]

The molecular basis for PCD in plant cells is still somewhat of an enigma. Though efforts made by researchers in recent years have revealed much of the molecular mechanisms of plant PCD, an incomplete picture of the process remains.[68] Many of the proteins known to be essential for animal PCD appear to be absent in plants, at least by available computational models that predict protein function based on primary sequence.[68] This presents the possibility that the pathways underlying plant PCD may have diverged substantially from apoptosis. Despite this, many continue to argue for the existence of apoptosis in plants.[68]

The current evidence indicates that plants do not possess caspases, which are key players in the mammalian apoptotic mechanism. Plants do possess metacaspases, which, though evolutionarily related to caspases, do not cleave at the same peptide sequence. Whether metacaspases function in plant PCD as caspases do in apoptosis is a matter of

debate amongst experts.[73,74] While there is evidence to suggest that metacaspases do participate in plant PCD, the exact role that these proteins play in vivo has not been well characterized.[75] Of note, one protein, tudor staphylococcal nuclease (TSN), has been identified as a conserved substrate between caspase-3 and metacaspase II during animal and plant PCD, respectively.[76]

Though plants presumably lack "true" caspases, plant genomes do appear to contain genes which code for products with caspase-like activity, meaning that the sequence cleaved by a caspase in animal cells is targeted by some other enzyme in plant cells. Furthermore, many of these putative caspase-like proteins have been implicated to play roles in plant PCD, though the majority of their proteolytic targets in vivo are largely unknown.[77]

In addition to the apparent absence of caspases, plants also do not possess BCL-2 family members, another family of proteins that are instrumental in regulating the apoptotic process. Interestingly, however, is that BCL-2-associated anthogene (BAG) proteins have been identified in plants.[78,79] As in animal cells, BAG proteins act as co-chaperones for Hsp70 and act to suppress cell death in plant cells.[80] It has been argued that the occurrence of the conservation of proteins which interact with BCL-2 family members in a non-animal lineage is suggestive of a conserved PCD program between plants and animals.

As plant PCD is accompanied by changes in mitochondrial morphology, loss of MMP, and the subsequent release of cytochrome c, the mitochondrion is also thought to play a role in plant PCD.[81] Further evidence for this role was demonstrated in a study wherein broken mitochondria were introduced into an *Arabidopsis* cell-free system.[82]

This resulted in nuclear alterations that are typical of PCD, including DNA laddering, high molecular weight fragmentation of DNA, and chromatin condensation, whereas the addition of purified cytochrome c had no noticeable effect on these features. The authors originally postulated that a $Mg^+$-dependent nuclease residing in the intermembrane space of the mitochondrion was the causative agent of the observed high molecular weight cleavage and chromatin condensation, though it has more recently been proposed that the dependency of this effect on the co-addition of cytosolic extract with the damaged mitochondria suggests that additional players are likely required.[81]

While comparisons between plant and animal cell death, in an attempt to parallel the two processes, are commonplace, it is indisputable that two organelles, which are absent in animals, participate in plant PCD. The first such organelle is the vacuole. The vacuole is believed to have evolved as a way for plants to combat infection by pathogens in the absence of a dedicated immune system.[83] Vacuole-mediated PCD can proceed by either a destructive or nondestructive route.[72] During the non-destructive variant, the vacuole fuses with the plasma membrane, and is dependent on the caspase-1-like activity of the proteasomal subunit PBA1. This fusion causes the discharge of vacuolar contents into the extracellular space, resulting in the indirect death of the cell, though the exact mechanisms of how the death of the cell is brought about have not yet been described.[72] This extracellular release of enzymes is believed to have evolved as a means of attacking bacterial cells which, in plants, do not enter the cells themselves, but instead proliferate in the extracellular space.[84] The destructive subtype of vacuole-mediated PCD is characterized by the intracellular collapse of the vacuolar membrane, which results in the release of hydrolytic enzymes, as well as the putative caspase-3-like protein VPE, into

the cytoplasm.[85] This mechanism is believed to have evolved as a way of combatting

viral infection, as viruses must enter the plant cell in order to reproduce. The intracellular

release of hydrolytic enzymes serves to eliminate viral material in an attempt to prevent

the virus from spreading to nearby cells. One consequence of the vacuolar membrane

collapse is the degradation of intracellular components, leading to the death of the cell.

The other, and perhaps more obvious, plant-specific organelle involved in PCD is

the chloroplast. In animals, the primary source of reactive oxygen species during normal

metabolism is the mitochondrion. While plants do have functional mitochondria, the

chloroplast serves as the primary producer of ROS in plants.[86] The chloroplast generates

large quantities of ROS by two metabolic pathways: the photosynthetic pathway and the

chlorophyll synthesis pathway.[86] The photosynthetic pathways take place in the thylakoid

membranes of the chloroplast via the reaction centers Photosystem I (PSI) and

Photosystem II (PSII). PSI is necessary for the light-dependent generation of superoxide

radicals ($O_2^-$), which are then utilized to generate $H_2O_2$. PSII is involved in the generation

of singlet oxygen ($^1O_2$).[86] Consistent with the hypothesis that chloroplast-produced ROS

contributes to plant PCD, the absence of light prevents the accumulation of $H_2O_2$ in the

chloroplast and dramatically slows the rate of PCD in *Nicotiana tabacum* leaves.[87] The

opposite effect is seen during exposure to intense light, which can induce PCD in plants

by the surplus of ROS produced by overactive photosynthetic machinery.[88] Another

significant source of ROS that originates from the chloroplasts are the intermediates of

chlorophyll breakdown products.[89]

In addition to being a significant producer of ROS, the chloroplast is thought to play roles in plant PCD which may exist independently of its function in ROS generation. In a mechanism which highly resembles the release of cytochrome c from the mitochondrion, cytochrome f has been documented to be released from the thylakoid membranes of the chloroplast into the cytosol during plant PCD.[90–92] The similarity between these two processes has led to the hypothesis that the cytochrome f may play a role similar to cytochrome c during plant PCD. Further support for this hypothesis was established when purified cytochrome f was demonstrated to induce DNA laddering and caspase-3-like activity.[90]

### Unicellular Organisms

***Eukaryotes*** In unicellular eukaryotes, PCD has been most-thoroughly studied in the budding yeast *Saccharomyces cerevisiae*. Yeast cells have been observed to undergo PCD in response to a number of different environmental stressors, including $H_2O_2$, acetic acid, ethanol, osmotic pressure, UV irradiation, heat, exposure to heavy metals, and various chemical treatments.[93,94]

Yeast cells undergoing PCD exhibit a number of phenotypes which are associated with PCD in animals. These features include chromatin condensation, accumulation of ROS, externalization of phosphatidylserine, loss of plasma membrane integrity, loss of mitochondrial membrane potential, and the formation of the mitochondrial outer membrane pore.[95] DNA degradation can be visualized during PCD in *S. cerevisiae* cells by TUNEL staining. Interestingly, DNA laddering is not observed when genomic DNA from PCD cells is run on a agarose gel; rather, a distinct "smear" is commonly noted.[96–98] While the physiological significance of these observations have not been explored in

detail, it has been proposed that the absence of DNA laddering can be attributed to the chromatin structure of *S. cerevisiae*, as little to no linker regions exist between nucleosomes in this organism.[99]

The molecular basis for PCD *in S. cerevisiae* has been characterized in some detail, though much is still to be learned.[95] Several orthologs of PCD participants from other phyla have been identified in yeast (

Table 1). Notably, several of the key players in apoptosis, such as p53, MDM2, BCL-2 family members, and caspases, do not appear to be present in yeast. Of note, however, is that several *S. cerevisiae* proteins with putative caspase-like activity have been described. Also worth noting is the presence of a single metacaspase in *S. cerevisiae*.

Table 1

Molecular contributors to PCD in Budding Yeast

| Yeast Protein | Orthologs | Functions in *S. cerevisiae* PCD |
| --- | --- | --- |
| Nma111 | HtrA2/Omi | Pro-apoptotic. Serine protease which aggregates in the nucleus during PCD.[100] Cleaves Bir1 through unknown mechanisms.[101] |
| Bir1 | IAP | Anti-apoptotic. Localized to nucleus and cytoplasm. Cleaved by Nma111 during PCD.[102] |
| Ndi1 | Aif/AMID | Pro-apoptotic. Mitochondrial nuclease which trasnlocates to nucleus upon loss of MMP. Contributes to degradation of nuclear DNA.[103,104] |
| Nuc1 | EndoG | Pro-apoptotic. Mitochondrial nuclease which trasnlocates to nucleus upon loss of MMP. Contributes to degradation of nuclear DNA.[105] |

| Yeast Protein | Orthologs | Functions in *S. cerevisiae* PCD |
|---|---|---|
| Esp1 | Separin | Pro-apoptotic. Putative caspase-like seperase. During PCD, cleaves Mcd1, a cohesin that is responsible for maintenance, and repair, of DNA. A small C-terminal peptide fragment of Mcd1 translocates to the mitochondria during PCD, leading to the loss of mitochondrial membrane potential.[106] |
| Kex1 | Caspase-like | Likely pro-apoptotic. Serine carboxypeptidase. Loss of this gene results in decreased caspase-like activities in response to PCD-inducing stress.[107] |
| Cyclin C | Cyclin C | Pro-apoptotic. Localizes to the mitochondria after stress, causing hyperfission and nuclear fragmentation.[108] |
| Ybh3 | Bax | Pro-apoptotic. Putatively Bax-like. Following stress, translocates to the mitochondrion, where it induces MMP.[109] |
| Yca1 | Metacaspases | Pro-apoptotic. Believed to be a central mediator of yeast PCD, but the mechanisms by which it performs this function remains largely unknown.[94] |
| TatD | TatD | Pro-apoptotic. Endo-/exo-nuclease which contributes to DNA degradation.[110] |
| Dnm1 | Drp1 | Pro-apoptotic. Interacts with Fis1 during PCD to induce mitochondrial fission.[111] |
| Fis1 | Fis1 | Pro-apoptotic. Interactions with Dnm1 during PCD to induce mitochondrial fission.[111] |

Note: This table has been adapted, in part, from Strich[112].

Similarly to what is observed in animal cells, cytochrome c is released from the mitochondria during yeast PCD. In spite of this shared PCD feature between animals and yeast, an APAF-1 ortholog has yet to be identified in yeast. As APAF-1 provides the link between cytochrome c and apoptosis in animals, the lack of a detectable APAF-1 ortholog in yeast casts doubt as to whether the liberation of cytochrome c from the mitochondrion serves a function in yeast PCD.

In addition to yeast, PCD processes have also been documented in other unicellular eukaryotes. As the signaling mechanisms for these observed PCD programs have yet to be characterized in detail, they will not be discussed here, though their occurrences are reviewed elsewhere.[113,114] Of note, however, is a study in which a putative p53-like protein was discovered to be present in the slime mold *Entamoeba histolytica.*[115]

*Bacteria* Intriguingly, a number of apoptotic characteristics have been observed during bacterial PCD. These include plasma membrane depolarization, externalization of phosphatidylserine, DNA degradation, accumulation of ROS, cell shrinkage, and DNA condensation.[116]

The molecular basis of PCD in bacteria is poorly characterized.[116] Following is a discussion of the most-studied bacterial PCD pathway, which involves the *mazEF* pathway. *mazEF* is widely distributed throughout bacteria, so it functions as an excellent starting point for discussions centering around bacterial PCD.

*mazEF* is an example of a toxin-antitoxin (TA) system. TA systems are quite prevalent in bacterial cells, and serve to initiate cell death when certain conditions are met.[117] A TA system is composed of two or more genes which exist as an part of an operon. The two fundamental components of a generalized TA system include a toxin, which is lethal to the cell, and the corresponding antitoxin, which blocks the lethal effects of the toxin.[118] Because the two genes are nested within a single operon, they are expressed simultaneously. As a result of being co-expressed with its antitoxin, the toxin has little to no effect on the cell. The *mazEF* system is an example of a Type II TA system, which is classified as a system in which both the toxin and the antitoxin are

proteins, and the antitoxin inhibits the functions of the toxin directly via protein-protein interactions.[119]

The involvement of the *mazEF* system in bacterial PCD is well established, though the exact role that this system plays during PCD has not been fully elucidated.[116] Under normal conditions, the antitoxin, mazE, is bound to the toxin, mazF. In response to a PCD-inducing stress, mazE is degraded, freeing mazF from its inhibitor. Once activated, mazF, an endoribonuclease, is believed to selectively inhibit the bulk synthesis of proteins by cleaving mRNA transcripts at ACA sequences.[120] Under this model, the translation of cell death inhibitors is inhibited, while the translation of cell death effectors remains unaffected.[119] Interestingly, mazF-induced cell death can be annulled by exogenous expression of the mazE antitoxin, but only up to a certain point.[121] This indicates that the PCD process is reversible up to the so-called "point of no return".[116]

The execution steps of bacterial PCD following initiation by mazF are not well characterized.[116] The loss of plasma membrane polarization and subsequent degradation of the cell wall is under the control of the holin-like CidA and antiholin-like LrgA endogenous bacterial proteins.[116] Holins were first characterized as bacteriophage-encoded proteins which coordinate the membrane permeabilization and lysis of infected bacterial cells at specified times.[122] Holins achieve this by oligomerizing in the bacterial membrane to form a pore. Notably, however, is the observation that endogenous holin- and antiholin-like genes are widespread throughout bacteria.[116]

While the exact mechanism underlying the interaction between CidA and LrgA remains unknown, both proteins localize to the bacterial plasma membrane. Under non-PCD conditions, LrgA associates with CidA and inhibits its pro-PCD function.[123]

Furthermore, during bacterial PCD, the association of LrgA and CidA is disrupted, through unidentified mechanisms. Following this, CidA monomers oligomerize to form a pore in the plasma membrane, resulting in the depolarization of the cell.[123]

Utilizing this information, a tentative model for stress-induced bacterial PCD has been proposed.[116] In this scenario, the bacterial TA and SOS responses are activated in response to stress, causing the cell to enter a quiescent state. During this period, the bacterial cell attempts to cope with the stress. An inability of the cell to adequately cope results in the inactivation of LrgA. This inactivation allows for CidA to oligomerize, resulting in the loss of plasma membrane integrity and depolarization of the cell. The end result of this process is nuclease activation, which causes DNA fragmentation, as well as the recruitment and activation of extracellular peptidoglycan (PG) hydrolases, which results in cell wall degradation and subsequent autolysis.[116]

**Uncertainty in the Conservation of PCD**

**Arguments for the Conservation of PCD** The phrase "similar but different" has been used to comparatively describe the animal and plant PCD processes.[68] Similar idioms have been employed to relate unicellular PCD to that of animals.[124,125] The abundance of shared features during PCD in distantly-related phyla, such as phosphatidylserine externalization, genomic fragmentation, and the accumulation of reactive oxygen species, has led to the hypothesis that similar mechanisms underlie PCD in distantly-related organisms.

Some evidence from transgenic studies is consistent with the notion of a conserved PCD program between phylogenetic lineages. Here, an emphasis will be placed on those involving the BCL-2 family members. Heterologous expression of the

animal anti-apoptotic BCL-2 and Bcl-X$_L$ proteins in plant cells was sufficient to attenuate the plant PCD response.[126] Furthermore, overexpression of pro-apoptotic Bax promotes rapid cell death in transgenic plant cells.[127] Intriguingly, the Bax protein localizes to both the mitochondrion and chloroplast when expressed in plants, leading to the hypothesis that a yet-undiscovered Bax-like protein may exist in plants.[127]

As in plants, the heterologous expression of Bax from animals in yeast cells results in cell death with features of apoptosis.[128] Similarly, the expression of anti-apoptotic BCL-X$_L$ or BCL-2 in yeast cells attenuated PCD in response to stress.[129,130] Moreover, co-expression of Bax with BCL-X$_L$ or BCL-2 in yeast cells attenuated the lethal effect of Bax.[130–132] Of interest, heterologous BCL-X$_L$ or human BCL-2 can functionally replace the anti-PCD effects of endogenous yeast Fis1, suggesting that Fis1 may act in a similar manner to that of the antiapoptotic BCL-2 family members.[111]

Perhaps the most intriguing results involving the heterologous expression of Bax come from a recent study in bacteria.[133] When expressed in *E. coli*, the pro-apoptotic BCL-2 effectors Bax and Bak oligomerize to the bacterial plasma membrane to induce cell death and lysis.[133] Additionally, a mutated form of Bax, which causes defective apoptosis in animal cells, was also unable to induce PCD in bacterial cells.[133] Moreover, the co-expression of Bax with constitutively-active Bid repressed bacterial cell death. Finally, replacement of the endogenous holin in lambda bacteriophage resulted in functional bacteriophage particles.[133] The results of this study demonstrate a previously-undiscovered mechanistic link between bacterial cell death and the loss of MOMP during apoptosis. A number of parallels have been drawn between the mechanisms of the BCL-2 family members in animals and the mechanisms of the holin-antiholin system in

bacteria.[116] Namely, both systems function in mediating membrane disruption by controlling the formation of a pore, are regulated by homologous proteins, and control a terminal point of no return during cell death, as proteins downstream of pore formation (caspases in animals, peptidoglycan hydrolases in bacteria) carry out the execution of cell death.[116]

**Arguments Against a Conserved PCD** Although the current number of PCD studies that have been conducted in non-animal organisms are small in comparison to the number of studies carried out in animal systems, a review of the current literature suggests that substantial variation is likely to underlie the molecular mechanisms by which phylogenetically-distant organisms initiate and execute PCD. This concept is supported by the sparsity of discernable animal PCD homologs in non-animal lineages.

The presence of shared PCD features in distinct lineages does not, in and of itself, sufficiently support the claim that the processes are similar at a mechanistic level. A phenotype is brought about primarily as the result of genes which encode molecular pathways that produce the phenotype. Importantly, the phenotypes that result from genomic variation, not the genes themselves, are acted on by natural selection.[134] As such, natural selection may drive the evolution of distinct genes to produce the same phenotype. This concept of convergent evolution establishes that shared phenotypes between phylogenetically-distant species cannot necessarily be attributed to a conservation of the underlying molecular pathways.[134] Indeed, the observation that plants, in the presumed absence of true caspases, have evolved to maintain the observed "caspase-like" activities may be an example of this. Moreover, the emerging contributions of the chloroplast, a plant-specific organelle, during plant PCD indicates

that at least some of the mechanisms of plant PCD are distinct from PCD in organisms

that lack chloroplasts. While these observations do support the notion that PCD processes

are distinct, at least in some ways, between phyla, the degree to which these differences

exist between such organisms has yet to be fully determined, and the majority of

arguments for and against the conservation of PCD pathways throughout the tree of life

are founded in the context of an incomplete understanding of PCD in these organisms.

Furthermore, another important observation to be made is that, even amongst

animals, in which PCD is considered to be well conserved, the molecular mechanisms of

various PCD processes are very distinct from one another; PCD in mammals is, at a

molecular level, very different from PCD in non-mammal animals.[135] This observation

raises questions as to how related mammalian PCD might be from even more diverged

organisms, such as plants.

**On a Potential Conservation of PCD**

The discovery that the individual cells of a multicellular organism can, in a

genetically-dependent fashion, initiate and execute their own death represented a shifting

paradigm in the field of biology. The additional observation that the self-induced death of

damaged or aged cells provides a benefit to the organism as a whole, in a seemingly

altruistic manner, was a paradigm shift in its own right. Hence, it is unsurprising that

many of our own ideas and notions of PCD have been influenced by the emphasis that

has historically been placed on the apoptotic process.[134] The discovery that organisms

which are phylogenetically distant from animals also utilize PCD led to the proposition

that perhaps the mechanisms of cell death are not a recent evolutionary development, but

were rather inherited from a common ancestor. The revelation that PCD also occurs in

unicellular eukaryotes, and even in prokaryotes, is in agreement with this postulation.

When subsequent studies revealed that unicellular organisms utilize PCD at a population

level, parallels were immediately drawn between the apparent social aspects of PCD in

unicellular organisms and the cooperative nature of PCD in multicellular organisms.

Namely, that both processes provide a benefit to surrounding cells. This led to the

successive prediction that the social aspects of PCD between animals and unicellular

organisms are evolutionarily linked, and thus, may be evolutionarily conserved at the

molecular level.[134] This idea of a "core" molecular mechanism by which cells bring about

their own destruction has pervaded the scientific community for many years, and has had

an extraordinary impact on the way that PCD is perceived.[134] There are several points

that require attention with regard to the notion of a conserved PCD program.

**On the Emergence of PCD in Evolutionary History** Though it is unclear as to

how and when PCD evolved, the observation that contemporary bacterial species undergo

PCD has led to the speculation that PCD may have emerged prior to the rise of

eukaryotes. Thus, much of our understanding of the early PCD processes comes from

studies in bacteria. As described above, our current conception of PCD indicates that

PCD serves a collective purpose in many, if not all, unicellular populations.

The initial findings of a role for cell death in multicellular organisms, as well as

the additional discovery of its genetic basis, led to a so-called "fascination for function",

in which PCD became widely viewed within an implicit conceptual framework of both

purpose and design.[134] The observations that unicellular organisms also undergo forms of

death that are genetically-controlled, and that the group, rather than the organism as a

whole, receives the benefit of the individual cell's death, led researchers to apply the same

reasoning used to explain PCD in multicellular organisms to explain PCD in unicellular organisms.[134] Stated another way, the discoveries of the beneficial effects to neighboring cells in both unicellular and multicellular organisms was used as a way to relate the two processes. As a consequence of this ambiguous logic, PCD became subliminally perceived to have emerged when it became required to perform the evident functional roles within a unicellular community, and that this collective role of PCD may have given rise to the cooperative nature of PCD observed in more complex multicellular organisms.[134] While such a notion does make sense on some intuitive level, and, as a result, is easily digested by the majority of those in the field of biological sciences, this line of reasoning raises an important problem when viewed from the perspective of evolutionary theory: such a view implies that a complex set of genes and molecular pathways facilitating the induction of cell death emerged precisely when and where it was required in ancestral organisms.[134]

The assumption that the functions of PCD observed in contemporary organisms were the driving forces behind the initial emergence of PCD fails to take into account a core principle of evolutionary theory: descent with modification.[134] For instance, there are examples in which a complete dissociation exists between the functional advantage that a given set of genes currently confers to an organism and the initial reason for which these genes emerged and were maintained. Thus, rather than the perspective that "PCD was selected for because of the benefits that it provides", a more evolutionary sound viewpoint may be that "PCD was adapted to provide the benefit that it is observed to provide today". When the evolution of PCD is framed in such a manner, the point becomes clear that an understanding of the functional aspects and selective pressures of

PCD in early evolution are required to comprehend the molecular contrivances of ancient PCD programs.[134] Moreover, an understanding of both the acquisition of the ability to undergo PCD, as well as the selective pressures which maintained and modified the functional aspects of PCD, are necessary in order to relate the functional and molecular aspects of PCD between unicellular and multicellular organisms.

**On the Conservation of PCD Programs**

The notion that PCD programs are not directly selected for, but are instead a "tolerated" as byproducts of other selective pressures, is a prominent hypothesis for the emergence of PCD in early life and maintenance of PCD throughout evolutionary history.[134] Under this model, PCD is postulated to have evolved through a series of stages, wherein gene products that are under selection for advantageous traits they provide develop additional functions to participate in PCD. Thus, any and all gene products that confer a trait that is beneficial to the organism could potentially be adapted to participate in PCD. Because natural selection acts at multiple levels of organization, one implication of the multi-step model of PCD evolution is that the selective pressures facilitating the major transitions from unicellularity to complex multicellularity likely played a role in shaping PCD processes in these emerging organisms.

**Co-Selection at the Molecular Level** One way in which this idea might manifest is at the molecular level. The majority of known genes are pleiotropic, meaning that the expression of the gene has an effect on multiple phenotypes. Thus, a phenotype that is not directly acted on by natural selection, assuming that the trait is not maladaptive, and may be co-selected with another phenotype that is directly selected for. In the context of PCD, this may have involved the emergence of PCD functions in gene products which

participate in essential cellular processes. In agreement with this, many of the genes encoding processes such as cell cycle progression and metabolism are involved in PCD. An important implication of this, however, is that natural selection exerts distinct selective pressures on different organisms/groups/populations, depending on which adaptations most increase fitness. This is generally observed by environmental pressures. Hence, the features selected for in one group will not be exactly identical to those selected for in a different group. As a consequence, the genetic mutations that produce the phenotypes under selective pressure will be different between any two groups. Moreover, the random nature of genetic mutations dictates that, even between two closely-related groups, some variation of the phenotype selected for will be exist, implying that the function of the co-selected phenotype may also be altered between these two groups. Thus, selective pressures exerted on essential cellular functions will successively shape the genomic sequences which carry out these functions. This will, in turn, have a substantial effect on the PCD functions of that pleiotropic gene.

  **Co-Selection at the Cellular Level** A second way by which the effect of co-selection may have resulted in the divergence of PCD processes is in the context of intracellular structure. The mitochondria and chloroplasts are both eukaryotic organelles which arose as the result of endosymbiotic interactions between bacterial cells and their primitive eukaryotic hosts. The traits selected for in these cases were the abilities to generate large quantities of energy. The indispensable role of the mitochondrion has been well established in animal PCD. Characteristics similar to that of its animal counterpart have been observed during plant PCD, though it has yet to be shown that the plant mitochondrion plays the same essential role during PCD. Furthermore, as described

above, the presence of chloroplasts in plants and their absence in animals, in conjunction with the purported involvement of the chloroplast during plant PCD, implies that distinct mechanisms exist in the ways by which plants and animals carry out PCD.

In addition, the prokaryotic origin of these organelles likely had considerable ramifications for the evolution of eukaryotic PCD. One such implication of the early host-endosymbiont relationship is that many of the aspects of eukaryotic PCD may be artifacts of bacterial PCD pathways. It is intriguing to note that, in both plants and animals, proteins involved in PCD may have been inherited from the prokaryotic endosymbiont. In animals, this is exemplified by the mitochondrial serine protease HtrA2, a homolog of the bacterial stress tolerance protein HtrA.[137] Additionally, the plant homolog of a different bacterial protein, LrgB, was recently identified in *A. thaliana*. Though a role for bacterial LrgB has yet to be described, it exists as a constituent of the same operon as the antiholin-like LrgA. In *Arabidopsis*, LrgB localizes to the chloroplast and has been implicated to play a role in plant PCD.[138]

**A Case for the Characterization of PCD in Unicellular Organisms**

Ultimately, a detailed comparison of the molecular pathways underlying PCD requires a detailed understanding of the participants of each. With regard to PCD pathways, the sole process that has been thoroughly-described at the molecular level is apoptosis. Even still, much remains to be understood regarding the molecular mechanisms of apoptosis.

That the ability to undergo PCD is widespread throughout the evolutionary tree suggests the early development of PCD programs in evolutionary history.[134] Because of this, studies into the molecular basis of PCD in simpler unicellular organisms have the

potential to shed light on the molecular basis of PCD in an evolutionary context. Thus, it is here argued that an understanding of the molecular mechanisms which underlie PCD in unicellular organisms is foundational to understanding how PCD processes evolved in other life forms.

The results of a recent genomic study utilizing bioinformatics suggests that there may exist a "core" PCD program that has been conserved in unicellular eukaryotes.[139] The presence of such a mechanism amongst representative ancestral organisms may have great potential in shedding light on the enigmatic nature of the evolution and diversification of PCD. Moreover, such a mechanism has the invaluable potential to identify aspects of PCD that are conserved among higher organisms, such as plants and animals, which may subsequently lead to identification of divergent mechanisms which underlie these processes.

### *Chlamydomonas reinhardtii* and Programmed Cell Death

Green algae occupy a unique position on the phylogenetic tree. Believed to be among the closest representative ancestors of higher plants, green algae serve as a primary model for studying processes in such organisms.[140,141] This is, in part, because green algae possess functional chloroplasts. However, in addition, many plant-specific genes are found in green algae, making them an even more attractive organism in which to study plant processes.[142]

Of note, several photosynthetic green algal species, including *Micrasterias denticulata*, *Chlorella saccharophila*, *Dunaliella tertiolecta*, *Chlamydomonas reinhardtii, and Volvox carteri* have been documented to undergo PCD in response to various stressors.[91,143–147] PCD in such organisms exhibits features resembling animal apoptosis,

including DNA laddering, externalization of phosphatidylserine, vacuolation, degradation of intracellular components, DNA fragmentation, and accumulation of reactive oxygen species.[148] Intriguingly, as is observed in land plants, PCD in *C. saccharophila* is accompanied by the release of cytochrome f from the chloroplast, suggesting that this characteristic of plant PCD may have developed early in the divergence of PCD between plants and animals.[91]

If green algae represent prime candidates for studies in PCD, *Chlamydomonas* arguably serves as a prime green algal candidate for such studies. *C. reinhardtii* is an extremely well-studied model organism with a genome that has been sequenced and annotated.[149] Additionally, *C. reinhardtii* has been used to study a diverse array of biological processes, including motility[150], ciliary structure and function[151], lipid synthesis[152], genetics of the chloroplast[153,154], photosynthesis[155], heterotrophic metabolism[156,157], stress response and acclimation[158], and biofuel production.[159,160] Furthermore, a particularly novel aspect of *C. reinhardtii* is the ability to form rudimentary multicellular structures.[161] This, in conjunction with the fact that *C. reinhardtii* shares remarkable genomic similarities with its multicellular cousin, *V. carteri*, has led to the use of *Chlamydomonas* in laboratory studies seeking to understand the evolution of multicellularity.[162]

*Chlamydomonas* serves as a superb model organism for eukaryotic cell biology due to its functional conservation with plants, animals, and other unicellular eukaryotes.[149] These relationships have led to the light-hearted and colloquial coinage of light-hearted nicknames such as "planimal" or "green yeast" to describe *Chlamydomonas*. Moreover, the ease of culturing, rapid growth rate, and advantages in genetic

manipulation have established this organism as the preferred model for many forward genetic screens and other high-throughput studies.[163–165]

Adding to the utility of *C. reinhardtii* as a model is the fact that more advanced genetic and molecular tools are continually being developed for use in this organism.[166–169] Targeted DNA editing in *Chlamydomonas* is now plausible, and techniques such as clustered regularly interspaced palindromic repeats (CRISPR) and zing-finger nucleases (ZfNs) have recently been described for use in *C. reinhardtii*.[170,171] Moreover, the development of other molecular methods such as RNA interference (RNAi) and transactivator-like effectors (TALEs), as well as the creation of a *C. reinhardtii* knockout mutant library, have allowed for even further possibilities in the field of reverse genetics.[172–174]

One process which has not been well characterized in *Chlamydomonas* is PCD. *C. reinhardtii* has been observed to undergo PCD in response to a number of environmental stressors. Interestingly, it appears that many of the typical phenotypes observed during mammalian apoptosis are also observed during *C. reinhardtii* PCD (Table 2).

Table 2

Stress-induced PCD in *C. reinhardtii* and the resulting features

| Stressor | PCD Phenotypes |
| --- | --- |
| UV radiation[175,176] | Cell shrinkage, vacuolation, chromatin fragmentation, DNA fragmentation, and phosphatidylserine externalization. |
| Acetic Acid[177] | Accumulation of ROS, decrease in photosynthesis, DNA TUNEL-positive nuclei, DNA laddering. |
| Mastoparan[178] | Cell shrinkage, vacuolation, cytoplasm shrinkage and separation from the wall, nuclear condensation, dramatic production of ethylene and NO gas. |

| | |
|---|---|
| Heat[147,179–181] | ROS accumulation, DNA laddering, and phosphatidylserine externalization. |
| Menadione[182] | Decrease in photosynthesis, ROS accumulation, upregulation of antioxidant enzymes, loss of MMP, DNA fragmentation, caspase-3-like activity. |
| Hydrogen peroxide[183,184] | DNA fragmentation, ROS accumulation, caspase-3-like activity, and upregulation of antioxidant enzymes. |
| Potassium chloride[185] | Cell shrinkage, distorted chloroplast, vacuolation, accumulation of ROS, upregulation of antioxidant enzymes, loss of MMP, DNA fragmentation, nuclear disruption |

The phenotypic similarities between *C. reinhardtii* PCD and mammalian apoptosis suggest that common molecular mechanisms underlie the two processes. As is true of PCD in non-animal organisms, however, is that much of the molecular machinery that is known to be essential for apoptosis, such as caspases or p53, appears to be absent from the predicted *C. reinhardtii* proteome. Furthermore, though several putative contributors to *C. reinhardtii* PCD have been identified (Table 3), detailed signaling pathways underlying this process have yet to be elucidated. As a result, the extent to which *Chlamydomonas* PCD and mammalian apoptosis are similar is currently unknown.

Table 3

Putative molecular contributors to PCD in *Chlamydomonas reinhardtii*

| Contributor | Method of Detection | Results |
|---|---|---|
| DAD1[176] | RT-PCR | Gene expression was downregulated during PCD. |
| APAF-1[176] | Western blot | Increase in protein abundance during PCD |
| Caspase-1-like[178] | Enzymatic inhibitor and fluorescent substrate | Early application of inhibitors reduced the percent death of the population. Increase in enzymatic activity throughout PCD. |

| Caspase-3-like[178] | Enzymatic inhibitor | Early application of inhibitors reduced the percent death of the population. |
| Caspase-3-like[183] | Western blot, enzymatic inhibitor, fluorescent substrate, and measurement of substrate (PARP-1) cleavage. | During PCD, higher levels of caspase-3-like protein, higher levels of caspase-3-like activity, which were decreased in the presence of the inhibitor, and increased cleavage of PARP-1. |
| Metacaspase-2[184] | RT-PCR | Upregulated during PCD. |
| AIF[185] | Western blot | During PCD, cytoplasmic AIF levels increased, which were correlated with a change in mitochondrial membrane potential |
| PIG8[147] | Northern blot | Upregulated during PCD. |

## Objectives

As very little is known regarding the molecular underpinnings of PCD in *C. reinhardtii*, the overarching purpose of this work is to shed light on the genetic and molecular basis of PCD in *C. reinhardtii*. In order to achieve this goal, the following objectives are proposed.

Objective 1: *C. reinhardtii* proteins which participate in PCD will be predicted by comparing the sequences of all *C. reinhardtii* proteins to the protein sequences which contribute to PCD in other organisms. Using this information, several *C. reinhardtii* proteins which show high sequential similarity to a known PCD protein will be selected for further analysis.

Objective 2: To determine if each of the selected proteins is important in *C. reinhardtii* PCD, a reverse genetics approach will be undertaken. Mutant strains lacking a

selected putative PCD protein will be subjected to PCD-inducing conditions and assayed for changes in typical PCD phenotypes.

**Significance**

There are both direct and indirect benefits to an understanding of the molecular mechanisms underlying *C. reinhardtii* PCD.

**Direct benefits** *C. reinhardtii* has extensive potential for industry use. As a result of producing abnormally large amounts of dihydrogen during photosynthesis, ethanol through carbohydrate fermentation, and methane through anaerobic digestion, *C. reinhardtii* is currently being investigated as a potential source of clean biofuels as an alternative to fossil fuels.[186–188] Additionally, other aspects of the microalgal metabolism, such as the production of carotenoids, long-chain polyunsaturated fatty acids, and vitamins, have instigated their use in the production of food and animal feed additives.[189] One area of developmental research in the biofuel industry aims to increase the quantity of algal-derived biofuels by altering the culture conditions in which the organism is grown. In particular, stressful conditions have been shown to dramatically influence the quantities of industry-usable compounds produced by microalgae.[190] A thorough understanding of the molecular basis of PCD in *C. reinhardtii* has the potential to facilitate the development of PCD-resistant algal strains. This could allow for the culturing of *C. reinhardtii* under stressful conditions not otherwise possible, and thus may contribute to the commercial utilization of *C. reinhardtii* or other algal strains.

*Chlamydomonas* is also used as a biomarker for various environmental studies which seek to determine the toxicity of various compounds.[191–195] Many of the molecular markers that are assayed for in these studies are typical markers of PCD. As such, an

understanding of PCD in *C. reinhardtii* could potentially help to elucidate specific and relevant information regarding the molecular basis underlying the toxicity of a given compound.

**Indirect Benefits** Over 700 million years of evolution separate the chlorophytes (unicellular green algae) and streptophytes (non-chlorophyte green algae and land plants).[196] As such, unicellular green algae represent a highly unique phylogenetic group. Because of the unique position that *Chlamydomonas* occupies on the phylogenetic tree, findings in this organism are often easily translated to other phylogenetic clades. Understanding PCD in *C. reinhardtii* is likely to have several benefits in this regard.

As green algae pre-date multicellular plants, *C. reinhardtii* is often used as a model for physiological processes in higher plants.[149] This seems particularly relevant in light of the fact that over 33% of the proteins predicted from the *A. thaliana* genome have not been characterized in terms of their biological roles.[197] Understanding the PCD process in plants is a chief goal of many researchers, as the development of stress-resistant crops is essential for expanding the space on which it is possible to grow crops, increasing crop yield, and sustaining a food source in an environment which is continually growing warmer.

In addition, the phylogenetic position of *C. reinhardtii* is advantageous to evolutionary studies seeking to understand how PCD evolved and diversified in early life forms. As a unicellular eukaryote with photosynthetic machinery, *Chlamydomonas* is set apart from many of the traditionally-used models for PCD in unicellular eukaryotes, such as *S. cerevisiae*. Furthermore, given that the chlorophyte-streptophyte divergence occurred after the divergence of plants and animals, *C. reinhardtii* also serves as an

excellent transitionary model with the potential to yield novel insights into other

unicellular eukaryotes, higher plants, and animals.[198] As such, not only could *C.*

*reinhardtii* could aid in understanding how PCD evolved in early life, but it could also

help to understand how PCD diversified between distinct lineages.

## CHAPTER II

## Prediction of PCD Genes in *Chlamydomonas reinhardtii*

**Introduction**

The rapid advancement of biotechnological advancements in recent years has triggered a precipitous outpouring of molecular data. This surge in the quantity and production rate of available data has resulted in the emergence of a new field of study: bioinformatics. In short, bioinformatics can be described as the interface between computing and molecular biology.[198] In bioinformatics research, computational systems are used to analyze large amounts of biological data, most of which would be impossibly arduous to analyze manually. As computers are capable of handling and analyzing large quantities of information with astounding speed and efficiency, they are the ideal instruments on which to carry out the analysis of such data. The increasing abundance of molecular data, as well as the rapid advances made in computer science, have allowed for the development of new questions, hypotheses, and analyses in the field of biology.

As a consequence of the large amounts of data produced by modern molecular studies, databases dedicated to the curation of biological data serve to gather, organize, and store information produced by researchers.[199] A biological database usually specializes in one or several types of data, and a variety of databases exist for different types of biological data. Some databases collect sequence data, others gather data regarding the three-dimensional structure of biomolecules, and still others assemble systematic data to generate large interaction networks.[199]

A prominent area of research in the field of bioinformatics is sequence analysis.[200] The computational analysis of biological sequences has wide-reaching implications for the field of biology. For example, given a fully sequenced genome, the individual genes which comprise the genome can be predicted. From any of these putative gene sequences, a corresponding transcript sequence can be predicted. The transcript sequence can be used to predict the amino acid sequence of the resulting protein. The amino sequence can be used to predict a number of functional characteristics of the protein, including three-dimensional structure, functional domains, and active sites.

Functional predictions of an uncharacterized protein are derived from similarities to proteins of known biological function. The evolutionary conservation of peptide function is often accompanied by the conservation of amino acids key to that function.[201] Thus, the functions of an indeterminate protein can be predicted by comparing its constituent amino acid sequence with the amino acid sequences of functionally-characterized proteins. The most common method to search for sequential similarities between different proteins is to align the amino sequences with one another.[202] Each alignment tool works slightly differently, as each method utilizes different algorithms to align the amino and calculate the strength of the alignment. The most widely utilized tool for sequence alignment is the Basic Local Alignment Search Tool (BLAST).[200,202] BLAST works by searching for regions of similarity between two sequences and moves outward in a bidirectional manner from the initial alignment until the terminus of one of the sequences is reached or until the regions of similarity end.[203]

The Gene Ontology (GO) project is a database of gene product annotations, which seeks to represent the current knowledge of how genes function at multiple levels.[204] In

the GO database, gene products are assigned one or more functional annotations, called GO terms. GO terms fall into one of three broad, and exclusive, categories: the biological processes (BP) in which the protein participates, the cellular components (CC), or where in the cell the protein localizes to, or the molecular functions (MF) that are carried out by the protein.[204]

The three GO categories are structured as distinct hierarchical trees. In these acyclic graphs, each term can be related to one or more "parent" terms, which are broader in scope, or "child" terms, which are narrower in scope.[205] Each of the three categories is the top parent node of its own graph, and is the broadest annotation of its constituent child nodes, which, in turn, are the parent nodes of their own constituent child nodes, and so forth (Figure 2). In addition, a single node can be considered the child of several parent nodes within its own graph, exemplifying the fact that a GO term can be relevant in a variety of different biological contexts.[205] One benefit to being able to view GO terms in such a hierarchal manner is that it enables the visualization of the GO terms assigned to a dataset in as broad or narrow a biological context as desired.[204]
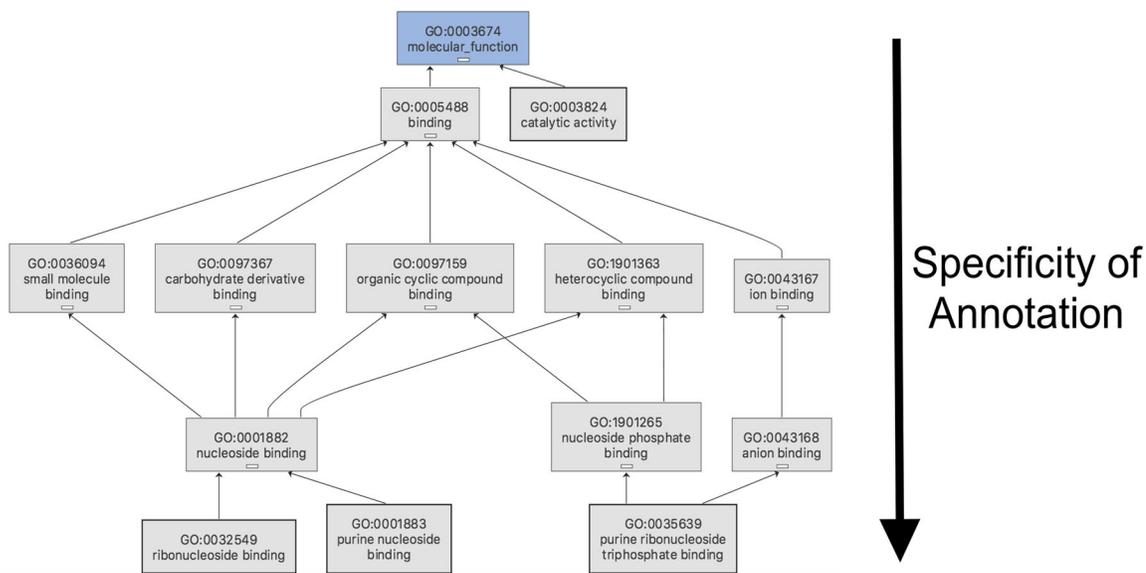
Figure 2. Example of the hierarchal classification system in Gene Ontology. The organization of GO terms can be visualized as vertically-connected nodes, depicting an ordered relationship. The parent nodes, located at a higher position on this diagram, are composed of one or more child nodes. In turn, each child node is the subset of at least one parent node. This example was created using BLAST2GO.

GO terms for an uncharacterized protein sequence can be predicted using a variety of available tools.[205] One such tool is BLAST2GO, a software that utilizes the results of a BLASTp alignment, in conjunction with other computational tools, in order to predict the GO term(s) for a query sequence.[206] The advancements in the speed and processing power of computers allows for the batch processing of many sequences simultaneously, and tools like BLAST2GO are often used to predict GO terms for large datasets that consist of many sequences.

The standardized BLAST2GO workflow for functional prediction of a protein sequence consists of several steps. First, a BLASTp search is carried out, using the protein(s) to be characterized as the query. Second, GO mapping is performed, wherein the GO terms associated with each of the subject sequences to which the protein of interest is aligned are retrieved. Third, GO terms are assigned to the protein of interest by

applying a mathematical formula called the annotation rule (AR). Briefly, the AR assigns GO terms to the protein of interest while accounting for 1) the strength of the alignment between the uncharacterized protein and the subject sequence, and 2) the strength of the GO annotation assigned to the subject. It then assigns the GO term of the lowest node which falls above a user-defined threshold.

One potentially useful feature of BLAST2GO is that it assigns a level to each term in the GO database.[207] In this system, each of the three GO categories are the first level of their own graph, child nodes of these categories are assigned to the second level, and so forth. Using Figure 2 as an example, "molecular function" would be assigned level 1, both "binding" and "catalytic activity" would be assigned level 2, and "small molecule binding", "carbohydrate binding", "organic cyclic compound binding", "heterocyclic compound binding", and "ion binding" would be assigned level 3. In this way, the level system is intended to estimate the relative specificity of a GO term by informing the user of what level the term belongs to.[207]

Since the *C. reinhardtii* genome was first published in 2007, it has been sequenced and reanalyzed a number of times. Each new release contains notable improvements in annotation and specificity.[208] This continued improvement in quality makes *C. reinhardtii* an excellent organism for studies in bioinformatics. A total of 19,526 amino acid sequences are predicted to be coded for by the current version of the *C. reinhardtii* genome (v5.5).

As noted in the previous chapter, exceedingly little is known regarding the genetic basis for PCD in *C. reinhardtii*. A survey of the literature suggests that no large-scale study to predict the *C. reinhardtii* PCD proteins has been conducted. As such, the

purpose of this study was to predict, on a large scale, the proteins which participate in *C. reinhardtii* PCD using sequential similarities to annotated PCD proteins from other organisms. Furthermore, we sought to characterize these potential PCD proteins by assigning GO terms to the dataset.

**Materials and Methods**

Figure 3 provides a visual overview of the methods that were utilized in this section of the project. Detailed computational methods can be found in the appendices. Unless otherwise noted, all management, analysis, and visualization of the data used in this study was carried out using the bash and R languages.
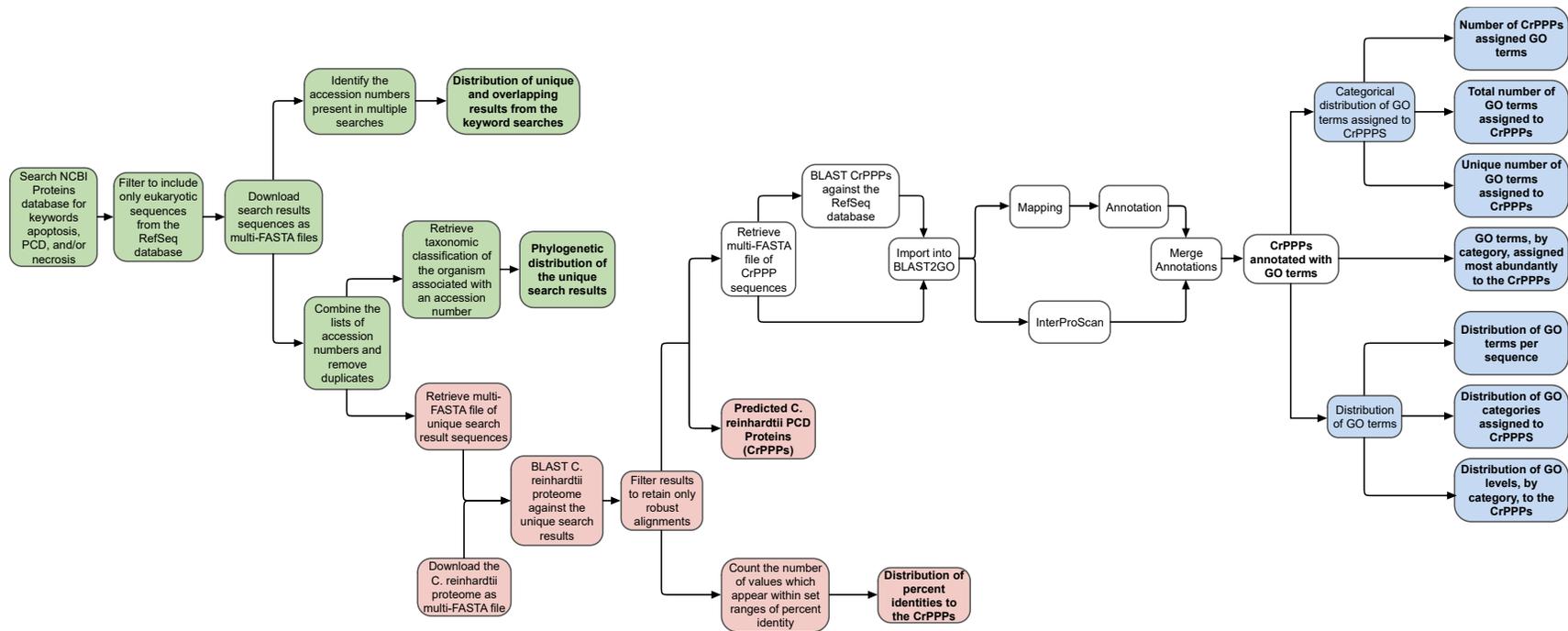
Figure 3. Overview of the methods used to predict and characterize *C. reinhardtii* proteins which participate in PCD. Nodes with bolded text represent output of the pipeline. Green nodes represent steps used to retrieve subject sequences, red nodes represent steps used to align the *C. reinhardtii* proteome against the custom PCD database, and blue nodes represent the steps used to characterize the predicted PCD proteins in *C. reinhardtii*.

**Retrieval of Annotated PCD Sequences** Using the National Center for

Biotechnology Information (NCBI) web page (https://www.ncbi.nlm.nih.gov), the

RefSeq Protein database was queried for amino acid sequences that were annotated with

the keywords "apoptosis", "necrosis", and/or "programmed cell death". The sequences

coded for by eukaryotic genomes were saved locally as three separate multi-FASTA files,

where each file contained the results of a single keyword search. The accession numbers

were extracted from each of the multi-FASTA files using the "Find" and "Replace"

functions in the text editing software TextWrangler. These three lists of accession

numbers were saved as separate text files. To determine the overlapping results between

the three keyword searches, the R function "reduce" was used to identify the accession

numbers which were present in more than one of the lists of accession numbers.

In order to create a single list of results from all three keyword searches, the three

files of accession numbers were combined into a single file. All duplicate accession

numbers (resulting from items that were returned in multiple keyword searches) were

removed from this combined file using the command line. To confirm that this operation

removed the correct number of entries, the number of accessions that were removed from

the file of combined search results was cross-referenced with the number of overlapping

items identified using the "reduce" command above.

To convert the combined list of accession numbers into a multi-FASTA file of

amino acid sequences, a custom E-Utilities script was written to retrieve the amino acid

sequence that corresponded with each accession number from the NCBI Entrez database.

To confirm that all sequences were processed by the E-Utilities script, the number of

sequences in the multi-FASTA file was cross-referenced with the number of lines in the list of combined, non-redundant accession numbers.

**Alignment of the *C. reinhardtii* Proteome to the Annotated PCD Sequences** A complete list of the predicted amino acid sequences coded for by the *C. reinhardtii* genome (v5.5) was downloaded as a multi-FASTA file from the Phytozome web page (https://phytozome.jgi.doe.gov/pz/portal.html). To access more computing power, an EC2 instance was created using Amazon Web Services (AWS) web page (https://aws.amazon.com). The multi-FASTA file of the *C. reinhardtii* proteome was uploaded to the AWS instance. Additionally, the multi-FASTA file of sequences from the combined keyword search results, obtained above, was also uploaded to the instance and formatted into a BLASTp database using the pre-installed BLAST+ software (version 2.5.0).

To identity *C. reinhardtii* sequences with similarities to one or more entries from the combined search results, a BLASTp protocol was carried out, in which the *C. reinhardtii* proteome was used to query the custom database of annotated PCD proteins. To facilitate any future use of BLASTp alignments in different formats, the raw output of BLASTp was specified to be in archive format (.asn). Using the blastformatter tool in the BLAST+ software, the archive-format file was used to create a new file of the BLAST results in tab-delimited format. Both formats of the BLASTp output were downloaded locally. To filter out the weaker matches, any alignments which had a bitscore <100 or a percent identity <25% were removed from the tab-delimited BLASTp results. From this list of robust alignments, the identifiers of the *C. reinhardtii* query sequences were parsed and saved into a new file.

To retrieve the amino acid sequences of the *C. reinhardtii* queries that aligned robustly with a sequence from the database of PCD proteins, the list of *C. reinhardtii* identifiers was uploaded into Phytozome as a new list. This list was then used to retrieve and download the corresponding amino acid sequences as a multi-FASTA file. These sequences were termed the *C. reinhardtii* predicted PCD proteins (CrPPPs).

**Assignment of GO Terms to the Predicted *C. reinhardtii* PCD Proteins** In order to predict functional characteristics of the CrPPPs, BLAST2GO was used to assign GO terms to the CrPPP dataset. The BLAST2GO software (version 4.1.9) was installed locally, and the multi-FASTA file of CrPPPs was imported into the program. The multi-FASTA file of CrPPPs was uploaded to the previously-launched AWS instance. A BLASTp search was conducted within the instance, using the CrPPPs to query the RefSeq protein database. The BLASTP output file (.xml2) was downloaded locally and imported into the BLAST2GO software. In order to annotate the CrPPPs with GO terms, the mapping and annotation functions of BLAST2GO were applied to the dataset. These functions utilize the imported BLASTp results and the InterProScan servers to assign GO terms to the CrPPPs.

**Results and Discussion**

In this study, the amino acid sequences from the *C. reinhardtii* predicted proteome were aligned, using BLASTp, with protein sequences from the RefSeq database that were annotated to be involved in several types of PCD. Based on sequence similarity, *C. reinhardtii* proteins that aligned strongly with one of the annotated PCD proteins were predicted to participate in *C. reinhardtii* PCD. Finally, in order to predict characteristics of the CrPPPs, these sequences were assigned GO terms using the BLAST2GO software.

**Retrieval of Annotated PCD Sequences** The keyword search results of the RefSeq Protein database are summarized in Figure 4a. The keywords "apoptosis", "necrosis", and "programmed cell death" searches returned 46,200, 33,249, and 12,238 results, respectively. Of these, 79,647 entries were the results of only a single search, 5,112 entries were results of both the "apoptosis" and "programmed cell death" searches, 866 entries were results of both the "apoptosis" and "necrosis" searches, 9 entries were results of both the "programmed cell death" and "necrosis" searches, and 22 entries were results of all three searches.

A)                                          B)
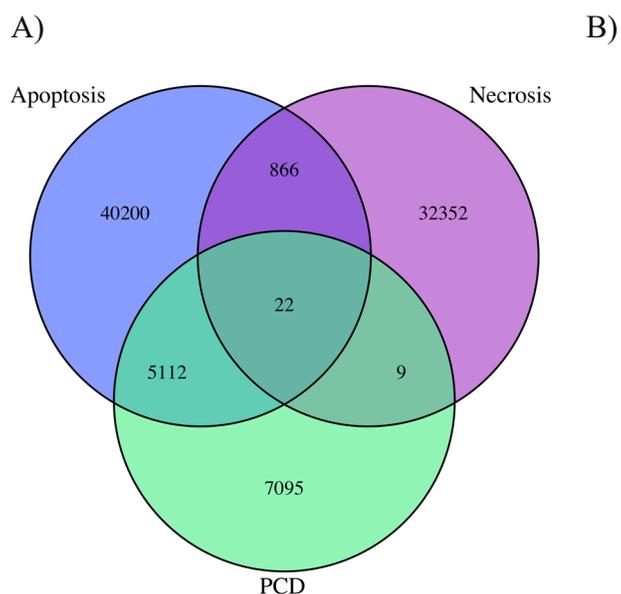


Figure 4. Composition of the Custom PCD Database. A) Unique and overlapping results of the three ENTREZ searches. Each circle represents the results of a single keyword search and is composed of non-overlapping and overlapping regions with other circles. Areas of a circle which do not overlap with another circle represent results which are unique to a single keyword search. Areas which are overlaps between two or more circles represent results that appeared in those searches. The number within an area designates how many sequences fall into that category. B) Phylogenetic distribution of the combined NCBI search results. In this pie chart, the area of each slice represents the relative number of non-redundant results contributed by a single phylogenetic group.

After compiling the search results into a single file and removing the duplicate items, 85,656 unique amino acid sequences remained. Of these sequences, the vast

majority (90%) of sequences came from animals, while 3% originated from protists, 3% came from fungi, and 4% were contributed by plants (Figure 4b). Though heavily skewed towards animals, this phylogenetic distribution is consistent with the observation that sequences annotated only with the keyword apoptosis, an animal-specific process, and/or necrosis, a process best characterized in animals, accounted for approximately 85% of the combined search results (Figure 4a).

Taken together, these results support the notion that a knowledge gap currently exists between PCD in animals and PCD in other, phylogenetically distinct organisms. This difference in understanding has the potential to severely limit the ability to formulate hypotheses regarding the molecular basis for PCD in non-animals. Since homology-based functional predictions of protein sequences are currently dependent on a degree of sequential conservation between the sequences being compared, and because the sequential similarity of proteins tends to decrease with evolutionary distance, it follows that such strategies are likely to be much less effective when comparisons are drawn between phylogenetically distant organisms.

**Alignment of the *C. reinhardtii* Proteome with the Annotated PCD Sequences**

The output of the BLASTp alignments between the *C. reinhardtii* proteome and the PCD database consisted of 908,445 alignments in total. Filtration of the results to include only the alignments with a percent identity >25% and a bit score of >100 resulted in the removal of 806,009 alignments (88.72%). The remaining 102,436 alignments were saved as a new file (henceforth referred to as the "filtered results"). A count of the unique query identifiers in the filtered results file revealed that 2,389 unique CrPPPs matched strongly with one or more protein sequences from the custom PCD database. Among the filtered

results, the majority of the CrPPPs possessed a percent identity of 25-50% with the

aligned subject protein (Figure 5). One potential explanation for the abundance of

alignments with a low percent identity is the profusion of subject sequences which are

phylogenetically distant from *C. reinhardtii* (Figure 4b).
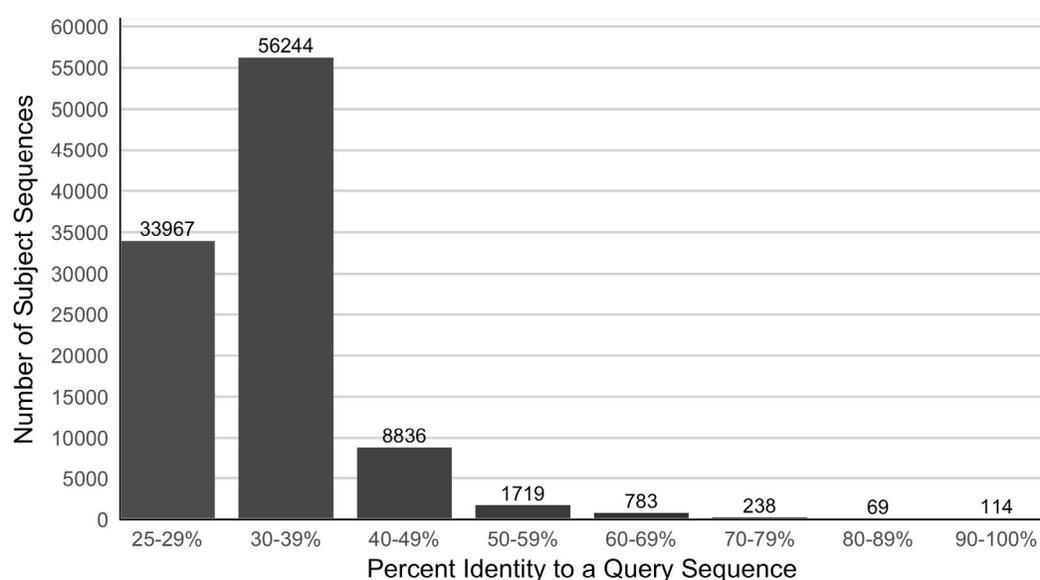


Figure 5. Percent identities of the filtered BLASTp alignments between the *C. reinhardtii* proteome and the PCD database. The height of each bar represents the number of alignments which possess a percent identity within the indicated range specified on the x-axis.

**Assignment of GO Terms to the *C. reinhardtii* Predicted PCD Proteins** A

single GO term consists of a GO name and the corresponding GO identifier. The output

of BLAST2GO gives any and all predicted GO terms assigned to each of the input amino

acid sequences. Furthermore, the software classifies each assigned GO term as a

biological process, cellular component, or molecular function. All GO terms assigned to

the CrPPPs were exported as a tab-delimited text file. Reformatting of the BLAST2GO

output file provided the number of CrPPPS that were assigned a GO term from each

category, the total number of GO terms from each category that were assigned to the *C.*

*reinhardtii* dataset, and the number of different (unique) GO terms that were assigned to

the dataset (Table 4).

Table 4

Gene Ontology annotations assigned to the *C. reinhardtii* predicted PCD proteins

|  | Number of CrPPPs Assigned GO Terms | Total Number of GO Terms Assigned to CrPPPs | Unique Number of GO Terms Assigned to CrPPPs |
|---|---|---|---|
| Biological Process (BP) | 1663 | 2628 | 639 |
| Cellular Component (CC) | 726 | 1203 | 212 |
| Molecular Function (MF) | 2173 | 3945 | 455 |

Note: "Unique" GO terms are those which were only counted once, regardless of how many times they were assigned to the CrPPPs.

Of the 2,389 CrPPPs, BPs, CCs, and MFs were assigned to 69.6%, 30.4%, and

90% of the CrPPPs, respectively (Table 4). For each of the three GO categories, the total

number of GO terms assigned to the CrPPPs exceeded the number of CrPPPs that were

assigned a GO term. This indicated that, for each GO category, a subset of the CrPPPs

were assigned multiple GO terms. To determine the frequency of GO terms assignments,

the number of GO terms assigned to each CrPPP was determined for the entire dataset

(Figure 6). The majority of CrPPPs (85.6%) were assigned between one and four GO

terms. Only 91 sequences (3.7%) were not assigned a GO term by BLAST2GO.
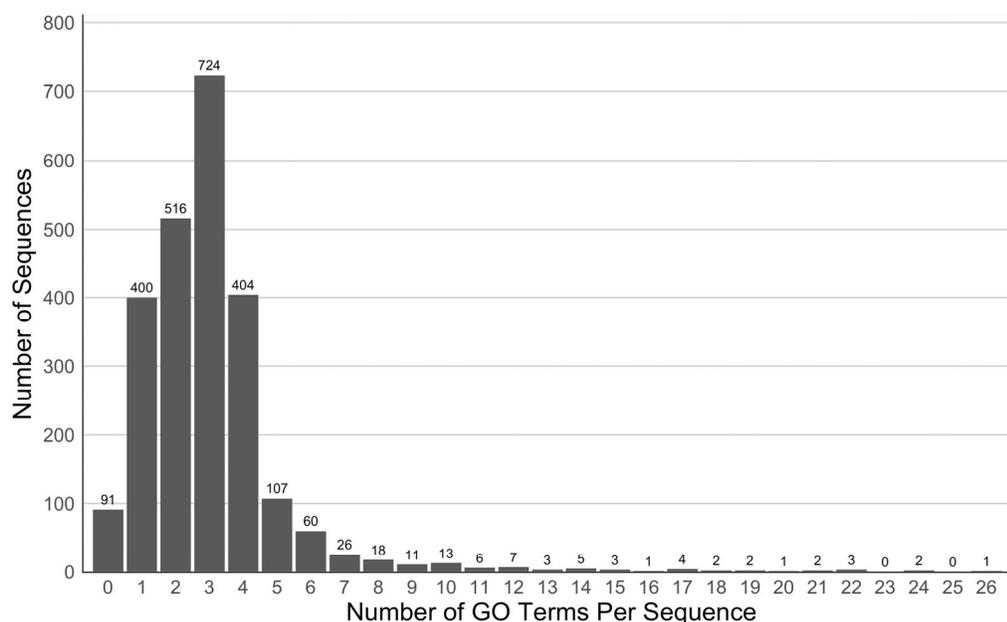
Figure 6. Frequencies of GO terms assigned to the *C. reinhardtii* dataset. The height of each bar represents the number of CrPPPs that were assigned the specified number of GO terms, found on the x-axis. One sequence, which was assigned 67 GO terms, was omitted from this figure.

To determine the extent of the overlap between GO categories assigned to the CrPPPs, the number of sequences assigned GO terms from each category was determined (Figure 7). It was found that approximately 20% of the CrPPPs were assigned at least one GO term from each of the three GO categories. Additionally, almost half of the CrPPPs that were assigned an annotation were assigned at least one GO term from both BPs and MFs. This data is in agreement with the observation that terms from these two categories were assigned more abundantly to the CrPPPs than terms from cellular components (Table 4).
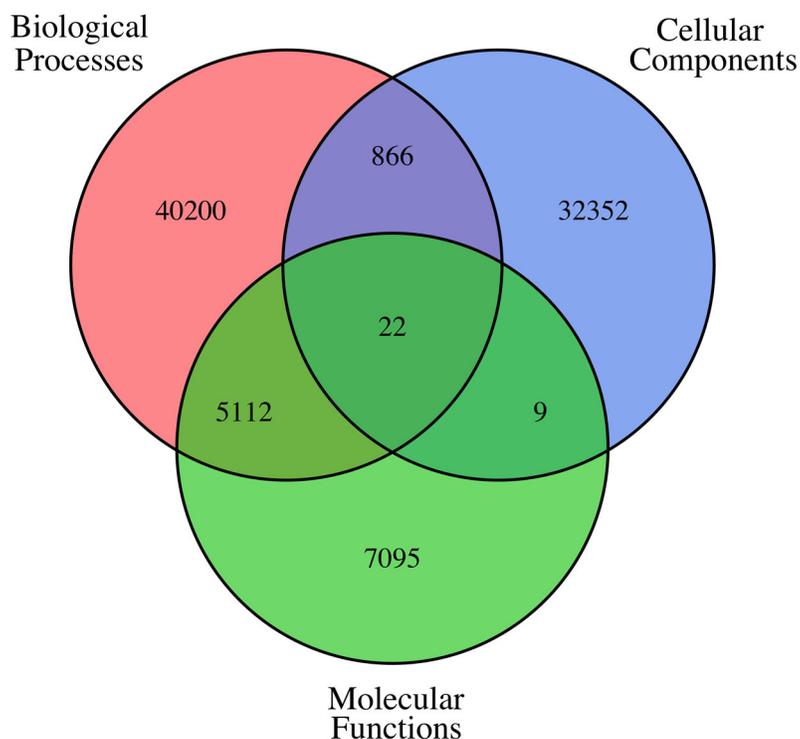
Figure 7. Overlap of annotations from the three GO categories assigned to the *C. reinhardtii* dataset. Each circle represents the number of sequences that were assigned GO terms from one of the three categories: BP, CC, or MF. Positions of overlap represent sequences which were assigned GO terms from more than one category.

To qualitatively estimate the specificity of the annotations assigned to the CrPPPs for each GO category, the number of sequences assigned a GO term at each BLAST2GO level was quantified (Figure 8). The majority of the annotations across many of the GO levels are MFs, which is in agreement with the previous observation that more MF terms were assigned to the CrPPPs than the other two GO categories (Table 4).

Figure 8. Distribution of annotation levels from each of the three GO categories assigned to the *C. reinhardtii* dataset. Categories, designated by color, are grouped by GO level, which is found on the x-axis. The height of each bar reflects the number of sequences that were assigned GO terms from a given level.

To identify the groups of GO terms that were assigned more abundantly to the

CrPPPs, the distribution of GO terms within each category was determined (Figure 9a-

c).While each of these most abundant terms are useful in a number of different biological

contexts, and thus are not specific to cell death, that these terms were assigned to the

CrPPPs may shed light on the mechanisms by which PCD is induced in *C. reinhardtii*. Of

the CrPPPs assigned terms from the BPs, a number were assigned terms related to

oxidation-reduction processes and cell redox homeostasis (Figure 9a). This did not come

as a surprise, given the important roles that ROS plays during PCD. In addition, several

BP terms assigned to the CrPPPs were related to signaling and gene regulation (Figure

9a). These include translation, DNA-templated regulation of transcription, transport,

signal transduction, intracellular signal transduction, transmembrane transport, and protein phosphorylation.

A number of interesting terms from MFs were assigned to the CrPPPs (Figure 9b), including those which involved zinc, calcium, and general metal binding. These typically act as cofactors for various enzymes within the cell. Of note, however, is that $Ca^{2+}$ plays a number of roles during PCD.[209] A number of biomolecule-binding terms were also assigned, including DNA and nucleic acid binding, as well as protein binding. Finally, two key signaling terms, protein kinase activity and oxidoreductase activity were also assigned to the CrPPPs.

The CrPPPs are predicted to localize to a diverse array of subcellular locations (Figure 9c). Interestingly, many of the sequences are predicted to be integral transmembrane proteins and/or localize to a membrane. In addition, many of the CrPPPs were predicted to localize to regions of the cell that are known to be important in the cell death, including the nucleus, cytoplasm, mitochondrion, and endoplasmic reticulum. Interestingly, 14 CrPPPs are predicted to localize to the chloroplast, which is predicted to be involved in several plant-specific subroutines during PCD. These CrPPPs may provide an excellent basis for exploring this possibility.
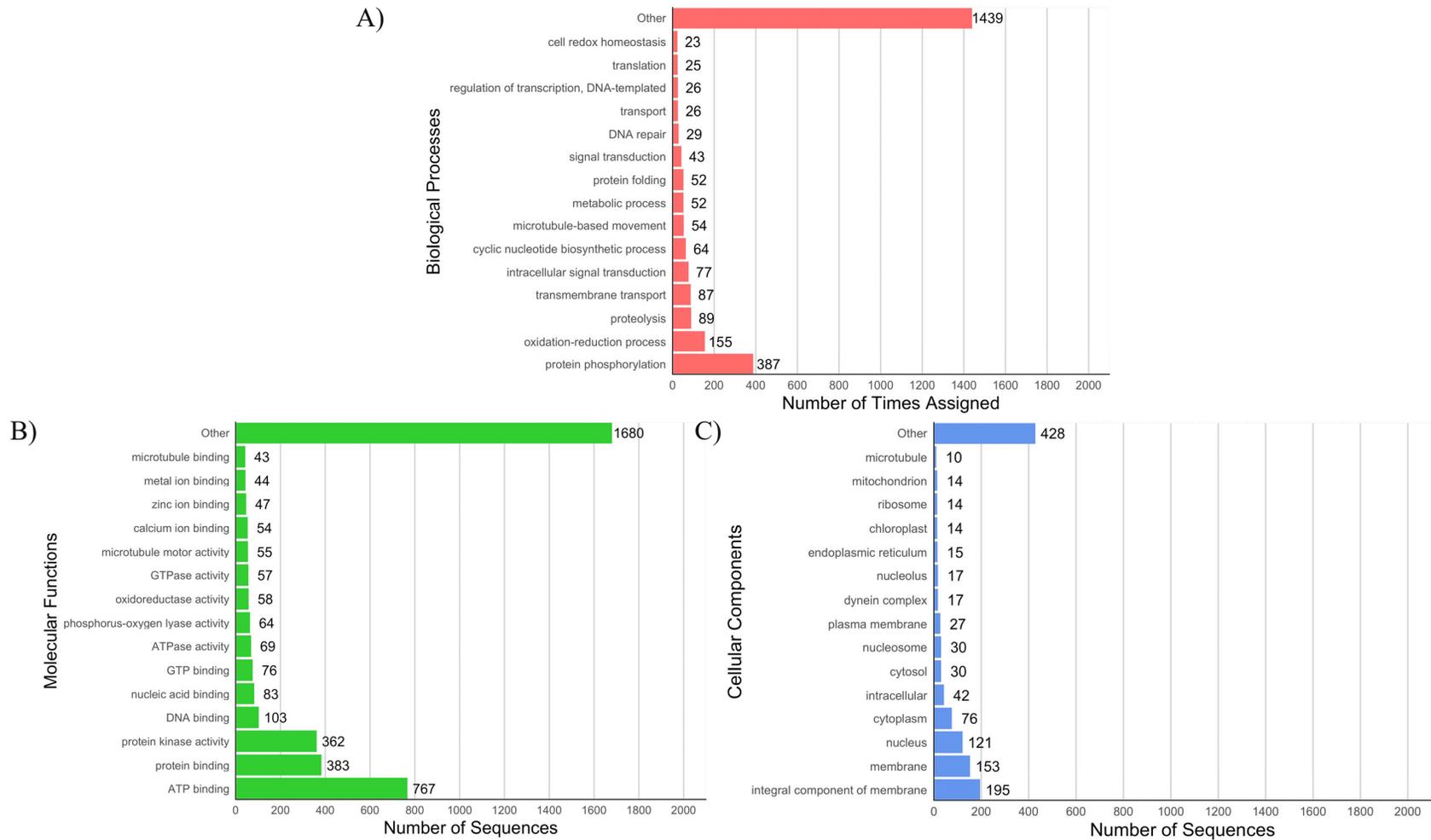
Figure 9. Distribution of the most abundant GO terms from each category assigned to the *C. reinhardtii* dataset. For the sake of conciseness, only the 15 most abundant terms are displayed for each category.

**Selection of CrPPPs of Interest** In order to facilitate future studies seeking to characterize PCD in *C. reinhardtii*, several genes of interest were selected from the list of CrPPPs (Table 5). In order to create this list, the file of BLASTp alignments between the *C. reinhardtii* proteome and the PCD database was manually explored. Using prior knowledge of the various PCD pathways described in the literature, primarily those observed in animals, plants, and yeast, subjects which were recognized as PCD proteins were identified. The CrPPPs which aligned with the subject(s) were added to the list of CrPPPs of interest.

**Selection of CrPPPs for Further Study** From the list of CrPPPs, several entries were selected for further study regarding a potential role during *C. reinhardtii* PCD. While no stringent criteria were in place for this selection process, several factors were considered. First, as a more robust alignment increases the probability of homologous function between two proteins, alignments with higher bitscores, e-values, query coverages, and/or percent identities were preferred. Second, as a central premise for the use of *C. reinhardtii* as a model for studies in PCD, it was presented in Chapter 1 that *C. reinhardtii* serves as a model for transitionary stages between distant phylogenetic groups. As such, the proteins selected for study were chosen as a mixture of proteins which were either conserved between multiple phylogenies, or were limited to a subgroup of organisms. Finally, as it was known that the study to follow would utilize mutant strains ordered from a mutant library, the last criterion used in the selection of CrPPPs to study further was whether a mutant strain with a disrupted allele of the gene coding for the selected protein was available.

Table 5

Predicted contributors to *C. reinhardtii* PCD

| query | bitscore | evalue | qcovs | pident | sacc | subject name | Organism |
|---|---|---|---|---|---|---|---|
| Cre01.g004350.t1.2 | 210 | 4.37E-56 | 80 | 26.63 | NP_001158150.1 | programmed cell death 6-interacting protein | *Mus musculus* |
| Cre01.g036350.t1.1 | 434 | 3.32E-131 | 64 | 43.197 | NP_191292.1 | Ca(2+)-ATPase isoform 11 (ACA11) | *Arabidopsis thaliana* |
| Cre01.g037000.t1.1 | 138 | 3.15E-34 | 91 | 25.155 | NP_011312.1 | Killer expression defective protein 1 (KEX1) | *Saccharomyces cerevisiae* |
| Cre01.g037000.t2.1 | 135 | 6.32E-34 | 91 | 25.604 | NP_011312.1 | Killer expression defective protein 1 (KEX1) | *Saccharomyces cerevisiae* |
| Cre01.g050150.t1.1 | 109 | 9.16E-26 | 54 | 32.889 | NP_015154.1 | Old yellow enzyme 3 (OYE3) | *Saccharomyces cerevisiae* |
| **Cre01.g052300.t1.1** | **156** | **1.09E-41** | **39** | **42.439** | **NP_005216.1** | **E2F-1** | ***Homo sapiens*** |
| **Cre01.g061807.t1.1** | **147** | **7.12E-43** | **84** | **37.5** | **NP_062254.2** | **Bax inhibitor 1** | ***Rattus norvegicus*** |
| | **156** | **2.37E-46** | **93** | **42.5** | **NP_199523.1** | **Bax inhibitor 1** | ***Arabidopsis thaliana*** |
| **Cre02.g085650.t1.1** | **202** | **2.05E-61** | **99** | **33.943** | **NP_009498.1** | **Deoxyribonuclease Tat-D** | ***Saccharomyces cerevisiae*** |
| Cre02.g097400.t1.2 | 179 | 1.03E-57 | 98 | 54.658 | NP_173985.1 | Eukaryotic translation initiation factor 5A-1 | *Arabidopsis thaliana* |
| Cre02.g108400.t1.2 | 137 | 9.91E-43 | 97 | 59.813 | NP_001335.1 | Defender against death 1 (DAD-1) | *Homo sapiens* |
| | 125 | 5.20E-38 | 97 | 51.429 | NP_174500.1 | Defender against death 1 (DAD-1) | *Arabidopsis thaliana* |
| Cre02.g116500.t1.1 | 125 | 7.77E-30 | 78 | 25.909 | NP_011312.1 | Killer expression defective protein 1 (KEX1) | *Saccharomyces cerevisiae* |

Continued

| query | bitscore | evalue | qcovs | pident | sacc | subject name | Organism |
|---|---|---|---|---|---|---|---|
| Cre02.g117500.t1.2 | 241 | 2.14E-71 | 77 | 31.827 | NP_194642.1 | Hexokinase 1 | *Arabidopsis thaliana* |
| Cre02.g145100.t1.1 | 628 | 0 | 75 | 43.804 | NP_191292.1 | Ca(2+)-ATPase isoform 11 (ACA11) | *Arabidopsis thaliana* |
| Cre03.g149500.t1.1 | 100 | 1.43E-22 | 55 | 42.149 | NP_567225.1 | Programmed cell death protein 2 (PDCD2) | *Arabidopsis thaliana* |
| | 104 | 2.78E-24 | 80 | 34.921 | NP_113826.1 | Programmed cell death protein 2 (PDCD2) | *Rattus norvegicus* |
| Cre03.g153800.t1.1 | 333 | 2.90E-111 | 67 | 55.797 | NP_973794.1 | GPI transamidase | *Arabidopsis thaliana* |
| Cre03.g163500.t1.2 | 147 | 6.32E-36 | 34 | 40.367 | NP_564287.1 | Executer 2 | *Arabidopsis thaliana* |
| | 133 | 1.87E-31 | 27 | 45.395 | NP_849488.1 | Executer 1 | *Arabidopsis thaliana* |
| Cre03.g180650.t1.1 | 501 | 6.11E-160 | 97 | 31.54 | NP_014276.1 | Nuclear mediator of apoptosis (NMA111) | *Saccharomyces cerevisiae* |
| Cre03.g184700.t1.2 | 103 | 1.13E-23 | 55 | 42.667 | NP_014840.4 | Metacaspase-1 | *Saccharomyces cerevisiae* |
| | 248 | 6.95E-78 | 99 | 35.16 | NP_178051.1 | Metacaspase-5 | *Arabidopsis thaliana* |
| | 246 | 3.36E-77 | 99 | 36.117 | NP_178052.1 | Metacaspase-4 | *Arabidopsis thaliana* |
| | 162 | 2.45E-45 | 78 | 37.269 | NP_173092.1 | Metacaspase-8 | *Arabidopsis thaliana* |
| | 109 | 3.35E-26 | 35 | 40.667 | NP_201229.1 | Metacaspase-3 | *Arabidopsis thaliana* |
| | 107 | 3.24E-25 | 35 | 37.333 | NP_171719.2 | Metacaspase-1 | *Arabidopsis thaliana* |
| | 102 | 1.77E-23 | 35 | 37.584 | NP_001031711.1 | Metacaspase-2 | *Arabidopsis thaliana* |

Continued

| query | bitscore | evalue | qcovs | pident | sacc | subject name | Organism |
|---|---|---|---|---|---|---|---|
| Cre03.g184700.t2.1 | 206 | 2.87E-62 | 98 | 32.843 | NP_178051.1 | Metacaspase-5 | *Arabidopsis thaliana* |
| | 204 | 2.82E-61 | 99 | 33.333 | NP_178052.1 | Metacaspase-4 | *Arabidopsis thaliana* |
| | 124 | 2.19E-31 | 77 | 33.878 | NP_173092.1 | Metacaspase-8 | *Arabidopsis thaliana* |
| Cre03.g206800.t1.1 | 156 | 1.03E-40 | 67 | 32.242 | NP_198006.1 | Sugar transport protein 13 | *Arabidopsis thaliana* |
| Cre03.g209393.t1.1 | 241 | 1.51E-74 | 82 | 39.25 | NP_015154.1 | Old yellow enzyme 3 (OYE3) | *Saccharomyces cerevisiae* |
| | 240 | 1.73E-74 | 79 | 39.535 | NP_012049.1 | Old yellow enzyme 2 (OYE2) | *Saccharomyces cerevisiae* |
| Cre03.g210513.t1.1 | 245 | 1.14E-76 | 86 | 39.348 | NP_015154.1 | Old yellow enzyme 3 (OYE3) | *Saccharomyces cerevisiae* |
| | 244 | 4.53E-76 | 84 | 39.744 | NP_012049.1 | Old yellow enzyme 2 (OYE2) | *Saccharomyces cerevisiae* |
| Cre04.g217922.t1.1 | 268 | 4.78E-78 | 36 | 38.953 | NP_083045.4 | Synoviolin | *Mus musculus* |
| Cre04.g226850.t1.2 | 257 | 1.55E-79 | 65 | 47.02 | NP_015171.1 | Proteinase A | *Saccharomyces cerevisiae* |
| Cre05.g232200.t1.2 | 223 | 2.46E-64 | 64 | 35.802 | NP_013586.1 | Internal NADH dehydrogenase (NDI-1) | *Saccharomyces cerevisiae* |
| Cre05.g242350.t1.2 | 359 | 8.50E-107 | 95 | 27.984 | NP_191292.1 | Ca(2+)-ATPase isoform 11 (ACA11) | *Arabidopsis thaliana* |
| Cre06.g257500.t1.2 | 312 | 1.09E-106 | 96 | 62.605 | NP_001154216.1 | General regulatory factor 1 (GRF1) | *Arabidopsis thaliana* |
| | 300 | 4.20E-103 | 96 | 62.343 | NP_113791.1 | 14-3-3 protein epsilon | *Rattus norvegicus* |
| | 276 | 1.24E-93 | 96 | 57.917 | NP_036611.2 | 14-3-3 protein gamma | *Homo sapiens* |
| | 271 | 9.49E-92 | 93 | 60.173 | NP_647539.1 | 14-3-3 protein beta/alpha | *Homo sapiens* |

| query | bitscore | evalue | qcovs | pident | sacc | subject name | Organism |
|---|---|---|---|---|---|---|---|
| | 268 | 3.02E-90 | 96 | 57.083 | NP_003396.1 | 14-3-3 protein eta | *Homo sapiens* |
| | 263 | 2.38E-88 | 93 | 58.009 | NP_037143.2 | 14-3-3 protein zeta/delta | *Rattus norvegicus* |
| | 258 | 2.75E-86 | 91 | 58.407 | NP_037185.1 | 14-3-3 protein theta | *Rattus norvegicus* |
| Cre06.g271200.t1.2 | 100 | 1.21E-21 | 54 | 27.607 | NP_001277999.1 | Apoptosis-inducing factor 3 (AIF3) | *Mus musculus* |
| | 100 | 1.30E-21 | 54 | 27.607 | NP_780387.2 | Apoptosis-inducing factor 3 (AIF3) | *Mus musculus* |
| Cre06.g274500.t1.2 | 121 | 2.84E-32 | 64 | 45.062 | NP_001186793.1 | BCL2L2-PABPN1 | *Homo sapiens* |
| Cre06.g279400.t1.2 | 166 | 2.23E-43 | 77 | 28.423 | NP_011312.1 | Killer expression defective protein 1 (KEX1) | *Saccharomyces cerevisiae* |
| Cre06.g303900.t1.1 | 128 | 1.19E-33 | 47 | 37.705 | NP_014641.2 | ADIPOR-like receptor IZH2 | *Saccharomyces cerevisiae* |
| Cre08.g362750.t1.2 | 177 | 1.81E-51 | 78 | 31.933 | NP_012327.1 | Mitochondrial nuclease (NUC1) | *Saccharomyces cerevisiae* |
| Cre08.g367650.t1.2 | 117 | 4.08E-28 | 86 | 29.487 | NP_001269876.1 | Apoptotic protease-activating factor 1 (APAF-1) | *Mus musculus* |
| | 117 | 5.94E-28 | 86 | 29.487 | NP_033814.2 | Apoptotic protease-activating factor 1 (APAF-1) | *Mus musculus* |
| Cre08.g367800.t1.2 | 196 | 1.63E-60 | 76 | 41.736 | NP_849368.1 | Nudix hydrolase 7 | *Arabidopsis thaliana* |
| **Cre08.g371052.t1.1** | **265** | **1.53E-72** | **42** | **34.632** | **NP_001424.3** | **Inositol-requiring enzyme 1 (IRE-1)** | ***Homo sapiens*** |
| Cre08.g384900.t1.2 | 107 | 1.49E-25 | 61 | 35.176 | NP_001320044.1 | cysteine-rich receptor-like protein kinase 20 | *Arabidopsis thaliana* |
| | 107 | 3.38E-25 | 97 | 27.679 | NP_190172.1 | cysteine-rich receptor-like protein kinase 4 | *Arabidopsis thaliana* |

Continued

| query | bitscore | evalue | qcovs | pident | sacc | subject name | Organism |
|---|---|---|---|---|---|---|---|
| | 102 | 8.74E-24 | 61 | 34.171 | NP_849425.1 | cysteine-rich receptor-like protein kinase 5 | *Arabidopsis thaliana* |
| Cre09.g388850.t1.1 | 608 | 0 | 80 | 42.907 | NP_191292.1 | Ca(2+)-ATPase isoform 11 (ACA11) | *Arabidopsis thaliana* |
| Cre09.g415800.t1.2 | 142 | 9.98E-36 | 85 | 33.051 | NP_001186421.1 | Programmed cell death protein 4 (PDCD4) | *Homo sapiens* |
| | 142 | 1.19E-35 | 85 | 33.051 | NP_663314.1 | Programmed cell death protein 4 (PDCD4) | *Homo sapiens* |
| | 142 | 1.62E-35 | 85 | 33.051 | NP_055271.2 | Programmed cell death protein 4 (PDCD4) | *Homo sapiens* |
| Cre10.g428750.t1.1 | 357 | 6.71E-99 | 98 | 26.78 | NP_055791.1 | Programmed cell death protein 11 (PDCD11) | *Homo sapiens* |
| Cre11.g475350.t1.1 | 134 | 4.93E-37 | 91 | 34.146 | NP_064709.2 | Anamorsin | *Homo sapiens* |
| Cre11.g480060.t1.2 | 145 | 1.30E-38 | 37 | 36.607 | NP_001156411.1 | Apoptosis-enhancing nuclease (AEN) | *Mus musculus* |
| Cre12.g483550.t1.2 | 107 | 1.75E-24 | 33 | 29.339 | NP_973794.1 | GPI transamidase | *Arabidopsis thaliana* |
| Cre12.g496650.t1.2 | 188 | 1.10E-58 | 78 | 43.304 | NP_181023.1 | Fatty acid 2-hydroxylase 1 (FAH1) | *Arabidopsis thaliana* |
| | 176 | 4.78E-54 | 80 | 40.351 | NP_193819.1 | Fatty acid 2-hydroxylase 2 (FAH2) | *Arabidopsis thaliana* |
| Cre12.g505350.t1.2 | 663 | 0 | 81 | 45.574 | NP_191292.1 | Ca(2+)-ATPase isoform 11 (ACA11) | *Arabidopsis thaliana* |
| **Cre12.g517350.t1.2** | **109** | **2.28E-30** | **72** | **48.837** | **NP_001117399.1** | **Lesion-stimulating disease 1 (LSD-1)** | ***Arabidopsis thaliana*** |
| Cre12.g517451.t1.1 | 153 | 9.16E-41 | 46 | 36.364 | NP_014840.4 | Metacaspase-1 | *Saccharomyces cerevisiae* |
| | 248 | 1.65E-76 | 47 | 49.167 | NP_194241.3 | Metacaspase-2 | *Arabidopsis thaliana* |

| query | bitscore | evalue | qcovs | pident | sacc | subject name | Organism |
|---|---|---|---|---|---|---|---|
| | 236 | 6.58E-73 | 55 | 44.561 | NP_171719.2 | Metacaspase-1 | *Arabidopsis thaliana* |
| | 214 | 1.90E-64 | 55 | 41.319 | NP_201229.1 | Metacaspase-3 | *Arabidopsis thaliana* |
| | 105 | 6.89E-24 | 29 | 39.474 | NP_178051.1 | Metacaspase-5 | *Arabidopsis thaliana* |
| | 102 | 7.40E-23 | 29 | 37.584 | NP_178052.1 | Metacaspase-4 | *Arabidopsis thaliana* |
| Cre12.g556228.t1.1 | 109 | 2.45E-26 | 89 | 31.045 | NP_001185625.1 | Apoptosis-inducing factor 2 (AIF2) | *Homo sapiens* |
| Cre13.g587000.t1.2 | 116 | 5.96E-27 | 77 | 27.551 | NP_062790.1 | Apoptosis-antagonizing transcription factor | *Mus musculus* |
| Cre13.g588550.t1.2 | 231 | 1.07E-73 | 98 | 43.182 | NP_190808.1 | Syntaxin-122 | *Arabidopsis thaliana* |
| Cre16.g664050.t1.2 | 186 | 4.34E-50 | 50 | 30.256 | NP_171714.2 | Vascular-associated death 1 (VAD1) | *Arabidopsis thaliana* |
| Cre16.g664050.t2.1 | 186 | 4.42E-50 | 50 | 30.256 | NP_171714.2 | Vascular-associated death 1 (VAD1) | *Arabidopsis thaliana* |
| **Cre16.g674050.t1.2** | **108** | **1.82E-26** | **94** | **30.368** | **NP_671713.1** | **p53-induced gene 3 (PIG3)** | *Homo sapiens* |
| Cre16.g681750.t1.1 | 611 | 0 | 75 | 43.237 | NP_191292.1 | Ca(2+)-ATPase isoform 11 (ACA11) | *Arabidopsis thaliana* |
| Cre16.g681750.t2.1 | 611 | 0 | 78 | 43.237 | NP_191292.1 | Ca(2+)-ATPase isoform 11 (ACA11) | *Arabidopsis thaliana* |
| Cre16.g685901.t1.1 | 670 | 0.00E+00 | 99 | 38.842 | NP_076054.1 | Exportin-2 | *Mus musculus* |
| Cre16.g691552.t1.1 | 164 | 8.69E-44 | 88 | 27.626 | NP_013586.1 | Internal NADH dehydrogenase (NDI-1) | *Saccharomyces cerevisiae* |
| Cre17.g727300.t1.2 | 167 | 3.10E-47 | 97 | 31.095 | NP_015154.1 | Old yellow enzyme 3 (OYE3) | *Saccharomyces cerevisiae* |

Continued

| query | bitscore | evalue | qcovs | pident | sacc | subject name | Organism |
|---|---|---|---|---|---|---|---|
| | 158 | 6.03E-44 | 97 | 30.457 | NP_012049.1 | Old yellow enzyme 2 (OYE2) | *Saccharomyces cerevisiae* |
| Cre17.g746597.t1.1 | 111 | 3.90E-25 | 85 | 25.813 | NP_011312.1 | Killer expression defective protein 1 (KEX1) | *Saccharomyces cerevisiae* |
| Cre19.g750547.t1.1 | 201 | 8.77E-57 | 77 | 36.34 | NP_013586.1 | Internal NADH dehydrogenase (NDI-1) | *Saccharomyces cerevisiae* |
| Cre19.g750547.t2.1 | 198 | 1.23E-55 | 81 | 35.28 | NP_013586.1 | Internal NADH dehydrogenase (NDI-1) | *Saccharomyces cerevisiae* |
| Cre19.g750547.t3.1 | 198 | 7.23E-56 | 70 | 38.872 | NP_013586.1 | Internal NADH dehydrogenase (NDI-1) | *Saccharomyces cerevisiae* |
| Cre19.g750547.t4.1 | 202 | 5.01E-57 | 81 | 35.98 | NP_013586.1 | Internal NADH dehydrogenase (NDI-1) | *Saccharomyces cerevisiae* |

Note: qseqid = *Chlamydomonas* query identifier (Phytozome), qcovs = percent of the *C. reinhardtii* query that is covered by the subject, pident = percent identity of the subject to the *Chlamydomonas* query, sacc = accession number of the subject sequence. Alignments are grouped together when either the query sequences are alternative transcripts of the same gene, or when multiple proteins from the PCD database aligned with a single *C. reinhardtii* protein. Bolded alignments represent the *C. reinhardtii* queries which were selected for further study.

**Analysis and Discussion of the CrPPPs Selected for Further Study** For each of the chosen CrPPPs, further predictive analyses were carried out to confirm that the robust BLASTp alignments were representative of functional similarities between the *C. reinhardtii* protein and its homologs in other organisms. For each selected *C. reinhardtii* protein, the homologous amino acid sequences from *H. sapiens*, *A. thaliana*, and *S. cerevisiae* were retrieved, if present, from the NCBI database. Notably, though a role in PCD has been established for the homolog from least one of these organisms, the existence of such a role may have yet to be determined in one or more of the homologs from other organisms. Such unconfirmed roles will be declared for each sequence in the following discussion. All retrieved sequences were imported into the locally-installed Geneious software and aligned using the included ClustalW alignment tool.[210,211] Subsequently, the domains, features, and annotations of all sequences were predicted using the InterProScan plugin available for Geneious.[212,213] To support the prediction that the *C. reinhardtii* protein and its homologs are similar in function, the presence and relative locations of the annotations assigned by InterProScan were qualitatively compared between sequences. The domains for the sequences were assigned individually, and thus independently of the amino acid alignments, for each individual sequence. Because the protein domains were predicted independently of the amino acid alignments, the shared presence and alignment of the domains between a *C. reinhardtii* protein and one or more of its homologs corroborates the prediction that the protein's function may be conserved in *C. reinhardtii*.

***Bax inhibitor 1*** BI-1is an ER transmembrane protein that acts to repress cell death in response to a number of different stressors.[214] Despite this well-established role, little is known regarding the mechanisms by which BI-1 attenuates PCD. BI-1 was initially identified by a functional screen in yeast, in which a human cDNA expression library was cloned into a Bax-expressing strain of *S. cerevisiae.*[215] Subsequent experiments revealed that BI-1 co-immunoprecipitates with the anti-apoptotic BCL-2 and BCLX$_L$ proteins, linking BI-1 with a critical anti-apoptotic pathway in animals.[215] BI-1 is also believed to achieve its anti-PCD functions through two additional mechanisms: the reduction of $Ca^{2+}$ levels in the endoplasmic reticulum and the negative regulation of the IRE-1 pro-PCD signaling mechanisms.[214]

BI-1 is well-conserved in animals, plants, and yeast. In *Arabidopsis*, ER-mediated PCD can be accelerated by disruption of the *bi-1* gene, and attenuated by overexpression of *bi-1*.[216] In mammalian cells, BI-1 acts early in the early stages of the adaptive response to ER stress by directly interacting with the stress sensor IRE-1.[217] In plants, BI-1 expression is controlled by the transcription factor bZIP60.[218] As plant IRE-1 mediates the transcriptional activity of bZIP60, the activation provides another link between BI-1 and IRE-1.

A strong case has been presented for the conserved function of BI-1 in distinct lineages. The expression of mammalian Bax in plant and yeast cells rapidly induced PCD, whereas the overexpression of BI-1 was sufficient to repress Bax-induced death.[215,219] When mammalian BI-1 was overexpressed in Bax-expressing plant cells, death was repressed.[219] Moreover, the co-expression of plant BI-1 with mammalian Bax in yeast cells resulted in the attenuation of Bax-induced cell death.[220] Finally, transgenic

yeast cells expressing either *Drosophila* or *Arabidopsis* BI-1 were markedly more resistant to PCD induced by both $H_2O_2$ and heat stress.[221] Taken together, the results of these studies are highly indicative of a conserved role for BI-1 in the repression of cell death.

Given the widespread presence of BI-1 in eukaryotic systems, it has the potential to serve as an excellent starting point for studying phylogenetically-conserved mechanisms of PCD in *C. reinhardtii*. The alignment of *C. reinhardtii* BI-1 to its homologs in *H. sapiens* (accession: AAU29521), *A. thaliana* (accession: AAG35727), and *S. cerevisiae* (accession: P48558.1) is shown in Figure 10. Each of the four sequences is predicted to contain cytoplasmic domains, transmembrane domains, and non-cytoplasmic, presumably luminal, domains. While each of the aligned BI-1 homologs possesses four cytoplasmic domains, the BI-1 sequences from *C. reinhardtii*, *A. thaliana*, and *S. cerevisiae* are predicted to contain seven transmembrane domains and four luminal domains. BI-1 from *H. sapiens* was only predicted to contain 6 transmembrane domains and three luminal domains. The biological significance of this observation is unknown, but it may indicate a divergence of human BI-1 from the other BI-1 proteins. Cumulatively, these results indicate a high degree of conservation between *C. reinhardtii* BI-1 and its homologs from other phyla, suggesting that the anti-PCD function of BI-1 may be conserved in *C. reinhardtii*.
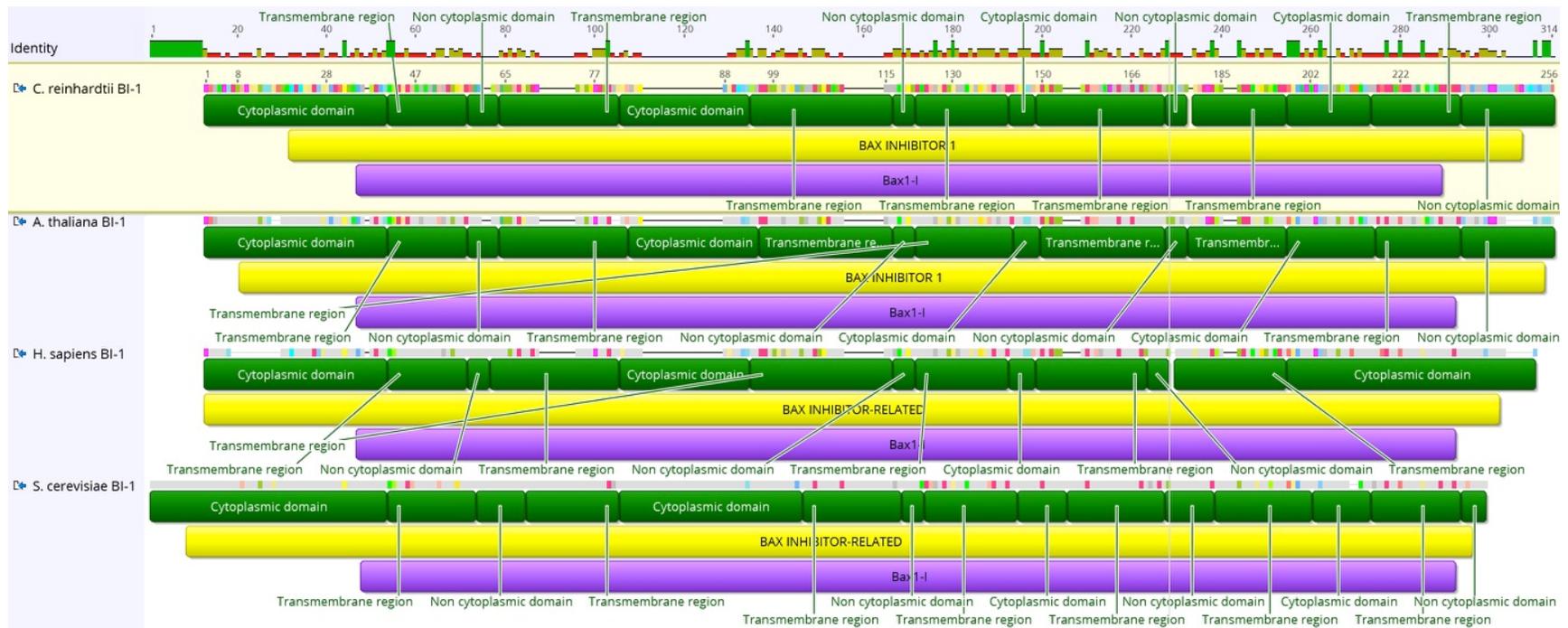
Figure 10. Alignment and annotation of the BI-1 proteins from *C. reinhardtii*, plants, humans, and yeast. The graph at the top of the figure represents the conservation of amino acids at each residue. Each protein has the results of the alignment, with conserved residues highlighted in color, and the annotations that were assigned to the proteins by InterProScan. The color of each annotation denotes its source, where yellow = Panther, purple = Pfam, and green = Phobius.

*Inositol-requiring enzyme 1* IRE-1 is an ER transmembrane protein which acts to mediate PCD in response to the accumulation of unfolded or misfolded proteins within the ER.[222] The accrual of such aberrant proteins is referred to as ER stress, and the cell attempts to correct this condition via the unfolded protein response (UPR).[223] IRE-1 is composed of a luminal region, a transmembrane domain, and a cytosolic region. The cytosolic region of IRE-1 contains two functional domains: a serine/threonine kinase domain and an endoribonuclease domain.[223] Though IRE-1 was initially believed to promote cell survival, recent studies have instead demonstrated that IRE-1 acts as a molecular switch during ER stress, and can act to either promote cell survival or to induce PCD.[222]

During the initial phase of ER stress, IRE-1 activates the UPR in an attempt to restore protein homeostasis.[224] At the onset of stress, IRE-1 monomers dimerize or oligomerize within the ER membrane and are activated by trans-autophosphorylation events.[223] Activated IRE-1 attempts to adapt to the stress by two primary mechanisms, which are carried out simultaneously.[223] Firstly, IRE-1 initiates a process known as Regulated IRE-1-Dependent Decay (RIDD), which selectively degrades the transcripts of ER-translocating proteins to reduce the protein-folding demand within the ER.[224] Secondly, the RNase domain of IRE-1 alternatively splices the bZIP transcription factors Xpb-1, bZIP60, and Hac1 in animals, plants, and yeast, respectively.[223] The activated bZIP transcription factor localizes to the nucleus, where it induces the transcription of ER quality control components.[223] In animals, if the cell is unable to restore ER protein homeostasis, the RIDD pathway switches targets and initiates the apoptotic process by the degradation of anti-apoptotic pre-miRNAs, one of which binds and prevents the

translation of caspase-2 mRNA.[224] This degradation allows for the upregulation of caspase-2, which then promotes the activation of the apoptotic program.

Importantly, recent studies indicate that, even in animals, the IRE-1 signaling network remains largely undiscovered, and the IRE-1-caspase-2 axis is likely only one of several mechanisms by which IRE-1 can initiate PCD in animal cells.[222,223] Further clouding the matter is that the pathways by which IRE-1 induces PCD in animals, plants, and yeast are likely to have diverged.[223] As such, multiple and possibly unique mechanisms may underlie IRE-1-mediated cell death in plants and yeast as well.[223,224]

As IRE-1 is conserved in eukaryotes, with potentially divergent PCD functions in different lineages, *C. reinhardtii* IRE-1 may represent a transitionary form of the protein. As such, the characterization of the role of IRE-1 during *C. reinhardtii* PCD may facilitate the discovery of previously unknown IRE-1 mechanisms in other organisms. The alignment of *C. reinhardtii* IRE-1 to homologous sequences from *H. sapiens* (accession: NP_001424), *A. thaliana* (accession: NP_565419), and *S. cerevisiae* (accession: NP_011946.1) is shown in Figure 11. Interestingly, several features of *C. reinhardtii* IRE-1 are unique compared to those of its homologs. First, both a transmembrane domain and a cytosolic domain are predicted to be present in IRE-1 from *H. sapiens*, *A. thaliana*, and *S. cerevisiae*. Neither domains were predicted to be present in *C. reinhardtii* IRE-1. Instead, a non-cytoplasmic domain is predicted to span the entire protein. Though the biological implications of this prediction are unknown and somewhat puzzling, it may indicate that *C. reinhardtii* IRE-1 is not a transmembrane protein. Characterization of IRE-1 in *C. reinhardtii* will be necessary to confirm this prediction. In addition, *C. reinhardtii* IRE-1 is larger than any of the other IRE-1 proteins and is

predicted to possess a zinc finger domain at the C-terminus of the peptide. Also of note is that a quinoprotein alcohol dehydrogenase-like domain is predicted to be present in *C. reinhardtii*, humans, and yeast, but not in plants. A search of the literature indicates that a role for this domain within IRE-1 has not been described in any of the other species, and its function remains unknown. Consistent with the known functions of IRE-1, all sequences are predicted to possess kinase and RNase domains. Cumulatively, these results indicate that there is a high degree of conservation between *C. reinhardtii* IRE-1 and its homologs from other phyla, suggesting that the function of IRE-1 as a mediator of PCD may be conserved in *C. reinhardtii*. Notably, the predictions that 1) *C. reinhardtii* IRE-1 is not a transmembrane protein, and 2) *C. reinhardtii* IRE-1 possess a unique a zinc finger domain at the C-terminal region may indicate additional and novel functions for *C. reinhardtii* IRE-1.

Figure 11. Alignment and annotation of the IRE-1 proteins from *C. reinhardtii*, plants, humans, and yeast. The graph at the top of the figure represents the conservation of amino acids at each residue. Each protein has the results of the alignment, with conserved residues highlighted in color, and the annotations that were assigned to the proteins by InterProScan. The color of each annotation denotes its source, where yellow = Panther, purple = Pfam, dark green = Phobius, maroon = PrositeProfiles, and olive green = SuperFamily.

***E2F*** The E2F family of proteins is conserved in animals and plants. In mammals, the E2F family comprises eight genes, *e2f1–8*, which code for a total of 9 distinct proteins.[225] The *e2f3* gene codes for both E2F3a and E2F3b by the utilization of alternative promoters. Traditionally, members of the E2F family are best known for their role in the regulation of cell cycle progression. Interestingly, studies demonstrate that one family member in particular, E2F1, can act as a key mediator of the apoptotic process. Pro-survival signals, such as those from the phosphatidylinositol 3 kinase (PI3K)–protein kinase B (Akt) and epidermal- growth-factor receptor (EGFR)–Ras pathways, inhibit E2F1-mediated apoptosis, whereas E2F1-mediated apoptosis suppresses such pro-survival signals. That E2F1 can induce both cell proliferation and apoptosis intuitively seems paradoxical, but such a mechanism may represent a centralized control system by which the cell decides to live or die when faced with stress.[225]

All E2F proteins contain a DNA-binding domain.[226] E2F1–5 possess an additional transactivation domain that enables activation of gene expression. Also, E2F1-6 contain a dimerization domain, through which the interaction with members of the dimerization-partner (DP) family members is achieved. The interaction with DP-family proteins enables these E2F proteins to bind DNA and act as transcriptional regulators.[227] E2F7 and E2F8 do not interact with DP family members, but bind DNA as either homodimers or E2F7-E2F8 heterodimers.[225]

The transcriptional activity of E2F family members is mediated by several mechanisms. The best understood involves the association of E2F with the retinoblastoma protein (pRb). This interaction both directly inhibits the ability of E2F to

transactivate, but also causes the recruitment of various chromatin modifiers and remodeling factors to the promoter regions of E2F target genes.[226]

E2F1 can induce apoptosis in a p53 -dependent or -independent manner.[228,229] In the p53-dependent manner, E2F1 controls the expression of the p53-activating kinases ataxia telangiectasia mutated (ATM) and checkpoint kinase 2 (Chk2).[230–232] In addition to promoting the activation of p53, E2F1 can also directly promote the expression of p53, as well as pro-apoptotic cofactors of p53, such as apoptosis stimulating protein of p53 1 and 2 (ASSP1-2), junction-mediating and -regulatory protein (JMY), and tumor protein p53-inducible nuclear protein 1 (TP53INP1).[226,233,234] Furthermore, E2F1 is also capable of promoting apoptosis independently of its transcriptional abilities by directly binding p53.[235]

E2F1 can also induce apoptosis independently of p53, primarily by upregulating pro-apoptotic genes and suppressing the expression of pro-survival signals. The pro-apoptotic genes upregulated by E2F include those which code for APAF1, caspases -3, -7, and -8, the BCL-2 effectors BAK and BOK, the BH3-only proteins BAD and BID, PIG8, and the p53 family member p73.[225,226,236] The survival signals that are repressed by E2F1 during apoptosis include those mediated by the transcription factor nuclear factor-kB (NF-kB) or by BCL-2 and its family member myeloid cell leukemia 1 (MCL-1).[237]

The mechanisms by which the apoptotic activity of E2F1 is regulated are not well understood. Thus far, the only pro-apoptotic cofactor of E2F1 that has been identified is JAB1, which interacts with E2F1 via a marked box domain, a domain within E2F1 which distinguishes it from other E2F family members.[238,239] An additional effector which

modulates the apoptotic activity of E2F1 is apoptotic inhibitor 5 (API5), which inhibits E2F1-mediated apoptosis.[240]

While E2F family members are conserved in some fungal lineages, these proteins are absent in yeast.[227] Though E2F family members are present in plants, their involvement in plant PCD has not been explored in detail. Interestingly, PCD in response to the DNA-damaging agent bleomycin is attenuated by E2F overexpression in tobacco, suggesting an anti-PCD role for E2F in plant PCD.[241]

Because E2F family members are conserved in both animals and plants, understanding the role of E2F in *C. reinhardtii* may aid in identifying features of PCD that are conserved between plants and animals. Furthermore, it is also predicted that *C. reinhardtii* E2F may also reveal features of PCD that have diverged from animals. This prediction stems from the observation that, in animals, E2F1 is capable of inducing apoptosis independently of p53. Of particular interest is the ability of E2F1 to activate p53 target genes to induce apoptosis. At least one pro-PCD gene targeted by both p53 and E2F1, PIG8, has been identified in both *C. reinhardtii* and animals. Furthermore, PIG8 epxpression is upregulated during *C. reinhardtii* PCD. As an ortholog of p53 has yet to be identified in *C. reinhardtii*, it may be the case that PIG8 is activated by E2F during *C. reinhardtii* PCD. In an even broader sense, we present the possibility that, in the absence of p53, *C. reinhardtii* E2F may act as a central mediator of PCD.

The alignment of *C. reinhardtii* E2F with E2F family members from *H. sapiens* (accessions: Q01094, Q14209, O00716, Q16254, Q15329, O75461, Q96AV8, A0AVK6) and *A. thaliana* (accessions: Q9FNY0.1, Q9FV71.1, Q9FV70.1, Q9LFQ9.1, Q8LSZ4.1, Q8RWL0.1) are shown in Figure 12 and Figure 13, respectively. While these alignments

do little to distinguish family members within the same organism, the two broad classes of E2F members become apparent when visualized in this manner. While some family members possess a heterodimerization domain for interaction with the DP regulators, others do not. *H. sapiens* possesses two E2F proteins, E2F7 and E2F8, which are not predicted to contain this domain, while *A. thaliana* possesses three such members: E2FD, E2FE, and E2FF. *C. reinhardtii* E2F aligned more robustly with E2F1-6 in humans, suggesting similarities to one or more of these proteins. Interestingly, when aligned with E2F family members from *Arabidopsis*, the DP dimerization domain of *C. reinhardtii* E2F did not align with the DP dimerization domain predicted in the *A. thaliana* E2FA-C. The implications of this result, if any, are not known. Cumulatively, these results suggest that a high degree of conservation exists between *C. reinhardtii* E2F and E2F family members from animals and plants. Importantly, however, the high degree of conservation between the E2F family members renders a more detailed analysis of the results from these studies not possible.
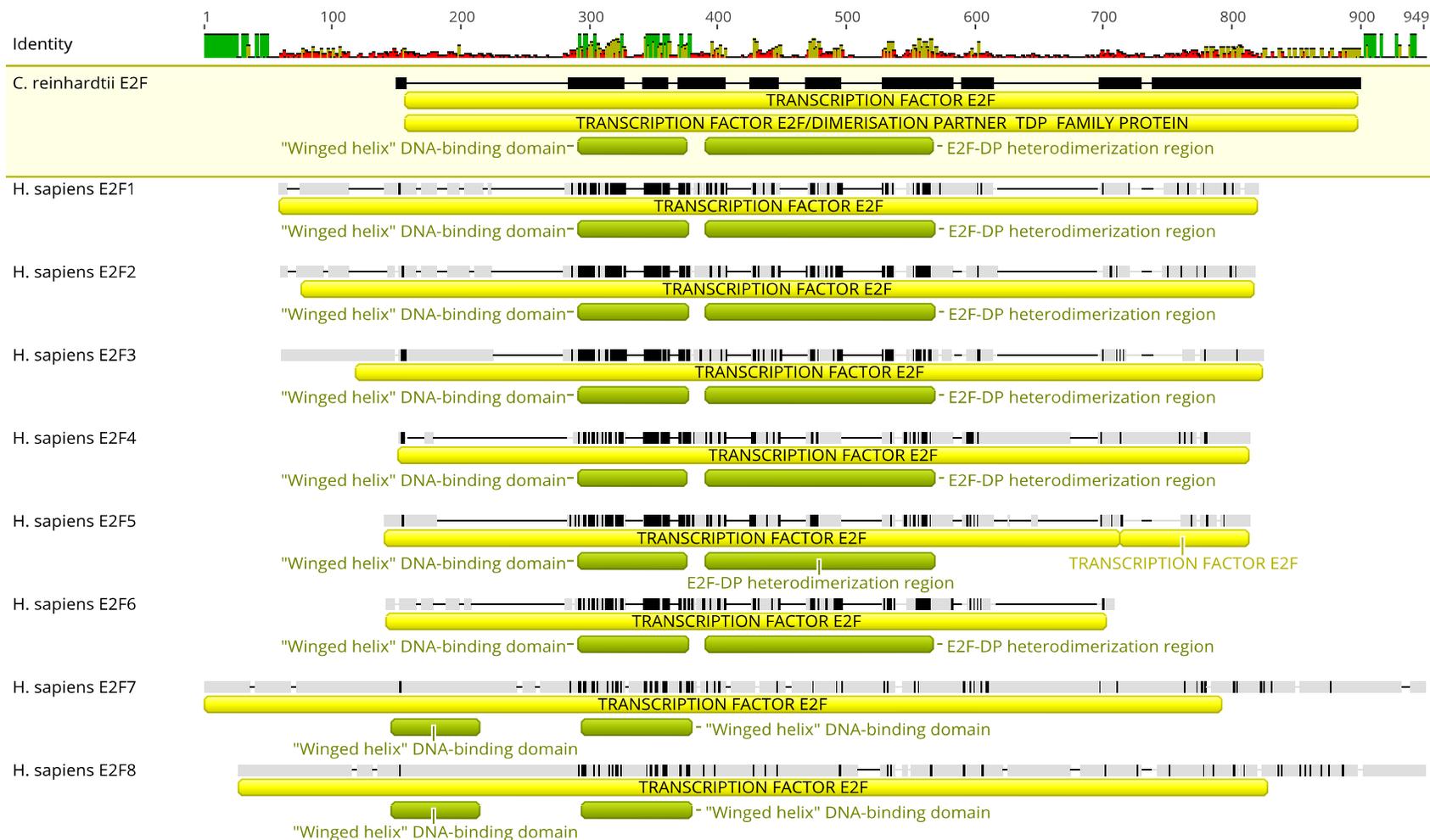
Figure 12. Alignment and annotation of the E2F proteins from *C. reinhardtii* and humans. The graph at the top of the figure represents the conservation of amino acids at each residue. Each protein has the results of the alignment, with conserved residues highlighted in color, and the annotations that were assigned to the proteins by InterProScan. The color of each annotation denotes its source, where yellow = Panther, purple = Pfam, and olive green = SuperFamily.
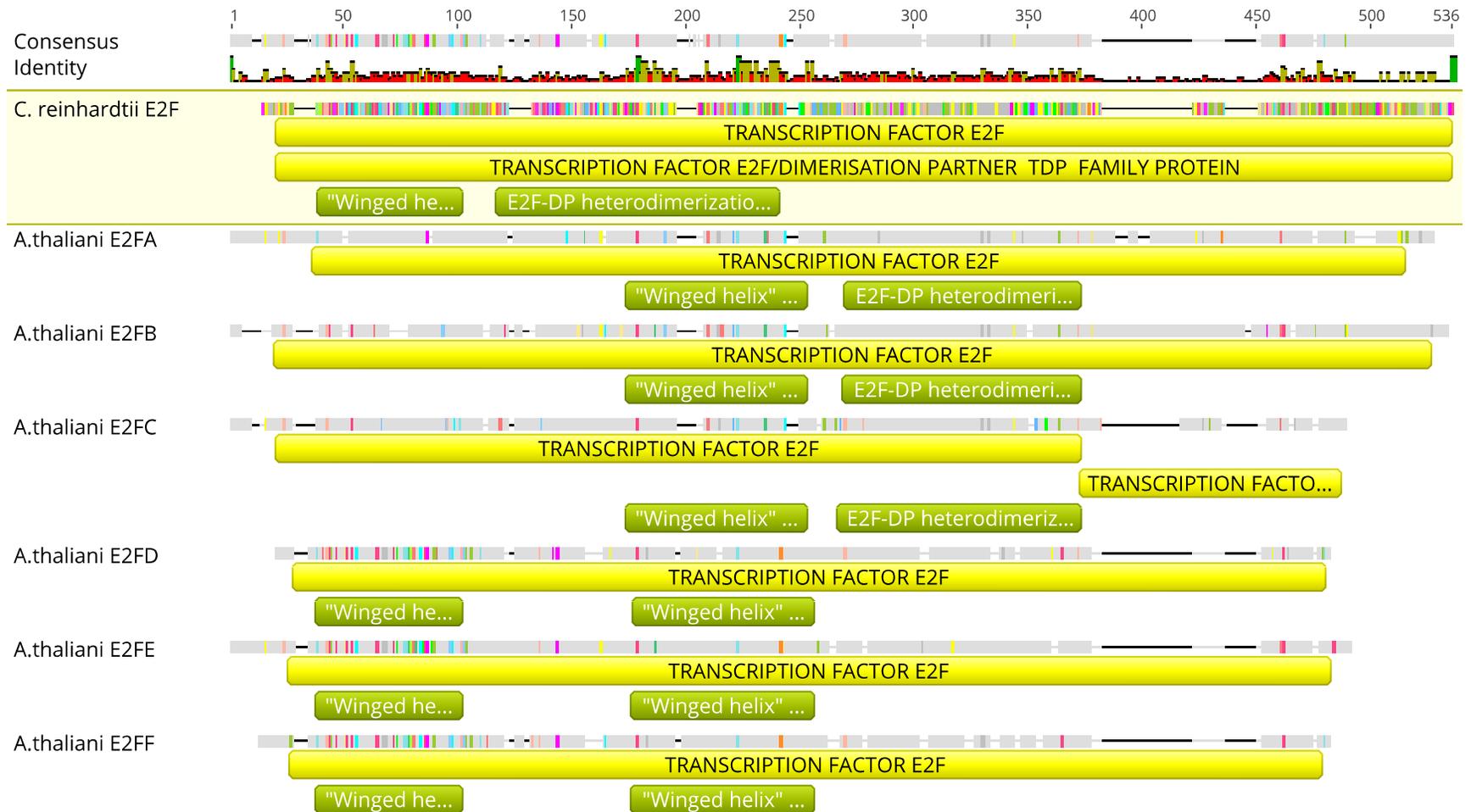
Figure 13. Alignment and annotation of the E2F proteins from *C. reinhardtii* and plants. The graph at the top of the figure represents the conservation of amino acids at each residue. Each protein has the results of the alignment, with conserved residues highlighted in color, and the annotations that were assigned to the proteins by InterProScan. The color of each annotation denotes its source, where yellow = Panther, purple = Pfam, and olive green = SuperFamily.

**Tat-D**  TatD was first described in *E. coli* as a component of the twin arginine translocation (TAT) operon[242] Unlike TatA, TatB, and TatC, which localize to the bacterial plasma membrane to form a complex necessary for the transport of proteins from the cytoplasm to the periplasm, TatD was determined to be a cytoplasmic protein with $Mg^{2+}$-dependent single-stranded DNase activity.[243,244] Because of this, it was proposed that TatD is unlikely to function in protein export with the other components of the *Tat* operon.[243]

A search for apoptotic nucleases in *Caenorhabditis elegans* identified CRN-2, a homolog of bacterial TatD.[245] Using an RNAi-based approach, repression of *crn-2* expression resulted in a delayed PCD during development, suggesting a potential role for CRN-2 in *C. elegans* PCD. Moreover, knockdown of *crn-2* expression in the background of a mutation in *ced-3*, the *C. elegans* homolog of caspase-9, resulted in a significant inhibition of cell death. Interestingly, however, an enhanced TUNEL-positive phenotype, a common marker for cell death, was observed when *crn-2* was knocked down in a background strain harboring a mutation in *cps-6*, the *C. elegans* homolog of *endoG*.[245] Cumulatively, the results of this study may indicate that *crn-2* in *C. elegans* plays a pro-PCD role during PCD, but additional research into the matter is required.

In *S. cerevisiae*, TAT-D was shown to be an endo-/exonuclease which plays dual roles in the degradation of genomic DNA.[110] TAT-D in *S. cerevisiae* acts as an endonuclease by non-specifically cleaving dsDNA and as an exonuclease by excising DNA in the 3'-5' direction. In agreement with a role for *S. cerevisiae* TAT-D in PCD, yeast strains overexpressing TAT-D exhibited an increased rate apoptosis when treated with $H_2O_2$.[110] In contrast to the results obtained from *C. elegans*, an *S. cerevisiae* mutant

strain lacking functional TAT-D exhibited a resistance to PCD in response to $H_2O_2$.

Interestingly, despite the contrasting survival results, strains of *C. elegans* and *S.*

*cerevisiae* lacking TAT-D/CRN-2 exhibited an enhancement of TUNEL-positive

phenotypes.[110]

Studies in bacteria revealed that *E. coli* TatD is a $Mg^{2+}$-dependent 3'-5'

exonuclease.[243] In addition, this study also found that *E. coli* TatD exhibits a strong

specificity for single-stranded DNA and RNA, and possesses a very low affinity for

dsDNA, indicating that the endonuclease role for this enzyme in animals and yeast may

not be conserved in bacteria. Similarly to the results observed in *S. cerevisiae*, a strain of

*E. coli* deficient in TatD was found to be sensitive to treatment with low levels of

$H_2O_2$.[243]  While this could suggest that TatD may contribute to PCD in bacteria, the

authors postulate that TatD is a DNA repair enzyme, and that the decreased survival may

be due to a role for TatD in DNA damage accumulated by exposure to $H_2O_2$.

TAT-D is widely conserved throughout all kingdoms of life, with homologs in

animals, plants, fungi, and prokaryotes.[243] Despite this, little is known regarding the role

of TatD during PCD. Moreover, investigations of the nuclease in different organisms

suggest that the function of TAT-D may have diverged between phyla. As such, we

reasoned that understanding the role of TAT-D during *C. reinhardtii* PCD may further

facilitate an understanding of how PCD is carried out in both prokaryotes and eukaryotes.

The alignment of *C. reinhardtii* TAT-D to homologs from *H. sapiens* (accession:

AAH64964.1), *A. thaliana* (accession: NP_190807.3), and *S. cerevisiae* (accession:

P34220) is shown in Figure 14. Each of the four TAT-D proteins were predicted to be

metallo-dependent hydrolases of the TAT-D DNase family. These results indicate a high

degree of conservation between *C. reinhardtii* TAT-D and its homologs from other phyla,

suggesting that the pro-PCD functions of TAT-D may be conserved in *C. reinhardtii.*
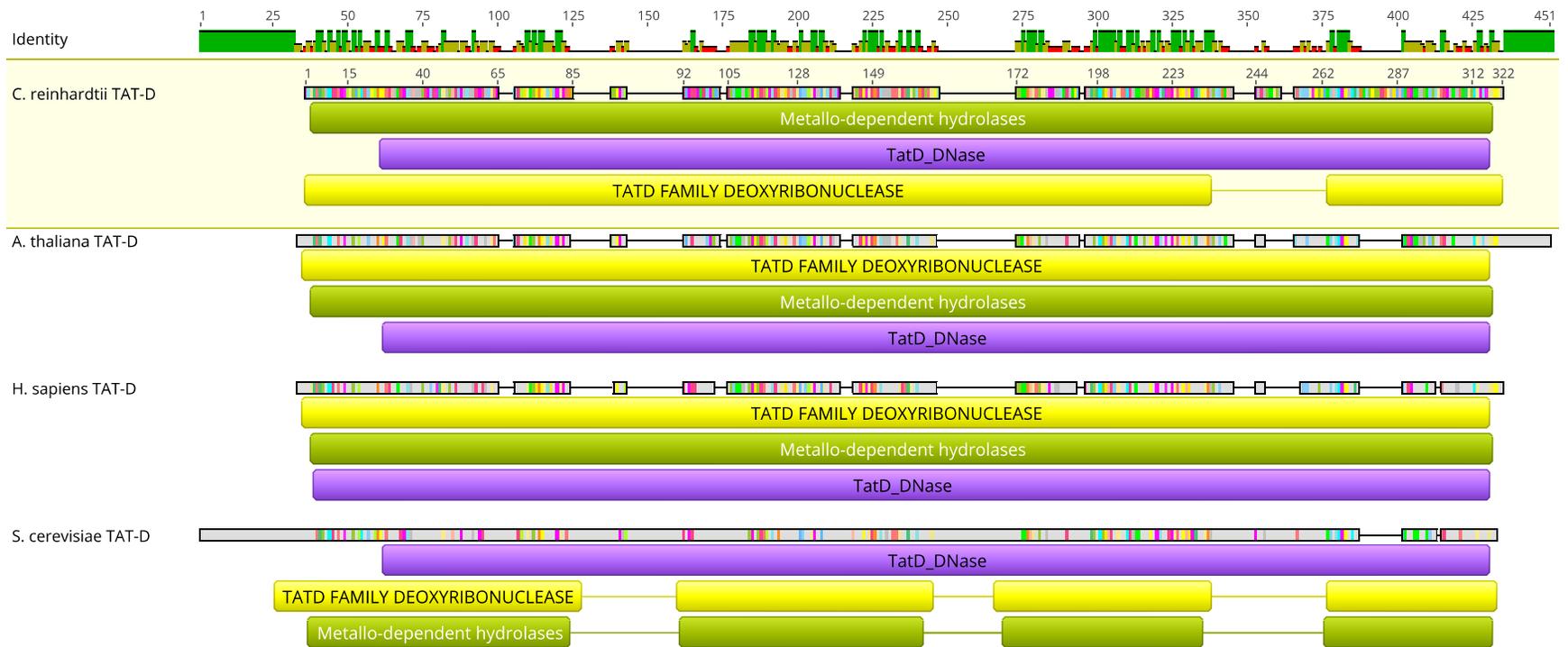
Figure 14. Alignment and annotation of the TAT-D proteins from *C. reinhardtii*, plants, yeast, and humans. The graph at the top of the figure represents the conservation of amino acids at each residue. Each protein has the results of the alignment, with conserved residues highlighted in color, and the annotations that were assigned to the proteins by InterProScan. The color of each annotation denotes its source, where yellow = Panther, purple = Pfam, and olive green = SuperFamily.

***Lesion-stimulating disease 1*** Lesion-stimulating disease 1 (LSD-1) is a negative regulator of cell death in plants.[246] Early studies into LSD-1 revealed that *lsd-1* loss-of-function in *Arabidopsis* results in a distinctive "runaway" cell death phenotype. This phenotype is characterized by the inability of leaf cells to restrict the propagation of cell death following an initial stress stimulus.[246,247] This uncontrolled death comes about as the result of accumulated ROS, including superoxide ions and hydrogen peroxide.[88,247–249] The elevated levels of ROS observed in *Arabidopsis lsd-1* mutants are the result of lowered antioxidant enzyme activity, suggesting that LSD-1 acts as a positive regulator of antioxidant machinery during PCD.

Despite the long-standing recognition of LSD-1 as an important repressor of PCD in plants, the mechanisms by which LSD-1 attenuates PCD have remained largely undiscovered.[246] The runaway death phenotype of *lsd-1* mutants is dependent on enhanced disease susceptibility 1 (ESD-1) and phytoalexin deficient 4 (PAD-4), as loss-of-function mutations in either *eds-1* or *pad-4* in the *lsd-1* mutant strain attenuate the runaway cell death phenotype. Moreover, EDS-1 and PAD-4 are believed to function antagonistically to LSD-1, as both produce effects which contrast LSD-1 in response to stress.[249] The LSD-1/EDS-1/PAD-4 regulatory system has been well established, and may represent a core mechanism by which PCD is regulated in the cell.[249]

LSD-1 contains three zinc(Zn)-finger motifs, which mitigate a protein's interactions between proteins, DNA, and/or RNA. The LSD-1 Zn-finger domains have been categorized as C2C2-type, which are distinct for some transcription factors.[250] As such, LSD-1 has been implicated to function as a transcriptional regulator during PCD.[246]

In addition to functioning as a putative transcriptional regulator, a yeast two-hybrid screen revealed ten distinct proteins with which LSD-1 may interact. At least one of these proteins, metacaspase-1, is a positive regulator of plant PCD, and interacts with the second and third Zn-finger domains of LSD-1.[251] One proposed model postulates that, under unstressed conditions, LSD-1 is bound to the inactive metacaspase-1 zymogen, suppressing its conversion into the active protease.[252] In response to a PCD-inducing stress, pro-PCD signaling causes the dissociation of LSD-1 from metacaspase-1, and subsequent maturation and activation of metacaspase-1 brings about PCD.[252]

In response to the meager understanding of the molecular and cellular functions of LSD-1, a very recent study sought to characterize the protein in greater detail.[249] While it should be noted that this analysis did not examine the role of LSD-1 during PCD, the obtained results establish the extraordinarily broad implications of LSD-1 in the plant cell. Namely, this study identifies a previously-unpredicted function of LSD-1 as a scaffolding protein which interacts with proteins that are engaged in a diverse array of molecular pathways. These include methylation, ubiquitination, cell cycle control, cell wall formation, gametogenesis, and embryo development. Furthermore, the authors demonstrate that the association of LSD-1 with proteins from these pathways is dependent on the redox state of the cell, and that oxidative stress has a strong effect on the proteins that LSD-1 interacts with. Moreover, this study also confirmed the long-standing prediction that LSD-1 acts as a transcriptional regulator that translocates from the cytoplasm to the nucleus by directly interacting with the nuclear pore complex protein nucleoporin autopeptidase.

As LSD-1 is conserved only in plants, it likely represents a divergent component of PCD that is plant-specific. As such, it may prove a useful subject for the study of plant-specific PCD processes. Furthermore, the cumulative observations that LSD-1 1) interacts with proteins from a number of essential plant PCD processes, 2) interacts with a distinct set of proteins during oxidative stress, 3) directly regulates gene transcription, and 4) is essential for the control and negative regulation of PCD, may collectively implicate LSD-1 as a primary point of integration for pro- and anti-PCD signaling. If true, this would allude to LSD-1 as a central mediator of the life and death decision-making process in plants. In addition, the conservation of LSD-1 in plants and *C. reinhardtii* suggests that the acquisition of LSD-1 was an early event in the evolution of plants. Because of this, *C. reinhardtii* LSD-1 may serve as a prime subject for understanding the early divergence of PCD between plants and other eukaryotes.

The alignment of *C. reinhardtii* LSD-1 its homolog in *A. thaliana* (accession: OAP01040) is shown in A. Both proteins were predicted to contain LSD-1-type zinc finger domains. Perplexingly, the putative LSD-1 protein in *C. reinhardtii* was predicted to be LOL-1. A search of the literature revealed that LSD-1-like 1 (LOL-1) is evolutionarily related to LSD-1 and has been implicated to function as a positive regulator of PCD in plants.[253] Similar to LSD-1, LOL-1 possesses three zinc finger domains, indicating that it may also function as a transcriptional regulator. It has been proposed that LSD-1 and LOL-1 function antagonistically to regulate plant PCD.[253] In addition, the discovery of other LOL proteins in plants presents even further possibilities for how the LSD-1 pathways might be regulated in *Arabidopsis*.[246]

When *A. thaliana* LOL-1 (accession: Q93ZB1.1) was integrated into the

alignment of LSD-1 from *C. reinhardtii* (Figure 15B), the *C. reinhardtii* protein aligned

more robustly with LSD-1 (42% identity) than with LOL-1 (38% identity). Furthermore,

InterProScan functional annotation did not annotate the LOL-1 protein as LOL-1,

indicating that this annotation system may not be sufficient to accurately distinguish

between the two closely-related proteins. Cumulatively, these results indicate that LSD-1

is highly conserved between *C. reinhardtii* and *A. thaliana*, suggesting that the anti-PCD

function of LSD-1 may be conserved in *C. reinhardtii*. However, the possibility that the

*C. reinhardtii* protein functions as LOL-1 rather than LSD-1 could not be excluded by

these analyses.

Figure 15. Alignment and annotation of the LSD-1 homologs from C. reinhardtii and plants. A) *C. reinhardtii* LSD-1 aligned with LSD-1 from *A. thaliana*. B) *C. reinhardtii* LSD-1 aligned with LSD-1 and LOL-1 from *A. thaliana*. The graph at the top of each figure represents the conservation of amino acids at each residue. Each protein has the results of the alignment, with conserved residues highlighted in color, and the annotations that were assigned to the proteins by InterProScan. The color of each annotation denotes its source, where yellow = Panther, purple = Pfam, and light blue = TIGRFAM.

*p53-induced gene 3* p53-induced gene 3 (PIG3) was first identified as one of 13 previously-undiscovered genes that are upregulated by p53 prior to the onset of apoptosis.[254] In spite of the apparent specificity of p53 to animal lineages, PIG3 is well conserved, and orthologs of PIG3 exist in both plants and bacteria.[255] Of note, a search of the protein databases and literature suggests that PIG3 is not present in yeast. PIG3 is a quinone oxidoreductase which catalyzes the reduction of quinone into hydroxyquinone.[254] Though PIG3 is known to participate in the apoptotic process, the mechanisms by which it induces apoptosis are not well-characterized.

Following the synthesis of PIG3 during apoptosis, the treatment of cells with a quinone oxidoreductase inhibitor caused a significant reduction in ROS accumulation, as well as an extensive inhibition of apoptosis, suggesting that PIG3 positively regulates the apoptotic process by contributing to the accumulation of ROS.[254]

In addition to its role in the generation of ROS, PIG3 has also been shown to contribute to the apoptotic process by other mechanisms. PIG3 positively regulates apoptosis by directly interacting with the p53 inhibitor MDM2.[256] This association attenuates the MDM2-dependent ubiquination, and subsequent proteasomal degradation of, p53. Moreover, the PIG3-MDM2 association promotes the ubiquination and subsequent proteasomal degradation of MDM2.[256]

The robust BLASTp alignment of a *C. reinhardtii* protein with human PIG3 (Table 5) indicates that the two proteins may be homologous. Given that *PIG3* is also present in plants, a lineage presumed to lack p53, it presents a unique opportunity to investigate an aspect of PCD possessing both conserved and divergent characteristics.

Furthermore, as PIG3 is present in *C. reinhardtii*, but not in *S. cerevisiae*, it may

represent another aspect of PCD that has been lost in other unicellular lineages.

The alignment of PIG3 from *C. reinhardtii* with PIG3 from *H. sapiens* (Q53FA7)

and *A. thaliana* (NP_193889.1) is shown in Figure 16. Several domains are predicted to

be conserved across all three sequences, including C- and N-terminus-specific alcohol

dehydrogenase domains and an NAD(P)-binding domain with a Rossman-fold structure.

In addition, *C. reinhardtii* PIG3 is predicted to be a quinone oxidoreductase-like.

Cumulatively, these results suggest a high degree of conservation of PIG3 between the

phylogenetic clades.
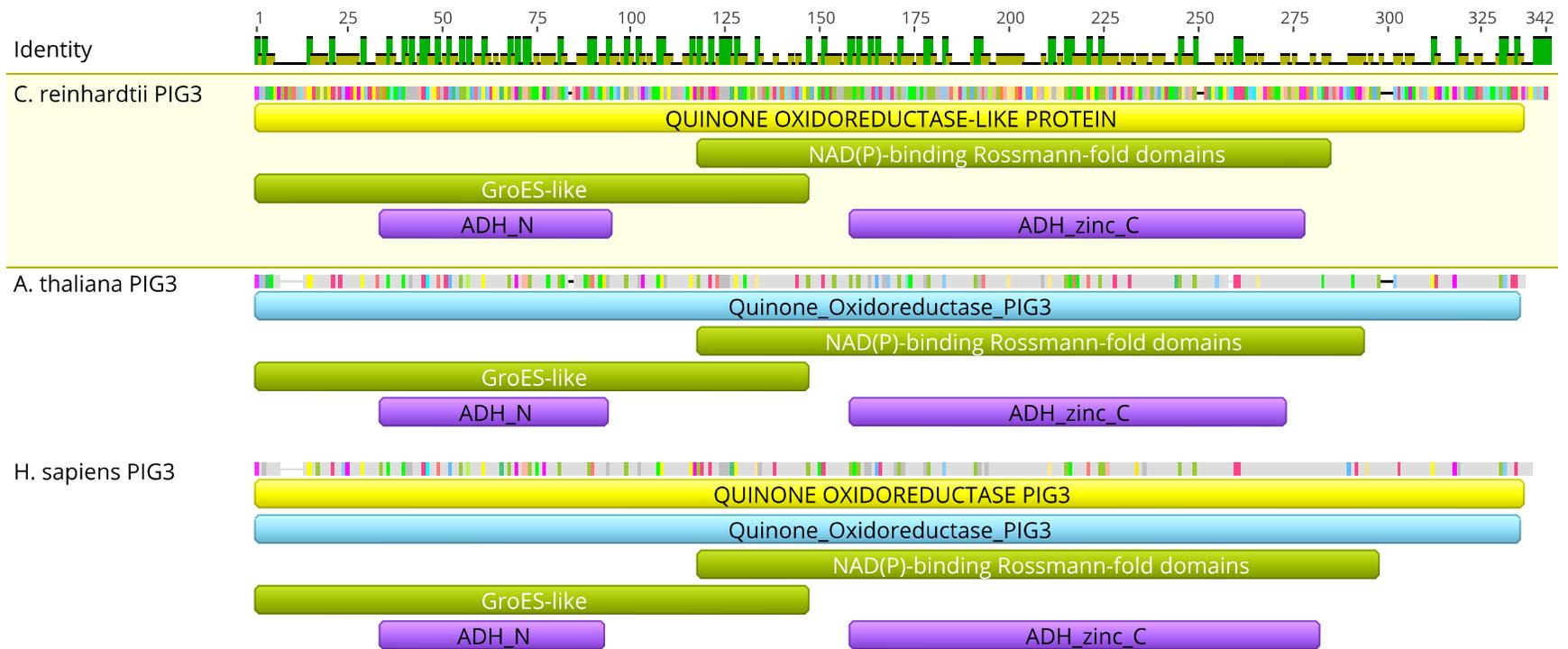
Figure 16. Alignment and annotation of the PIG3 proteins from C. reinhardtii, humans, and plants. The graph at the top of the figure represents the conservation of amino acids at each residue. Each protein has the results of the alignment, with conserved residues highlighted in color, and the annotations that were assigned to the proteins by InterProScan. The color of each annotation denotes its source, where yellow = Panther, olive green = Superfamily, purple = Pfam, and light blue = TIGRFAM.

**Summary**

Currently, very little is known about the molecular mechanisms by which *C. reinhardtii* initiates and executes PCD. As such, the first aim of this study was to predict on a large scale, *C. reinhardtii* proteins which participate in PCD. To achieve this, the *C. reinhardtii* proteome was aligned with protein sequences from other organisms that were annotated to be involved in PCD using BLASTp. After removing the less robust alignments from the BLASTp results, 2,389 *C. reinhardtii* query sequences remained. We believe that this list will serve as an excellent starting point for studies which seek to characterize the molecular basis of PCD in *C. reinhardtii*. To further facilitate such efforts, we characterized the CrPPPs by assigning predicted GO terms to each of the sequences using BLAST2GO.

The second aim of this study was to select several of the CrPPPs for further investigation regarding a role during *C. reinhardtii* PCD. In order to accomplish this goal, the BLASTp alignments of CrPPPs to the annotated PCD sequences was explored manually, and the CrPPPs which aligned strongly with a known PCD protein were used to create a list of proteins which we believe represent particularly strong candidates for further investigation. From this list of CrPPPs of interest, several were selected to be studied in this project. To validate that the robust BLASTp alignments represent similar characteristics between the known PCD subject proteins and the CrPPPs selected for further study, the amino acid sequences of the homologs from *H. sapiens*, *A. thaliana*, and *S. cerevisiae* were aligned with one another, the sequences were characterized using InterProScan annotations, and the presence/absence and locations of the annotations were compared between the *C. reinhardtii* protein and its homologs. Each of the CrPPPs

chosen for further study was predicted to be similar, both in domain composition and structure, to its homologs, and thus represent strong candidates for studies seeking to elucidate the molecular mechanisms of PCD in *C. reinhardtii* (Table 6).

Table 6

*C. reinhardtii* proteins selected for further study

| Protein | Conservation | Predicted PCD Role |
|---------|--------------|--------------------|
| BI-1 | Eukaryotes | Anti-PCD |
| IRE-1 | Eukaryotes | PCD Mediator |
| E2F | Animals, plants | Pro-PCD |
| TAT-D | Eukaryotes, prokaryotes | Pro-PCD |
| LSD-1 | Plants | Anti-PCD |
| PIG3 | Eukaryotes | Pro-PCD |

**CHAPTER III**

**Selection and Verification of *C. reinhardtii* Mutant Strains**

**Introduction**

Mutant strains are extremely powerful tools which facilitate the elucidation of gene function by examination of the phenotypes that result from the absence of a functional gene product. Collections of individual mutant strains are called mutant libraries, and can be produced or purchased by researchers interested in studying the function of a specific gene or a set of genes. Mutant libraries have revolutionized the study of gene function studies in bacteria, yeast, animals, and land plants.[257–261]

Though mutant libraries can be generated on a large scale by a number of different techniques, one of the most common means of mutant library construction is through random insertional mutagenesis, a process by which foreign DNA introduced into the cell and non-specifically integrated introduced into the genome. The insertional loci are mapped for each transformant by sequencing the genomic sequencing that flank either side of the cassette. In addition, a selection marker, such as an antibiotic resistance gene, is integrated into the cassette to positively-select for transformant cells, while inhibiting the growth of non-transformants, when cultured on growth medium containing the antibiotic.

Historically, reverse genetic studies in *C. reinhardtii* have been restricted by the lack of a comprehensive *Chlamydomonas* mutant library. In 2016, a commercially-available library of *C. reinhardtii* mutant strains was made available for researchers by the *Chlamydomonas* Library Project (CLiP).[173] The library was generated by electroporation-facilitated random insertional mutagenesis in the background strain cc-

4533 of *C. reinhardtii*. The insertional cassette used to generate the mutants contains a gene conferring resistance to paromomycin, an antibiotic to which *C. reinhardtii* is naturally susceptible.[262] As such, positive selection of *C. reinhardtii* transformants following mass mutagenesis was achieved by plating the transformation culture onto medium containing the antibiototic. Transformants colonies were isolated and organized on solid medium in a layout similar to that of a 384-well plate.[173]

The insertional locus of each colony was determined using a two-step combinatorial super-pooling method, in conjunction with large-scale sequencing.[173] The first pooling scheme combined all colonies from a single plate. The second pooling scheme combined colonies by plate position across all plates. The genomic sequences flanking the inserts of all colonies in each pool were identified by two large-scale sequencing techniques: *Chlamydomonas* MmeI-based insertion site sequencing (ChlaMmeSeq) and Linear and Exponential Amplification coupled with Paired-end Sequencing (LEAP-Seq).[173,262] Following the determination of genomic flanking sequences within in each pool, deconvolution algorithms were employed to assign the detected genomic flanking sequences to each of the individual colonies. Alignment of these flanking sequences with the *C. reinhardtii* genome revealed the insertional locus of the cassette for each colony.

As suggested by the curators of the CLiP library, the reported insertional locus of each mutant strain should be individually confirmed prior to using the strains in experiments. In the article describing the generation of the mutant library, the authors demonstrated that insertion sites mapped by the large-scale methods described above were accurate in approximately 75% of the available strains. Furthermore, the curators of

the library caution that any culture received by the library may contain a mixture of more than one mutant strain. As such, validation of the mapped insertional mutation is a necessary prerequisite to experimental utilization of any CLiP mutant strain.

In previous work, we used a large-scale approach to predict 2,389 *C. reinhardtii* proteins which participate in PCD. From this list of *C. reinhardtii* predicted PCD proteins, several were selected to be examined further for a role in *C. reinhardtii* PCD. The purpose of the current study is twofold. First, for each CrPPP that was previously selected for further study, we sought to select a mutant strain of *C. reinhardtii* with a disrupted gene coding for the selected protein. Second, for each of the chosen mutant strains, we sought to validate the insertion sites reported by the CLiP library.

**Materials and Methods**

**Acquisition of *C. reinhardtii* Strains and Primers** Mutant strains with an insert in one particular gene were chosen from the CLiP library (Table 7). To search for strains with an insert in the gene of interest, the gene locus was used to query the mutant collection from the CLiP website. If, for a given gene, multiple mutant strains were available, insertions that were 1) in the coding region of the predicted gene, and/or 2) had a high probability of being mapped correctly, were preferentially selected. The *C. reinhardtii* strains arrived by mail as slant cultures of growth medium. Upon the arrival of the ordered strains, an initial backstock of each strain was created.

Table 7

Insertional characteristics of *C. reinhardtii* strains with a mutation in the selected genes.

| gene | gene id | mutant_ID | chromosome | side | feat. | flanking_seq | conf |
|------|---------|-----------|------------|------|-------|--------------|------|
| *bi-1* | Cre01.g061807 | LMJ.RY0402.079236 | 1 | 3' | 3'UTR | GTCTGAGCGTGACTCGAAGGACCTTGTCTC | 73% |
| | **Cre01.g061807** | **LMJ.RY0402.114320** | **1** | **3'** | **intron** | **ATATTCAAGACAGCGCGCTGGCGACTTGCA** | **95%** |
| | **Cre01.g061807** | **LMJ.RY0402.114320** | **1** | **5'** | **intron** | **GTTGCCTGTGGATGTTGCCCAGCATTGACT** | **95%** |
| | Cre01.g061807 | LMJ.RY0402.127549 | 1 | 3' | intron | ACACACACCTACACACAAACACACACACAT | 73% |
| | Cre01.g061807 | LMJ.RY0402.147808 | 1 | 5' | intron | CGAGGACGTGACTTGGCAGCTGCAAGTCGC | 73% |
| *ire-1* | Cre08.g371052 | LMJ.RY0402.058154 | 8 | 3' | MSVs | CAAATCCTTGCACCACGTCAGCAGGTCCCG | 73% |
| | Cre08.g371052 | LMJ.RY0402.079723 | 8 | 3' | exon | TCTCCACGGCGGTCCTGGTTGGCTACGAGC | 58% |
| | Cre08.g371052 | LMJ.RY0402.098795 | 8 | 5' | intron | AGCAAAGCACATGAGGAAAGGGTTACTTAC | 58% |
| | **Cre08.g371052** | **LMJ.RY0402.122895** | **8** | **5'** | **MSVs** | **ATAGGACGGGGGTTGCATGGATTGGTACAG** | **95%** |
| | **Cre08.g371052** | **LMJ.RY0402.122895** | **8** | **3'** | **MSVs** | **CTGGCTACGCCTCCTCTTCGTTTCTAATGA** | **95%** |
| | Cre08.g371052 | LMJ.RY0402.209343 | 8 | 3' | MSVs | GGGTGAGTGAGAGCGTGTGTTTGAGTGTGT | 73% |
| *e2f* | **Cre01.g052300** | **LMJ.RY0402.074518** | **1** | **5'** | **exon** | **GGTCCTGGTCGTCGGGGTCCGCTCCCCGGT** | **95%** |
| | **Cre01.g052300** | **LMJ.RY0402.074518** | **1** | **3'** | **exon** | **CTACGGTGGAGGGGACGAGGACGACGAGG** | **95%** |
| | Cre01.g052300 | LMJ.RY0402.087995 | 1 | 5' | 3'UTR | GCCCTCTCCCCACTCCATGACTTGCATCTC | 95% |
| | Cre01.g052300 | LMJ.RY0402.087995 | 1 | 5' | 3'UTR | CAGGATTTTAGTGGGTAGCAGTCGTAAAGT | 95% |
| | Cre01.g052300 | LMJ.RY0402.128436 | 1 | 3' | intron | GCGCGTGCGAAACCCCTCGCTACATGGGAA | 95% |
| | Cre01.g052300 | LMJ.RY0402.128436 | 1 | 5' | intron | CACCATGCAAGGCTCACCTCAACCTTCCCT | 95% |
| | Cre01.g052300 | LMJ.RY0402.135078 | 1 | 3' | 3'UTR | TGGCTGGGTTGATAACGACCTTTTCCTGCC | 58% |
| | Cre01.g052300 | LMJ.RY0402.181547 | 1 | 5' | 3'UTR | GAGAAGAAGAAGCGGCTGGGCGGCGACATC | 73% |

| gene | gene id | mutant_ID | chromosome | side | feat. | flanking_seq | conf |
|------|---------|-----------|------------|------|-------|--------------|------|
| | Cre01.g052300 | LMJ.RY0402.190878 | 1 | 5' | 3'UTR | CACCACCACACTAACACCGTGTAACGACTC | 95% |
| | Cre01.g052300 | LMJ.RY0402.190878 | 1 | 3' | 3'UTR | GTCATGGAGTGGGGAGAGGGCTGATCAAAC | 95% |
| *tat-d* | Cre02.g085650 | LMJ.RY0402.046580 | 2 | 5' | intron | CATGCACCACTCACATGGGTAGGCCCGACT | 58% |
| | Cre02.g085650 | LMJ.RY0402.146555 | 2 | 5' | exon | CCGCTTGTACTGCACCGTGGGGTGTCACCC | 73% |
| | **Cre02.g085650** | **LMJ.RY0402.192095** | **2** | **3'** | **exon** | **CGCCACGGCCTCCAGGTCCGTTCTCCCTT** | **95%** |
| | **Cre02.g085650** | **LMJ.RY0402.192095** | **2** | **5'** | **exon** | **TCTTCGGCAACACCGAGCGAATGTTCTTTC** | **95%** |
| | Cre02.g085650 | LMJ.RY0402.195934 | 2 | 3' | exon | TCCTTGCACCGAGTGGGGTGACACCCCACG | 95% |
| | Cre02.g085650 | LMJ.RY0402.195934 | 2 | 5' | exon | GTTCGAGGACCACCCGGGGGGGACCGGAAGC | 95% |
| | Cre12.g517350 | LMJ.RY0402.076546 | 12 | 3' | 3'UTR | GCGATGATGGGCCAAGGAGCGATAGGGACA | 73% |
| *lsd-1* | **Cre12.g517350** | **LMJ.RY0402.114954** | **12** | **3'** | **exon** | **GGCACCCGGCACCCGTTGCACACGATTTGG** | **73%** |
| | Cre12.g517350 | LMJ.RY0402.131712 | 12 | 5' | 3'UTR | CTCCTCTGCATCTCATCATACCAGCGTCAC | 95% |
| | Cre12.g517350 | LMJ.RY0402.131712 | 12 | 5' | 3'UTR | ATGGGTACCAGCTATTCCATGTTGTGTAGG | 95% |
| | Cre12.g517350 | LMJ.RY0402.198243 | 12 | 5' | exon | GGCTCCATGGCCCCGCCTTCTCAAAGCCAC | 73% |
| | Cre12.g517350 | LMJ.RY0402.208237 | 12 | 3' | 5'UTR | AGTGTGAAGCTAATCGCGCAGGGGCTACAA | 73% |
| *pig3* | **Cre16.g674050** | **LMJ.RY0402.061331** | **16** | **5'** | **exon** | **CATCATGTAGGAGCCCCTGCAGGCAGGCAG** | **95%** |
| | **Cre16.g674050** | **LMJ.RY0402.061331** | **16** | **3'** | **exon** | **GTTTGGCCAGGGCAAGCTGCGCGTGGAGGT** | **95%** |
| | Cre16.g674050 | LMJ.RY0402.167637 | 16 | 3' | intron | TGGCCCCGGCCACTCGGTTCAGTGTCCATT | 58% |
| | Cre16.g674050 | LMJ.RY0402.197441 | 16 | 3' | exon | GTGGCCACTTGGCCGAAGCGCGTGCGCAGG | 58% |

NOTE: Bolded entries designate strains that were selected for further study. Each row represents a side of the cassette that was mapped by the CLiP library. As each inserted cassette has a genomic sequence flanking either side, a single mutant strain should have two rows: one for each side of the cassette inserted into the genome. An exception to this is the *lsd-1* mutant strain, in which only one side of the cassette was mapped. MSVs = multiple splice variants. Table adapted from the "basic mutant data table" file from the CLiP website (available at https://www.chlamylibrary.org/download).

For each mutant strain, the CLiP library suggests genomic primers to use for verification of the insertional locus. The library also provides the primer sequences which bind the insertional cassette and face outward towards the genome. Notably, these same primers were used to map the genomic flanking sequences. The suggested genomic primer sequences for each locus of interest, as well as the sequences for the cassette primers, were retrieved from the CLiP website (Table 8). All primers were ordered from Thermo Fisher. The genomic primers were rehydrated to a concentration of 20 µM, and the cassette primers were rehydrated to a concentration of 10 µM.

Table 8.

Cassette and genomic primers used for verification of the mutant strains.

| Primer Name | Primer sequence (5' to 3') |
| --- | --- |
| *bi-1* G1 | AACCTATCACAATCCGCTGG |
| *bi-1* G2 | CCCAACAGAACAAGCGGTAT |
| *ire-1* G1 | CAACAGTTGCGTAGCGTTGT |
| *ire-1* G2 | GGATTGTGGTTTCAGAGCGT |
| *e2f* G1 | CGCAGCTGTACCATCACTCA |
| *e2f* G2 | CCTGAACACACATGCCAAAC |
| *tat-d* G1 | GCTCAGGGGGTAGAATGTCA |
| *tat-d* G2 | GCTCTGGGGTAAGGGTAAGG |
| *lsd-1* G1 | TGCATAGTGTGCATGCGTAA |
| *lsd-1* G2 | AGAGGCGTGCTTGTGAAGAT |
| *pig3* G1 | TGGAAATAGCTGTCGCTGTG |
| *pig3* G2 | CTTGCCCTGCACACAATCTA |
| Cassette Primer (C1) | GCACCAATCATGTCAAGCCT |
| Cassette Primer (C2) | GACGTTACAGCACACCCTTG |

**Verification of Mutant Strains** As recommended by the curators of CLiP library, the verification of each mutant strain consisted of four steps: 1) procuration of individual colonies from the slant culture received to produce a clonal culture of cells, 2) PCR amplification of the insertional locus reported by the CLiP library using genomic primers targeting either side of the insertional locus, 3) PCR amplification of both cassette-genome junctions using the locus-specific genomic primers in combination with the cassette primers, and 4) sequencing of the cassette-genome junction PCR products.

*Obtaining Clonal Populations* In order to obtain a single colony from each strain, an inoculating loop was used to transfer a small number of cells from the stock slant to a fresh TAP plate, where an isolation streak was performed. The plates were allowed to incubate for three days. For each strain, five unique and well-isolated colonies were selected. Each of the selected colonies was picked onto a fresh TAP plate, where they were individually streaked to obtain small lawns of growth.

*Extraction of Crude Genomic DNA* For each strain, a small loopful of cells was taken from a single lawn and transferred into a PCR tube containing 50 µL of freshly-prepared 10 mM EDTA, pH 8.0. The tubes were vortexed for ten seconds. The samples were then boiled in a thermocycler at 100°C for 10 minutes, then cooled at 4° C for 1 minute. The samples were vortexed for 10 seconds and subsequently centrifuged at 1000 x g for 1 minute. The resulting supernatant, containing the crude genomic DNA, was aliquoted into fresh 1.5 mL centrifuge tubes.

*PCR Amplification of Insertion Sites*

*Prediction of Amplicon Sizes Using Genomic Primers* The predicted binding sites of all genomic primers were determined. The BLASTn feature in Phytozome was used to

align the G1 and G2 primer sequences recommended for each strain (Table 8) with the *C. reinhardtii* genome (v5.5). For each pair of genomic primers, the predicted amplicon size was calculated by taking the difference, in base pairs, of the genomic sites to which the primers aligned.

*Laboratory Methods* The PCR reactions were prepared as described by the CLiP website, with some modifications. Primer mixes of G1/G2 primers were prepared for each of the insertional loci by combining 50 µL of 20 µM G1 primer, 50 µL of 20 µM G2 primer, and 100 µL sterile water. A stock "PCR Solution" was prepared by mixing DMSO and sterile water in a 1.25:20.25 ratio.

For each pair of genomic primers, two separate PCR reactions were carried out. In one reaction, genomic DNA from the corresponding mutant strain was used as a template. In the other reaction, genomic DNA from the wild-type strain was used as a template. The PCR reactions were prepared by adding 21.5 µL PCR Solution, 2.5 µL G1/G2 primer mix, and 1 µL of the appropriate genomic DNA into an Illustra PuReTaq Ready-To-Go PCR Bead (GE Healthcare). The PCR reactions were carried out as described in Table 9.

Table 9.

PCR conditions for amplification of insertional loci and cassette-genome junctions

| Number of Repetitions | Temperature (°C) | Time |
| --- | --- | --- |
| Once | 95 | 5 minutes |
| 40 Cycles | 95 | 30 seconds |
| | 58 | 45 seconds |
| | 72 | 2 minutes |

| Number of Repetitions | Temperature (°C) | Time |
|---|---|---|
| Once | 72 | 10 minutes |
| Once | 10 | Hold |

The PCR products were immediately subjected to electrophoresis on a 1% agarose gel made with TAE and 0.00002% SYBR Safe (Thermo-Fisher). Electrophoresis was performed at 120V for 30 minutes, and the bands were visualized using the ChemiDoc XRS+ (Bio-Rad) imaging system, which had been outfitted with the XcitaBlue conversion screen (Bio-Rad) for compatibility with SYBR Safe.

### PCR Amplification of the Cassette-Genome Junctions

*Determination of the Cassette Primer Binding* Sites In order to amplify the cassette-genome junctions, the CLiP library provides primer sequences designed to bind either side of the cassette and face outwards towards the genome. However, the binding sites of the cassette primers on the cassette are not provided. As a portion of the cassette will be included in the PCR product of each cassette-genome junction, the binding sites of the cassette primers are necessary to accurately predict the size of the cassette-genome amplicons. In addition, knowledge of the cassette side that each primer binds is necessary to determine the proper cassette/genome primer combinations to use when amplifying the cassette-genome junctions.

The nucleotide sequence of the CIB1 cassette was retrieved from the CLiP website. The cassette sequence, along with the C1 and C2 primer sequences, were imported into the Geneious (v10.1.2) software. The binding sites of the C1 and C2 primers on the CIB1 cassette were identified using the "Design New Primers" function.

*Determination of Necessary Genomic/Cassette Primer Combinations* The CLiP website provides a document titled "Instructions for characterizing insertion sites by PCR" (https://www.chlamylibrary.org/help). The authors of this document provide an example of a PCR-based method to confirm that an insert was mapped correctly. In this example, the cassette-genome junctions were amplified using the cassette primer C1 in conjunction with the genomic primer G1, and the cassette primer C2 in conjunction with the genomic primer G2. Of note, the cassette can insert into the genome in one of two orientations: sense or antisense. In the example provided, however, a single mutant strain, and thus only one cassette orientation, was used.

In this study, when designing the PCR reactions to amplify the cassette-genome junctions, it was unclear if the orientation of the cassette would alter the combination of cassette/genome primers to be used. As such, for each mutant strain, the orientation of the cassette was determined. The orientation of the cassette, in conjunction with the binding locations of the cassette primers, were used to determine the appropriate combination of genomic and cassette primers to use.

For each mutant strain, the genomic sequences reported to flank each side of the cassette (Table 7) were aligned with the *C. reinhardtii* genome using the BLASTn feature on the Phytozome webpage. The location of the genomic flanking sequences relative to one another revealed the orientation of the insert in each strain. For each of the strains, the cassette orientation and the predicted binding sites of the genomic primers were used to determine the necessary cassette/genome primer combination needed to amplify the cassette-genome junctions.

*Prediction of Cassette-Genome Amplicon Sizes* Amplification of any cassette-genome junction would consist of two parts: the base-pairs contributed by the genome and the base-pairs contributed by the cassette. To determine the size of the genomic portion of the cassette-genome amplicons, the difference, in bp, of each flanking sequence location and the binding site of its corresponding genomic primer was calculated. The portion of the cassette-genome junction amplicon contributed by the cassette was determined by taking the difference of the cassette primer binding site and the terminal residue of the cassette. Finally, in order to predict the total size of each cassette-genome junction amplicon, the portion of the amplicon contributed by the genome and the portion of the amplicon contributed by the cassette were summed together.

*Laboratory Methods* Primer mixes of G1/C1 and G2/C2 were prepared by adding 100 µL of 10 µM cassette primer, 50 µL of 20 µM genomic primer, and 50 µL of water to a sterile 1.5 mL microcentrifuge tube. The PCR reactions were prepared by adding 21.5 µL PCR Solution, 2.5 µL G1/G2 primer mix, and 1 µL of the appropriate genomic DNA into a PCR tube containing a Ready-To-Go bead. The PCR reactions were allowed to proceed under the previous conditions (Table 9). Agarose electrophoresis and imaging of the PCR products were carried out as described above.

***Sequencing*** The PCR products resulting from the amplification of the cassette-genome junction were sent to Lone Star labs (http://www.lslabs.com) for sequencing. The G1/C1 products were sequenced using the C1 primer, while the G2/C2 products were sequencing using the C2 primer.

**Results and Discussion**

**Verification of Mutant Strains**

***Prediction of the Cassette Primer Binding Sites*** It was reported by the CLiP website that the cassette primers C1 and C2 bind opposite sides of the cassette and face outward towards the genome. To determine where each primer binds the cassette, the primer sequences were aligned to the cassette sequence using Geneious (Figure 17). These results confirmed that the C1 and C2 primers bind to opposite sides of the cassette and are oriented such that the 3' end of each primer faces the genome.

As the amplification of the cassette-genome junctions utilizes a primer that binds the cassette, a portion of that cassette will be included in the resulting PCR product. As such, in order to accurately predict the size of the expected PCR product, it is necessary to determine the number of base pairs that the cassette will contribute to the amplicon. To determine the number of base pairs from the cassette that would be included in the cassette-genome junction amplicons, the distance between the most interior residue of the cassette primer binding site and the most exterior residue of the cassette was determined. It was found that the innermost residue of the C1 primer binding site is 123 bp from the end of the cassette. Similarly, the most interior residue of the C2 primer binding site was found to be 134 bp from the side of the cassette that it binds to (Figure 17). These findings were used to accurately predict the expected amplicon sizes of the cassette-genome junction PCR products.

***Determination of the Cassette-Genome Primer Combinations*** Because it was unclear if the orientation of the cassette in the genome would alter the genomic/cassette primer combinations to use when amplifying the cassette-genome junctions, the

combination of primers necessary to amplify the cassette-genome junction was determined for each strain. This was achieved by deductively piecing together several pieces of information.

First, the relative locations of the suggested genomic primers revealed which genomic primer binds the genome upstream of the insertion site and which genomic primer binds the genome downstream of the insertion site. This indicated which primer would act as the forward primer during PCR, and which primer would as the reverse primer.

Second, the relative locations of the genomic sequences that flank each side of the insert were determined. At this point, it should be noted that the CLiP library informs users of the library which side of the cassette that each cassette primer binds (Table 7). However, the nomenclature system that is used labels one side of the cassette as the "5' side" and the other side of the cassette as the "3' side". In light of the fact that the cassette can insert into the genome in an inverse manner, thus causing the "5' side" to be oriented downstream of the "3' side", this nomenclature was found to be confusing. While such terminologies were not used while conducting the experiments herein, the knowledge that the genomic sequence flanking the so-called "5' side" of the cassette was obtained using the C1 primer and that the genomic sequence flanking the so-called "3' side" was obtained using the C2 primer proved useful.
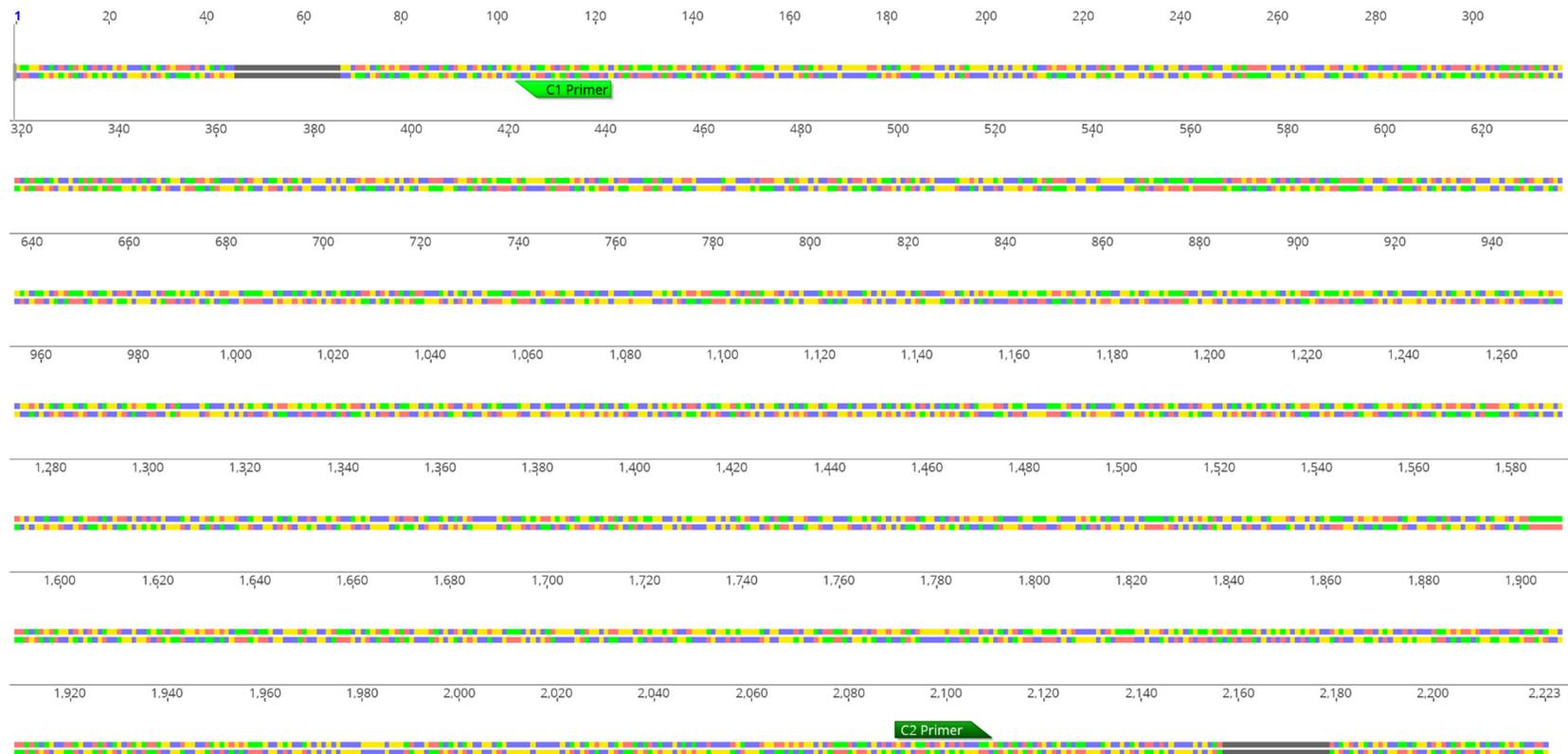
Figure 17. CIB1 Cassette with C1 and C2 Primer binding sites shown. The C1 primer one side of the cassette, while the C2 primer binds the opposite side of the cassette. The primers bind opposite strands of the cassette and face outward towards the genome.
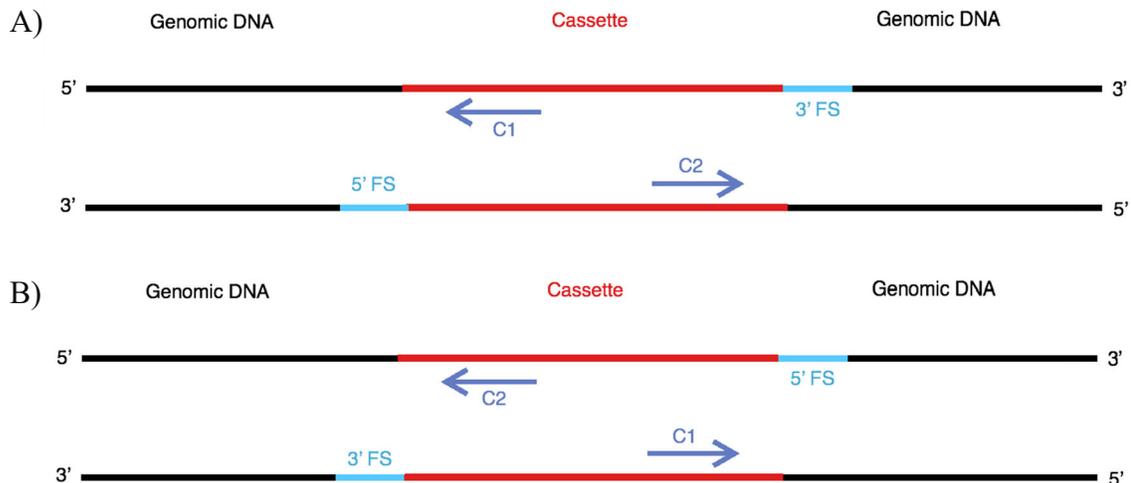
Figure 18. The effect of cassette orientation on the cassette primer positioning and elongation direction. With regard to the orientation of the cassette, two scanrios are possible: A) The cassette is not inverted during insertion to the *C. reinhardtii* genome. B) The cassette is inverted during insertion into the *C. reinhardtii* genome. Genomic sequences flanking either side of the cassette are shown in light blue.

Using the relative locations of the genomic sequences flanking the insert, in combination with the knowledge of which cassette primer was used to produce each flanking sequence (Table 7), the locations of the cassette primers relative to one another, and thus the orientation of the overall cassette, could be determined for each mutant strain. Understanding the location and directionality of each cassette primer allowed for determination of the proper cassette/genomic primer combinations to use for amplification of the cassette-genome junctions. For each of the mutant strains, regardless of the cassette orientation, it was found that the primer combinations G1/C1 and G2/C2 should be used when amplifying the cassette-genome junctions.
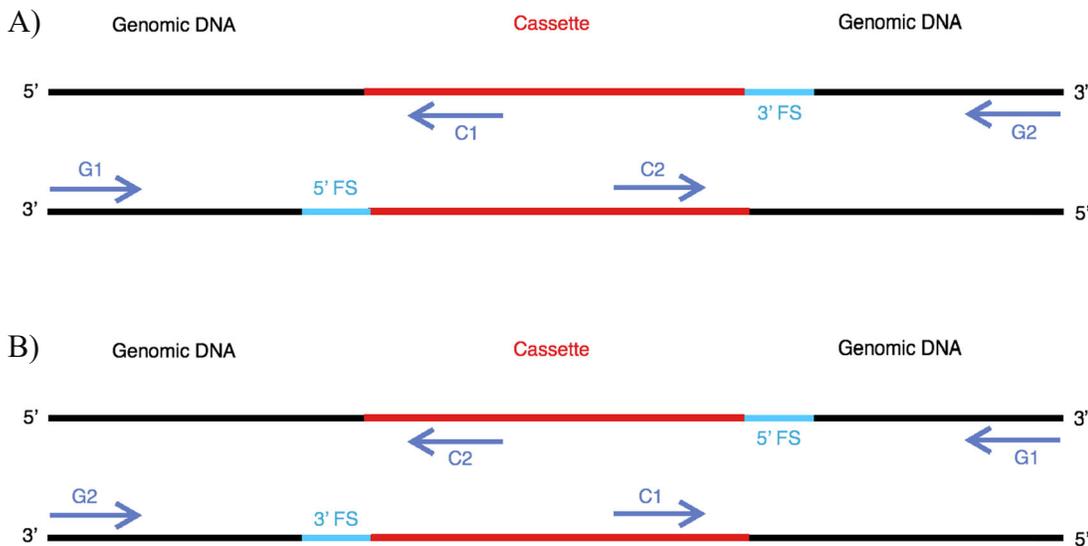
Figure 19. The effect of cassette orientation on the binding sites and elongation direction of the suggested genomic primers. Genomic sequences flanking either side of the cassette are shown in light blue.

### Prediction of PCR Results

*Qualitative Predictions* If the insertion loci of the mutant strains were mapped correctly by the CLiP library, several results were expected following PCR. First, amplification of the insertional loci using genomic primers G1/G2 would yield products of the expected size in the WT strain only. No products were expected to be present in the mutant strains, as the size of the insert was predicted to cause the reaction to fail. Second, amplification of the cassette-genome junctions using primer combinations G1/C1 and G2/C2 would yield products in the mutant strain only. No products were expected to be present in the WT strain, as the cassette primers would have no cassette to bind.

*Quantitative Predictions of Amplicon Sizes* The predicted sizes of all PCR products were determined. The expected amplicon sizes of the G1/G2 products were determined by taking the difference, in nucleotides, of the predicted binding sites of each primer. The expected amplicon sizes of the cassette-genome junctions, using the primer

combinations G1/C1 and G2/C2, were determined by 1) calculating the difference, in nucleotides, of the genomic primer binding site and the genomic sequence flanking the cassette in order to determine number of nucleotides contributed to the product by the genome, and 2) adding the nucleotides contributed to the amplicon by the cassette (Figure 17) to the expected size of the PCR product (Figure 20).
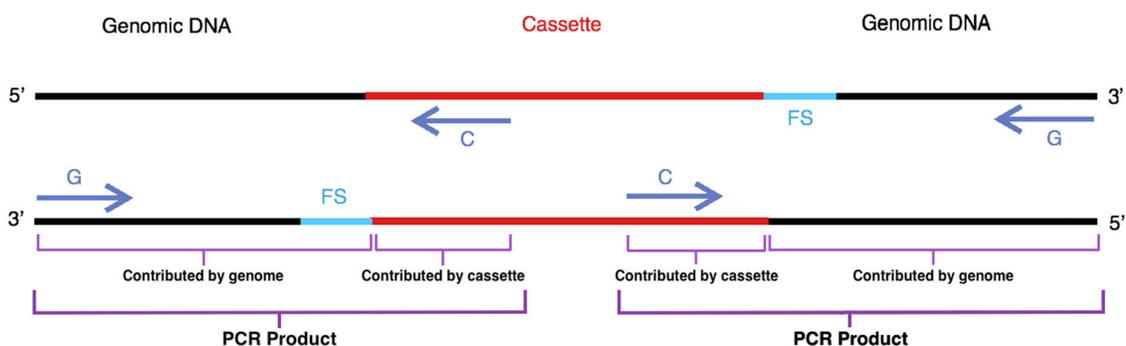


Figure 20. Visualization of the components of the two cassette-genome amplicons. For each PCR reaction, base pairs originating from both the cassette and the genome contribute to the overall size of the expected PCR product.

***Validation of Insert Mapping By Sequencing*** Observing products of the expected sizes in a mutant strain which are absent in the WT strain gives strong evidence for the claim that insert has been mapped correctly by the CLiP library. However, this method is not entirely conclusive, as it only demonstrates that the products are of the expected size; it does not establish that the products observed are the cassette-genome amplicons.

In order to show that the sequences obtained by amplifying the cassette-genome junctions are the same nucleotide sequences that compose the cassette-genome junction, the products from these PCR reactions were sent off for sequencing. The cassette primer that was used to amplify the cassette-genome junction was also used to sequence the PCR products. After the sequencing products were obtained, they were aligned with a simulation of the insertion site. This simulation was created for each mutant strain by

obtaining the gene sequence from Phytozome, the CIB1 cassette sequence from the CLiP

library (https://www.chlamylibrary.org/showCassette?cassette=CIB1), and inserting the

CIB1 sequence into the reported loci within the Geneious software. The sequencing

products from cassette-genome junction were also imported into Geneious and were

aligned with the simulation of the disrupted gene using discontiguous BLASTn. It was

expected that, if the amplification of the cassette-genome junctions was successful, the

sequencing products would align with their respective cassette-genome junctions in the

simulated insertion.

**Verification of Insertion Sites in Each Mutant Strain** In the following

paragraphs, the results of the mutant verification experiments will be presented and

discussed for each of the mutant strains.

***Verification of the bi-1 Mutant Strain*** Based on the alignment of the *bi-1* G1 and

*bi-1* G2 primers with the *C. reinhardtii* genome, it was predicted that PCR amplification

of the *bi-1* locus would yield a product of 1,420 bp using WT template DNA (Figure

21A). This prediction was supported, as a product slightly smaller than 1,500 bp was

observed (Figure 21C). Furthermore, it was expected that PCR amplification would not

yield a product in the *bi-1* mutant strain. This prediction was also supported, as no

product was visible in the *bi-1* mutant lane (Figure 21C).

With regard to the cassette-genome junctions, it was predicted that PCR

amplification using the *bi-1* G1/C1 and *bi-1* G2/C2 primer combinations would give

products of 896 bp and 766 bp, respectively. (Figure 21B). Both of these predictions were

supported, as PCR performed using *bi-1* mutant genomic DNA as the template yielded

products of the expected size with either the *bi-1* G1/C1 or the *bi-1* G2/C2 primer

combinations (Figure 21D). It was also predicted that PCR performed using the *bi-1* G1/C1 or the *bi-1* G2/C2 primer combinations would not produce a product when template DNA from the WT strain is used. As no product was visible in the WT lane, these predictions were also supported (Figure 21D).

Sequencing results of the cassette-genome junction PCR products from the *bi-1* mutant strain were only able to partially confirm the mapped insertional locus reported for the *bi-1* mutant strain (Figure 21E). While the sequenced *bi-1* G1/C1 product aligned well with the simulated insertion site in the *bi-1* gene, the sequencing reaction failed to obtain a sequence for the *bi-1* G2/C2 products.
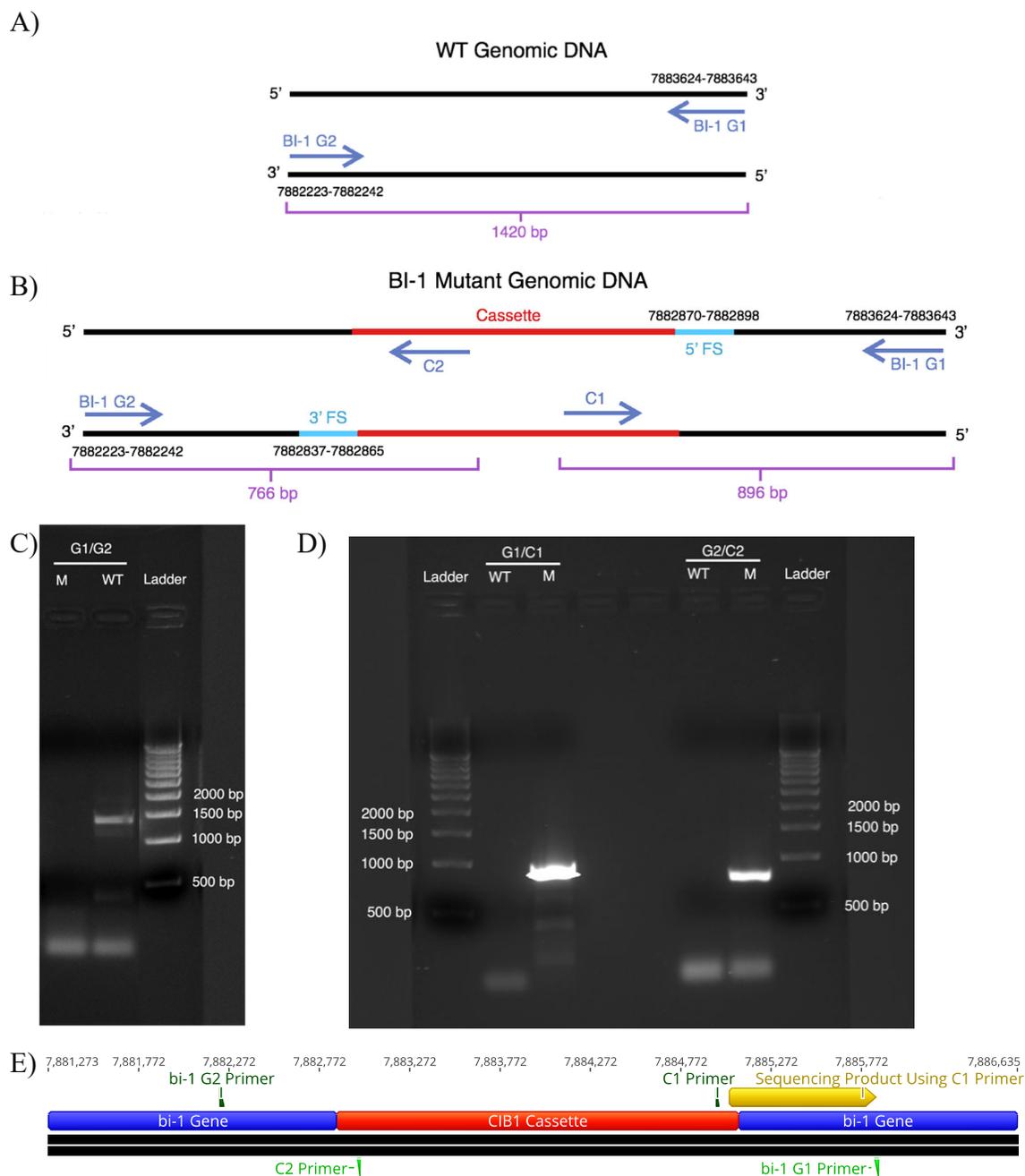
Figure 21. Verification of the *bi-1* mutant strain. A) Predicted binding sites and elongation direction of genomic primers *bi-1* G1 and *bi-1* G2 on WT genomic DNA based on the alignment of the primer sequences with the *C. reinhardtii* genome. These binding sites allowed for the prediction of the amplicon size when PCR is performed using WT genomic DNA. B) The genomic sequences reported to flank either side of the cassette (shown in light blue), the binding sites and elongation direction of all cassette and genomic primers, and the expected amplicon sizes when PCR is performed with *bi-1* G1/C1 and *bi-1* G2/C2 primer combinations with *bi-1* mutant genomic DNA. A 500 bp ladder is included for reference.  C) Results of PCR using the primer combination *bi-1*

G1 and *bi-1* G2 on both *bi-1* mutant (M) and WT genomic DNA. A 500 bp ladder is included for reference. D) Results of PCR using the primer combinations *bi-1* G1/C1 and *bi-1* G2/C2 on both *bi-1* mutant (M) and WT genomic DNA. A 500 bp ladder is also shown for reference. E) Alignment of the sequencing results from the *bi-1* mutant strain cassette-genome junctions with a simulation of the insertion site. The *C. reinhardtii bi-1* gene is shown in blue, the inserted CIB1 cassette is shown in red, the binding sites of both the genomic and the cassette primers are shown in green, and the sequencing product(s) is shown in gold.

***Verification of the ire-1 Mutant Strain*** Based on the alignment of the *ire-1* G1

and *ire-1* G2 primers with the *C. reinhardtii* genome, it was predicted that PCR

amplification of the *ire-1* locus would give a product of 1,457 bp in the WT strain (Figure

22A). Though a product was visible in the WT lane, the weak signal intensity indicates

poor amplification (Figure 22C). Additionally, it was expected that PCR amplification

would not yield a product in the *ire-1* mutant strain. This prediction was also supported,

as no product was visible in the *ire-1* mutant lane (Figure 22C).

It was also predicted that amplification of the cassette-genome junctions using

*ire-1* G1/C1 and *ire-1* G2/C2 would give products of 1,049 bp and 649 bp, respectively,

in the *ire-1* mutant strain (Figure 22B). Both of these predictions were supported, as PCR

performed using *ire-1* mutant genomic DNA as the template yielded products of the

expected size with either the *ire-1* G1/C1 or the *ire-1* G2/C2 primer combinations (Figure

22D). It was also predicted that PCR performed using the *ire-1* G1/C1 or the *ire-1* G2/C2

primer combinations would not produce a product when template DNA from the WT

strain is used, as no cassette would be present for the cassette primers to bind. As no

product was visible in the WT lane, these predictions were also supported (Figure 22D).

Sequencing results of the cassette-genome junction PCR products from the *ire-1*

mutant strain were only able to partially confirm the mapped insertional locus reported

for the *ire-1* mutant strain (Figure 21E). While the sequenced *ire-1* G2/C2 product

aligned well with the simulated insertion site in the *ire-1* gene, the sequencing reaction

failed to obtain a sequence for the *ire-1* G1/C1 products.
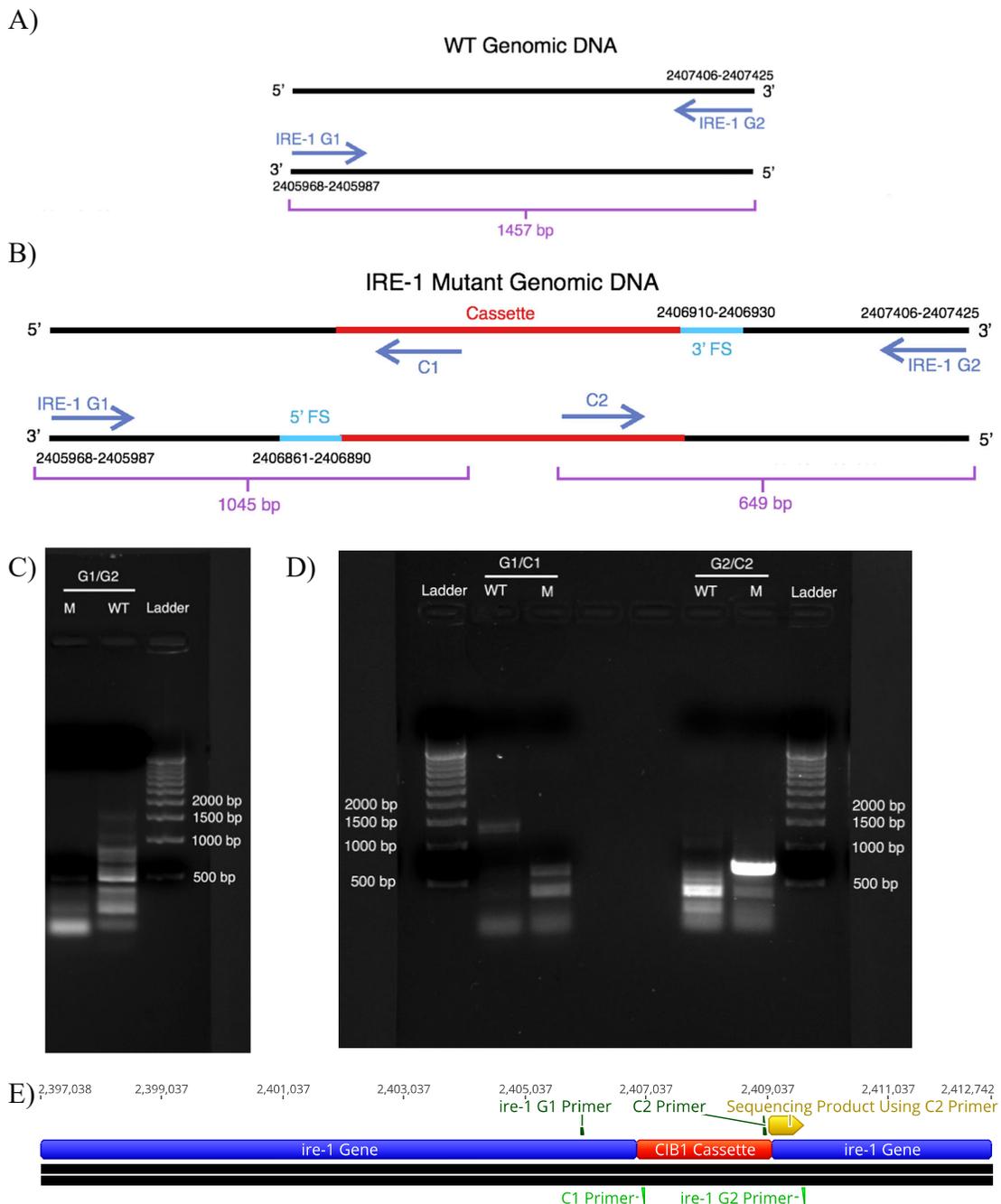


Figure 22. Verification of the *ire-1* mutant strain. A) Predicted binding sites and elongation direction of genomic primers *ire-1* G1 and *ire-1* G2 on WT genomic DNA based on the alignment of the primer sequences with the *C. reinhardtii* genome. These binding sites allowed for the prediction of the amplicon size when PCR is performed using WT genomic DNA. B) The genomic sequences reported to flank either side of the

cassette (shown in light blue), the binding sites and elongation direction of all cassette and genomic primers, and the expected amplicon sizes when PCR is performed with *ire-1* G1/C1 and *ire-1* G2/C2 primer combinations with *ire-1* mutant genomic DNA. A 500 bp ladder is included for reference. C) Results of PCR using the primer combination *ire-1* G1 and *ire-1* G2 on both *ire-1* mutant (M) and WT genomic DNA. A 500 bp ladder is included for reference. D) Results of PCR using the primer combinations *ire-1* G1/C1 and *ire-1* G2/C2 on both *ire-1* mutant (M) and WT genomic DNA. A 500 bp ladder is also shown for reference. E) Alignment of the sequencing results from the *ire-1* mutant strain cassette-genome junctions with a simulation of the insertion site. The *C. reinhardtii ire-1* gene is shown in blue, the inserted CIB1 cassette is shown in red, the binding sites of both the genomic and the cassette primers are shown in green, and the sequencing product(s) is shown in gold.

   ***Verification of the e2f Mutant Strain*** Based on the alignment of the *e2f* G1 and G2 primers with the *C. reinhardtii* genome, it was predicted that PCR amplification of the *e2f* locus would give a product of 1,380 bp in the WT strain (Figure 23A). This prediction was supported, as a product between 1,000 and 1,500 bp was observed (Figure 23C). Additionally, it was expected that PCR amplification would not yield a product in the *e2f* mutant strain. This prediction was also supported, as no product was visible in the *e2f* mutant lane (Figure 23C).

   It was also predicted that amplification of the cassette-genome junctions using *e2f* G1/C1 and *e2f* G2/C2 would give products of 636 bp and 991 bp, respectively, in the *e2f* mutant strain (Figure 23B). Both of these predictions were supported, as PCR performed using *e2f* mutant genomic DNA as the template yielded products of the expected size with either the *e2f* G1/C1 or the *e2f* G2/C2 primer combinations (Figure 23D). It was also predicted that PCR performed using the *e2f* G1/C1 or the *e2f* G2/C2 primer combinations would not produce a product when template DNA from the WT strain is used, as no cassette would be present for the cassette primers to bind. As no product was visible in the WT lane, these predictions were also supported (Figure 23D).

Sequencing results of the cassette-genome junction PCR products from the *e2f* mutant strain were only able to confirm the mapped insertional locus reported for the *e2f* mutant strain (Figure 21E). The sequences obtained from the *e2f* G1/C1 and *e2f* G2/C2 products aligned well with the simulated insertion site in the *e2f* gene.
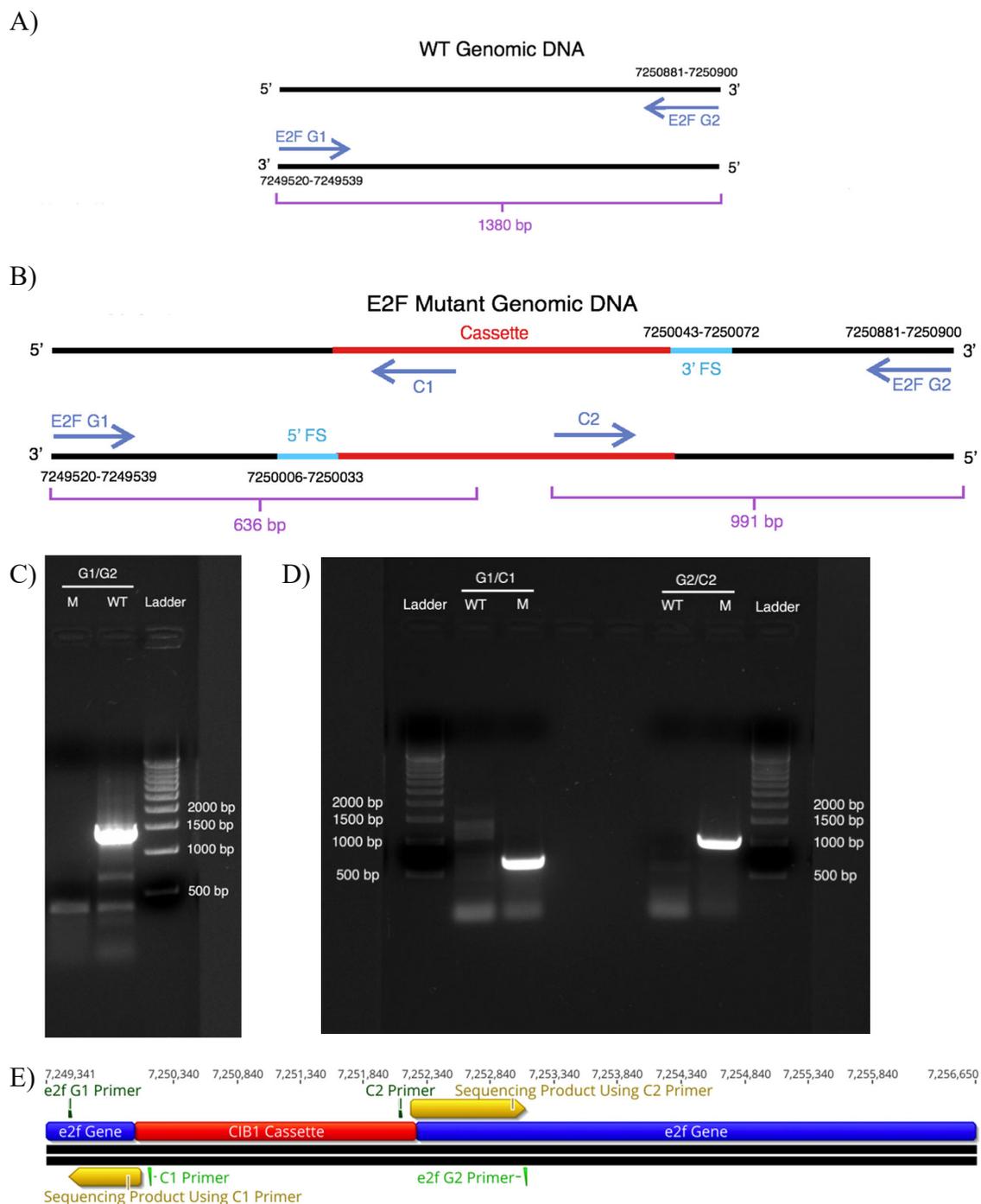
Figure 23. Verification of the *e2f* mutant strain. A) Predicted binding sites and elongation direction of genomic primers *e2f* G1 and *e2f* G2 on WT genomic DNA based on the alignment of the primer sequences with the *C. reinhardtii* genome. These binding sites allowed for the prediction of the amplicon size when PCR is performed using WT genomic DNA. B) The genomic sequences reported to flank either of the cassette (shown in light blue), the binding sites and elongation direction of all cassette and genomic primers, and the expected amplicon sizes when PCR is performed with *e2f* G1/C1 and *e2f* G2/C2 primer combinations with *e2f* mutant genomic DNA. C) Results of PCR using the

primer combination *e2f* G1 and *e2f* G2 on both *e2f* mutant (M) and WT genomic DNA. A 500 bp ladder is included for reference. A 500 bp ladder is also shown for reference. D) Results of PCR using the primer combinations *e2f* G1/C1 and *e2f* G2/C2 on both *e2f* mutant (M) and WT genomic DNA. A 500 bp ladder is also shown for reference. E) Alignment of the sequencing results from the *e2f* mutant strain cassette-genome junctions with a simulation of the insertion site. The *C. reinhardtii e2f* gene is shown in blue, the inserted CIB1 cassette is shown in red, the binding sites of both the genomic and the cassette primers are shown in green, and the sequencing product(s) are shown in gold.

*Verification of the tat-d Mutant Strain* Based on the alignment of the *tat-d* G1

and *tat-d* G2 primers with the *C. reinhardtii* genome, it was predicted that PCR

amplification of the *tat-d* locus would give a product of 1,410 bp in the WT strain (Figure

24A). This prediction was supported, as a product slightly smaller than 1,500 bp was

observed (Figure 24C). Additionally, it was expected that PCR amplification would not

yield a product in the *tat-d* mutant strain. This prediction was also supported, as no

product was visible in the *tat-d* mutant lane (Figure 24C).

It was also predicted that amplification of the cassette-genome junctions using *tat-d* G1/C1 and *tat-d* G2/C2 would give products of 1,082 bp and 577 bp, respectively, in

the *tat-d* mutant strain (Figure 24B). Both of these predictions were supported, as PCR

performed using *tat-d* mutant genomic DNA as the template yielded products of the

expected size with either the *tat-d* G1/C1 or the *tat-d* G2/C2 primer combinations (Figure

24D). It was also predicted that PCR performed using the *tat-d* G1/C1 or the *tat-d* G2/C2

primer combinations would not produce a product when template DNA from the WT

strain is used, as no cassette would be present for the cassette primers to bind. As no

product was visible in the WT lane, these predictions were also supported (Figure 24D).

Sequencing results of the cassette-genome junction PCR products from the *tat-d*

mutant strain were only able to partially confirm the mapped insertional locus reported

for the *tat-d* mutant strain (Figure 21E). While the sequenced *tat-d* G1/C1 product

aligned well with the simulated insertion site in the *tat-d* gene, the sequencing reaction

failed to obtain a sequence for the *tat-d* G2/C2 products.
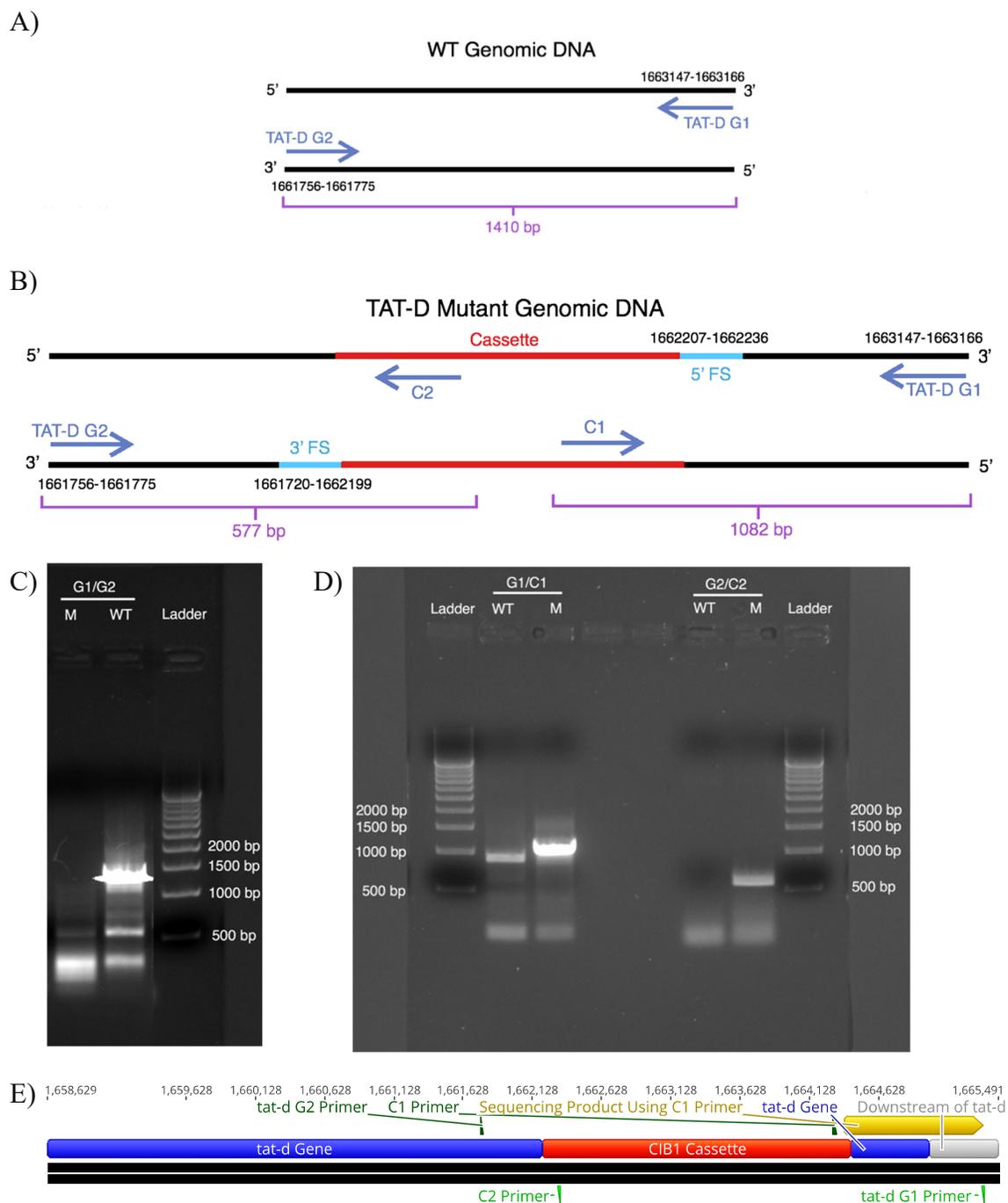
Figure 24. Verification of the *tat-d* mutant strain. A) Predicted binding sites and elongation direction of genomic primers *tat-d* G1 and *tat-d* G2 on WT genomic DNA based on the alignment of the primer sequences with the *C. reinhardtii* genome. These binding sites allowed for the prediction of the amplicon size when PCR is performed using WT genomic DNA. B) The genomic sequences reported to flank either side of the cassette (shown in light blue), the binding sites and elongation direction of all cassette and genomic primers, and the expected amplicon sizes when PCR is performed with *tat-d* G1/C1 and *tat-d* G2/C2 primer combinations with *tat-d* mutant genomic DNA. A 500 bp

ladder is included for reference.  C) Results of PCR using the primer combination *tat-d* G1 and *tat-d* G2 on both *tat-d* mutant (M) and WT genomic DNA. A 500 bp ladder is included for reference. D) Results of PCR using the primer combinations *tat-d* G1/C1 and *tat-d* G2/C2 on both *tat-d* mutant (M) and WT genomic DNA. A 500 bp ladder is also shown for reference. E) Alignment of the sequencing results from the *tat-d* mutant strain cassette-genome junctions with a simulation of the insertion site. The *C. reinhardtii e2f* gene is shown in blue, the inserted CIB1 cassette is shown in red, the binding sites of both the genomic and the cassette primers are shown in green, and the sequencing product(s) are shown in gold.

   ***Verification of the lsd-1 Mutant Strain*** Based on the alignment of the *lsd-1* G1 and *lsd-1* G2 primers with the *C. reinhardtii* genome, it was predicted that PCR amplification of the *lsd-1* locus would give a product of 1,941 bp in the WT strain (Figure 25A). This prediction was supported, as a product slightly smaller than 2,000 bp was observed (Figure 25C). Additionally, it was expected that PCR amplification would not yield a product in the *lsd-1* mutant strain. This prediction was also supported, as no product was visible in the *lsd-1* mutant lane (Figure 25C).

   It was also predicted that amplification of the cassette-genome junctions using *lsd-1* G1/C1 would yield a product of 1,122 bp in the *lsd-1* mutant strain (Figure 25B). This prediction was supported, as PCR performed using *lsd-1* mutant genomic DNA as the template yielded products of the expected size with the *lsd-1* G1/C1 primer combination (Figure 25D). It was also predicted that PCR performed using the *lsd-1* G1/C1 or the *lsd-1* G2/C2 primer combinations would not produce a product when template DNA from the WT strain is used, as no cassette would be present for the cassette primers to bind. No product was visible in the WT lane, indicating that this prediction was also supported (Figure 25D).

   For the amplification of the cassette-genome junctions, the *lsd-1* mutant strain represents a unique instance. For this particular strain, only one of the genomic sequences predicted to flank the cassette was reported (Table 7). This indicates that the CLiP library

were unable to produce the genomic sequence which flanks the other side of the cassette. With regards to the verification of the insertion site in the lsd-1 mutant strain, the lack of a genomic flanking sequence for this side of the cassette renders the prediction of the expected product size problematic. It could be assumed that the CLiP sequencing reaction for this side of the cassette simply failed. If this were the case, the product size of the amplicon could be predicted by assuming that the genomic sequence adjacent to the other flanking sequence flanked this side of the cassette.

Sequencing results of the cassette-genome junction PCR products from the *lsd-1* mutant strain were not able to confirm the presence of the insert in the *lsd-1* mutant strain. (Figure 21E).
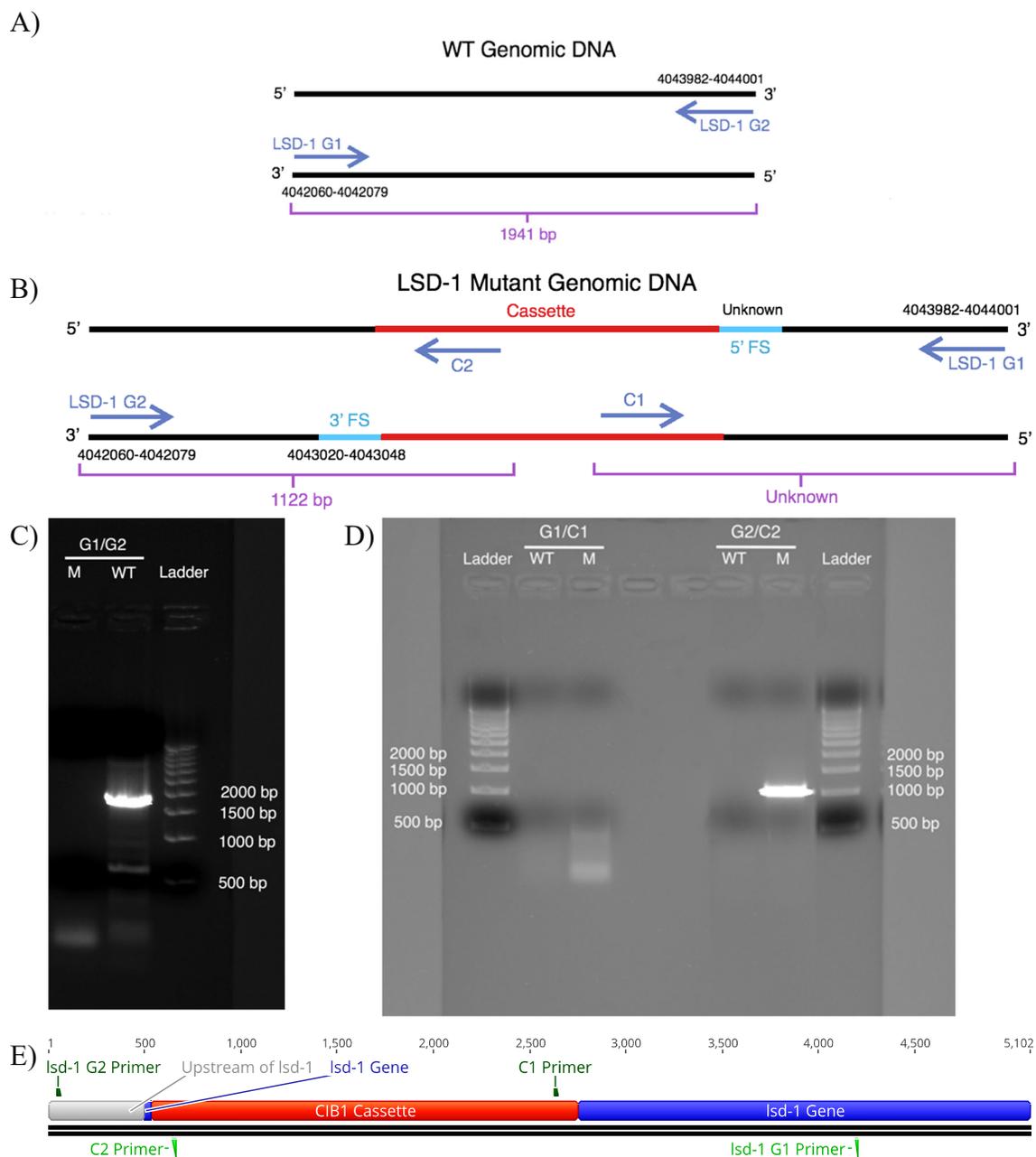
Figure 25. Verification of the *lsd-1* mutant strain. A) Predicted binding sites and elongation direction of genomic primers *lsd-1* G1 and *lsd-1* G2 on WT genomic DNA based on the alignment of the primer sequences with the *C. reinhardtii* genome. These binding sites allowed for the prediction of the amplicon size when PCR is performed using WT genomic DNA. B) The genomic sequences reported to flank either side of the cassette (shown in light blue), the binding sites and elongation direction of all cassette and genomic primers, and the expected amplicon sizes when PCR is performed with *lsd-1* G1/C1 and *lsd-1* G2/C2 primer combinations with *lsd-1* mutant genomic DNA. C) Results of PCR using the primer combination *lsd-1* G1 and *lsd-1* G2 on both *lsd-1* mutant (M) and WT genomic DNA. A 500 bp ladder is included for reference. D) Results of PCR using the primer combinations *lsd-1* G1/C1 and *lsd-1* G2/C2 on both *lsd-1* mutant

(M) and WT genomic DNA. A 500 bp ladder is also shown for reference. E) Alignment of the sequencing results from the *tat-d* mutant strain cassette-genome junctions with a simulation of the insertion site. The *C. reinhardtii e2f* gene is shown in blue, the inserted CIB1 cassette is shown in red, the binding sites of both the genomic and the cassette primers are shown in green, and the sequencing product(s) are shown in gold.

***Verification of the pig3 Mutant Strain*** Based on the alignment of the *pig3* G1 and *pig3* G2 primers with the *C. reinhardtii* genome, it was predicted that PCR amplification of the *pig3* locus would give a product of 1,517 bp in the WT strain (Figure 26A). This prediction was supported, as a product slightly smaller than 1,500 bp was observed (Figure 26C). Additionally, it was expected that PCR amplification would not yield a product in the *pig3* mutant strain. This prediction was also supported, as no product was visible in the *pig3* mutant lane (Figure 26C).

It was also predicted that amplification of the cassette-genome junctions using *pig3* G1/C1 and *pig3* G2/C2 would give products of 846 bp and 880 bp, respectively, in the *pig3* mutant strain (Figure 26B). Both of these predictions were supported, as PCR performed using *pig3* mutant genomic DNA as the template yielded products of the expected size with either the *pig3* G1/C1 or the *pig3* G2/C2 primer combinations (Figure 26D). It was also predicted that PCR performed using the *pig3* G1/C1 or the *pig3* G2/C2 primer combinations would not produce a product when template DNA from the WT strain is used, as no cassette would be present for the cassette primers to bind. As no product was visible in the WT lane, these predictions were also supported (Figure 26D).

Sequencing results of the cassette-genome junction PCR products from the *pig3* mutant strain were only able to partially confirm the mapped insertional locus reported for the *pig3* mutant strain (Figure 21E). While the sequenced *pig3* G2/C2 product aligned

well with the simulated insertion site in the *pig3* gene, the sequencing reaction failed to

obtain a sequence for the *pig3* G1/C1 products.
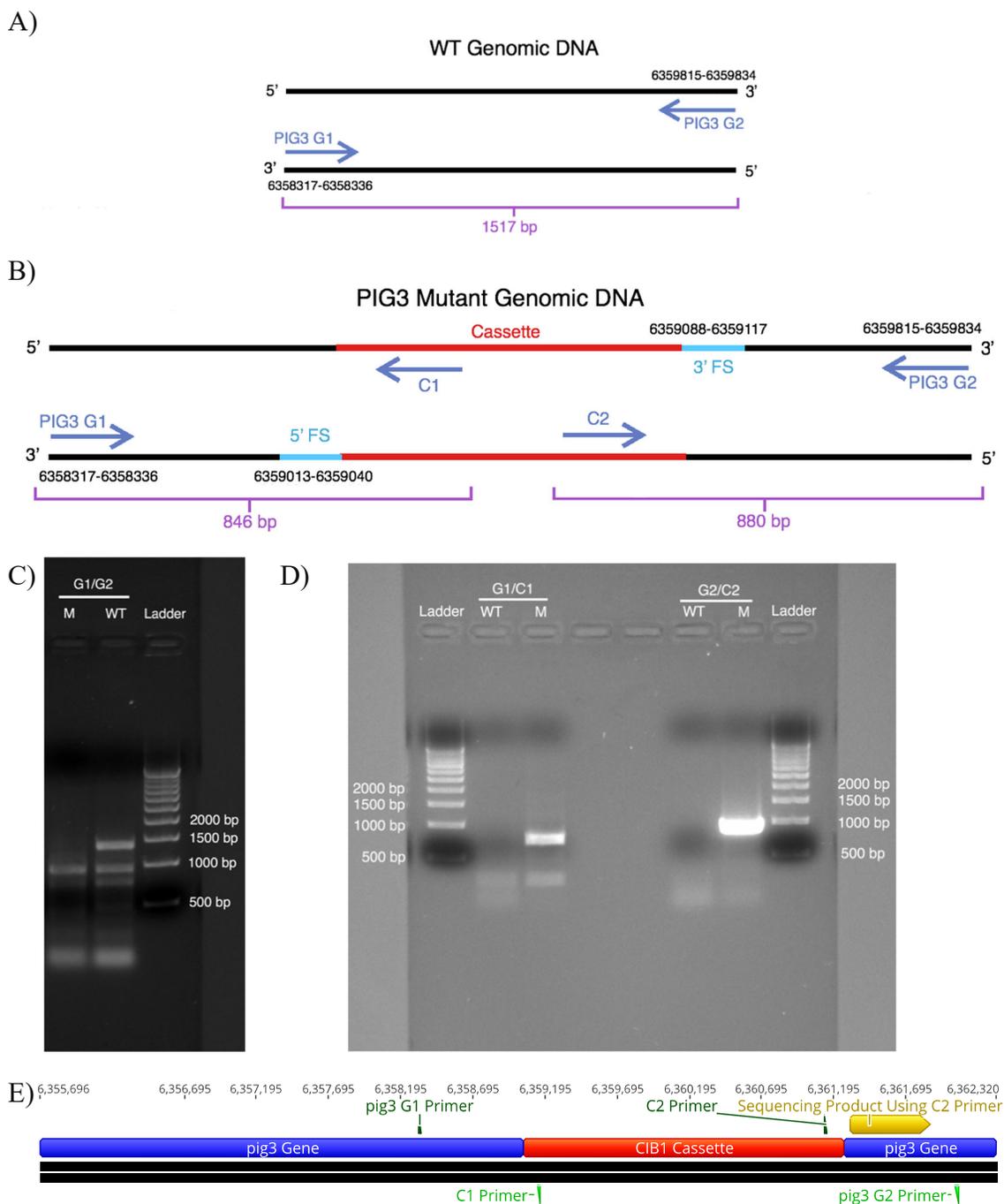


Figure 26. Verification of the *pig3* mutant strain. A) Predicted binding sites and elongation direction of genomic primers *pig3* G1 and *pig3* G2 on WT genomic DNA based on the alignment of the primer sequences with the *C. reinhardtii* genome. These

binding sites allowed for the prediction of the amplicon size when PCR is performed using WT genomic DNA. B) The genomic sequences reported to flank either side of the cassette (shown in light blue), the binding sites and elongation direction of all cassette and genomic primers, and the expected amplicon sizes when PCR is performed with *pig3* G1/C1 and *pig3* G2/C2 primer combinations with *pig3* mutant genomic DNA. C) Results of PCR using the primer combination *pig3* G1 and *pig3* G2 on both *pig3* mutant (M) and WT genomic DNA. A 500 bp ladder is included for reference. D) Results of PCR using the primer combinations *pig3* G1/C1 and *pig3* G2/C2 on both *pig3* mutant (M) and WT genomic DNA. A 500 bp ladder is also shown for reference. E) Alignment of the sequencing results from the *tat-d* mutant strain cassette-genome junctions with a simulation of the insertion site. The *C. reinhardtii e2f* gene is shown in blue, the inserted CIB1 cassette is shown in red, the binding sites of both the genomic and the cassette primers are shown in green, and the sequencing product(s) are shown in gold.

**Summary**

In the present work, strains of *C. reinhardtii*, each with an insertional cassette disrupting a selected protein-coding gene, were selected for further study regarding a predicted role in *C. reinhardtii* PCD. To validate the insertion sites mapped by the CLiP library, a three-fold, PCR-based approach was used. First, genomic primers designed to bind the genome ~1,000 bp upstream and downstream of the insertion site were used to amplify each putative insertional locus in both the corresponding mutant strain and the WT strain. Given that the size of the cassette should prohibit the PCR reaction from completing, it was predicted that a PCR product of the expected size would be present in the WT strain, but not in the mutant strain.

Second, the two cassette-genome junctions in each of the mutant strains were amplified by PCR. Each junction was amplified using one of the cassette primers and its corresponding genomic primer. It was expected that these reactions would yield products of the expected sizes in the mutant strains, but not in the WT strain, as no cassette would be present for the cassette primers to bind.

Finally, to confirm that the PCR products from the amplification of the cassette-genome junctions were sent off for sequencing to confirm that the observed products were indeed from the cassette genome-genome junctions. To confirm that the sequenced products from the G1/C1 and G2/C2 PCR reactions were of the cassette-genome junctions, a simulation of the insertion was created for each of the strains by inserting the CIB1 cassette nucleotide sequence, in its determined orientation, into the insertion site reported by the CLiP library. For each mutant strain, the sequences obtained from PCR amplification of the cassette-genome junctions, if any, were aligned with the simulated insertion site.

Table 10.

Summary of results obtained during verification of mutant strains.

| | Amplify Insertion Site with G1/G2 primers? | | Amplify Cassette-Genome Junctions? | | | | Sequencing Results Confirm Insertion? | |
| | | | G1/C1 Primers | | G2/C2 Primers | | | |
| | WT Strain | Mutant Strain | WT Strain | Mutant Strain | WT Strain | Mutant Strain | G1/C1 Product | G2/C2 Product |
|---|---|---|---|---|---|---|---|---|
| *bi-1* | Yes | No | No | Yes | No | Yes | Yes | Failed |
| *ire-1* | No | No | No | Yes | No | No | Failed | Yes |
| *e2f* | Yes | No | No | Yes | No | Yes | Yes | Yes |
| *tat-d* | Yes | No | No | Yes | No | Yes | Yes | Failed |
| *lsd-1* | Yes | No | No | No | No | Yes | Failed | Failed |
| *pig3* | Yes | No | No | Yes | No | Yes | Failed | Yes |

NOTE: "Failed" denotes instances where no sequence was obtained from the sequencing reaction of the cassette-genome PCR product.

**CHAPTER IV**

**Assessment of PCD Phenotypes in Mutant Strains of *C. reinhardtii***

**Introduction**

      **Characteristics of PCD** As discussed in Chapter 1, cells which have undergone PCD, or are in the process of undergoing PCD, exhibit a number of characteristics which distinguish them from non-PCD cells.

      *Metabolic Activity* The cessation of life is a rudimentary way by which death may be defined. Thus, the death of the cell is accompanied by the loss of features that are characteristic of living cells. As such, a classic method of distinguishing between live and dead cells is to assay for the presence of metabolic activity. In general, assays that test for the presence of metabolic activity utilize a probe that is cleaved by an enzyme, or class of enzymes, that can be used to indicate metabolic activity.[263] The probe is designed such that, when cleaved by the target enzyme(s), the resulting product is detectable by either fluorescence or colorimetric methods.[95]

      *Plasma Membrane Integrity* The plasma membrane (PM) is a universal feature of prokaryotic and eukaryotic cells. The PM functions as a physical barrier to separate the intracellular and extracellular environments. The PM is said to be intact when this barrier is preserved. The maintenance of PM integrity requires the input of energy. During PCD, the cessation of metabolic activity results in the termination of metabolic processes, which, in turn, causes the loss of PM integrity. Once compromised, the integrity of the PM cannot be restored, so this feature is useful as an indicator of cell death. Methods for detecting PM integrity generally rely on the use of exclusionary dyes that are not retained by cells possessing an intact plasma membrane. Such probes are generally either

colorimetric or fluorescent, and the intracellular accumulation of the probe can be determined using the appropriate detection method. In the case of fluorescent probes, the dye often will fluoresce under specific conditions. For example, the most often used indicator of PM integrity, propidium iodide, is a DNA intercalating agent that emits a signal that is enhanced significantly when bound to DNA.

*Phosphatidylserine Externalization* Phospholipids are a primary component of the plasma membrane. Under normal cellular conditions, PM phospholipids are distributed asymmetrically across the lipid bilayer; phosphatidylserine (PS) and phosphatidylethanolamine (PE) reside within the inner leaflet of the membrane, while phosphatidylcholine (PC) and sphingomyelin (SM) exist primarily in the outer leaflet of the PM.[264,265] Phospholipids are translocated from one layer of the PM to the other by enzymes known as phospholipid translocases. Phospholipid translocases come in three varieties: scramblases non-specifically and bi-directionally move phospholipids between the lipid layers, flippases specifically transport PS and PE from the outer leaflet to the inner leaflet, and floppases are believed to move PC from the inner to the outer leaflet of the PM.[266,267] The asymmetric nature of the phospholipid distribution across the PM is maintained primarily by the antagonistic actions of flippases/floppases and scramblases.[267]

Many instances of PCD are accompanied by the externalization of PS.[268–271] As such, PS externalization has long been used as a marker of PCD.[272] PS exposure serves to facilitate the elimination of dead or dying cells via macrophages, which possess PS receptors that facilitate the recognition of apoptotic cells.[48] The binding of PS to the PS receptor causes the engulfment and destruction of the apoptotic cell by the macrophage.[48]

The molecular mechanisms that lead to the externalization of PS have only recently begun to be explored in detail. As such, the signaling pathways that underlie PS externalization remain largely uncharacterized.[273] However, recent evidence indicates that PS exposure during PCD in animal systems may be result of the direct destruction of flippases and the activation of scramblases by activated caspase-3.[267]

***Accumulation of Reactive Oxygen Species*** The accumulation of reactive oxygen species (ROS) is a characteristic of PCD that is observed in all branches of life.[274,275] Small quantities of ROS are produced in cells as byproducts of normal metabolic processes, such as the production of ATP. Aside from being a byproduct of cellular metabolism, ROS also play important roles in various signaling pathways, such as those involved with transduction, phagocytosis and cell growth. However, as the name implies, ROS exhibit a tendency to react with a number of intracellular biomolecules, including proteins, lipids, and nucleic acids.[276–278] As the alteration of these biomolecules results in massive, and sometimes irreversible, cellular damage, elevated levels of ROS are highly deleterious to the cell. Cells utilize two primary strategies to circumvent the detrimental effects associated with high concentrations of ROS. First, intracellular ROS levels are tightly controlled via the synthesis of antioxidant enzymes, which serve to counteract the harmful effects of ROS.[279] Secondly, cellular processes that generate large amounts of ROS are localized to specific organelles, such as the mitochondrion, chloroplast, and endoplasmic reticulum, which function as sites of containment for ROS.[280]

The accumulation of ROS is a key attribute of PCD, and the accrual of these damaging compounds is, in part, what contributes to the destruction of the intracellular components during PCD.[280] PCD-associated accumulation of ROS occurs by two primary

mechanisms: the repression of antioxidant enzyme activities, and the release of ROS

from localized intracellular regions. In light of the fact that accumulation of ROS is a

universally-observed characteristic of PCD, this feature has been monitored extensively

in PCD studies.

*Colony Growth* The process of cell division requires energetic input. The

termination of metabolic activity at the onset of cell death results in the absence of energy

sources required for cell division. Consequently, the inability to divide is one way by

which cell death may be assessed.[95] Given that individual cells are not visible to the

naked eye, the amount of observable growth on solid medium is an approach that is

commonly used to assess cell death. A number of different techniques can be used to

assess cell viability by this approach, including the quantitation of colony forming units

or, in instances where individual colonies cannot be distinguished, the assessment of

growth density.[105,281–285]

*DNA Laddering* Fragmentation of genomic DNA is a hallmark characteristic of

PCD. DNA fragmentation is achieved through the activities of a number of different

nucleases.[286] Perhaps the most well-known nuclease that participates in the apoptotic

process is caspase-activated DNase (CAD). Under normal conditions, CAD is inhibited

via the binding of its repressor, inhibitor of CAD (iCAD).[286] Initiation of the apoptotic

program is accompanied by the activation of caspases -3 and -7, which target and cleave

iCAD. Following the degradation of its inhibitor, activated CAD promotes the

degradation of genomic DNA.[286] CAD-mediated DNA fragmentation is a progressive

phenotype that proceeds through two primary stages. First, chromosomal DNA is

fragmented into larger (50-100 kb) fragments.[286] Following this initial degradation, DNA

is cleaved at inter-nucleosomal regions, resulting in the formation of smaller fragments, approximately 180 bp in size.[287] Though CAD is the most-characterized nuclease associated with apoptosis, other nucleases that participate in the fragmentation of genomic DNA during PCD are endonuclease G, DNase II, DNase γ, and AIF.[288,288]

A classic technique for visualization of the early stages of DNA fragmentation during apoptosis is the DNA laddering assay, in which extracted genomic DNA is run on an agarose gel.[287] Following electrophoretic separation, distinct bands, a result of genomic cleavage between nucleosomes, are visible on the gel. The fragmentation pattern resembles the rungs of a ladder when visualized in this manner and, as such, is referred to as DNA "laddering".

**Characteristics of PCD in *C. reinhardtii*** When subjected to PCD-inducing conditions, *C. reinhardtii* cells exhibit many of the same characteristics observed in animal cells that are undergoing PCD. Despite the presence of phenotypic commonalities between animal PCD and *C. reinhardtii* PCD, the detailed mechanisms by which *C. reinhardtii* initiates and executes PCD have yet to be explored in detail.

In previous work, we utilized bioinformatics methods to predict *C. reinhardtii* proteins that participate in PCD on a large scale. We then selected several predicted *C. reinhardtii* gene products that, based on homology to known PCD proteins, we hypothesized to participate in *C. reinhardtii* PCD, either as facilitators or inhibitors (Table 6). We then selected several *C. reinhardtii* mutant strains from a library generated by random insertional mutagenesis. For each of the chosen mutant strains, the CLiP library reported that the insertion site of the cassette was located within the coding region

of one of the selected protein-coding genes. To validate this, we confirmed the reported insertion site in each mutant strain using PCR-based methods.

The purpose of this study is to utilize a reverse genetics approach to explore the potential roles(s) that the proteins of interest may play during *C. reinhardtii* PCD. Using time-course methodologies, we examined PCD phenotypes exhibited by each of the mutant strains and compared the results to those observed in the parent WT strain. We expected that, if a selected *C. reinhardtii* protein contributes to PCD, the disruption of the protein-coding gene would result in an altered PCD program in the mutant strain. Furthermore, we predicted that disruption of the PCD program may provide insight into the molecular role of the gene product during heat-induced PCD in *C. reinhardtii*.

**Materials and Methods**

**Culture Conditions** All *C. reinhardtii* cultures were maintained at a constant temperature of 25°C in an illuminated incubator set to a 14/10 hour light/dark cycle. Stock cultures were maintained on slants of solid TAP medium ($20.0 \times 10^{-3}$ M Tris, $7.00 \times 10^{-3}$ M $NH_4Cl$, $8.30 \times 10^{-4}$ M $MgSO_4 \cdot 7H_2O$, $4.50 \times 10^{-4}$ M $CaCl_2 \cdot 2H_2O$, $1.65 \times 10^{-3}$ M $K_2HPO_4$, $1.05 \times 10^{-3}$ M $KH_2PO_4$, $1.34 \times 10^{-4}$ M $Na_2EDTA \cdot 2H_2O$, $1.36 \times 10^{-4}$ M $ZnSO_4 \cdot 7H_2O$, $1.84 \times 10^{-4}$ M $H_3BO_3$, $4.00 \times 10^{-5}$ M $MnCl_2 \cdot 4H_2O$, $3.29 \times 10^{-5}$ M $FeSO_4 \cdot 7H_2O$, $1.23 \times 10^{-5}$ M $CoCl_2 \cdot 6H_2O$, $1.00 \times 10^{-5}$ M $CoCl_2 \cdot 6H_2O$, $4.44 \times 10^{-6}$ M $(NH_4)_6MoO_3$, 0.001% (v/v) $CH_3COOH$, 1.7% (w/v) agar). In order to maintain a sufficient number of cells to use for experiments, subcultures of the slant stocks were maintained on solid TAP plates. All subcultures were passaged a maximum of three times, after which a new subculture was streaked from the stock slant. Liquid cultures

were inoculated using cells obtained from a subcultured plate and gently agitated by constant aeration for 2-4 days prior to use.

**Analysis of Mutant Strains Under Normal Conditions** To identify strains with abnormalities which could interfere with the subsequent PCD experiments, unstressed cells were analyzed for irregularities in morphology and growth rate.

*Morphological Analysis* Liquid cultures of all strains were inoculated in R medium (1.6 x $10^{-4}$ M $H_3BO_3$, 3.5 x $10^{-5}$ M $ZnSO_4 \bullet 7H_2O$, 1.8 x $10^{-6}$ M $MnSO_4 \cdot H_2O$, 8.4 x $10^{-7}$ M $CoCl_2 \cdot 6H_2O$, 8.3 x $10^{-7}$ M, $Na_2MoO_4 \cdot 2H_2O$, 2.8 x $10^{-7}$ M $CuSO_4 \cdot 5H_2O$, 1.7 x $10^{-3}$ M Na Citrate $\cdot$ $2H_2O$, 3.7 x $10^{-5}$ M $FeCl_3 \cdot 6H_2O$, 3.6 x $10^{-4}$ M $CaCl_2 \cdot 2H_2O$, 1.2 x $10^{-3}$ M $MgSO_4 \cdot 7H_2O$, 3.7 x $10^{-3}$ M $NH_4NO_3$, 1.5 x $10^{-3}$ M $KH_2PO_4$, pH to 6.8 with 10% $K_2HPO_4 \cdot 3H_2O$, and 0.022 M $CH_3COONa$) and allowed to incubate under normal culturing conditions, as described above. After two days, samples of each strain were taken. To observe the morphologies of individual, unstressed cells, the samples were visualized using phase-contrast microscopy at 400X magnification. Additionally, the cells were qualitatively analyzed for differences in size, shape, and motility relative to the parent WT strain.

*Growth Curve Analysis* In order to compare the rate of growth between the mutant strains and the parent WT strain, 500 mL Erlenmeyer flasks containing 300 mL of liquid R medium were inoculated with cells taken from a plate subculture. Once a day, at the time of inoculation, a 500 μL aliquot was removed from each culture. Lugol's iodine was added to each of the samples at a 1:1 ratio and pipetted to mix. Hemacytometer counts were performed to determine the densities of the cultures at each time point.

**Analysis of Mutant Strains Under PCD-inducing Conditions**

*Induction of PCD by Heat Stress* Several days prior to the beginning of an experiment, liquid cultures of all strains to be used were started in a 250 mL Erlenmeyer flask containing 125 mL of liquid R medium. The flasks were allowed to incubate under normal culturing conditions, as described above. After several days, the cell density of each culture was estimated by hemacytometer counts. The volume of each culture needed to produce 5 mL of culture at a density of $4.0 \times 10^6$ cells/mL was calculated, and the appropriate volume of culture was aliquoted into a 15 mL conical tube. Cells were pelleted by centrifugation at 3,200 x g for 5 minutes. The supernatant was immediately decanted, and fresh R medium was added to each tube to the 5 mL mark. To resuspend the pellets, and also to allow the cells to acclimate to the new conditions, the tubes were allowed to incubate on a rotating shaker for 30-60 minutes on the benchtop under ambient light. Following this, 1 mL aliquots were placed into 1.5 mL microcentrifuge tubes.

To induce PCD, heat stress was applied to the experimental groups by submergence in a 42°C water bath for two hours. The control groups were left to incubate on the benchtop for the same period of time. In some cases, sampling was conducted after completion of the heat stress. In these situations, the heat stress tubes were removed from the water bath and placed with the control tubes until sampling was completed. The overhead lighting in the room was turned off and all cultures were illuminated by an overhead table lamp equipped with a 14W (60W equivalent) Soft White fluorescent bulb (Great Value). At the times specified for each experiment, samples were removed from all groups.

***Imaging*** All images from the microscopy assays were obtained using an AMScope MU140-CCD camera. For each sample, 5 µL of cells was loaded onto a hemacytometer. The entire grid and the surrounding area was visualized with a Nikon Eclipse E400 microscope at 40X magnification. Images of this field of view were captured using brightfield and epifluorescence microscopy using the AMLite software (AMScope, version 1.0.8718).

***Metabolic Activity Assay*** To identify cells with metabolic activity, the fluorescent dye fluorescein diacetate (FDA) was utilized to measure esterase activity. A heat stress was applied to cells as described above. Samples were taken at 30 minute intervals for 3 hours. At each sampling, 1 µL of 2.4 mM FDA was added to 100 µL of cells. The tubes were inverted several times to mix, and allowed to incubate in the dark at room temperature for 1 minute. Cells were loaded onto a hemacytometer, visualized, and imaged as described above.

***Plasma Membrane Integrity Assay*** To identify cells that had lost plasma membrane integrity, the fluorescent dye SYTOX Green (Life Technologies) was utilized.

A heat stress was carried out as described above. Samples were taken at 30 minute intervals for 3 hours. At each sampling, 1 µL SYTOX Green was added to 100 µL of culture. The tubes were allowed to incubate in the dark at room temperature for 2 minutes. Following the incubation period, cells were loaded onto a hemacytometer, visualized, and imaged as described above.

***Phosphatidylserine Externalization Assay*** To identify cells displaying externalized phosphatidylserine, the Alexa Fluor AnnexinV/Dead Cell Apoptosis Kit (Thermo Fisher) was utilized.

A heat stress was carried out as described above. Samples were taken at 60 minute intervals for 4 hours. During sampling, 100 μL of cells from each culture were aliquoted into 1.5 mL microcentrifuge tubes containing 100 μL ice-cold PBS. The samples were centrifuged at 16,100 x g for 30 seconds, and the supernatant was aspirated. The cells were then resuspended in 100 μL 1X AnnexinV binding buffer, and 25 μL from each tube was aliquoted into a fresh 1.5 mL microcentrifuge tube. To this, 2.5 μL of AnnexinV was added. The samples were allowed to incubate at room temperature for 15 minutes, whilst covered in aluminum foil to protect the cells from ambient light. Following the incubation period, cells were loaded onto a hemacytometer, visualized, and imaged as described above.

*ROS Accumulation Assay* To identify cells which had accumulated reactive oxygen species, the fluorescent probe 5-(and 6)-chloromethyl-2'7-dichlorodihydrofluorescein diacetate acetyl ester (CM-H$_2$CFDA; Thermo-Fisher) was used. Prior to the induction of PCD, 10 μL of CM-H2DCFDA was added to each culture. A heat stress was carried out, as described above. Samples were taken at 30 minute intervals for 2 hours. At the time of sampling, 20 μL of cells were removed from each tube and placed into a 1.5 μL microcentrifuge tube containing 1 mL of fresh R medium. The cells were centrifuged at 16,100 x g for 30 seconds, and the supernatant was aspirated until approximately 20 μL of liquid remained, along with the pellet. The cells were gently resuspended in the remaining liquid. From this, cells were loaded onto a hemacytometer and imaged as described above.

*Plating Assay* To assess the growth capabilities of each strain throughout the PCD process, cells were plated at different times. A heat stress was performed, as described

above. Samples were taken at 30 minute intervals for 3 hours. During sampling, 75 μL of each culture was taken, added to a sterile 1.5 mL microcentrifuge tube containing 75 μL fresh medium, and inverted several times to mix. The 150 μL suspension was plated onto solid TAP medium. The plates were allowed to incubate for 3 weeks. Following this, the plates were individually imaged using the ChemiDoc XRS+ (Bio-Rad) imaging system, which had been outfitted with the VersaDoc conversion plate (Bio-Rad).

*DNA Laddering Assay* To qualitatively determine the presence and extent of DNA laddering, genomic DNA was extracted using the DNeasy Plant Mini Kit (Qiagen) at designated time points throughout a heat stress. DNA laddering was visualized simultaneously in all samples using agarose gel electrophoresis.

A heat stress was carried out as described above, with some modifications.[180] It was previously determined that 20 mL of stationary phase culture is necessary to obtain quantities of DNA sufficient for the visualization of laddering on an agarose gel (data not shown). Additionally, a preliminary examination of the modified DNeasy protocol indicated that between forty-five and sixty minutes should be allotted to complete a single extraction from start to finish. The lengthy nature of this process added a complicating factor to the timeframe of the experiment, as samples were to be collected every thirty minutes. As such, the experiment was designed such that 1) each strain would be sampled at thirty minute intervals for the duration of the two hour heat stress, and 2) all samples could be collected and processed simultaneously.

Each culture was resuspended to 100 mL at a density of $1.0 \times 10^7$ cells/mL in fresh R medium. The resuspended cultures were distributed evenly into five 50 mL conical tubes, such that the culture volume in each tube was 20 mL. Each of the five

conical tubes was labeled with the time that the cells were to be heat-stressed: 0, 30, 60, 90, and 120 minutes. A timer was set for two hours, and the samples to be heat stressed for two hours were submerged in a 42°C water bath. At thirty minute increments, the tubes to be subjected to a heat-stress for the amount of time remaining on the timer were submerged in the hot water bath (e.g. samples to be heat-stressed for 90 minutes were submerged 30 minutes after the first samples were submerged, etc.). After two hours, all tubes were removed from the water bath.

Following the heat stress, genomic DNA was extracted from each of the samples using the Plant DNeasy kit according the manufacturer's instructions, with some modifications. The samples were centrifuged at 3,200 x g for 5 minutes, and the supernatant was decanted. Each pellet was resuspended in 400 µL of API buffer, containing 4 µL 100mg/mL RNase A, 20 µL 10% SDS, and 4 mg Proteinase K. The cells were gently pipetted to mix and transferred to a 15 mL conical tube. The cells were placed into a heating block set to 65°C for 10 minutes. Subsequently, 130 µL of Buffer P3 was added, and the tube was inverted to mix. The cells were placed in an ice bath for 5 minutes, then centrifuged for 5 minutes at 13,400 x g. The supernatant was transferred into the provided column within a 2 mL collection tube. After centrifugation at 16,100 x g for two minutes, the supernatant of the flow through was transferred to a fresh 1.5 mL microcentrifuge tube. The volume of the sample was determined, and 1.5 volumes Buffer AW1 were added. Next, 650 µL was transferred into the DNeasy Mini spin column within a 2 mL collection tube and centrifuged at 5,900 x g for one minute. This was repeated until all volume from the Buffer AW1 step was depleted. The column was transferred into a fresh collection tube and 500 µL of Buffer AW2 was added to the

column. The apparatus was centrifuged at 16,100 x g for two minutes. The column was transferred to a fresh 1.5 mL microcentrifuge tube. Fifty microliters of Buffer AE was pipetted onto the membrane of the column. After allowing to incubate on the benchtop for 5 minutes, the apparatus was centrifuged at 5,900 x g for 1 minute. To extract any remaining DNA from the column, another 50 μL Buffer AE was added to the membrane, allowed to incubate for 5 minutes, and centrifuged at 5,900 x g for 1 minute. The flow-through, containing the purified genomic DNA, was stored at -20°C when not in use.

The concentration of DNA in each sample was determined using a Qubit 2.0 (Thermo-Fisher), in conjunction with the Qubit dsDNA High Sensitivity (HS) Assay Kit according to the manufacturer's instructions. The 200X dsDNA HS Reagent was diluted to the working concentration using dsDNA HS buffer at a 1:200 ratio. A two-point calibration curve was made using the standards included in the dsDNA HS kit; ten microliters of each standard was diluted into 190 μL of 1X dsDNA HS Reagent to calibrate the standard curve for samples between 0 and 500 ng/mL For each sample of *C. reinhardtii* genomic DNA from the heat stress experiment, 5 μL of sample was diluted into 195 μL of 1X dsDNA HS Reagent, and the concentration of dsDNA was determined using the Qubit 2.0.

Following the quantification of dsDNA in each of the samples, a 1% agarose gel was prepared using 1X TAE and 0.00002% 10,000X SybrSafe (Invitrogen). To accurately assess the extent of DNA laddering, the quantity of DNA loaded into each of the wells needed be the same. In order to load the gel with the maximum amount of DNA possible, the sample with the lowest DNA concentration was identified. The maximum volume of the wells was predetermined to be approximately 30 μL. Thus, the quantity of

DNA from each sample that was added to the gel was equivalent to the quantity of DNA present in 30 µL of the least concentrated sample.

**_Post-Acquisition Processing and Analysis of Images_** All image analysis was performed using ImageJ. Any image processing was carried out in either ImageJ or Adobe Photoshop Lightroom 5.

_Microscopy Assays_ For each sample, a brightfield and fluorescence image of the same field of view was captured. The following steps were repeated for all image pairs. First, the brightfield and fluorescent images were non-destructively overlaid and cropped to include only the hemacytometer grid. The two resulting cropped images were saved as separate files. The number of particles in the cropped brightfield and fluorescent images were determined automatically using custom macros. The macros were written such that, for each image that was automatically scored, another image was created, in which the particles that were scored were overlaid with an outline. This allowed for the visualization of the particles that were scored by the automated counts. Each image was then manually verified to ensure that the automated scoring of the fluorescent and brightfield images was accurate. To validate the automated counts of the brightfield images, the brightfield image overlaid with particle outlines was examined. Cells which were not recognized as particles were scored, and this number was added to the automated count for the brightfield image. Non-cell particles which were scored were also identified, and this number was subtracted from the automated count for that image.

To validate the automated counts of the fluorescence field of view, an additional step was added. Because some images contained a fair amount of background fluorescence, it was difficult to determine by looking at a fluorescent image alone if a

fluorescent spot was originating from a cell, or was merely background. To address this issue, the automatically-scored fluorescent image, overlaid with the particle outlines, was overlaid with the brightfield image. This allowed for the visualization and manual scoring of 1) fluorescent spots originating from a cell, but which had not been scored by the automated methods, and 2) background fluorescent spots which had been scored as a fluorescent cell. Both of these instances were scored, and either added or subtracted from the automated particle counts for the fluorescent image.

For each pair of brightfield and fluorescent images, the corrected counts were then used to determine the percentage of fluorescent cells in each field of view.

*Plating Assay* An automated macro was written to determine the relative area that was occupied by visible growth on each plate after three weeks of growth. On plates where contamination was present, the contaminant colonies were removed, if possible, using the Spot Removal tool in Adobe Photoshop Lightroom 5. This method of removal was not possible in all instances of contamination, however, as contaminant growth overlapped with *C. reinhardtii* growth on some of the plates.

*DNA Laddering Assay* The gel image from the DNA laddering assay was altered in order to better visualize the bands. For the sake of consistency, all transformations were applied to the entire image rather than to specific portions of the gel. First, the black and white pixels of the gel were inverted in ImageLab (v5.2). The image was then exported as a TIFF file, which was subsequently imported into Adobe Photoshop Lightroom 5. The exposure, contrast, highlights, shadows, whites, blacks, and clarity settings were altered across the entire image until the DNA bands were most prominent.

***Statistical Analysis*** To determine if there was a significant effect of strain and/or time on the measured variable amongst the heat-stressed strains, the data from the microscopy and plating assays were analyzed using two-factor repeated measures analysis of variance (ANOVA) tests. Because there was interest as to whether there was a difference in the measured variable between each of the mutant strains and the parent WT strain at each time point throughout the heat stress, multiple comparison post-hoc contrasts were also performed between each mutant strain and the parent WT strain at each time point. The p-values reported from the post-hoc tests were adjusted by Holm's method, a variant of the Bonferroni correction.[289] While the Bonferroni method is used to minimize the rate of Type I (false positive) errors that are prone to occur when testing multiple hypotheses, it is more likely to generate Type II (false negative) errors when applied to many tests simultaneously. Holm's corrective method was developed to minimize the Type I error rate associated with testing multiple hypotheses simultaneously while also minimizing the Type II error rate associated with the Bonferroni correction.[290]

**Results and Discussion**

### Analysis of Unstressed Cells

Prior to analysis of the mutant strains during PCD, the cells were first analyzed under non-stress conditions.

***Morphological Analysis*** Visualization of each mutant strain by phase-contrast microscopy revealed that the *e2f*, *tat-d*, and *lsd-1* mutant strains exhibited a tendency to form palmelloids under unstressed conditions (Figure 27C-E). Attempts to disrupt the structures by either mechanical force or resuspension in 10 mM Hepes for 1 hour were unsuccessful in disturbing the aggregates. Furthermore, the initial formation of

palmelloids could not be prevented by culturing the cells in either TAP or R medium. Interestingly, not all cells in these cultures were integrated into palmelloids. Based on the qualitative assessment of morphological features of non-palmelloid cells from these strains, individual cells could be broadly categorized into one of two classes: cells that appeared normal, in terms of behavior, size and morphology, relative to the parent WT strain (Figure 27A), and so-called "giant" cells, which exhibited a size substantially larger than that of WT cells. Furthermore, these giant cells also displayed a degree of variation with regard to ciliary presence/number, as some cells were aciliate, while as many as six cilia were counted on some of the ciliated giant cells. Anecdotally, cells of the *e2f* mutant strain tended to form palmelloids with neat exclusivity. Very few individual cells were noted in this strain, and only the occasional giant cell was observed. In contrast, many of the observed non-palmelloid cells of the *tat-d* and *lsd-1* mutant strains displayed the giant phenotype.

At the time that these observations were made, it was known that the majority of the assays performed in this study would require cell scoring under low-magnification brightfield microscopy. As these assays mandated the ability to differentiate individual cells, and because discerning individual cells comprising a palmelloid is difficult and time-consuming to carry out by these methods, the *e2f*, *tat-d*, and *lsd-1* mutant strains were excluded from the remainder of the experiments in this study. On the other hand, the *bi-1*, *ire-1*, and *pig3* mutant cells appeared identical to the parent WT strain, in terms of morphology, size, and behavior, as determined by phase-contrast microscopy (Figure 27B,C,F). As visualization of the *bi-1*, *ire-1*, and *pig3* strains by microscopy indicated an absence of obvious physical abnormalities at the cellular level, these strains met the

requirements for further analysis under normal conditions. Importantly, the "normal" appearance of cells in these strains does not, in and of itself, exclude the possibility of an unintended effect of mutagenesis, as such an effect could be present without manifesting as an obvious physical phenotype.
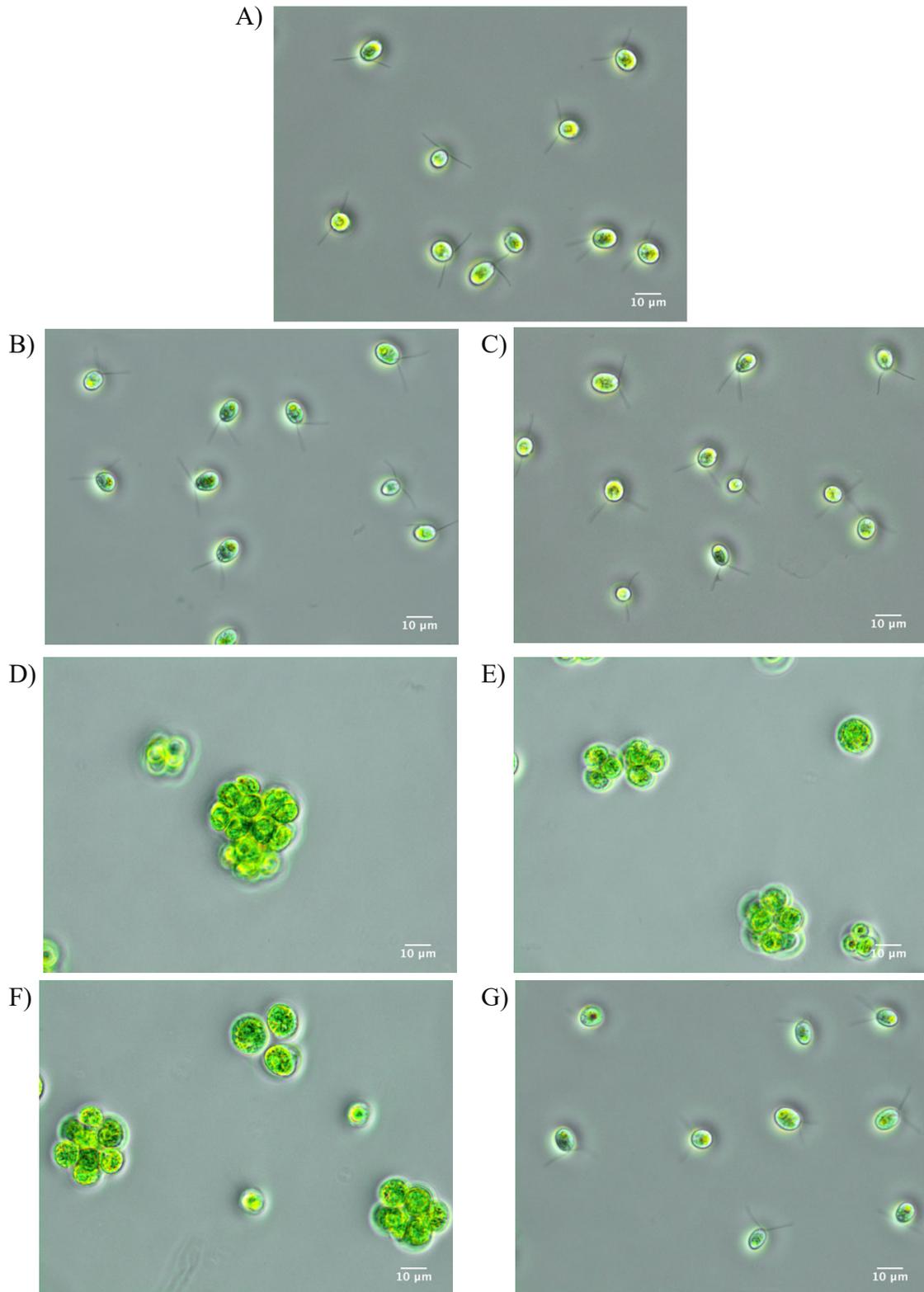
Figure 27. Morphologies of the *C. reinhardtii* strains utilized in this study. A) WT strain, B) *bi-1* mutant strain, C) *ire-1* mutant strain, D) *e2f* mutant strain, E) *tat-d* mutant strain, F) *lsd-1* mutant strain, G) *pig3* mutant strain.

***Growth Curve Analysis*** With this experiment, we sought to determine if the loss of *bi-1*, *ire-1*, or *pig3* would have an effect on the ability of *C. reinhardtii* to grow under normal culturing conditions. Though statistical analysis indicates that time had a significant effect on the mean cell density ($p = 5.84 \times 10^{-15}$), no significant difference in the mean daily cell density was observed between the parent WT strain and the *bi-1*, *ire-1*, and *pig3* mutant strains ($p > 0.05$). These results indicate that the rate of cell division in each of the mutant strains was approximately equivalent to that of the parent WT strain under normal culturing conditions (Figure 28). We interpret these results to suggest that the growth of *C. reinhardtii* is unaffected by the disruption of *bi-1*, *ire-1*, or *pig3* genes, implying that no metabolic pathways essential for cell division under mixotrophic conditions were affected by the disruption of these genes. Future studies utilizing these strains may wish to validate that the growth rate of each mutant strain is unaffected by the gene disruption under heterotrophic and photoautotrophic growth conditions as well.
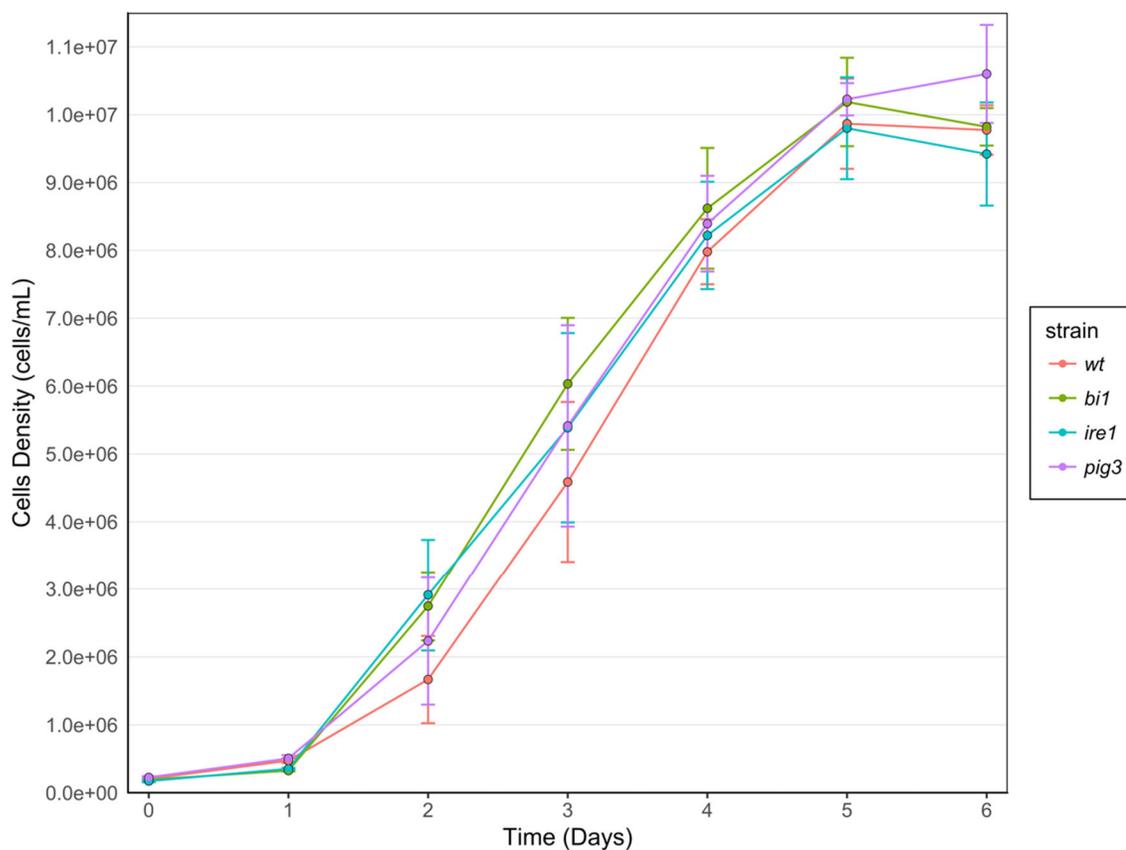
Figure 28. Growth curves of the *bi-1*, *ire-1*, and *pig3* mutant strains and parent WT strain. Plotted points represent the mean cell density of each strain at each time point (n=3). Error bars represent the ± standard error of each of the plotted points. Statistically significant differences were not detected between any of the mutant strains and the parent WT strain across all time points.

### Analysis of Strains Under PCD-Inducing Conditions

*Metabolic Activity Assay* The probe fluorescein diacetate (FDA) is one of the most frequently-used methods of identifying metabolically-active cells, particularly in microbial cells. Specifically, FDA enables the detection of cells exhibiting esterase activity, which is inferred to indicate metabolic activity. Following its entry into the cell, the acetate groups of FDA are non-specifically hydrolyzed by intracellular esterases, producing a standalone fluorescein.[291] This product can be detected using fluorescent methods, and can be visualized in individual cells using fluorescence microscopy. As

such, a cell that has accumulated a fluorescent product is inferred to be metabolically active.

With this experiment, we sought to determine if the loss of BI-1, IRE-1, or *PIG3* would have an effect on the loss of metabolic activity that occurs during PCD (Figure 29). Across all strains subjected to heat stress, the most drastic decrease in the percentage of FDA-positive cells occurred between 30 and 90 minutes. In all heat-treated strains, very few members of the population exhibited fluorescence at 120 minutes.

Statistical analysis indicated that time had a significant effect on the observed percentage of FDA-positive cells ($p < 2$ x $10^{-16}$). While the overall kinetics of the percentage of FDA-positive cells were similar for each the four strains throughout the heat stress procedure, there were several time points at which the percentage of fluorescent cells in one or more of the mutant strains was found to be significantly different than that of the parent WT strain. In particular, the percentage of fluorescent cells in the *bi-1* mutant strain was found to be significantly lower than that of the parent WT strain at 30 minutes ($p = 1.27$ x $10^{-8}$) and significantly higher at 90 minutes ($p = 4.10$ x $10^{-12}$). Additionally, in the *ire-1* mutant strain, the percentage of fluorescent cells was significantly lower than that of the parent WT strain at 60 minutes ($p = 1.12$ x $10^{-6}$).

Taken together, the results indicate that the overall kinetics of the decrease in the percentage of metabolically-active cells are not drastically altered in the absence of BI-1, IRE-1, or PIG3. In the parent WT strain, only a slight decrease in the percentage of fluorescent cells was observed after 30 minutes of heat stress; a very small portion of the population ceases metabolic activity at the onset of heat stress. The loss of IRE-1 or PIG3 did not exhibit a noticeable effect on this feature, though the more drastic decrease in the

percentage of metabolically-active cells in the *bi-1* mutant strain suggests that the loss of

BI-1 may accelerate the cessation of metabolic activity during the initial response to heat

stress. Furthermore, the steep decrease in the percentage of FDA-positive cells in all

strains between 30 and 90 minutes suggests that the vast majority of the population halts

metabolic activities during this time frame. Though this effect was enhanced in the *ire-1*

mutant strain at 90 minutes, the lack of a statistical difference in the percentage of

fluorescent cells between either the *bi-1* or the *ire-1* mutant strains and theparent  wild

type strain at 120 minutes suggests that the loss of either of these genes does not affect

the percentage of the population exhibiting metabolic activity at the conclusion of the two

hour heat stress.
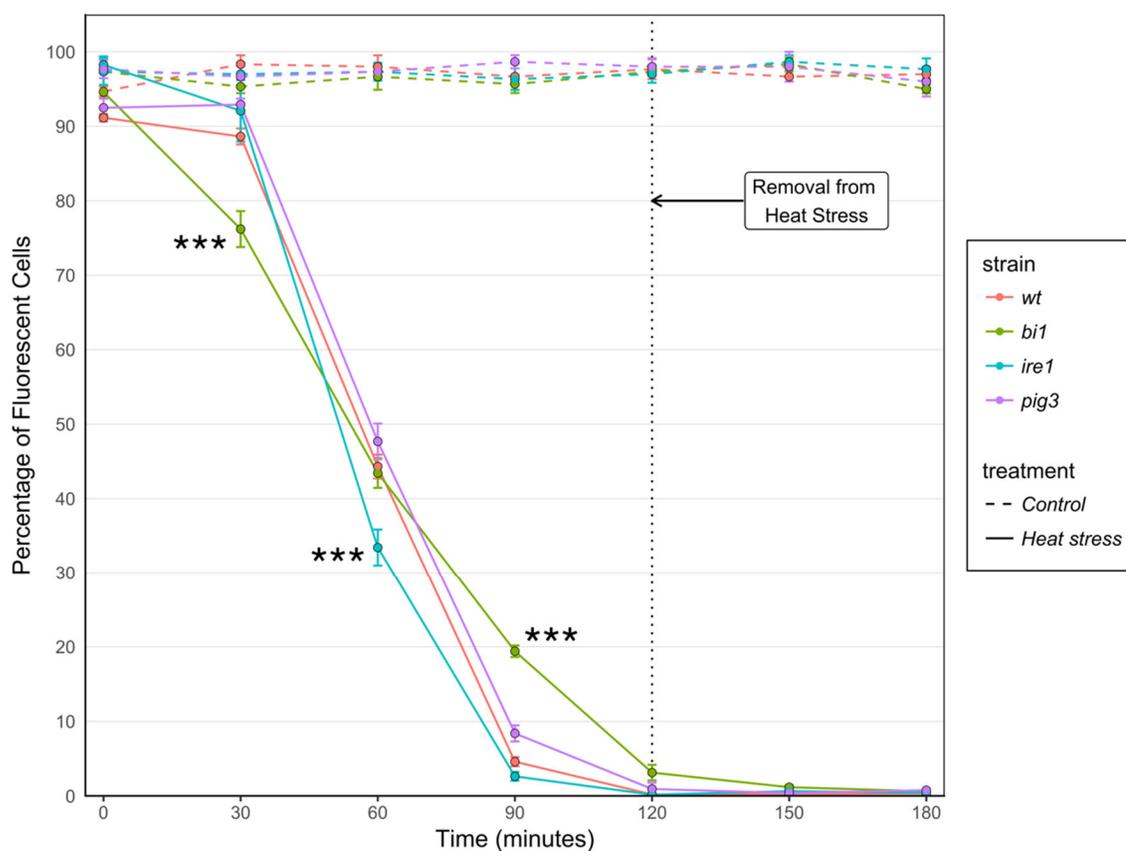


Figure 29. Stress-induced loss of metabolic activity in the *bi-1*, *ire-1*, and *pig3* mutant
strains compared to the parent WT strain. Plotted points represent the mean percentage of

fluorescent cells in each strain at each time point (n=3). Error bars represent the ± standard error of each strain at each time point. Statistical analysis was carried out using a two-way repeated measures analysis of variance (ANOVA). Post-hoc contrasts were carried out for each time point, wherein each mutant strain was compared to the parent WT strain. Statistical significance is denoted according to p-value; "***" indicates a significant difference when $p < 0.001$.

***Plasma Membrane Integrity Assay*** Here, the fluorescent probe SYTOX Green was used to visualize cells with compromised plasma membranes during PCD. SYTOX Green is a DNA intercalating agent which exhibits a >500 fold increase in fluorescence intensity when bound to DNA.[292,293] Because SYTOX Green is impermeable to cells with intact plasma membranes, living cells exclude the dye completely and are not stained.[293] Upon the loss of plasma membrane integrity at death, SYTOX enters the cell and binds DNA, resulting in a detectable fluorescent emission.[294] Thus, a cell which fluoresces is inferred to have lost plasma membrane integrity.

With this study, we sought to determine if the loss of *bi-1*, *ire-1*, or *pig3* would have an effect on the loss of plasma membrane integrity that occurs during PCD (Figure 30). The kinetics of the percentage of SYTOX Green-positive cells were similar between all four strains during the first 90 minutes of heat stress. After 90 minutes, the percentage of fluorescent cells continued to increase in the parent WT strain. In the *ire-1* and *pig3* mutant strains, the percentage of fluorescent cells increased only slightly between 90 and 120 minutes, then stayed relatively stable between 120 and 180 minutes. Intriguingly, the percentage of fluorescent cells in the *bi-1* mutant strain decreased dramatically after 90 minutes. One potential explanation for the decrease in the percentage of fluorescent cells could be that, because SYTOX Green binds DNA, and because DNA is being rapidly degraded in PCD (and potentially even more so in the *bi-1* strain), there is an insufficient quantity of DNA for SYTOX Green to bind.

Statistical analysis indicated that the percentage of SYTOX-positive cells was significantly affected by strain (p = 1.06 x 10$^{-14}$), time (p < 2 x 10$^{-16}$) and the interaction between strain and time (p = 1.18 x 10$^{-12}$). Though the kinetics of the percentage of fluorescent cells were similar between all strains within the first 90 minutes of heat stress, there were several time points at which the percentage of fluorescent cells in one or more of the mutant strains was found to be significantly different than that of the parent WT strain. The percentage of fluorescent cells in the *pig3* mutant strain was significantly lower than that of the parent WT strain at 60 minutes (p = 2.24 x 10$^{-5}$). The *bi-1* mutant strain showed a significantly lower percentage of fluorescent cells than the parent WT strain at 120 minutes (p < 2 x 10$^{-16}$), 150 minutes (p < 2 x 10$^{-16}$), and 180 minutes (p < 2 x 10$^{-16}$). The *ire-1* mutant strain showed a significantly lower percentage of fluorescent cells than the parent WT strain at 120 minutes (p = 7.7 x 10$^{-4}$), 150 minutes (p = 1.01 x 10$^{-12}$), and 180 minutes (p < 2 x 10$^{-16}$). Finally, the *pig3* mutant strain showed a significantly lower percentage of fluorescent cells than the parent WT strain at 120 minutes (p = 7.34 x 10$^{-11}$), 150 minutes (p = 4.7 x 10$^{-11}$), and 180 minutes (p < 2 x 10$^{-16}$).
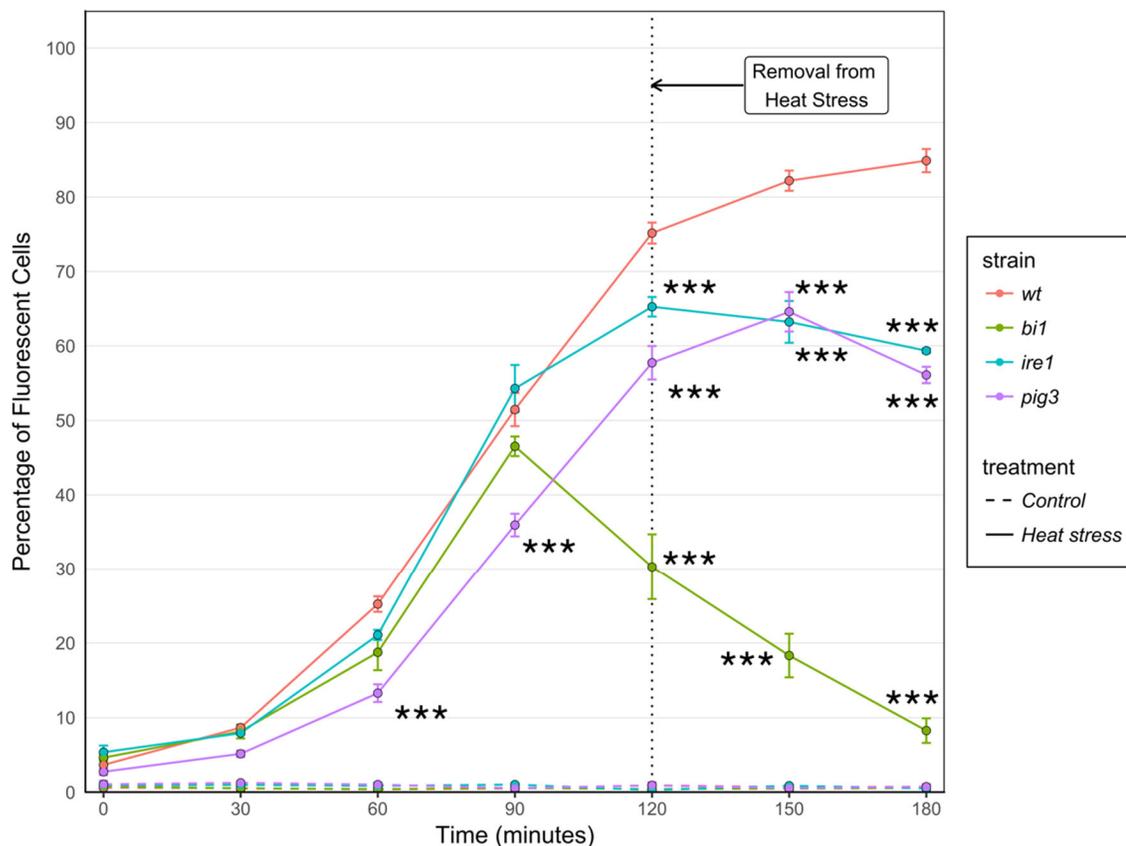
Figure 30. Loss of plasma membrane integrity in the *bi-1*, *ire-1*, and *pig3* mutant strains and parent WT strain during PCD. Plotted points represent the mean percentage of fluorescent cells in each strain at each time point (n=3). Error bars represent the ± standard error of each strain at each time point. Statistical analysis was carried out using a two-way repeated measures analysis of variance (ANOVA). Post-hoc contrasts were carried out for each time point, wherein each mutant strain was compared to the parent WT strain. Statistical significance is denoted according to p-value; "***" indicates a significant difference when $p < 0.001$.

The percentage of the population in the parent WT strain exhibiting a loss of plasma membrane integrity increased steadily during the first 30 minutes of heat stress. This rate of increase was slowed, but not stopped, by the removal of cells from heat stress conditions slowed this rate of increase.

The loss of BI-1 did have an effect on the loss of plasma membrane integrity, as the percentage of the fluorescent cells in the *bi-1* mutant strain was lower than that of the parent WT strain from 60 minutes onward. Unexpectedly, a drastic decrease in the

percentage of fluorescent cells was observed in the *bi-1* mutant strain at 120, 150, and 180 minutes, though the implications of this decrease are unknown. Of potential importance, the observed decrease was observed to begin prior to the removal the cells from the heat stress.

The loss of IRE-1 also appeared to have an effect of the loss of plasma membrane integrity, as the percentage of fluorescent cells in the *ire-1* mutant strain was slightly lower than that of parent WT strain in the early time points and significantly lower than the parent WT strain at later time points. Following the removal of the sample from heat stress conditions, the percentage of fluorescent cells decrease slightly.

The loss of PIG3 also appeared to have an effect on the loss of plasma membrane activity, as the percentage of fluorescent cells in the *pig3* mutant strain was lower than that of the parent WT strain across all time points. Furthermore, the rate of increase in the percentage of fluorescent cells in the *pig3* mutant strain was not as rapid as any of the other strains. Interestingly, as was observed in the parent WT strain, the percentage of fluorescent cells in the *pig3* mutant strain continued to increase following the removal from heat stress conditions, though began to decrease soon thereafter.

Taken together, it is not entirely clear what the results of this assay indicate with regards to plasma membrane integrity. In particular, the decrease in the percentage of fluorescent cells, most notable in the *bi-1* strain, is inconsistent with the current understanding of PCD processes in other organisms. Generally speaking, the loss of plasma membrane integrity is one of the few PCD features that reliably indicates the death of the cell. If the intracellular presence of a DNA-intercalating, membrane impermeable, dye indicates death, and if the percentage of cells that exhibit the presence

of that dye decreases at later time points, a direct interpretation of the results would indicate that, during the stress, fewer cells in the population are dead at the later time points compared to the early time points. Given that the current understanding of PCD indicates that a cell, once dead, cannot regain characteristics of life, this possibility is rejected here. Instead, it is thought that some other variable, unaccounted for in the experimental design, better explains the observed results. Specifically, we suspect that the fragmentation of the biomolecule to which SYTOX Green binds, DNA, may cause the observed decrease in the percentage of fluorescent cells at later time points. This is primarily based on the reasoning that, if the detectable fluorescence of SYTOX green is dependent on its ability to bind DNA, and the PCD process effects the DNA binding sites of SYTOX (because DNA is degraded during PCD), then the ability of SYTOX to reliably indicate cell death may be limited to the stages of cell death in which DNA has not yet been extensively fragmented.

*Phosphatidylserine Externalization Assay* The probe Annexin V was used to visualize cells displaying externalized phosphatidylserine to the outer leaflet of the plasma membrane. Annexin V is a cellular protein that binds phosphatidylserine directly.[295,296] The conjugation of Annexin V to a fluorophore, in this case AlexaFluor488, facilitates the visualization of cells bound by Annexin V by fluorescence microscopy.[62,272] As such, fluorescing cells are inferred to have externalized phosphatidylserine.

In this experiment, we sought to determine if the loss of BI-1, IRE-1, or PIG3 would affect phosphatidylserine exposure observed during PCD. At 60 minutes, nearly all cells were fluorescing, and no statistically-significant difference was detected between

the parent WT and any of the mutant strains. In all four strains, the percentage of fluorescent cells decreased after the 60 minute time point (Figure 31). Statistical analysis indicated that the percentage of Annexin V-positive cells was significantly affected by strain ($p = 1.83 \times 10^{-4}$), time ($p = 7.33 \times 10^{-11}$) and the interaction between strain and time ($p = 6.98 \times 10^{-3}$). The *bi-1* mutant strain showed a significantly lower percentage of fluorescent cells than the parent WT strain at 120 minutes ($p = 2.40 \times 10^{-14}$), 180 minutes ($p < 2 \times 10^{-16}$), and 240 minutes ($p < 2 \times 10^{-16}$). The *ire-1* mutant strain showed a significantly lower percentage of fluorescent cells than the parent WT strain at 180 minutes ($p = 3.53 \times 10^{-6}$) and 240 minutes ($p = 3.81 \times 10^{-8}$). Finally, the *pig3* mutant strain showed a significantly lower percentage of fluorescent cells than the parent WT strain at 180 minutes ($p = 1.11 \times 10^{-14}$) and 240 minutes ($p < 5.09 \times 10^{-10}$).
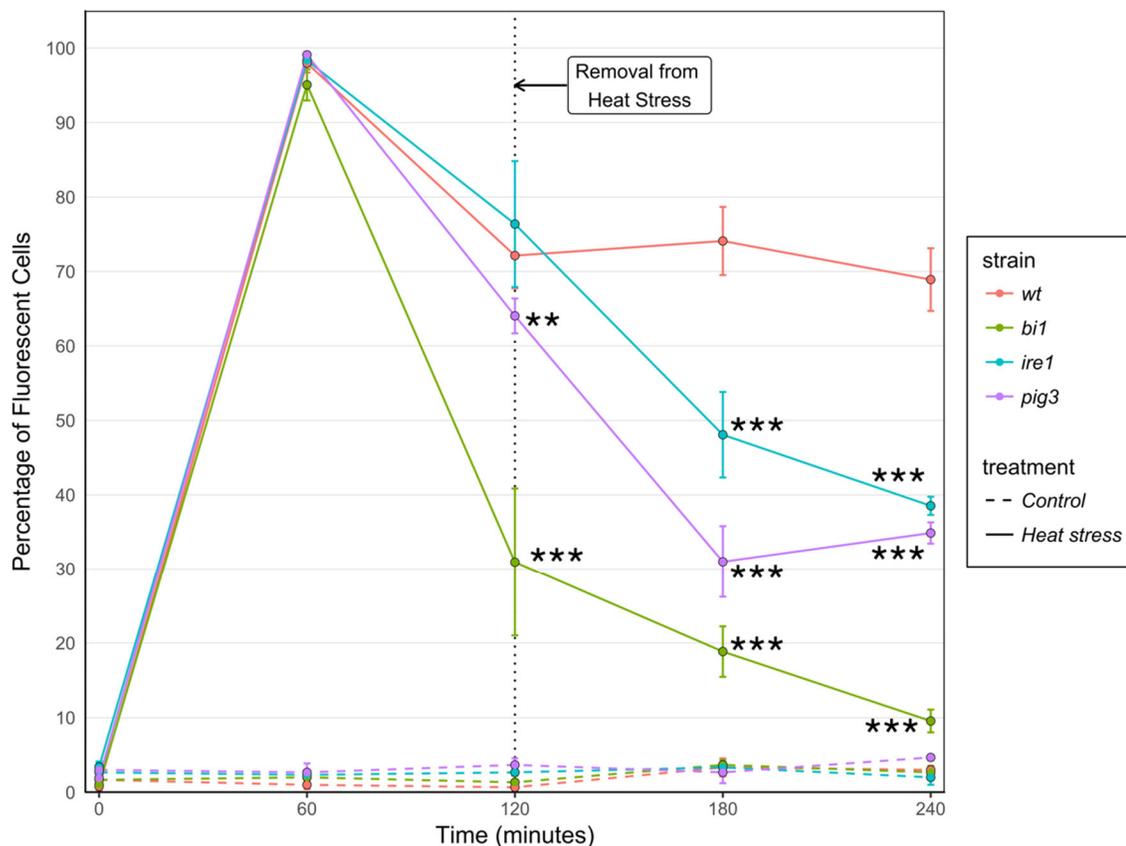
Figure 31. Phosphatidylserine externalization in the *bi-1*, *ire-1*, and *pig3* mutant strains compared to the parent WT strain during PCD. Plotted points represent the mean percentage of fluorescent cells in each strain at each time point (n=3). Error bars represent the ± standard error of each strain at each time point. Statistical analysis was carried out using a two-way repeated measures analysis of variance (ANOVA). Post-hoc contrasts were carried out for each time point, wherein each mutant strain was compared to the parent WT strain. Statistical significance is denoted according to p-value: "**" indicates a significant difference when $p < 0.01$.

Taken together, these results demonstrate a nearly population-wide externalization of phosphatidylserine in all strains within the first hour of heat stress, indicating that this initial onset of PS externalization is not effected by the loss of BI-1, IRE-1, or PIG3. As, the first sampling was not carried out until 60 minutes into the stress, future studies may seek to determine if the initial rate of increase of in the percentage of Annexin V positive cells is affected in the absence of BI-1, IRE-1, or PIG3. Strikingly, the percentage of Annexin V-positive cells decreases in all strains at subsequent time

points, indicating that the loss of BI-1, IRE-1, or PIG3 do not prevent this observed decrease. Furthermore, the rate of decrease in the percentage of fluorescent cells is significantly enhanced in the *bi-1*, *ire-1*, and *pig3* mutant strains, suggesting that the loss of these genes somehow accelerated this rate of decrease.

The biological implications for the results of the Annexin V assay are unclear. The observed decrease in the percentage of fluorescent cells indicates that, by some unknown mechanism, Annexin V is unable to bind externalized phosphatidylserine at the later time points.

*ROS Accumulation Assay* In this study, we utilized the fluorescent probe CM-$H_2$DCFDA to identify cells which have accumulated ROS during a PCD-inducing heat stress. CM-$H_2$DCFDA is a chloromethyl and acetate derivative of dihydrodichlorofluorescein ($H_2$DCFDA), a widely-used compound for the detection of intracellular reactive oxygen species. Thanks to the presence of lipophilic acetate groups, $H_2$DCFDA is able to freely enter the cell by passive diffusion.[297] Once in the cytosol, the acetate groups of $H_2$DCFDA are hydrolyzed nonspecifically by intracellular esterases, yielding the product 2',7' dichlorodihydrofluorescein ($H_2$DCF).[298] $H_2$DCF is oxidized to form the radical HDCF•, which is further oxidized to form the fluorescent product dichlorofluorescein DCF.[299] This study utilizes a variant of $H_2$DCFDA, CM-$H_2$DCFDA. The addition of two chloromethyl groups results in the final product of its intracellular processing to be CM-DCF, a fluorescent compound reported to possess higher retention levels in the cell compared to DCF by reducing the ability of the compound to diffuse passively into the extracellular space.[300]

In the context of the percentage of fluorescent cells, the kinetics between the mutant strains and the parent WT were substantially different (Figure 32). The parent WT strain displayed the highest percentage of fluorescent cells at 30 minutes, at which point ~90% of the cells were fluorescing. The percentage of fluorescent cells decreased at later time points in the parent WT strain. In contrast, only half of the cells were fluorescing at 30 minutes in the *bi-1* mutant strain, and ~25-30% of the cells were fluorescing in the *ire-1* and *pig3* mutant strains. In the *bi-1* mutant strain, the highest percentage of cells, ~70%, was observed at 90 minutes. In the *ire-1* and *pig3* mutant strains, the highest percentage of fluorescent cells, ~65-70%, was observed at 60 minutes. For each of the four strains, the percentage of fluorescent cells was drastically reduced at the 120 minute time point.

Statistical analysis indicated that the percentage of CM-H$_2$DCFDA -positive cells was significantly affected by time ($p = 7.05$ x $10^{-9}$) and the interaction between strain and time ($p = 1.2$ x $10^{-3}$). Several statistically-significant differences in the percentage of fluorescent cells were noted between the parent WT strain and one or more of the mutant strains throughout the heat stress. The *bi-1* mutant strain showed a significantly lower percentage of fluorescent cells than the parent WT strain at 30 minutes ($p < 2$ x $10^{-16}$) and a significantly higher percentage of fluorescent cells than the parent WT strain at 90 minutes ($p < 2$ x $10^{-16}$). The *ire-1* mutant strain showed a significantly lower percentage of fluorescent cells than the parent WT strain at 30 minutes ($p < 2$ x $10^{-16}$) and a significantly higher percentage of fluorescent cells than the parent WT strain at 90 minutes ($p < 2$ x $10^{-16}$).  The *pig3* mutant strain showed a significantly lower percentage of fluorescent cells than the parent WT strain at 30 minutes ($p < 2$ x $10^{-16}$) and a

significantly higher percentage of fluorescent cells than the parent WT strain at 90

minutes ($p < 2$ x $10^{-16}$). In the parent WT strain, the highest increase in the percentage of

fluorescent cells occurred between 0 and 30 minutes following heat stress. This is also

when the highest percentage of the population exhibiting ROS (90%) was observed in

this strain. Between 30 and 60 minutes, the percentage of cells exhibiting accumulated

ROS dropped dramatically. The loss of the *bi-1*, *ire-1*, or *pig3* genes had similar effects

on the percentage of fluorescent cells with regards to both the rate at which the

percentage of fluorescent cells increased, but also in terms of the maximum percentage of

the population that exhibited the accumulation of ROS. The initial accumulation of ROS

observed in the parent WT strain was attenuated in each of the three mutant strains (more

so in the *ire-1* and *pig3* mutant strains than the *bi-1* mutant strain). Furthermore, while

the percentage of the parent WT population exhibiting accumulated ROS decreased after

the 30 minute time point, the percentage of the *ire-1* and *pig3* mutant populations that

exhibited accumulated ROS did not peak until 60 minutes. In the *bi-1* mutant strain, the

percentage of cells exhibiting ROS accumulation did not reach its maximum until the 90

minute time point.
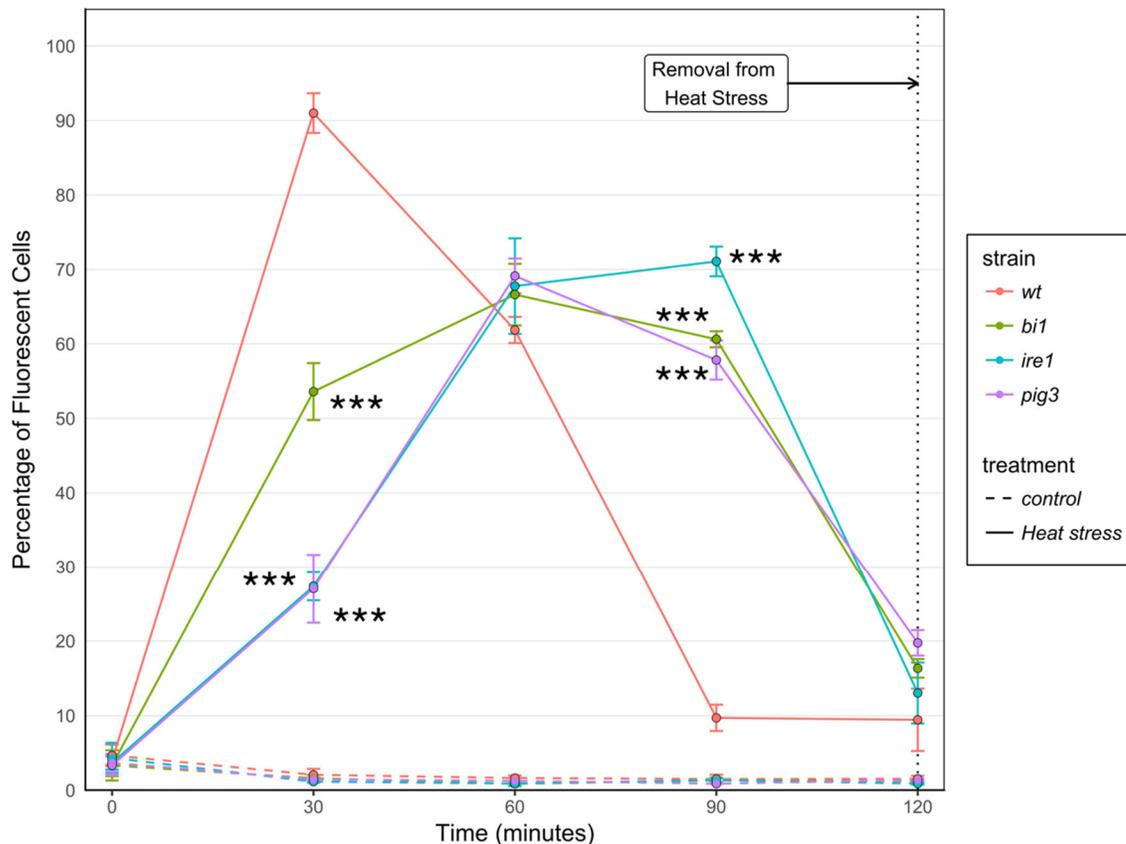
Figure 32. ROS accumulation in the *bi-1*, *ire-1*, and *pig3* mutant strains compared to the WT strain during PCD. Plotted points represent the mean percentage of fluorescent cells in each strain at each time point (n=3). Error bars represent the ± standard error of each strain at each time point. Statistical analysis was carried out using a two-way repeated measures analysis of variance (ANOVA). Post-hoc contrasts were carried out for each time point, wherein each mutant strain was compared to the parent WT strain. Statistical significance is denoted according to p-value; "***" indicates a significant difference when $p < 0.001$.

Regardless, at 120 minutes, the percentage of fluorescent cells in each of the strains dropped below 20%. Together, these results indicate that the loss of either BI-1, IRE-1, or PIG3 significantly reduced the capability to accumulate ROS in the cell, both in terms of timing and the percentage of cells exhibiting the accumulation of ROS.

***Plating Assay*** With this experiment, we sought to determine if the loss of either the *bi-1*, *ire-1*, or *pig3* genes would affect the area of vegetative growth on solid medium throughout and following a PCD-inducing heat stress.

In each of the four strains, the most drastic decrease in growth area occurred within the first 60 minutes of heat stress (Figure 33). Statistical analysis indicated that the growth area was significantly affected by time ($p < 2 \times 10^{-16}$) and the interaction between strain and time ($p = 1.29 \times 10^{-4}$). At only the 30 minute time point was there a significant difference between the growth area of the parent WT strain and one of the mutant strains. At 30 minutes, both the *bi-1* ($p = 7.72 \times 10^{-7}$) and *pig3* ($p < 2 \times 10^{-16}$) mutant strains showed a significantly greater growth area than that of the parent WT strain. Though the mean measured area of the *ire-1* mutant strain was higher than parent WT strain at 30 minutes, the difference was not statistically significant.

Of note, the mean area of the *pig3* mutant strain was higher than that observed in the parent WT strain between 30-120 minutes. Additionally, the mean growth area of the *bi-1* mutant strain was higher than that of the parent WT strain between 30-90 minutes. Finally, the mean area of the *ire-1* mutant strain was higher than the parent WT strain between 30-120 minutes.
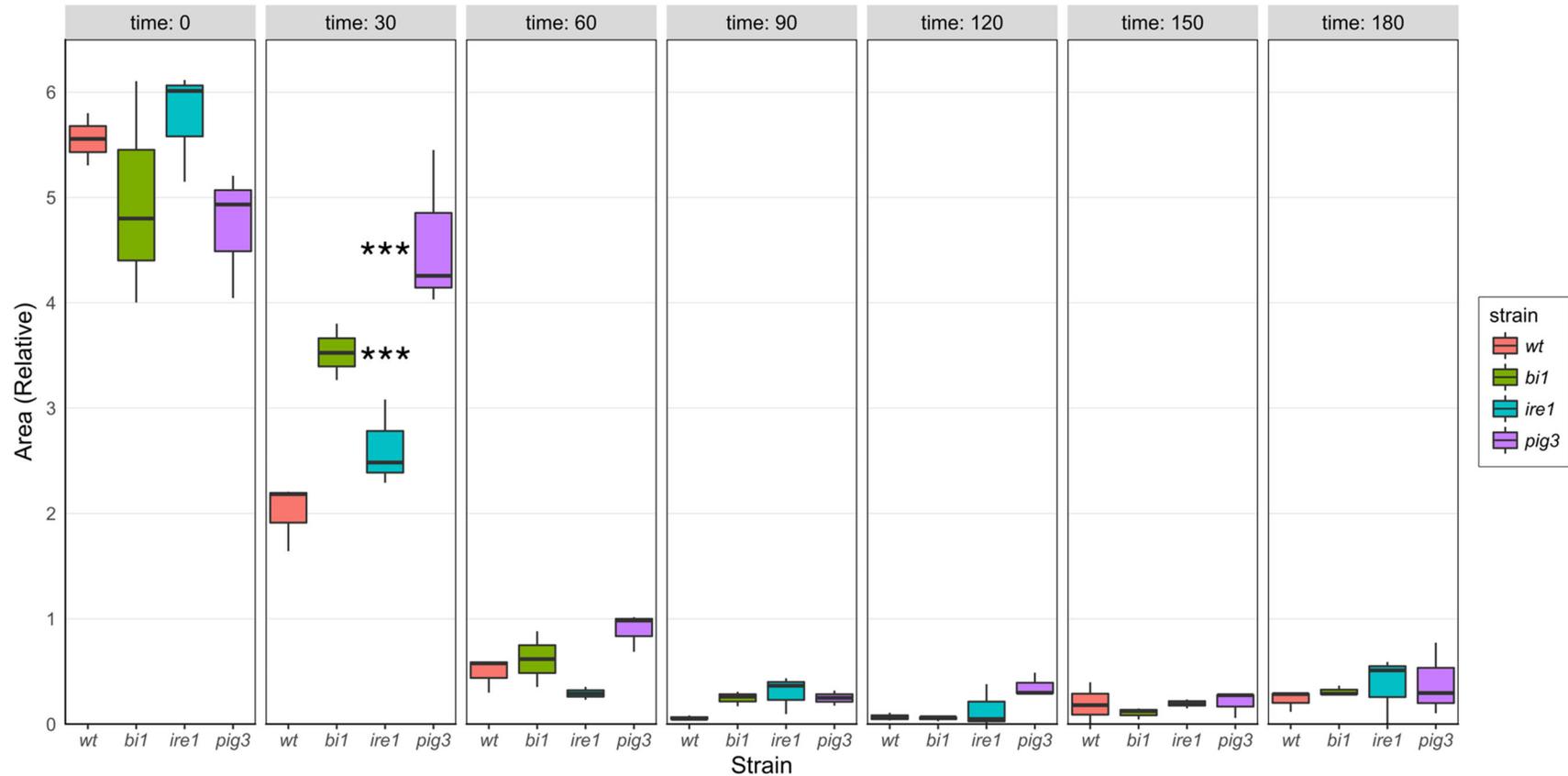
Figure 33. Relative growth area of the *bi-1*, *ire-1*, and *pig3* mutant strains during PCD. This boxplot depicts the relative growth areas measured for each strain at each time point throughout the heat stress (n=3). The boxes represent data between the 25th and 75th percentiles, solid horizontal line within each of the boxes represents the median, and the whiskers represent the minimum and maximum values. Statistical analysis was carried out using a two-way repeated measures analysis of variance (ANOVA). Post-hoc contrasts were carried out for each time point, wherein each mutant strain was compared to the parent WT strain. Statistical significance is denoted according to p-value: "***" indicates a significant difference when p < 0.001.

***DNA Laddering Assay*** In this experiment, we sought to determine if the loss of either BI-1, IRE-1, or PIG3 would affect either the timing or intensity of DNA laddering during PCD (Figure 34). In all three mutant strains, DNA laddering appeared more prominent than in the parent WT strain, suggesting that genomic fragmentation is enhanced in the absence of BI-1, IRE-1, or PIG3. The patterns of genomic DNA fragmentation were similar between the *ire-1* mutant strain and parent WT strain in that genomic fragmentation was not clearly present until 120 minutes. However, the laddering in the *ire-1* mutant strain was more prominent than that observed in the parent WT strain at the 120 minute time point. This could potentially indicate that the absence of IRE-1 enhances the occurrence of DNA fragmentation without expediting the process. In the *bi-1* and *pig3* mutant strains, the onset of DNA fragmentation occurred rapidly between the 60 and 90 minute time points, potentially indicating that the absence of BI-1 or *PIG3* enhances both the rate and occurrence of DNA fragmentation in *C. reinhardtii*. Notably, only one trial of this assay was conducted, and these results will require verification by future experiments.
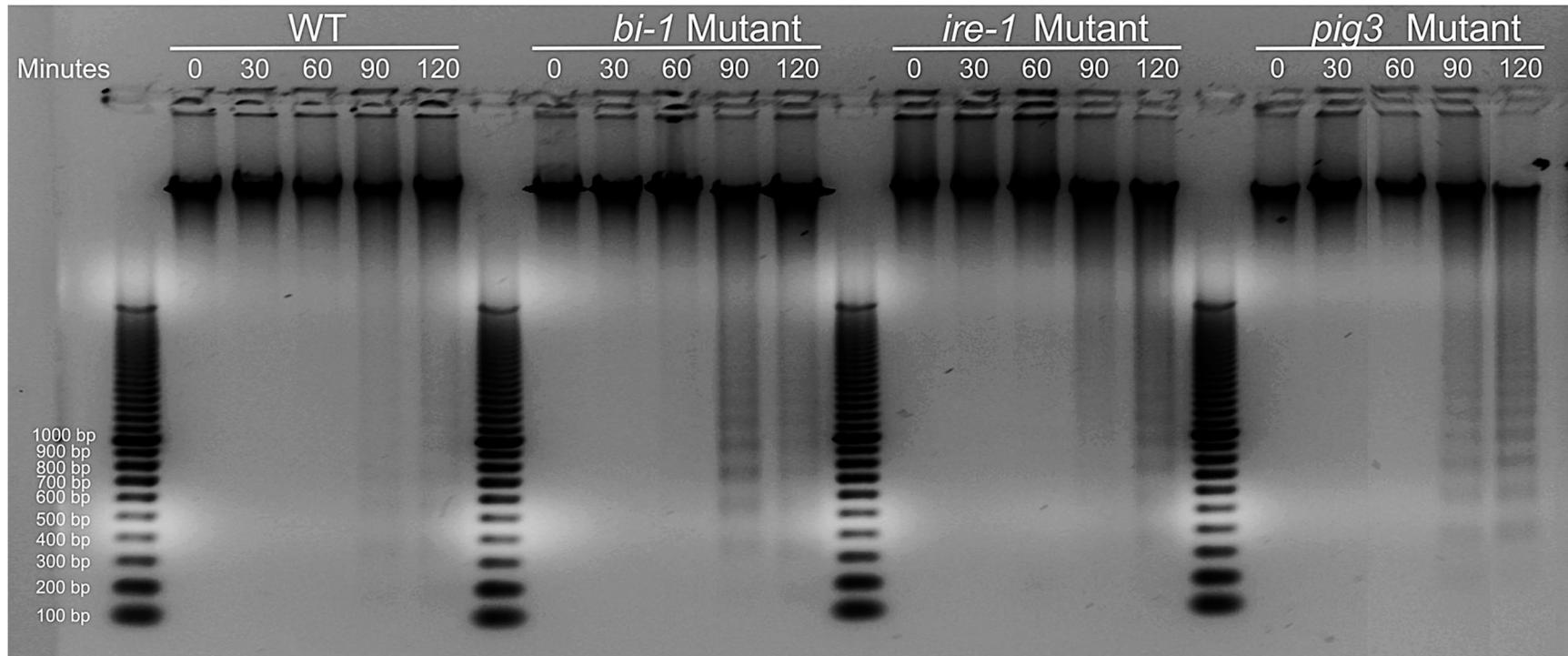
Figure 34. DNA laddering in the *bi-1*, *ire-1*, and *pig3* mutant strains during PCD. Each lane contains purified genomic DNA from a single strain at a single time point (n=1). Samples are grouped first by strain, then by time point.

**Summary**

In previous studies, *C. reinhardtii* proteins that participate in PCD were predicted using a large-scale, homology-based, approach. From this list of *C. reinhardtii* predicted PCD proteins (CrPPPs), we selected several proteins to investigate further for a role in *C. reinhardtii* PCD. We hypothesized that each of these selected proteins participates in *C. reinhardtii* PCD, and we sought to test this hypothesis utilizing a reverse genetics approach. To this end, mutant strains of *C. reinhardtii*, each harboring an insertional mutation in one of the selected protein-coding genes, were ordered from the CLiP library and the reported insertion site was validated for each strain.

The purpose of this study was to test the hypothesis that each of these proteins participates in PCD in *C. reinhardtii*. To accomplish this, a reverse genetics approach was utilized, wherein we assayed for typical PCD features in each of the mutant strains following exposure to a PCD-inducing heat stress. We predicted that *C. reinhardtii* strains deficient in a single selected protein would exhibit an alteration of one or more PCD phenotypes relative to the parent WT strain. Notably, morphological analysis of the *e2f*, *tat-d*, and *lsd-1* mutant strains exposed a tendency of these strains to form multicellular aggregates known as palmelloids under normal culturing conditions. As such structures are not conducive for the manual scoring of cells by microscopy, these strains were excluded from the study. A summary of the results for the *bi*-1, *ire*-1, and *pig3* mutant strains is shown in Table 11.

Table 11

Comparison of PCD features in each mutant strain to the parent WT strain.

| Feature | *bi-1* Mutant Strain | *ire-1* Mutant Strain | *pig3* Mutant Strain |
| --- | --- | --- | --- |
| Loss of Metabolic Activity (Figure 29) | Altered | Altered | Approximately the same |
| Loss of Plasma Membrane Integrity (Figure 30) | Attenuated after 120 minutes | Attenuated after 120 minutes | Attenuated after 60 minutes |
| PS Externalization (Figure 31) | Attenuated after 120 minutes | Attenuated after 180 minutes | Attenuated after 120 minutes |
| ROS Accumulation (Figure 32) | Delayed and attenuated | Delayed and attenuated | Delayed and attenuated |
| Cell Division Ability (Figure 33) | Greater area at 30 minutes | Approximately the same throughout | Greater area at 30 minutes |
| DNA Laddering (Figure 34) | More rapid, enhanced | Delayed | Enhanced |

***bi-1* Mutant Strain** Bax inhibitor-1 (BI-1) is a conserved ER-transmembrane protein that acts to repress PCD in distinct phylogenetic clades. Given the anti-PCD functions of BI-1 in other organisms, it was expected that enhanced PCD features would be observed in a *C. reinhardtii* strain lacking functional BI-1 (Table 6).

Of the experiments performed in this study, only the DNA laddering assay provided strong support for the prediction that a *C. reinhardtii* strain lacking functional BI-1 would exhibit enhanced PCD features (Figure 34). That DNA laddering in the *bi-1* mutant strain was observed to be more prominent at 90 minutes than the parent WT strain at any of the time points, suggests that DNA laddering is enhanced in the absence of BI-1. Additionally, DNA laddering in the *bi-1* mutant strain is less prominent at 120 minutes compared to 90 minutes, suggesting that the DNA fragmentation process as a whole may be expedited.

The other results obtained from this study were either inconclusive or did not support the prediction that the *bi-1* mutant strain would exhibit enhanced PCD features. While a greater decrease in the percentage of metabolically active cells was observed in the *bi-1* mutant strain after 30 minutes of heat stress, this trend was not observed at later time points. In fact, a greater percentage of cells in the *bi-1* mutant strain was observed to be metabolically active at both the 90 and 120 minute time points relative to the parent WT strain (Figure 29). Unexpectedly, the percentage of *bi-1* mutant cells exhibiting a loss of plasma membrane integrity was slightly lower than that of the parent WT strain within the first 90 minutes of heat stress (Figure 30). Perplexingly, a prominent decrease in the percentage of SYTOX Green-positive cells was observed in the *bi-1* mutant strain after 90 minutes of heat stress. Intuitively, it seems unlikely that cells regain plasma membrane integrity, as this feature is thought to be irreversible and has been proposed to be one of the most reliable marker for cell death.[95] Given that DNA intercalating agents such as SYTOX Green bind DNA, we predict that the observed decrease in the percentage of fluorescent cells may be due to the extensive fragmentation of DNA at these later time points. Importantly, this prediction is consistent with the extensive DNA fragmentation observed in the *bi-1* mutant strain from the results of the DNA laddering assay (Figure 34). If true, this explanation of the SYTOX Green results would be in agreement with the prediction that the loss of BI-1 would augment PCD in *C. reinhardtii*, though it would simultaneously call into question the efficacy of SYTOX green as an indicator of membrane permeability in the presence of extensive DNA degradation. Unexpectedly, ROS accumulation in the *bi-1* mutant strain was attenuated during heat-induced PCD (Figure 32). As BI-1 acts to inhibit the formation of the MPTP, a key

contributor to the ROS accumulation observed during apoptosis, it was expected that that

the absence of BI-1 would result in a more robust accumulation of ROS during PCD.

Also surprisingly, a larger growth area was observed for the *bi-1* mutant strain compared

to the parent WT strain at 30, 60, and 90 minutes, suggesting a higher number of viable

cells at these time points (Figure 33).

    *ire-1* **Mutant Strain** Inositol-requiring enzyme 1 (IRE-1) is a phylogenetically-

conserved ER transmembrane protein that acts as both a stress sensor and primary

regulator of PCD. In light of the PCD-modulating role of IRE-1 in other organisms, it

was expected that attenuated PCD features would be observed in a *C. reinhardtii* strain

lacking functional IRE-1 (Table 6).

    Of the experimental results obtained in this study, some were consistent with the

prediction that a *C. reinhardtii* strain lacking functional IRE-1 would exhibit attenuated

PCD features. A lower percentage of cells exhibiting a loss of plasma membrane integrity

was observed in the *ire-1* mutant strain relative to the parent WT strain at 120, 150, and

180 minutes following subjection to heat stress (Figure 30). Additionally, ROS

accumulation in the *ire-1* mutant strain is both delayed and decreased relative to the

parent WT strain (Figure 32). Finally, when plated at various time points during the heat

stress, the area of growth in the *ire-1* mutant strain was greater than that of the parent WT

strain at all times assayed, with the exception of the 60 minute time point (Figure 33).

    The results of the other experiments performed were either inconclusive or did not

support the prediction that PCD would be attenuated in the *ire-1* mutant strain. Namely,

the percentage of metabolically-active cells was only slightly reduced in the *ire-1* mutant

strain relative to that of the parent WT strain at 60 and 90 minutes into the heat stress

(Figure 29). Whether these differences are relevant in a biological context is unclear. Additionally, the results of the DNA laddering assay suggest that DNA fragmentation may be enhanced in the absence of IRE-1, as the DNA laddering observed in the *ire-1* mutant strain is more prominent at 120 minutes than any of the parent WT lanes (Figure 34). Unexpectedly, however, DNA laddering in the *ire-1* mutant strain was not visible at 90 minutes. As faint DNA laddering is visible in the parent WT strain at this time, the results of this experiment may indicate that DNA laddering is delayed in the absence of IRE-1.

*pig3* **Mutant Strain** p53-induced gene 3 (PIG3) is a quinone oxidoreductase that is conserved in plants and animals. In mammals, PIG3 expression is activated by the apoptotic regulator, p53. In light of the pro-PCD functions of PIG3 in animals, it was expected that attenuated PCD features would be observed in a *C. reinhardtii* strain lacking functional PIG3 (Table 6).

Of the experiments performed in this study, some of the results obtained were in agreement with the prediction that a *C. reinhardtii* strain lacking functional PIG3 would exhibit diminished PCD features. The percentage of cells exhibiting a loss of plasma membrane integrity was lower in the *pig3* mutant strain relative to the parent WT strain across all time points (Figure 30). Moreover, the percentage of cells exhibiting ROS accumulation was both delayed and diminished in the *pig3* mutant strain relative to the parent WT strain (Figure 32). Additionally, when cells were plated at various time points during and after the heat stress, the mean area of growth observed in the *pig3* mutant strain was consistently higher than that of the parent WT strain (Figure 33).

Some of the experiments carried out in this study either did not support the prediction that PCD features would be attenuated in the *pig3* mutant strain or were inconclusive. The percentage of cells exhibiting metabolic activity in the *pig3* mutant strain was slightly less than that of the parent WT strain during the first 90 minutes of heat stress (Figure 29). Whether this observation is relevant in a biological context is unknown. DNA laddering was more prominent in the *pig3* mutant strain than in the parent WT mutant strain at both 90 and 120 minutes, suggesting that the loss of PIG3 enhances DNA laddering in *C. reinhardtii* (Figure 34).

**CHAPTER V**

**Project Summary, Study Limitations, and Future Studies**

**Project Summary**

In a broad sense, the preceding work was carried out in order to elucidate some of the molecular mechanisms that underlie programmed cell death (PCD) in *Chlamydomonas reinhardtii*. An understanding of PCD in *C. reinhardtii* is beneficial for two primary reasons. First, despite the fact that PCD processes are observed throughout the tree of life, exceedingly little is known regarding the molecular foundations of PCD in non-animal organisms. The poor characterization of such a widespread and fundamental process limits our understanding of basic biological processes, both at the cellular and organismal level. We argue that an understanding of PCD in unicellular organisms, many of which are widely regarded to represent ancestral life forms, is crucial to understanding PCD processes in other organisms. This is because 1) at least some aspects of PCD may be conserved between a given multicellular organism and the unicellular ancestor, and 2) as speciation of unicellular organisms gave rise to the vast array of biological diversity observed today, findings in unicellular organisms can often be translated to higher organisms of common descent.

As a second benefit of understanding PCD in *C. reinhardtii*, we present that *C. reinhardtii* is an ideal organism in which to study conserved aspects of PCD, as the relatedness of *C. reinhardtii* to both animals and higher plants makes findings in this organism likely relevant to these systems. In this way, *Chlamydomonas* is set apart from traditional models of unicellular PCD, such as *Saccharomyces*. Furthermore, the utility of

*C. reinhardtii* for industrial use means that understanding PCD in this organism is highly likely to benefit the areas of industry in which it is utilized, such as biofuel and nutrition.

The purpose of this chapter is to summarize the work that was described in previous sections of this thesis, discuss potential limitations to these experiments, and present future studies for consideration.

**Prediction and Selection of Putative PCD Proteins in *C. reinhardtii*** The goals of this work were to 1) use computational methods to predict proteins in *C. reinhardtii* that participate in PCD on a large scale, and 2) select several proteins for further investigation regarding a role in *C. reinhardtii* PCD. The results of these experiments are summarized in Table 5 and Table 6, and suggest that some mechanisms of PCD in other organisms may be conserved in *C. reinhardtii*.

**Selection and Verification of *C. reinhardtii* Mutant Strains** In order to investigate a role for the selected gene products during *C. reinhardtii* PCD, we sought to employ a reverse-genetics approach, in which mutant strains of *C. reinhardtii*, each deficient in one of the selected proteins, would be utilized. We ordered the mutant strains, one for each selected CrPPP (Table 6), from the *Chlamydomonas* Library Project (CLiP) library of mutant strains. Prior to experimental utilization of the selected strains, the reported insertional locus was verified for each strain using methods recommended by the CLiP library. This process consisted of three steps, the results of which are summarized in Table 11.

Finally, to confirm that the PCR products from the amplification of the cassette-genome junctions were sent off for sequencing to confirm that the observed products were indeed from the cassette genome-genome junctions. To confirm that the sequenced

products from the G1/C1 and G2/C2 PCR reactions were of the cassette-genome

junctions, a simulation of the insertion was created for each of the strains by inserting the

CIB1 cassette nucleotide sequence, in its determined orientation, into the insertion site

reported by the CLiP library. For each mutant strain, the sequences obtained from PCR

amplification of the cassette-genome junctions, if any, were aligned with the simulated

insertion site.

### Comparative Analysis of PCD Features in *C. reinhardtii* Mutant Strains

Following the verification procedures for the mutant strains, we tested the hypotheses

that each of the selected proteins participates in *C. reinhardtii* PCD. To this end, a reverse

genetics approach was utilized, wherein each of the mutant strains was assayed for

typical PCD features in response to a PCD-inducing heat stress. It was predicted that *C.

reinhardtii* strains deficient in a single selected protein would exhibit an alteration in one

or more of the PCD phenotypes measured relative to the parent WT strain. A synopsis of

the results for the *bi-1*, *ire-1*, and *pig3* mutant strains is shown in Table 11. To

summarize, the results obtained in this study demonstrate that each of the mutant strains

exhibited one or more PCD phenotypes which were distinct from that of the parent WT

strain. These findings suggest that BI-1, IRE-1, and PIG3 may participate in PCD *C.

reinhardtii*.

### Limitations and Future Studies

**Database Construction** Not all proteins that are involved in PCD were included

in the database used to predict PCD proteins in *C. reinhardtii*. Given that the molecular

bases for recently-discovered forms of PCD, such as pyroptosis or ferroptosis, are poorly

understood, we chose to exclude amino acid sequences from such PCD subtypes from the

database. When the underlying molecular pathways for these processes are characterized in more detail, future studies might attempt to integrate these forms of PCD into the prediction of PCD participants in *C. reinhardtii*.

We also opted to exclude amino acid sequences of non-eukaryotic origin from the database, as the total number of eukaryotic sequences alone exceeded 90,000 search results. Given the possibility that some aspects of PCD in non-eukaryotic organisms may be conserved in *C. reinhardtii*, future studies that seek to further elucidate the molecular basis of *C. reinhardtii* may construct a database using amino acid sequences of bacterial, archaeal, and/or viral origin.

It should also be noted that some of the sequences included in the PCD database may not actually participate in PCD. A keyword search from the NCBI webpage returns all amino acid sequences from the searched databases that contain the keyword in the entry's GenPept page. However, the presence of a keyword in a GenPept entry does not guarantee the involvement of the protein in that process. Hence, it is conceivable that a protein, which is not involved in PCD, contains one of the search terms in its GenPept entry. However, because the filtered BLASTp results consisted of over 100,000 alignments between the CrPPPs and the database proteins, manual confirmation of a role in PCD for every subject sequence was not feasible. One reason we suspect that the database in this study may contain superfluous entries is that, in searches of the RefSeq database, we opted to include search results from all eukaryotic organisms in the database. The basis of PCD in many of these organisms remains largely uncharacterized, indicating that the annotations assigned to such organisms may be predictive only. Thus, while the strategy to include amino acid sequences from all eukaryotes in the search

results seemed advantageous because of the extensive nature of the resulting database, we suspect that this comprehensiveness may have come at the cost of stringency with regard to a verified role during PCD for the amino acid sequences that originated from uncharacterized genomes. As an alternative approach, it may be beneficial to construct future databases using only sequences from organisms in which PCD has been extensively characterized.

In light of the limitations discussed above, we propose that the UniProt database may be more suitable than the RefSeq database for exploratory studies such as this.[301,302] The UnitProt webpage (www.uniprot.org) provides descriptive titles for each entry, including all known synonyms and alternative names. In addition, the user-friendly webpage also possesses extensive search capabilities, which allows the user to perform a keyword search using a number of different parameters, and subsequently filter the search results using those same parameters. As an additional feature, the UniProt webpage provides the Gene Ontology terms accompanying each entry, along with the associated annotation score, allowing the user to assess the robustness of the annotation based on the evidence used to assign the term. Together, these features of the UniProt database may be a great benefit to future studies which seek to 1) create a more customized database, and 2) quickly and efficiently analyze individual entries.

**Prediction of PCD Proteins in *C. reinhardtii*** The computational strategy used to predict the *C. reinhardtii* proteins that participate in PCD utilized the BLASTp alignment tool to identify similarities between a *C. reinhardtii* protein and a protein in the custom database.

***Methods of Detecting Similarity*** Other tools, which utilize distinct algorithms to detect sequential similarity, may be utilized as an alternative to BLASTp. One of the most recent tools that have been developed for the detection of sequence similarities is Hmmer, which uses an approach known as probabilistic modeling to detect subtle signatures that are shared between a query amino acid sequence and individual protein families.[303,304] In some ways, this technique offers a sensitivity that is superior to methods that utilize a one-to-one approach of sequence alignment, as it allows for the detection of residues conserved within protein families, which may not be detectable by traditional pairwise alignment methods. Until recently, a limitation of computational abilities and the complex calculations associated with probabilistic modeling restricted the suitability of Hmmer to the analysis of only a few sequences.[305] However, the rapid advancements in the capacities of computational systems, in conjunction with the performance improvements seen in the latest version, Hmmer3, have expanded the usage of this tool to larger datasets.[305]

Another alignment tool, PSI-BLAST, utilizes an approach that is somewhat similar to Hmmer to detect distantly-related sequences. To detect such homologies, PSI-BLAST utilizes an iterative alignment strategy, wherein the user designates meaningful alignments following each iteration.[306] Aligned positions are subsequently used to query the database a second time, following which the user again designates alignments that are deemed to be meaningful. This process is repeated until no new subjects emerge in the alignments. One disadvantage of PSI-BLAST is that, because of the requirement for manual designation of meaningful alignments, it is not suitable for the analysis of more than a few sequences. Moreover, this technique introduces an additional potential for

human error, as the tedious nature of denoting meaningful alignments has the potential to result in the misidentification of a meaningful alignment. This will, in turn, affect the alignments that are reported in subsequent iterations of the BLASTp procedure.

*Sequential and Functional Similarity of Amino Acid Sequences* Alignment tools are, alone, unable to accurately predict functional similarities between two amino acid sequences. Many examples exist of proteins that, despite homology, perform vastly different functions in distinct organisms.[201] As such, the alignment of a query protein to a subject protein does not, in and of itself, guarantee similar functions and/or characteristics for the two proteins.[201] Secondly, many instances of unrelated proteins that share similar or identical functions in the cell have also been described. Therefore, the lack of an alignment between query and subject sequences cannot be used to infer that the two proteins are functionally dissimiliar.[201] Given the limitations of sequence alignment tools, further comparative bioinformatics analyses, such as the comparison of domain composition/location, should be utilized to supplement the results of an alignment between two amino acid sequences.

*Elucidation of Conserved Pathways* In light of the apparent role of the selected CrPPPs during *C. reinhardtii* PCD, it stands to reason that other components of the PCD pathways in which these proteins participate may also be conserved in *C. reinhardtii*. Future studies seeking to identify conserved aspects of PCD in *C. reinhardtii* may wish to take a predictive computational approach to addressing this question. One bioinformatics resource that may be useful in identifying conserved aspects of PCD in *C. reinhardtii* is through functional association networks. Tools such as the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING; https://string-db.org/) database may be ideal for

such purposes.[307] STRING is a web-based database of known and predicted protein-protein interactions. To use the database, the user produces the name(s) or amino acid sequence(s) of one or more proteins and specifies the source organism in which to search for interactions. From this, STRING produces an interaction network of the predicted and known proteins with which the searched protein interacts. Proteins are represented as nodes, and associations between proteins are designated by strings that connect two nodes.

Using one of the CrPPPs from this study as an example, it can be supposed that a future study might wish to identify components of the IRE-1 PCD pathways that are conserved in *C. reinhardtii*. First, known interactors of IRE-1 in another organism, suppose *Homo sapiens*, can be identified using the STRING database. Importantly, IRE-1 participates in processes other than PCD, and it is likely that not all proteins that interact with IRE-1 in *H. sapiens* will be of relevance to PCD. As such, PCD roles for these interactors should be confirmed at some point. Once proteins that interact with IRE-1 in *H. sapiens* have been identified, these amino acid sequences can be used to query the *C. reinhardtii* proteome. This will yield *C. reinhardtii* homologs of proteins that interact with IRE-1 in *H. sapiens*. If, in *H. sapiens*, these proteins participate in PCD, the *C. reinhardtii* homologs may serve as candidate proteins to study further in *C. reinhardtii* PCD. Moreover, if a *C. reinhardtii* homolog of the IRE-1 interactor is found to participate in PCD, further studies could be conducted to determine if the interaction between the two proteins is conserved in *C. reinhardtii*.

**Verification of CLiP Mutant Strains** In this study, mutant strains generated by random insertional mutagenesis were procured from the CLiP library. Several techniques,

as recommended by the curators of the CLiP library, were utilized to confirm the insertional locus reported for each strain.

*Analysis of Unexpected Results* In the instance where one or more unexpected results are obtained during the validation of CLiP mutant strains, an integrated analysis of all results is necessary to accurately predict the underlying cause for the unexpected results. As an example, the amplification of the *ire-1* insertional locus was unsuccessful when WT genomic DNA was supplied as the template for PCR amplification (Figure 22C). Given that PCR reactions performed subsequently relied on at least one component from this reaction, the results of these experiments may be useful to deduce the causative agent responsible for this failure. Indeed, the amplification of the cassette-genome junctions of the *ire-1* mutant strain indicates that cassette-genome junction amplified by the *ire-1* G1/ C1 primer combination was also unsuccessful (Figure 22D).

Taking into consideration the failure of both reactions that utilized *ire-1* G1 primer, we suspect that the *ire-1* G1 primer may be the cause for the failure of both PCR reactions. This prediction can be further supported by ruling out other possible explanations. One such explanation for this result might be that the template genomic DNA from the *ire-1* mutant strain was of poor quality. However, the successful amplification of the other cassette-genome junction in the *ire-1* mutant strain suggests genomic DNA of reasonable quality. An alternative explanation might employ the reasoning that the inability to amplify the *ire-1* G1/C1 cassette-genome junctions may be a result of the cassette primer. However, other PCR reactions that also utilized the C1 primer were successful in almost all instances. Moreover, the only other instance in which the amplification of a G1/C1 cassette-genome junction was unsuccessful (the *lsd-1*

mutant strain; Figure 25D) was correlated with the lack of a reported genomic sequence flanking the C1 side of the cassette by the CLiP library (Table 7). These observations suggest that, in the case of the *lsd-1* mutant strain utilized for this study, the unsuccessful amplification of the G1/C1 cassette-genome junction can potentially be attributed to a truncation of the cassette during insertion, which may have resulted in the loss of the C1 primer binding site. However, a flanking sequence was reported by the CLiP library for the *ire-1* mutant strain, indicating that such a truncation did not occur in the *ire-1* mutant strain.

With alternative explanations for the failure to amplify the *ire*-1 G1/C1 cassette-genome junction in the *ire-1* mutant strain ruled out, the hypothesis that the *ire-1* G1 primer was the reason for the failed PCR reactions seems reasonable. In this way, the results of the individual validation experiments, in conjunction with the insertional information reported for the strain, may potentially explain the unexpected results of one or more PCR reactions.

**Comparative Analysis of PCD Features in Mutant Strains** In this study, a reverse genetics approach was taken to determine if the loss of a selected protein-coding gene would cause an alteration of PCD phenotypes in *C. reinhardtii*. To achieve this, strains of *C. reinhardtii* were subjected to a PCD-inducing heat stress, and PCD features of mutant strains lacking one of the selected proteins were compared to the phenotypes of the parent WT strain.

***Unintended Effects of Mutagenesis*** In a general sense, the implicit logic that underlies the utilization of mutant strains to elucidate gene function is that, if a gene product participates in a given process, the loss/disruption of that gene should affect this

process in one way or another. If the gene product positively or negatively regulates the process, the loss of the gene should attenuate or augment the process, respectively. Given that even the simplest biological systems are exceptionally complex, a rudimentary "if this, then that" logic is, in and of itself, an oversimplified way of interpreting the results obtained from mutant strains. There are a number of factors that may confound the results obtained from mutant strains. In particular, a primary concern to consider when utilizing mutant strains is that changes in the observed phenotype may not necessarily indicate a role for the disrupted gene in PCD.

The expression of "off-target effects" is a phrase commonly used in molecular biology and refers collectively to any and all unintended consequences of a method or treatment. It is unclear whether this term can be applied to a strain generated by random mutagenesis, as such a technique, by definition, lacks a specific target. Nevertheless, any effects that were unintended by the creators of the strain(s) are factors that have the potential to alter the measured phenotype of a mutant strain. As such, these unintended effects, as they are referred to here, require consideration by anyone utilizing such strains. Several plausible unintended effects are discussed below.

*Pleiotropic Genes* The first confounding factor in the use of mutant strains stems from the pleiotropic nature of certain genes. A gene is said to be pleiotropic when it has an effect on multiple phenotypes. As such, the alteration and/or disruption of a pleiotropic gene might effect a phenotype in indirect and unpredictable ways.

As an example of how this notion might play out, consider the mitochondrial protein cytochrome c. As outlined in Chapter 1, cytochrome c plays an important role in the execution of apoptosis in animals. For the sake of this example, one can suppose that

cytochrome c is a completely uncharacterized protein, with the exception of some recent evidence which indicates that cytochrome c may play a role in positively-regulating apoptosis. Given only this information, one might hypothesize that murine cells deficient in cytochrome c will exhibit an extended lifespan as a result of an attenuated apoptotic process. To test this hypothesis, a researcher examines the lifespan of murine cells lacking cytochrome c and finds that, contrary to what was predicted, cells lacking cytochrome c instead exhibit a lifespan that is significantly shortened. Furthermore, when mouse embryos are engineered to express a non-functional variant of cytochrome c, the animals die prematurely during embryonic development. Based solely on these results, the researcher might conclude that cytochrome c may not positively-regulate apoptosis, as was previously predicted. Instead, the observation that cells which lack cytochrome c are more susceptible to undergoing apoptosis suggests that cytochrome c may actually function as an anti-apoptotic factor.

Abandoning the hypothetical example, the current knowledge of the apoptotic process indicates that the conclusion reached by the researcher is incorrect. It is known that cytochrome c is a pleiotropic gene which acts not only as a positive regulator of apoptosis, but also plays a well-established role in oxidative phosphorylation during the synthesis of ATP.[308] In fact, the metabolic role of cytochrome c was established long before the discovery that cytochrome c also plays a role in apoptosis. However, given that the researcher in the example above is not aware of cytochrome c's participation in oxidative phosphorylation, the conclusion that the scientist reaches is not an illogical one.

This example is actually an extensively-adapted account of a study which demonstrated that mice embryos lacking cytochrome c are resistant to stress-induced

apoptosis.[309] In the real study, because the metabolic role of cytochrome c was already very well-known, it was not difficult for the authors to deduce that the premature death of cytochrome c-deficient embryos could likely be attributed to defective oxidative phosphorylation, rather than an anti-apoptotic role for cytochrome c in PCD.

The purpose of the example was to demonstrate, in the context of PCD, that phenotypes exhibited by mutant strains harboring an altered gene are not always a reliable indicator of the gene's function in the process under investigation. This notion is of particular importance in the context of pleiotropic genes. In the case of cytochrome c, an alternative explanation for a result which seemed to contrast the hypothesis was readily available. This represents an entirely different scenario than the current study, as the function(s) of disrupted genes are largely or entirely uncharacterized in *C. reinhardtii*. Similarly to the fictionalized cytochrome c example, some or all of the selected *C. reinhardtii* gene products may play roles in normal cellular processes. As such, it is not known if the hypothetical non-PCD roles for these gene products were disrupted by the insertional mutation. Though the growth curves of the *bi-1*, *ire-1*, and *pig3* mutant strains did not indicate the presence of any metabolic defects in these strains (Figure 28), the plausibility of a non-PCD role for these genes cannot be ruled out by this assay alone. The possibility of pleiotropic genes is an important factor to consider, as the disruption of a non-PCD pathway might have an indirect effect on the PCD phenotypes.

*Gene Deletion and Genome Mutation* A recent study in *S. cerevisiae* demonstrated that the loss of a single gene may result in the evolution of other portions of the genome.[310] Notably, this phenomenon is not unique to yeast mutant strains, but has also been well documented in cancerous cells, wherein the development of an initial

mutation, termed a driver mutation, increases the likelihood of additional mutations, termed passenger mutations, developing at other genomic locations.[311,312]

As this study utilized mutant strains of *C. reinhardtii*, it seems appropriate to note the observations made in *S. cerevisiae*. The occurrence of such passenger mutations in mutant *C. reinhardtii* strains has yet to be explored. However, if the principle that an initial mutation drives genome evolution can be generalized to apply to all organisms, genome evolution may present a confounding factor to studies that utilize mutant strains of *C. reinhardtii*, as this evolution may generate strains that are divergent from the parent WT strain in multiple genes.[313] Given that the degree of genomic variation in *C. reinhardtii* lab strains suggests that the baseline rate of evolution for this organism may be relatively high to begin with, an initial mutation could conceivably exacerbate the evolutionary propensity for *C. reinhardtii* even further.[314,315]

*Cassette Cleavage and Integration* Transformation in *C. reinhardtii* is a notoriously difficult task.[167] The complications that arise during the transformation of *C. reinhardtii* have been primarily attributed to the abundance of nucleases that are present in *Chlamydomonas* cells. These nucleases target exogenous genetic material with high efficiency, making the insertion of an intact cassette into the *C. reinhardtii* genome a challenging endeavor.[262] Cleavage of the construct may result in a truncated cassette, which may or may not effectively disrupt the gene in which the cassette inserts.[173] Additionally, because the cleavage of the construct causes fragmentation of the cassette, nuclease activity might also result in the insertion of a cassette fragment into another region of the genome. This additional insertion could conceivably have an effect on PCD phenotypes in a manner that is independent of the mapped mutation. Moreover, were an

additional cassette fragment present in the genome, it might not be detected by the large-scale mapping methods used by the CLiP library, as this strategy requires the presence of the primer binding site.[173]

*Detecting Unintended Effects of Mutagenesis* Given the potential unintended effects that may be present in *C. reinhardtii* mutant strains, future studies that utilize such strains should seek to accomplish two primary goals. First, to rule out additional factors that may alter the phenotypes exhibited by a mutant strain, it should be determined that the mutagenesis technique, as well as the mutation itself, are absent of unintended genomic effects. Second, to confirm that the observed phenotypes are due to the disruption of the gene of interest, it should also be demonstrated that the phenotypes observed in the mutant strain are linked to the disrupted allele.

Whole-genome sequencing is likely the most direct method for identifying unintended genomic effects of mutagenesis.[162] Sequencing the genome of the mutant strain in question and comparing it to the genome of the parent WT strain would allow the researcher to 1) identify regions of the mutant genome which have diverged from the parent WT strain, and 2) identify and map any fragments of the cassette which may have inserted elsewhere in the genome of the mutant strain.

As an alternative to whole genome sequencing, Southern blotting may be utilized to detect the presence of cassette fragments that have inserted into the genome.[316] To utilize a Southern blot for this purpose, nucleotide probes, conjugated to a detectable reporter, should be designed for various regions of the insertional cassette. The genome of the mutant strain would then be fragmented with specified combinations of restriction endonucleases.[317] The resulting fragments would be separated by electrophoresis on an

agarose gel and transferred to a membrane, which would be treated with the pre-designed

cassette probes. The target cassette sequences would be visualized using a reporter

conjugated to the cassette probe.[317] Given that the genome is fragmented randomly, it

would be expected that, if a cassette fragment was inserted into a different portion of the

genome than the mapped cassette, it would be likely be located within a fragment of a

different molecular weight than that of the mapped cassette. Thus, if a given target

sequence is present in more than one portion of the genome, it would be expected that

multiple bands for the target sequence would be visualized on the blot.[317]

The downside of this approach compared to whole genome sequencing is that,

because the probes are specific to only a portion of the cassette, several probes would

need to be designed and validated for specificity to the cassette. Moreover, the Southern

blot is only useful for the detection of additional cassette sequences that have not been

mapped by the CLiP library and does not address the possibility of genomic mutation as a

result of the insertion. Thus, whole genome sequencing may represent a more thorough

approach to detecting unintended effects of mutagenesis, as both unmapped insertion

sites and genomic evolution might be detected by this method.

*Linking the Mutant Phenotype to the Disrupted Gene* To validate that a phenotype

observed in the mutant strain is a result of the disrupted gene, and not an unintended

effect of the mutagenesis, a useful strategy might be to correlate the presence of the

mutant allele to the altered phenotype.

One characteristic of *C. reinhardtii* that makes it particularly suitable for genetic

studies is its ability to alternate between haploid and diploid states.[318] Under standard

laboratory conditions, *C. reinhardtii* exists in a haploid state and undergoes mitotic cell

division. When specific conditions are met, *C. reinhardtii* cells may undergo sexual

reproduction. Importantly, *Chlamydomonas* cells may be classified by their mating type,

either + or -, and sexual reproduction can only occur between two cells of opposite

mating type.[318] To summarize the process, sexual reproduction involves fusion of the two

parental strains to form the zygospore, a diploid cell encased in a thick cell wall. The

zygospore can be induced to meiotically germinate, which results in four haploid

progeny. The daughter cells exist as a closely-oriented cluster for a brief period of time,

termed a tetrad, and physical separation of these progeny with a precise instrument

facilitates the formation of four clonal colonies that are distinct from one another.[318]

This characteristic of *C. reinhardtii* can be utilized to correlate the presence of a

mutant allele to an altered phenotype.[318,319] In the case where a mutant strain exhibits a

phenotype that is different than that of the parent WT strain, crossing the mutant strain

with the WT strain of the opposite mating type will produce four progeny strains, which

can be described as $X^+$, $X^-$, $Y^+$, and $Y^-$, where "X" and "Y" denote the two allelic variants

of the gene of interest, and "+" and "-" denote the mating type. Using this study as an

example, the CLiP mutant strains were generated in a – mating type background strain.

Therefore, mating a CLiP mutant with a + mating type strain would, as an end product,

produce the mutant allele in both the + and – genetic backgrounds. The differential

phenotype that was detected in the initial mutant strain can be assayed again, this time

using all progeny of the cross. If the differential phenotype is the result the mutant allele,

this phenotype should be present in all strains which possess this allele and not present in

all strains that do not possess the allele. Ideally, multiple backcrosses should be

performed in order to confirm that the altered phenotype follows the mutant allele through several generations.[319]

Another method that can be used to link the disruption of a gene to a particular phenotype is to use multiple mutant strains for the same gene of interest.[95] One can obtain multiple mutant strains with different mutations in the same gene and determine if these strains exhibit similar results for the phenotype in question. As is true of the backcross experiments, this strategy seeks to establish a correlation between the disruption of a specific gene and the presence of an altered phenotype.

An additional technique, called the mutant rescue, is another way by which a mutant allele and altered phenotype can be correlated.[320] This strategy seeks to link the disrupted allele to the mutant phenotype by demonstrating that a normal version of the gene prevents the occurrence of the altered phenotype. With this technique, the WT allele of the disrupted gene is expressed in the mutant strain.[321] If the altered phenotype is caused by the disrupted gene, the expression of the WT allele should restore the WT phenotype, thus "rescuing" the WT phenotype.[322]

### General Comments on the Assessment of PCD in C. reinhardtii

*Validation of Fluorescent Probes* Though the fluorescent probes used in this study have been previously utilized to assess PCD in *C. reinhardtii*, many have yet to be validated as accurate indicators of the presumed phenotype in this system. Attributing the signal of a probe to the purported phenotype in the cell would require direct validation by detailed biochemical experiments. An alternative approach which might facilitate the validation of individual probes is the combinatorial use of fluorescent probes that test for the same phenotype. Though this would not directly validate that a probe accurately and

specifically indicates a particular phenotype, it would demonstrate that distinct probes, which possess unique chemical compositions but test for the same phenotype, cause the same cells to fluoresce.

*Scoring of Fluorescent Cells* The assays which utilized fluorescent probes were carried out by manually scoring cells which were recognized as fluorescent under a hemacytometer. Though this method was useful for determining the percentage of fluorescent cells within the population, the time-consuming nature of the imaging process reduced the sampling frequency possible within a given period of time. Additionally, the analysis of captured images was also an approach with low throughput. The determination of the percentage of fluorescent cells within a population might be expedited by use of an automated system, such as a flow cytometer.

As an alternative to determining the percentage of fluorescent cells within a sample, a fluorescence plate reader could be used to measure the total fluorescence emitted by a sample. This approach may be particularly useful for preliminarily determining the overall kinetics of a phenotype of interest. The resulting data might then be used to determine the optimal times to measure that particular phenotype. One drawback to this approach is that changes in the total fluorescence intensity in a sample cannot necessarily be attributed to an increase in the percentage of fluorescent cells within the population, as brighter fluorescence from a select subpopulation could also cause a stronger signal.

*Population-Level Analysis* Of importance, the experiments conducted in this study utilized a population-level approach to assess PCD in *C. reinhardtii*. In the assays that utilized fluorescent probes, this approach can be seen in light of the fact that we 1)

detected fluorescence on a cell-to-cell basis using a binary "fluorescent/not fluorescent" scoring methods, 2) calculated the percentage of the cells in the sample that were fluorescing, and 3) compared the percentage of fluorescent cells in each mutant strain to the percentage of fluorescent cells in the parent WT strain for each time point in order to determine if the loss of a selected protein had an effect on the detected phenotype. The assays that did not use fluorescent probes also took a population approach to assessing PCD in *C. reinhardtii*. Genomic fragmentation was determined by qualitative assessment of DNA laddering of all cells in a sample. Additionally, the viability assay assessed the capacity for cell division by measuring the total area of all colonies present on each plate.

An alternative strategy by which PCD features might be assessed is by the analysis of individual cells rather than the population as a whole. For the studies that utilize fluorescent probes, quantitative fluorescence microscopy might be used to determine if a treatment has an effect on the intensity of fluorescence emitted from individual cells. Also, genomic fragmentation can be assessed in individual cells by methods utilizing fluorescence microscopy, such as the comet or TUNEL assays. Finally, the viability assay can be performed by scoring the number of colony-forming units, rather than measuring the total growth area of the population.

***Discussion of Results from the Annexin V Assay*** One experiment, the phosphatidylserine externalization assay, yielded results that were somewhat ambiguous (Figure 31). Somewhat unsurprisingly, the maximum percentage of the population exhibiting PS externalization peaked within the first sixty minutes of heat stress for each of the strains. Unexpectedly, however, was that the percentage of fluorescent cells decreased at subsequent time points. Moreover, a statistically-significant difference in the

percentage of Annexin V-positive cells was detected between the mutant strains and the parent WT strain at these later time points, indicating that the observed differences may be relevant in a biological context. A search of the literature did not provide a precedence for these results, and the only instances in which Annexin V staining is described to be reversible are those in which the cell, having not yet committed to PCD, returns to a metabolically-active and proliferative state.[323]

The decrease in the percentage of Annexin V-positive cells appears to begin at some point between one and two hours, during which time the cells remain subjected to heat stress. Intuitively, the notion that the cells would return to a metabolically-active state during persistent exposure to such adverse environmental conditions seems an unlikely possibility. The results of the other experiments performed in this study, when taken together, are also not indicative of a return to metabolic activity during the PCD-inducing stress (Figure 29, Figure 33). Furthermore, the observation that the decrease in the percentage of Annexin V-positive cells occurs at around the same time as changes in other PCD features indicates that this decrease may be relevant in the context of PCD. Consistent with this observation, the results of the metabolic activity (Figure 29), PM integrity (Figure 30), and viability (Figure 33) assays indicate that, at the population level, some pro-death event may be occurring between thirty and ninety minutes of heat stress.

Using these observations, we postulated that, as an alternative explanation to the reacquisition of metabolic activity, the observed decrease in the percentage of Annexin V positive cells may be a result of PS degradation during *C. reinhardtii* PCD, such that it is no longer bound by Annexin V. Though a brief search of the literature did not

immediately return studies that report the occurrence of such a phenomenon during PCD, several observations may support this prediction.

*Phosphatidylserine Decarboxylases and PCD* First, we inquired if a precedence for PS degradation is documented in the literature. We found that PS is regularly synthesized and degraded as a part of normal cellular metabolism.[265] Though the mechanisms of these processes appear to differ slightly between animals and yeast, the degradation of PS in both organisms is carried out by PS decarboxylase (PSD) enzymes. Notably, PSDs are localized to specific organelles within the cell; animal PSDs reside in the mitochondrial intermembrane space, anchored to the inner mitochondrial membrane, while PSDs in yeast localize to the mitochondrial and Golgi complex/vacuole membranes.[265]

From these findings, we reasoned that, since PCD involves the non-specific degradation of intracellular components, a subsequent release of PSDs from the associated organelle(s) may explain the decreased percentage of PS-externalizing cells observed at later time points during *C. reinhardtii* PCD. Alternatively, we also reasoned that, if a PSD in *C. reinhardtii* was localized to the mitochondrial intermembrane space, the formation of the mitochondrial permeability transition pore (MPTP) might be an another mechanism by which the PSD could be delocalized.

To corroborate the plausibility of an involvement of a *C. reinhardtii* PSD in the observed decrease in Annexin V-positive cells, we sought to determine if a PSD is predicted to be present in the *C. reinhardtii* proteome. To this end, a BLASTp search was carried out within the Phytozome webpage, wherein the amino acid sequences of PSDs from *H. sapiens* (AAH01482.1) and *S. cerevisiae* (AJT32496.1, AJS27943.1) were used

to query the predicted *C. reinhardtii* proteome (v5.5). A single predicted PSD in *C. reinhardtii* (Cre02.g119600.t1.1) was returned as a hit (results not shown). The possibility that this *C. reinhardtii* protein is responsible for the decrease in the percentage of Annexin V-positive cells at later stages of heat stress may be an avenue for future studies to investigate.

*Autophagy and Autophagic Cell Death* Of note, our research into PSDs led us to an another intriguing discovery: that PSDs catalyze the conversion of PS to another phospholipid, phosphatidylethanolamine (PE). While we initially predicted that this conversion might itself explain the decrease in Annexin V-positive cells, a study using artificial lipid bilayers demonstrated that Annexin V also possesses a high affinity for PE within the plasma membrane.[324] Given that Annexin V binds both PS and PE with high efficacies, we reasoned that the observed decrease in the percentage of Annexin V-positive cells likely cannot be explained solely by the intermembrane conversion of PS to PE. Importantly, however, a search of the literature for a role of PSDs during PCD led us to discover that high intracellular levels of PE are required for a process known as autophagy in both yeast and animal cells.[325]

As a brief overview, the term autophagy broadly refers to several distinct physiological processes that result in the degradation of unnecessary cytosolic components within the lysosome.[326] In this discussion, the term autophagy will be used in reference to macroautophagy, a specific type of autophagy. As a point of clarification, other forms of autophagy, such as chaperone-mediated autophagy and microautophagy, do exist.[327] These alternative variants of autophagy should not be confused with macroautophagy in the following discussion.

Autophagy plays an integral part of everyday cellular function, primarily by ensuring the protection of the cell from the detrimental consequences of normal metabolism, such as damaged mitochondria or aggregated proteins.[326] Autophagy is also important in the cellular response to non-metabolic stressors. In the early response to perturbations of cellular homeostasis, lysosomal degradation of autophagic substrates increases. In this context, autophagy plays a crucial cytoprotective role in promoting the adaptive recovery of homeostasis and ultimate survival of the cell during stress.[327] Intriguingly, the participants of autophagy do not act solely in a cytoprotective manner. In a contrasting role to that observed in the initial cellular stress response, autophagic components can also mediate the death of the cell. Consequently, such forms of PCD are referred to as "autophagic cell death".[327]

Importantly, the contrasting roles of autophagy during stress can be distinguished by the pharmacogenic intervention of autophagic machinery.[47] Such an inhibition promotes cell death in instances where autophagy acts in cytoprotection, while the impedance of cell death by pharmacogenic intervention of autophagy indicates a pro-PCD role for autophagy.[48] This approach might be utilized to test the hypothesis that autophagic machinery plays a role in *C. reinhardtii* PCD.

*Tentative Models for Heat-Induced PCD in C. reinhardtii* The evidence presented in the preceding discussions suggests that several possible mechanisms may explain the decrease in the percentage of Annexin V positive cells observed in this study.

In the first model, the execution of PCD causes the degradation of intracellular organelles. This results in the subsequent delocalization of PSDs, which then convert membrane PS to PE. As an alternative model, the release of PSDs may not be caused by

the degradation of organelles, but rather by the formation of the MPTP, which causes the release of mitochondrial intermembrane space components into the cytosol. As each of these scenarios take place during the execution stage of PCD, both models implicate the execution of PCD as the cause of the decrease in the percentage of Annexin V-positive cells.

An additional model can also be conceived, in which the decrease in the percentage of Annexin V-positive cells is indicative of a switch between the pro-survival functions of autophagy and the pro-death mechanisms of autophagic cell death. Under this model, *C. reinhardtii* cells initiate autophagy in the early stages of heat stress and maintain metabolic activity (Figure 29) and plasma membrane integrity during this period (Figure 30). PS is externalized during this initial response to the heat stress and may eventually be converted into PE. Given that Annexin V has a high affinity for both PS and PE, this conversion could very well remain undetected by the methods utilized in this study. The reversal of Annexin V-positivity at later time points might be explained by the translocation of PE into the intracellular space. However, given that 1) intracellular PE participates in autophagy, a pro-survival process, and 2) the decrease in the percentage of Annexin V-positive cells occurs within approximately the same time frame as PCD phenotypes in the other assays performed, this prediction may not hold true.

**Concluding Remarks**

We conducted the preceding work with the intention to shed light on the molecular mechanisms by which PCD occurs in *C. reinhardtii*. As an ancillary benefit, we deem that this work will also facilitate additional studies into the molecular basis *C. reinhardtii* PCD. The results of these studies contribute to the diminutive list of *C.*

*reinhardtii* proteins that may participate in PCD. The presence of homologs in *C. reinhardtii* to PCD proteins of diverse lineages has strong implications for the evolution of PCD in higher organisms, as it is in agreement with the hypothesis that some aspects of PCD have been conserved throughout evolutionary history. As a result, the results of this work may aid in the establishment of *C. reinhardtii* as a novel system in which to study PCD. Though we consider this work to be merely a drop in the proverbial bucket, the findings obtained from these studies generate many new biological questions..

While differences in PCD phenotypes were detected between each of the *C. reinhardtii* mutant strains and the parental WT strain, the implications of these findings with respect to PCD at the cellular level have yet to be fully determined. Although the purpose of this study was to determine some of the molecular basis of PCD in *C. reinhardtii*, a more detailed understanding of this process at the cellular level may help to frame the results of this study in a broader context.

# REFERENCES

1. Elmore, S. Apoptosis: A Review of Programmed Cell Death. *Toxicol. Pathol.* **35,** 495–516 (2007).

2. Galluzzi, L. *et al.* Essential versus accessory aspects of cell death: recommendations of the NCCD 2015. *Cell Death Differ.* **22,** 58–73 (2015).

3. Lockshin, R. A. & Williams, C. M. Programmed cell death—II. Endocrine potentiation of the breakdown of the intersegmental muscles of silkmoths. *J. Insect Physiol.* **10,** 643–649 (1964).

4. Diamantis, A., Magiorkinis, E., Sakorafas, G. H. & Androutsos, G. A Brief History of Apoptosis: From Ancient to Modern Times. *Oncol. Res. Treat.* **31,** 702–706 (2008).

5. Kerr, J. F., Wyllie, A. H. & Currie, A. R. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. *Br. J. Cancer* **26,** 239–257 (1972).

6. Evan, G. I. & Vousden, K. H. Proliferation, cell cycle and apoptosis in cancer. *Nature* **411,** 342–348 (2001).

7. Sulston, J. E. & Horvitz, H. R. Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. *Dev. Biol.* **56,** 110–156 (1977).

8. Ellis, H. M. & Horvitz, H. R. Genetic control of programmed cell death in the nematode C. elegans. *Cell* **44,** 817–829 (1986).

9. White, K. *et al.* Genetic control of programmed cell death in Drosophila. *Science* **264,** 677–683 (1994).

10. Lowe, S. W. & Lin, A. W. Apoptosis in cancer. *Carcinogenesis* **21,** 485–495 (2000).

11. Immunohistochemical evidence for apoptosis in Alzheimer's di... : NeuroReport. *LWW* Available at:

http://journals.lww.com/neuroreport/Fulltext/1994/12000/Immunohistochemical_evide nce_for_apoptosis_in.31.aspx. (Accessed: 23rd May 2017)

12. Lev, N., Melamed, E. & Offen, D. Apoptosis and Parkinson's disease. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **27,** 245–250 (2003).

13. Brill, A., Torchinsky, A., Carp, H. & Toder, V. The Role of Apoptosis in Normal and Abnormal Embryonic Development. *J. Assist. Reprod. Genet.* **16,** 512–519 (1999).

14. Print, C. G. & Loveland, K. L. Germ cell suicide: new insights into apoptosis during spermatogenesis. *BioEssays* **22,** 423–430

15. Jacobson, M. D., Weil, M. & Raff, M. C. Programmed Cell Death in Animal Development. *Cell* **88,** 347–354 (1997).

16. Vaux, D. L. & Korsmeyer, S. J. Cell death in development. *Cell* **96,** 245–254 (1999).

17. Hautegem, T. V., Waters, A. J., Goodrich, J. & Nowack, M. K. Only in dying, life: programmed cell death during plant development. *Trends Plant Sci.* **20,** 102–113 (2015).

18. Pellettieri, J. & Alvarado, A. S. Cell Turnover and Adult Tissue Homeostasis: From Humans to Planarians. *Annu. Rev. Genet.* **41,** 83–105 (2007).

19. Allocati, N., Masulli, M., Di Ilio, C. & De Laurenzi, V. Die for the community: an overview of programmed cell death in bacteria. *Cell Death Dis.* **6,** e1609 (2015).

20. Albuquerque, P. & Casadevall, A. Quorum sensing in fungi – a review. *Med. Mycol.* **50,** 337–345 (2012).

21. Miller, M. B. & Bassler, B. L. Quorum sensing in bacteria. *Annu. Rev. Microbiol.* **55,** 165–199 (2001).

22. Nedelcu, A. M., Driscoll, W. W., Durand, P. M., Herron, M. D. & Rashidi, A. On the paradigm of altruistic suicide in the unicellular world. *Evol. Int. J. Org. Evol.* **65,** 3–20 (2011).

23. Tan, I. S. & Ramamurthi, K. S. Spore formation in Bacillus subtilis. *Environ. Microbiol. Rep.* **6,** 212–225 (2014).

24. Al-Hinai, M. A., Jones, S. W. & Papoutsakis, E. T. The Clostridium Sporulation Programs: Diversity and Preservation of Endospore Differentiation. *Microbiol. Mol. Biol. Rev.* **79,** 19–37 (2015).

25. Shimkets, L. J. Intercellular signaling during fruiting-body development of Myxococcus xanthus. *Annu. Rev. Microbiol.* **53,** 525–549 (1999).

26. O'Connor, K. A. & Zusman, D. R. Development in Myxococcus xanthus involves differentiation into two cell types, peripheral rods and spores. *J. Bacteriol.* **173,** 3318–3333 (1991).

27. Muñoz-Dorado, J., Marcos-Torres, F. J., García-Bravo, E., Moraleda-Muñoz, A. & Pérez, J. Myxobacteria: Moving, Killing, Feeding, and Surviving Together. *Front. Microbiol.* **7,** (2016).

28. Fets, L., Kay, R. & Velazquez, F. Dictyostelium. *Curr. Biol.* **20,** R1008–R1010 (2010).

29. Katz, E. R. Dictyostelium: Evolution, Cell Biology, and the Development of Multicellularity. By Richard H Kessin. *Q. Rev. Biol.* **77,** 453–454 (2002).

30. Srinivasan, S., Alexander, H. & Alexander, S. The Dictyostelium fruiting body – a structure of cells and cellulose. *Trends Cell Biol.* **10,** 315 (2000).

31. Webb, J. S., Givskov, M. & Kjelleberg, S. Bacterial biofilms: prokaryotic adventures in multicellularity. *Curr. Opin. Microbiol.* **6,** 578–585 (2003).

32. Das, T., Sehar, S. & Manefield, M. The roles of extracellular DNA in the structural integrity of extracellular polymeric substance and bacterial biofilm development. *Environ. Microbiol. Rep.* **5,** 778–786 (2013).

33. Montanaro, L. *et al.* Extracellular DNA in biofilms. *Int. J. Artif. Organs* **34,** 824–831 (2011).

34. Allesen-Holm, M. *et al.* A characterization of DNA release in Pseudomonas aeruginosa cultures and biofilms. *Mol. Microbiol.* **59,** 1114–1128 (2006).

35. Grande, R. *et al.* Extracellular DNA in Helicobacter pylori biofilm: a backstairs rumour. *J. Appl. Microbiol.* **110,** 490–498 (2011).

36. González-Pastor, J. E. Cannibalism: a social behavior in sporulating Bacillus subtilis. *FEMS Microbiol. Rev.* **35,** 415–424 (2011).

37. Claverys, J.-P. & Håvarstein, L. S. Cannibalism and fratricide: mechanisms and raisons d'être. *Nat. Rev. Microbiol.* **5,** 219–229 (2007).

38. Greene, D. A., Stevens, M. J., Obrosova, I. & Feldman, E. L. Glucose-induced oxidative stress and programmed cell death in diabetic neuropathy. *Eur. J. Pharmacol.* **375,** 217–223 (1999).

39. Davidson, J. F., Whyte, B., Bissinger, P. H. & Schiestl, R. H. Oxidative stress is involved in heat-induced cell death in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci.* **93,** 5116–5121 (1996).

40. Chen, Y.-R., Wang, X., Templeton, D., Davis, R. J. & Tan, T.-H. The Role of c-Jun N-terminal Kinase (JNK) in Apoptosis Induced by Ultraviolet C and γ Radiation

DURATION OF JNK ACTIVATION MAY DETERMINE CELL DEATH AND PROLIFERATION. *J. Biol. Chem.* **271,** 31929–31936 (1996).

41. Abramov, A. Y., Scorziello, A. & Duchen, M. R. Three Distinct Mechanisms Generate Oxygen Free Radicals in Neurons and Contribute to Cell Death during Anoxia and Reoxygenation. *J. Neurosci.* **27,** 1129–1138 (2007).

42. Greenberg, J. T. & Yao, N. The role and regulation of programmed cell death in plant–pathogen interactions. *Cell. Microbiol.* **6,** 201–211

43. Roos, W. P. & Kaina, B. DNA damage-induced cell death by apoptosis. *Trends Mol. Med.* **12,** 440–450 (2006).

44. Ryter, S. W. *et al.* Mechanisms of Cell Death in Oxidative Stress. *Antioxid. Redox Signal.* **9,** 49–89 (2006).

45. Boyce, M. & Yuan, J. Cellular response to endoplasmic reticulum stress: a matter of life or death. *Cell Death Differ.* **13,** 363–373 (2006).

46. Reape, T. J. & McCabe, P. F. Apoptotic-like regulation of programmed cell death in plants. *Apoptosis* **15,** 249–256 (2010).

47. Galluzzi, L. *et al.* Molecular mechanisms of cell death: recommendations of the Nomenclature Committee on Cell Death 2018. *Cell Death Differ.* **25,** 486–541 (2018).

48. Galluzzi, L. *et al.* Molecular definitions of cell death subroutines: recommendations of the Nomenclature Committee on Cell Death 2012. *Cell Death Differ.* **19,** 107–120 (2012).

49. Ichim, G. & Tait, S. W. G. A fate worse than death: apoptosis as an oncogenic process. *Nat. Rev. Cancer* **16,** 539–548 (2016).

50. Tait, S. W. G. & Green, D. R. Mitochondria and cell death: outer membrane permeabilization and beyond. *Nat. Rev. Mol. Cell Biol.* **11,** 621–632 (2010).

51. Kroemer, G., Galluzzi, L. & Brenner, C. Mitochondrial membrane permeabilization in cell death. *Physiol. Rev.* **87,** 99–163 (2007).

52. Susin, S. A. *et al.* Molecular characterization of mitochondrial apoptosis-inducing factor. *Nature* **397,** 441–446 (1999).

53. Daugas, E. *et al.* Mitochondrio-nuclear translocation of AIF in apoptosis and necrosis. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **14,** 729–739 (2000).

54. Li, L. Y., Luo, X. & Wang, X. Endonuclease G is an apoptotic DNase when released from mitochondria. *Nature* **412,** 95–99 (2001).

55. Haupt, S., Berger, M., Goldberg, Z. & Haupt, Y. Apoptosis - the p53 network. *J. Cell Sci.* **116,** 4077–4085 (2003).

56. Kastenhuber, E. R. & Lowe, S. W. Putting p53 in Context. *Cell* **170,** 1062–1078 (2017).

57. Bortner, C. D. & Cidlowski, J. A. A necessary role for cell shrinkage in apoptosis. *Biochem. Pharmacol.* **56,** 1549–1559 (1998).

58. Deschesnes, R. G., Huot, J., Valerie, K. & Landry, J. Involvement of p38 in Apoptosis-associated Membrane Blebbing and Nuclear Condensation. *Mol. Biol. Cell* **12,** 1569–1582 (2001).

59. Enari, M. *et al.* A caspase-activated DNase that degrades DNA during apoptosis, and its inhibitor ICAD. *Nature* **391,** 43–50 (1998).

60. Ly, J. D., Grubb, D. R. & Lawen, A. The mitochondrial membrane potential (Δψm) in apoptosis; an update. *Apoptosis* **8,** 115–128 (2003).

61. Simon, H.-U., Haj-Yehia, A. & Levi-Schaffer, F. Role of reactive oxygen species (ROS) in apoptosis induction. *Apoptosis* **5,** 415–418 (2000).

62. Koopman, G. *et al.* Annexin V for flow cytometric detection of phosphatidylserine expression on B cells undergoing apoptosis. *Blood* **84,** 1415–1420 (1994).

63. Galluzzi, L. & Kroemer, G. Necroptosis: A Specialized Pathway of Programmed Necrosis. *Cell* **135,** 1161–1163 (2008).

64. Christofferson, D. E. & Yuan, J. Necroptosis as an alternative form of programmed cell death. *Curr. Opin. Cell Biol.* **22,** 263–268 (2010).

65. Bergsbaken, T., Fink, S. L. & Cookson, B. T. Pyroptosis: host cell death and inflammation. *Nat. Rev. Microbiol.* **7,** 99–109 (2009).

66. Dixon, S. J. *et al.* Ferroptosis: An Iron-Dependent Form of Nonapoptotic Cell Death. *Cell* **149,** 1060–1072 (2012).

67. Gozuacik, D. & Kimchi, A. Autophagy and Cell Death. *Curr. Top. Dev. Biol.* **78,** 217–245 (2007).

68. Dickman, M., Williams, B., Li, Y., de Figueiredo, P. & Wolpert, T. Reassessing apoptosis in plants. *Nat. Plants* **3,** 773–779 (2017).

69. Doorn, W. G. van & Woltering, E. J. Many ways to exit? Cell death categories in plants. *Trends Plant Sci.* **10,** 117–122 (2005).

70. van Doorn, W. G. *et al.* Morphological classification of plant cell deaths. *Cell Death Differ.* **18,** 1241–1246 (2011).

71. van Doorn, W. G. Classes of programmed cell death in plants, compared to those in animals. *J. Exp. Bot.* **62,** 4749–4761 (2011).

72. Hara-Nishimura, I. & Hatsugai, N. The role of vacuole in plant cell death. *Cell Death Differ.* **18,** 1298–1304 (2011).

73. Carmona-Gutierrez, D., Fröhlich, K.-U., Kroemer, G. & Madeo, F. Metacaspases are caspases. Doubt no more. *Cell Death Differ.* **17,** 377–378 (2010).

74. Enoksson, M. & Salvesen, G. S. Metacaspases are not caspases – always doubt. *Cell Death Differ.* **17,** 1221 (2010).

75. Tsiatsiani, L. *et al.* Metacaspases. *Cell Death Differ.* **18,** 1279–1288 (2011).

76. Sundström, J. F. *et al.* Tudor staphylococcal nuclease is an evolutionarily conserved component of the programmed cell death degradome. *Nat. Cell Biol.* **11,** 1347–1354 (2009).

77. Bozhkov, P. V. & Lam, E. Green death: revealing programmed cell death in plants. *Cell Death Differ.* **18,** 1239–1240 (2011).

78. Li, Y. & Dickman, M. Processing of AtBAG6 triggers autophagy and fungal resistance. *Plant Signal. Behav.* **11,** e1175699 (2016).

79. Williams, B., Kabbage, M., Britt, R. & Dickman, M. B. AtBAG7, an Arabidopsis Bcl-2-associated athanogene, resides in the endoplasmic reticulum and is involved in the unfolded protein response. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 6088–6093 (2010).

80. Li, Y., Williams, B. & Dickman, M. Arabidopsis B-cell lymphoma2 (Bcl-2)-associated athanogene 7 (BAG7)-mediated heat tolerance requires translocation, sumoylation and binding to WRKY29. *New Phytol.* **214,** 695–705 (2017).

81. Reape, T. J., Brogan, N. P. & McCabe, P. F. Mitochondrion and Chloroplast Regulation of Plant Programmed Cell Death. in *Plant Programmed Cell Death* 33–53 (Springer, Cham, 2015). doi:10.1007/978-3-319-21033-9_2

82. Balk, J., Chew, S. K., Leaver, C. J. & McCabe, P. F. The intermembrane space of plant mitochondria contains a DNase activity that may be involved in programmed cell death. *Plant J. Cell Mol. Biol.* **34,** 573–583 (2003).

83. Jones, J. D. G. & Dangl, J. L. The plant immune system. *Nature* **444,** 323–329 (2006).

84. Hatsugai, N. *et al.* A novel membrane fusion-mediated plant immunity against bacterial pathogens. *Genes Dev.* **23,** 2496–2506 (2009).

85. Hatsugai, N. *et al.* A plant vacuolar protease, VPE, mediates virus-induced hypersensitive cell death. *Science* **305,** 855–858 (2004).

86. Asada, K. Production and Scavenging of Reactive Oxygen Species in Chloroplasts and Their Functions. *Plant Physiol.* **141,** 391–396 (2006).

87. Liu, Y. *et al.* Chloroplast-generated reactive oxygen species are involved in hypersensitive response-like cell death mediated by a mitogen-activated protein kinase cascade. *Plant J.* **51,** 941–954 (2007).

88. Mühlenbock, P. *et al.* Chloroplast signaling and LESION SIMULATING DISEASE1 regulate crosstalk between light acclimation and immunity in Arabidopsis. *Plant Cell* **20,** 2339–2356 (2008).

89. Van Aken, O. & Van Breusegem, F. Licensed to Kill: Mitochondria, Chloroplasts, and Cell Death. *Trends Plant Sci.* **20,** 754–766 (2015).

90. Wang, H. *et al.* Induction of caspase-3-like activity in rice following release of cytochrome-f from the chloroplast and subsequent interaction with the ubiquitin-proteasome system. *Sci. Rep.* **4,** 5989 (2014).

91.Zuppini, A., Gerotto, C., Moscatiello, R., Bergantino, E. & Baldan, B. Chlorella saccharophila cytochrome f and its involvement in the heat shock response. *J. Exp. Bot.* **60,** 4189–4200 (2009).

92.Peters, J. S. & Chin, C. Evidence for cytochrome f involvement in eggplant cell death induced by palmitoleic acid. *Cell Death Differ.* **12,** 405–407 (2005).

93.Carmona-Gutierrez, D. *et al.* Apoptosis in yeast: triggers, pathways, subroutines. *Cell Death Differ.* **17,** 763–773 (2010).

94.Madeo, F. *et al.* Caspase-dependent and caspase-independent cell death pathways in yeast. *Biochem. Biophys. Res. Commun.* **382,** 227–231 (2009).

95.Carmona-Gutierrez, D. *et al.* Guidelines and recommendations on yeast cell death nomenclature. *Microb. Cell* **5,** 4–31

96.Madeo, F. *et al.* Oxygen Stress: A Regulator of Apoptosis in Yeast. *J. Cell Biol.* **145,** 757–767 (1999).

97.Granot, D., Levine, A. & Dor-Hefetz, E. Sugar-induced apoptosis in yeast cells. *FEMS Yeast Res.* **4,** 7–13 (2003).

98.Kitagaki, H., Araki, Y., Funato, K. & Shimoi, H. Ethanol-induced death in yeast exhibits features of apoptosis mediated by mitochondrial fission pathway. *FEBS Lett.* **581,** 2935–2942 (2007).

99.Sharon, A., Finkelstein, A., Shlezinger, N. & Hatam, I. Fungal apoptosis: function, genes and gene function. *FEMS Microbiol. Rev.* **33,** 833–854 (2009).

100.Fahrenkrog, B., Sauder, U. & Aebi, U. The S. cerevisiae HtrA-like protein Nma111p is a nuclear serine protease that mediates yeast apoptosis. *J. Cell Sci.* **117,** 115–126 (2004).

101. Fahrenkrog, B. Nma111p, the pro-apoptotic HtrA-like nuclear serine protease in Saccharomyces cerevisiae: a short survey. *Biochem. Soc. Trans.* **39,** 1499–1501 (2011).

102. Walter, D., Wissing, S., Madeo, F. & Fahrenkrog, B. The inhibitor-of-apoptosis protein Bir1p protects against apoptosis in S. cerevisiae and is a substrate for the yeast homologue of Omi/HtrA2. *J. Cell Sci.* **119,** 1843–1851 (2006).

103. Li, W. *et al.* Yeast AMID Homologue Ndi1p Displays Respiration-restricted Apoptotic Activity and Is Involved in Chronological Aging. *Mol. Biol. Cell* **17,** 1802–1811 (2006).

104. Wissing, S. *et al.* An AIF orthologue regulates apoptosis in yeast. *J. Cell Biol.* **166,** 969–974 (2004).

105. Büttner, S. *et al.* Endonuclease G regulates budding yeast life and death. *Mol. Cell* **25,** 233–246 (2007).

106. Yang, H., Ren, Q. & Zhang, Z. Cleavage of Mcd1 by Caspase-like Protease Esp1 Promotes Apoptosis in Budding Yeast. *Mol. Biol. Cell* **19,** 2127–2134 (2008).

107. Hauptmann, P. & Lehle, L. Kex1 protease is involved in yeast cell death induced by defective N-glycosylation, acetic acid, and chronological aging. *J. Biol. Chem.* **283,** 19151–19163 (2008).

108. Cooper, K. F., Khakhina, S., Kim, S. K. & Strich, R. Stress-induced nuclear-to-cytoplasmic translocation of cyclin C promotes mitochondrial fission in yeast. *Dev. Cell* **28,** 161–173 (2014).

109. Büttner, S. *et al.* A yeast BH3-only protein mediates the mitochondrial pathway of apoptosis. *EMBO J.* **30,** 2779–2792 (2011).

110. Qiu, J., Yoon, J.-H. & Shen, B. Search for apoptotic nucleases in yeast: role of Tat-D nuclease in apoptotic DNA degradation. *J. Biol. Chem.* **280,** 15370–15379 (2005).

111. Fannjiang, Y. *et al.* Mitochondrial fission proteins regulate programmed cell death in yeast. *Genes Dev.* **18,** 2785–2797 (2004).

112. Strich, R. Programmed Cell Death Initiation and Execution in Budding Yeast. *Genetics* **200,** 1003–1014 (2015).

113. Deponte, M. Programmed cell death in protists. *Biochim. Biophys. Acta* **1783,** 1396–1405 (2008).

114. Shemarova, I. V. Signaling mechanisms of apoptosis-like programmed cell death in unicellular eukaryotes. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **155,** 341–353 (2010).

115. Mendoza, L. *et al.* Ehp53, an Entamoeba histolytica protein, ancestor of the mammalian tumour suppressor p53. *Microbiol. Read. Engl.* **149,** 885–893 (2003).

116. Bayles, K. W. Bacterial programmed cell death: making sense of a paradox. *Nat. Rev. Microbiol.* **12,** 63–69 (2014).

117. Unterholzner, S. J., Poppenberger, B. & Rozhon, W. Toxin–antitoxin systems. *Mob. Genet. Elem.* **3,** (2013).

118. Page, R. & Peti, W. Toxin-antitoxin systems in bacterial growth arrest and persistence. *Nat. Chem. Biol.* **12,** 208–214 (2016).

119. Ramisetty, B. C. M., Natarajan, B. & Santhosh, R. S. mazEF-mediated programmed cell death in bacteria: "What is this?" *Crit. Rev. Microbiol.* **41,** 89–100 (2015).

120. Kumar, S. & Engelberg-Kulka, H. Quorum sensing peptides mediating interspecies bacterial cell death as a novel class of antimicrobial agents. *Curr. Opin. Microbiol.* **21,** 22–27 (2014).

121. Amitai, S., Yassin, Y. & Engelberg-Kulka, H. MazF-mediated cell death in Escherichia coli: a point of no return. *J. Bacteriol.* **186,** 8295–8300 (2004).

122. Wang, I. N., Smith, D. L. & Young, R. Holins: the protein clocks of bacteriophage infections. *Annu. Rev. Microbiol.* **54,** 799–825 (2000).

123. Ranjit, D. K., Endres, J. L. & Bayles, K. W. Staphylococcus aureus CidA and LrgA Proteins Exhibit Holin-Like Properties ▿. *J. Bacteriol.* **193,** 2468–2476 (2011).

124. Tatischeff, I. *et al.* Inhibition of multicellular development switches cell death of Dictyostelium discoideum towards mammalian-like unicellular apoptosis. *Eur. J. Cell Biol.* **80,** 428–441 (2001).

125. Peeters, S. H. & de Jonge, M. I. For the greater good: Programmed cell death in bacterial communities. *Microbiol. Res.* **207,** 161–169 (2018).

126. Chen, S. & Dickman, M. B. Bcl-2 family members localize to tobacco chloroplasts and inhibit programmed cell death induced by chloroplast-targeted herbicides. *J. Exp. Bot.* **55,** 2617–2623 (2004).

127. Lacomme, C. & Santa Cruz, S. Bax-induced cell death in tobacco is similar to the hypersensitive response. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 7956–7961 (1999).

128. Ligr, M. *et al.* Mammalian Bax triggers apoptotic changes in yeast. *FEBS Lett.* **438,** 61–65 (1998).

129. Saraiva, L., Silva, R. D., Pereira, G., Gonçalves, J. & Côrte-Real, M. Specific modulation of apoptosis and Bcl-xL phosphorylation in yeast by distinct mammalian protein kinase C isoforms. *J. Cell Sci.* **119,** 3171–3181 (2006).

130. Chen, S.-R., Dunigan, D. D. & Dickman, M. B. Bcl-2 family members inhibit oxidative stress-induced programmed cell death in Saccharomyces cerevisiae. *Free Radic. Biol. Med.* **34,** 1315–1325 (2003).

131. Greenhalf, W., Stephan, C. & Chaudhuri, B. Role of mitochondria and C-terminal membrane anchor of Bcl-2 in Bax induced growth arrest and mortality in Saccharomyces cerevisiae. *FEBS Lett.* **380,** 169–175 (1996).

132. Gross, A. *et al.* Biochemical and genetic analysis of the mitochondrial response of yeast to BAX and BCL-X(L). *Mol. Cell. Biol.* **20,** 3125–3136 (2000).

133. Pang, X. *et al.* Active Bax and Bak are functional holins. *Genes Dev.* **25,** 2278–2290 (2011).

134. On the origin, evolution, and nature of programmed cell death: a timeline of four billion years. *Publ. Online 20 March 2002 Doi101038sjcdd4400950* **9,** (2002).

135. Fuchs, Y. & Steller, H. Programmed Cell Death in Animal Development and Disease. *Cell* **147,** 742–758 (2011).

136. Spiess, C., Beil, A. & Ehrmann, M. A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein. *Cell* **97,** 339–347 (1999).

137. Yamaguchi, M. *et al.* Loss of the plastid envelope protein AtLrgB causes spontaneous chlorotic cell death in Arabidopsis thaliana. *Plant Cell Physiol.* **53,** 125–134 (2012).

138. Nedelcu, A. M. Comparative genomics of phylogenetically diverse unicellular eukaryotes provide new insights into the genetic basis for the evolution of the programmed cell death machinery. *J. Mol. Evol.* **68,** 256–268 (2009).

139. Lewis, L. A. & McCourt, R. M. Green algae and the origin of land plants. *Am. J. Bot.* **91,** 1535–1556 (2004).

140. Leliaert, F., Verbruggen, H. & Zechman, F. W. Into the deep: new discoveries at the base of the green plant phylogeny. *BioEssays News Rev. Mol. Cell. Dev. Biol.* **33,** 683–692 (2011).

141. Bowman, J. L., Floyd, S. K. & Sakakibara, K. Green Genes—Comparative Genomics of the Green Branch of Life. *Cell* **129,** 229–234 (2007).

142. Jiménez, C. *et al.* Different ways to die: cell death modes of the unicellular chlorophyte Dunaliella viridis exposed to various environmental stresses are mediated by the caspase-like activity DEVDase. *J. Exp. Bot.* **60,** 815–828 (2009).

143. Zuppini, A., Andreoli, C. & Baldan, B. Heat Stress: an Inducer of Programmed Cell Death in Chlorella saccharophila. *Plant Cell Physiol.* **48,** 1000–1009 (2007).

144. Darehshouri, A., Affenzeller, M. & Lütz-Meindl, U. Cell death upon H2O2 induction in the unicellular green alga Micrasterias. *Plant Biol.* **10,** 732–745 (2008).

145. Berges, J. A. & Falkowski, P. G. Physiological stress and cell death in marine phytoplankton: Induction of proteases in response to nitrogen or light limitation. *Limnol. Oceanogr.* **43,** 129–135 (1998).

146. Nedelcu, A. M. Evidence for p53-like-mediated stress responses in green algae. *FEBS Lett.* **580,** 3013–3017 (2006).

147. Kasuba, K. C., Vavilala, S. L. & D'Souza, J. S. Apoptosis-like cell death in unicellular photosynthetic organisms — A review. *Algal Res.* **12,** 126–133 (2015).

148. Merchant, S. S. *et al.* The Chlamydomonas Genome Reveals the Evolution of Key Animal and Plant Functions. *Science* **318,** 245–250 (2007).

149. Pazour, G. J. *et al.* Chlamydomonas IFT88 and its mouse homologue, polycystic kidney disease gene tg737, are required for assembly of cilia and flagella. *J. Cell Biol.* **151,** 709–718 (2000).

150. Yang, P. *et al.* Radial spoke proteins of Chlamydomonas flagella. *J. Cell Sci.* **119,** 1165–1174 (2006).

151. Merchant, S. S., Kropat, J., Liu, B., Shaw, J. & Warakanont, J. TAG, you're it! Chlamydomonas as a reference organism for understanding algal triacylglycerol accumulation. *Curr. Opin. Biotechnol.* **23,** 352–363 (2012).

152. Maul, J. E. *et al.* The Chlamydomonas reinhardtii plastid chromosome: islands of genes in a sea of repeats. *Plant Cell* **14,** 2659–2679 (2002).

153. Grossman, A. R. *et al.* Chlamydomonas reinhardtii at the crossroads of genomics. *Eukaryot. Cell* **2,** 1137–1150 (2003).

154. Minagawa, J. & Tokutsu, R. Dynamic regulation of photosynthesis in Chlamydomonas reinhardtii. *Plant J. Cell Mol. Biol.* **82,** 413–428 (2015).

155. Brueggeman, A. J. *et al.* Activation of the carbon concentrating mechanism by CO2 deprivation coincides with massive transcriptional restructuring in Chlamydomonas reinhardtii. *Plant Cell* **24,** 1860–1875 (2012).

156. Magneschi, L. *et al.* A mutant in the ADH1 gene of Chlamydomonas reinhardtii elicits metabolic restructuring during anaerobiosis. *Plant Physiol.* **158,** 1293–1305 (2012).

157. Hemme, D. *et al.* Systems-Wide Analysis of Acclimation Responses to Long-Term Heat Stress and Recovery in the Photosynthetic Model Organism Chlamydomonas reinhardtii. *Plant Cell Online* tpc.114.130997 (2014). doi:10.1105/tpc.114.130997

158. Grossman, A. R. *et al.* Multiple facets of anoxic metabolism and hydrogen production in the unicellular green alga Chlamydomonas reinhardtii. *New Phytol.* **190,** 279–288 (2011).

159. Scranton, M. A., Ostrand, J. T., Fields, F. J. & Mayfield, S. P. Chlamydomonas as a model for biofuels and bio-products production. *Plant J. Cell Mol. Biol.* **82,** 523–531 (2015).

160. Ratcliff, W. C. *et al.* Experimental evolution of an alternating uni- and multicellular life cycle in *Chlamydomonas reinhardtii. Nat. Commun.* **4,** 2742 (2013).

161. Prochnik, S. E. *et al.* Genomic Analysis of Organismal Complexity in the Multicellular Green Alga Volvox carteri. *Science* **329,** 223–226 (2010).

162. Hilton, L. K. *et al.* A Forward Genetic Screen and Whole Genome Sequencing Identify Deflagellation Defective Mutants in Chlamydomonas, Including Assignment of ADF1 as a TRP Channel. *G3 GenesGenomesGenetics* **6,** 3409–3418 (2016).

163. Schierenbeck, L. *et al.* Fast forward genetics to identify mutations causing a high light tolerant phenotype in Chlamydomonas reinhardtii by whole-genome-sequencing. *BMC Genomics* **16,** 57 (2015).

164.Cagnon, C. *et al.* Development of a forward genetic screen to isolate oil mutants in the green microalga Chlamydomonas reinhardtii. *Biotechnol. Biofuels* **6,** 178 (2013).

165.Lauersen, K. J., Kruse, O. & Mussgnug, J. H. Targeted expression of nuclear transgenes in Chlamydomonas reinhardtii with a versatile, modular vector toolkit. *Appl. Microbiol. Biotechnol.* **99,** 3491–3503 (2015).

166.Mussgnug, J. H. Genetic tools and techniques for Chlamydomonas reinhardtii. *Appl. Microbiol. Biotechnol.* **99,** 5407–5418 (2015).

167.Jinkerson, R. E. & Jonikas, M. C. Molecular techniques to interrogate and edit the Chlamydomonas nuclear genome. *Plant J. Cell Mol. Biol.* **82,** 393–412 (2015).

168.Rosales-Mendoza, S., Paz-Maldonado, L. M. T. & Soria-Guerra, R. E. Chlamydomonas reinhardtii as a viable platform for the production of recombinant proteins: current status and perspectives. *Plant Cell Rep.* **31,** 479–494 (2012).

169.Shin, S.-E. *et al.* CRISPR/Cas9-induced knockout and knock-in mutations in *Chlamydomonas reinhardtii*. *Sci. Rep.* **6,** 27810 (2016).

170.Sizova, I., Greiner, A., Awasthi, M., Kateriya, S. & Hegemann, P. Nuclear gene targeting in Chlamydomonas using engineered zinc-finger nucleases. *Plant J. Cell Mol. Biol.* **73,** 873–882 (2013).

171.Schmollinger, S. *et al.* Dissecting the heat stress response in Chlamydomonas by pharmaceutical and RNAi approaches reveals conserved and novel aspects. *Mol. Plant* **6,** 1795–1813 (2013).

172.Gao, H. *et al.* TALE activation of endogenous genes in Chlamydomonas reinhardtii. *Algal Res.* **5,** 52–60 (2014).

173. Li, X. *et al.* An Indexed, Mapped Mutant Library Enables Reverse Genetics Studies of Biological Processes in Chlamydomonas reinhardtii[OPEN]. *Plant Cell* **28,** 367–387 (2016).

174. Moharikar, S., D'Souza, J. S., Kulkarni, A. B. & Rao, B. J. Apoptotic-Like Cell Death Pathway Is Induced in Unicellular Chlorophyte Chlamydomonas Reinhardtii (chlorophyceae) Cells Following Uv Irradiation: Detection and Functional Analyses1. *J. Phycol.* **42,** 423–433 (2006).

175. Moharikar, S., D'Souza, J. S. & Rao, B. J. A homologue of the defender against the apoptotic death gene (dad1 )in UV-exposed Chlamydomonas cells is downregulated with the onset of programmed cell death. *J. Biosci.* **32,** 261–270 (2007).

176. Zuo, Z., Zhu, Y., Bai, Y. & Wang, Y. Acetic acid-induced programmed cell death and release of volatile organic compounds in Chlamydomonas reinhardtii. *Plant Physiol. Biochem. PPB Société Fr. Physiol. Végétale* **51,** 175–184 (2012).

177. Yordanova, Z. P., Woltering, E. J., Kapchina-Toteva, V. M. & Iakimova, E. T. Mastoparan-induced programmed cell death in the unicellular alga Chlamydomonas reinhardtii. *Ann. Bot.* **111,** 191–205 (2013).

178. Durand, P. M., Choudhury, R., Rashidi, A. & Michod, R. E. Programmed death in a unicellular organism has species-specific fitness effects. *Biol. Lett.* **10,** 20131088 (2014).

179. Durand, P. M., Rashidi, A. & Michod, R. E. How an organism dies affects the fitness of its neighbors. *Am. Nat.* **177,** 224–232 (2011).

180. Kay, P., Choudhury, R., Nel, M., Orellana, M. V. & Durand, P. M. Multicolour flow cytometry analyses and autofluorescence in chlorophytes: lessons from programmed

cell death studies in Chlamydomonas reinhardtii. *J. Appl. Phycol.* **25,** 1473–1482 (2013).

181. Sirisha, V. L., Sinha, M. & D'Souza, J. S. Menadione-induced caspase-dependent programmed cell death in the green chlorophyte Chlamydomonas reinhardtii. *J. Phycol.* **50,** 587–601 (2014).

182. Vavilala, S. L., Gawde, K. K., Sinha, M. & D'Souza, J. S. Programmed cell death is induced by hydrogen peroxide but not by excessive ionic stress of sodium chloride in the unicellular green alga Chlamydomonas reinhardtii. *Eur. J. Phycol.* **50,** 422–438 (2015).

183. Murik, O., Elboher, A. & Kaplan, A. Dehydroascorbate: a possible surveillance molecule of oxidative stress and programmed cell death in the green alga Chlamydomonas reinhardtii. *New Phytol.* **202,** 471–484 (2014).

184. Vavilala, S. L., Sinha, M., Gawde, K. K., Shirolikar, S. M. & D'Souza, J. S. KCl induces a caspase-independent programmed cell death in the unicellular green chlorophyte Chlamydomonas reinhardtii (Chlorophyceae). *Phycologia* **55,** 378–392 (2016).

185. Rupprecht, J. From systems biology to fuel—Chlamydomonas reinhardtii as a model for a systems biology approach to improve biohydrogen production. *J. Biotechnol.* **142,** 10–20 (2009).

186. Choi, S. P., Nguyen, M. T. & Sim, S. J. Enzymatic pretreatment of Chlamydomonas reinhardtii biomass for ethanol production. *Bioresour. Technol.* **101,** 5330–5336 (2010).

187. Mahdy, A., Mendez, L., Ballesteros, M. & González-Fernández, C. Enhanced methane production of Chlorella vulgaris and Chlamydomonas reinhardtii by hydrolytic enzymes addition. *Energy Convers. Manag.* **85,** 551–557 (2014).

188. Spolaore, P., Joannis-Cassan, C., Duran, E. & Isambert, A. Commercial applications of microalgae. *J. Biosci. Bioeng.* **101,** 87–96 (2006).

189. Cheng, D. & He, Q. Assessment of Environmental Stresses for Enhanced Microalgal Biofuel Production – An Overview. *Front. Energy Res.* **2,** (2014).

190. Soto, C., Hellebust, J. A., Hutchinson, T. C. & Sawa, T. Effect of naphthalene and aqueous crude oil extracts on the green flagellate Chlamydomonas angulosa. I. Growth. *Can. J. Bot.* **53,** 109–117 (1975).

191. Stevenson, L. M. *et al.* Environmental Feedbacks and Engineered Nanoparticles: Mitigation of Silver Nanoparticle Toxicity to Chlamydomonas reinhardtii by Algal-Produced Organic Compounds. *PLOS ONE* **8,** e74456 (2013).

192. Beauvais-Flück, R., Slaveykova, V. I. & Cosio, C. Cellular toxicity pathways of inorganic and methyl mercury in the green microalga Chlamydomonas reinhardtii. *Sci. Rep.* **7,** 8034 (2017).

193. Lavoie, M., Sabatier, S., Garnier-Laplace, J. & Fortin, C. Uranium accumulation and toxicity in the green alga Chlamydomonas reinhardtii is modulated by pH. *Environ. Toxicol. Chem.* **33,** 1372–1379 (2014).

194. Taylor, N. S. *et al.* Molecular toxicity of cerium oxide nanoparticles to the freshwater alga Chlamydomonas reinhardtii is associated with supra-environmental exposure concentrations. *Nanotoxicology* **10,** 32–41 (2016).

195. Becker, B. Snow ball earth and the split of Streptophyta and Chlorophyta. *Trends Plant Sci.* **18,** 180–183 (2013).

196. Kourmpetis, Y. A. I., van Dijk, A. D. J., van Ham, R. C. H. J. & ter Braak, C. J. F. Genome-Wide Computational Function Prediction of Arabidopsis Proteins by Integration of Multiple Data Sources1[C][W][OA]. *Plant Physiol.* **155,** 271–281 (2011).

197. Hedges, S. B., Blair, J. E., Venturi, M. L. & Shoe, J. L. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evol. Biol.* **4,** 2 (2004).

198. Luscombe, N. M., Greenbaum, D. & Gerstein, M. What is bioinformatics? A proposed definition and overview of the field. *Methods Inf. Med.* **40,** 346–358 (2001).

199. Baxevanis, A. D. The importance of biological databases in biological discovery. *Curr. Protoc. Bioinforma.* **Chapter 1,** Unit 1.1 (2009).

200. Pearson, W. R. An Introduction to Sequence Similarity ("Homology") Searching. *Curr. Protoc. Bioinforma. Ed. Board Andreas Baxevanis Al* **0 3,** (2013).

201. Pearson, W. R. Protein Function Prediction: Problems and Pitfalls. *Curr. Protoc. Bioinforma.* **51,** 4.12.1-8 (2015).

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

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

204. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25,** 25–29 (2000).

205. Khodiyar, V. K., Dimmer, E. C., Huntley, R. P. & Lovering, R. C. Fundamentals of Gene Ontology Functional Annotation. in *Knowledge-Based Bioinformatics* 169–208 (Wiley-Blackwell, 2010). doi:10.1002/9780470669716.ch8

206. Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21,** 3674–3676 (2005).

207. Götz, S. *et al.* High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* **36,** 3420–3435 (2008).

208. Blaby, I. K. & Blaby-Haas, C. E. Genomics and Functional Genomics in <Emphasis Type="Italic">Chlamydomonas reinhardtii</Emphasis>. in *Chlamydomonas: Molecular Genetics and Physiology* 1–26 (Springer, Cham, 2017). doi:10.1007/978-3-319-66365-4_1

209. Zhivotovsky, B. & Orrenius, S. Calcium and cell death mechanisms: a perspective from the cell death community. *Cell Calcium* **50,** 211–221 (2011).

210. Kearse, M. *et al.* Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28,** 1647–1649 (2012).

211. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22,** 4673–4680 (1994).

212. Quevillon, E. *et al.* InterProScan: protein domains identifier. *Nucleic Acids Res.* **33,** W116-120 (2005).

213. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinforma. Oxf. Engl.* **30,** 1236–1240 (2014).

214. Henke, N. *et al.* The ancient cell death suppressor BAX inhibitor-1. *Cell Calcium* **50,** 251–260 (2011).

215. Xu, Q. & Reed, J. C. Bax inhibitor-1, a mammalian apoptosis suppressor identified by functional screening in yeast. *Mol. Cell* **1,** 337–346 (1998).

216. Watanabe, N. & Lam, E. BAX inhibitor-1 modulates endoplasmic reticulum stress-mediated programmed cell death in Arabidopsis. *J. Biol. Chem.* **283,** 3200–3210 (2008).

217. Lisbona, F. *et al.* BAX inhibitor-1 is a negative regulator of the ER stress sensor IRE1alpha. *Mol. Cell* **33,** 679–691 (2009).

218. Iwata, Y., Fedoroff, N. V. & Koizumi, N. Arabidopsis bZIP60 is a proteolysis-activated transcription factor involved in the endoplasmic reticulum stress response. *Plant Cell* **20,** 3107–3121 (2008).

219. Kawai-Yamada, M., Jin, L., Yoshinaga, K., Hirata, A. & Uchimiya, H. Mammalian Bax-induced plant cell death can be down-regulated by overexpression of Arabidopsis Bax Inhibitor-1 (AtBI-1). *Proc. Natl. Acad. Sci.* **98,** 12295–12300 (2001).

220. Kawai, M., Pan, L., Reed, J. C. & Uchimiya, H. Evolutionally conserved plant homologue of the Bax Inhibitor-1 (BI-1) gene capable of suppressing Bax-induced cell death in yeast1. *FEBS Lett.* **464,** 143–147 (1999).

221. Chae, H.-J. *et al.* Evolutionarily conserved cytoprotection provided by Bax Inhibitor-1 homologs from animals, plants, and yeast. *Gene* **323,** 101–113 (2003).

222. Wu, J., He, G.-T., Zhang, W.-J., Xu, J. & Huang, Q.-B. IRE1α Signaling Pathways Involved in Mammalian Cell Fate Determination. *Cell. Physiol. Biochem.* **38,** 847–858 (2016).

223. Chen, Y. & Brandizzi, F. IRE1: ER stress sensor and cell fate executor. *Trends Cell Biol.* **23,** (2013).

224. Maurel, M., Chevet, E., Tavernier, J. & Gerlo, S. Getting RIDD of RNA: IRE1 in cell fate regulation. *Trends Biochem. Sci.* **39,** 245–254 (2014).

225. Polager, S. & Ginsberg, D. E2F - at the crossroads of life and death. *Trends Cell Biol.* **18,** 528–535 (2008).

226. Bracken, A. P., Ciro, M., Cocito, A. & Helin, K. E2F target genes: unraveling the biology. *Trends Biochem. Sci.* **29,** 409–417 (2004).

227. Cao, L. *et al.* The ancient function of RB-E2F Pathway: insights from its evolutionary history. *Biol. Direct* **5,** 55 (2010).

228. Hsieh, J. K., Fredersdorf, S., Kouzarides, T., Martin, K. & Lu, X. E2F1-induced apoptosis requires DNA binding but not transactivation and is inhibited by the retinoblastoma protein through direct interaction. *Genes Dev.* **11,** 1840–1852 (1997).

229. Tiemann, F. & Hinds, P. W. Induction of DNA synthesis and apoptosis by regulated inactivation of a temperature-sensitive retinoblastoma protein. *EMBO J.* **17,** 1040–1052 (1998).

230. Berkovich, E. & Ginsberg, D. ATM is a target for positive regulation by E2F-1. *Oncogene* **22,** 161–167 (2003).

231. Powers, J. T. *et al.* E2F1 uses the ATM signaling pathway to induce p53 and Chk2 phosphorylation and apoptosis. *Mol. Cancer Res. MCR* **2,** 203–214 (2004).

232. Rogoff, H. A., Pickering, M. T., Debatis, M. E., Jones, S. & Kowalik, T. F. E2F1 induces phosphorylation of p53 that is coincident with p53 accumulation and apoptosis. *Mol. Cell. Biol.* **22,** 5308–5318 (2002).

233. Chen, D., Padiernos, E., Ding, F., Lossos, I. S. & Lopez, C. D. Apoptosis-stimulating protein of p53-2 (ASPP2/53BP2L) is an E2F target gene. *Cell Death Differ.* **12,** 358–368 (2004).

234. Fogal, V. *et al.* ASPP1 and ASPP2 are new transcriptional targets of E2F. *Cell Death Differ.* **12,** 369–376 (2005).

235. Hsieh, J.-K. *et al.* Novel function of the cyclin A binding site of E2F in regulating p53-induced apoptosis in response to DNA damage. *Mol. Cell. Biol.* **22,** 78–93 (2002).

236. Sung, Y. H. *et al.* Ei24, a novel E2F target gene, affects p53-independent cell death upon ultraviolet C irradiation. *J. Biol. Chem.* **288,** 31261–31267 (2013).

237. Stanelle, J. & Pützer, B. M. E2F1-induced apoptosis: turning killers into therapeutics. *Trends Mol. Med.* **12,** 177–185 (2006).

238. Hallstrom, T. C. & Nevins, J. R. Specificity in the activation and control of transcription factor E2F-dependent apoptosis. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 10848–10853 (2003).

239. Hallstrom, T. C. & Nevins, J. R. Jab1 is a specificity factor for E2F1-induced apoptosis. *Genes Dev.* **20,** 613–623 (2006).

240. Morris, E. J. *et al.* Functional Identification of Api5 as a Suppressor of E2F-Dependent Apoptosis In Vivo. *PLOS Genet.* **2,** e196 (2006).

241. Smetana, O., Široký, J., Houlné, G., Opatrný, Z. & Chabouté, M.-E. Non-apoptotic programmed cell death with paraptotic-like features in bleomycin-treated plant cells is

suppressed by inhibition of ATM/ATR pathways or NtE2F overexpression. *J. Exp. Bot.* **63,** 2631–2644 (2012).

242. Sargent, F. *et al.* Overlapping functions of components of a bacterial Sec-independent protein export pathway. *EMBO J.* **17,** 3640–3650 (1998).

243. Chen, Y.-C., Li, C.-L., Hsiao, Y.-Y., Duh, Y. & Yuan, H. S. Structure and function of TatD exonuclease in DNA repair. *Nucleic Acids Res.* **42,** 10776–10785 (2014).

244. Müller, M. & Klösgen, R. B. The Tat pathway in bacteria and chloroplasts (review). *Mol. Membr. Biol.* **22,** 113–121 (2005).

245. Parrish, J. Z. & Xue, D. Functional genomic analysis of apoptotic DNA degradation in C. elegans. *Mol. Cell* **11,** 987–996 (2003).

246. Lam, E. Controlled cell death, plant survival and development. *Nat. Rev. Mol. Cell Biol.* **5,** 305–315 (2004).

247. Jabs, T., Dietrich, R. A. & Dangl, J. L. Initiation of Runaway Cell Death in an Arabidopsis Mutant by Extracellular Superoxide. *Science* **273,** 1853–1856 (1996).

248. Mateo, A. *et al.* LESION SIMULATING DISEASE 1 is required for acclimation to conditions that promote excess excitation energy. *Plant Physiol.* **136,** 2818–2830 (2004).

249. Czarnocka, W. *et al.* The dual role of LESION SIMULATING DISEASE 1 as a condition-dependent scaffold protein and transcription regulator. *Plant Cell Environ.* **40,** 2644–2662 (2017).

250. Takatsuji, H. Zinc-finger transcription factors in plants. *Cell. Mol. Life Sci. CMLS* **54,** 582–596 (1998).

251. Coll, N. S. *et al.* Arabidopsis Type I Metacaspases Control Cell Death. *Science* **330,** 1393–1397 (2010).

252. Lam, E. & Zhang, Y. Regulating the reapers: activating metacaspases for programmed cell death. *Trends Plant Sci.* **17,** 487–494 (2012).

253. Epple, P., Mack, A. A., Morris, V. R. F. & Dangl, J. L. Antagonistic control of oxidative stress-induced cell death in Arabidopsis by two related, plant-specific zinc finger proteins. *Proc. Natl. Acad. Sci.* **100,** 6831–6836 (2003).

254. Polyak, K., Xia, Y., Zweier, J. L., Kinzler, K. W. & Vogelstein, B. A model for p53-induced apoptosis. *Nature* **389,** 300–305 (1997).

255. Porté, S. *et al.* Three-dimensional Structure and Enzymatic Function of Proapoptotic Human p53-inducible Quinone Oxidoreductase PIG3. *J. Biol. Chem.* **284,** 17194–17205 (2009).

256. Jin, M. *et al.* PIG3 Regulates p53 Stability by Suppressing Its MDM2-Mediated Ubiquitination. *Biomol. Ther.* **25,** 396–403 (2017).

257. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2,** 2006.0008 (2006).

258. Winzeler, E. A. *et al.* Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science* **285,** 901–906 (1999).

259. Venken, K. J. T. *et al.* MiMIC: a highly versatile transposon insertion resource for engineering Drosophila melanogaster genes. *Nat. Methods* **8,** 737–743 (2011).

260. Varshney, G. K. *et al.* A large-scale zebrafish gene knockout resource for the genome-wide study of gene function. *Genome Res.* **23,** 727–735 (2013).

261. Alonso, J. M. *et al.* Genome-wide insertional mutagenesis of Arabidopsis thaliana. *Science* **301,** 653–657 (2003).

262. Zhang, R. *et al.* High-Throughput Genotyping of Green Algal Mutants Reveals Random Distribution of Mutagenic Insertion Sites and Endonucleolytic Cleavage of Transforming DNA. *Plant Cell* **26,** 1398–1409 (2014).

263. Kwolek-Mirek, M. & Zadrag-Tecza, R. Comparison of methods used for assessing the viability and vitality of yeast cells. *FEMS Yeast Res.* **14,** 1068–1079 (2014).

264. van Meer, G., Voelker, D. R. & Feigenson, G. W. Membrane lipids: where they are and how they behave. *Nat. Rev. Mol. Cell Biol.* **9,** 112–124 (2008).

265. Leventis, P. A. & Grinstein, S. The distribution and function of phosphatidylserine in cellular membranes. *Annu. Rev. Biophys.* **39,** 407–427 (2010).

266. Balasubramanian, K. & Schroit, A. J. Aminophospholipid asymmetry: A matter of life and death. *Annu. Rev. Physiol.* **65,** 701–734 (2003).

267. Nagata, S., Suzuki, J., Segawa, K. & Fujii, T. Exposure of phosphatidylserine on the cell surface. *Cell Death Differ.* **23,** 952–961 (2016).

268. Brouckaert, G. *et al.* Phagocytosis of Necrotic Cells by Macrophages Is Phosphatidylserine Dependent and Does Not Induce Inflammatory Cytokine Production. *Mol. Biol. Cell* **15,** 1089–1100 (2004).

269. Krysko, D. V. *et al.* Macrophages use different internalization mechanisms to clear apoptotic and necrotic cells. *Cell Death Differ.* **13,** 2011–2022 (2006).

270. Wang, Q. *et al.* Pyroptotic cells externalize eat-me and release find-me signals and are efficiently engulfed by macrophages. *Int. Immunol.* **25,** 363–372 (2013).

271. Nagata, S., Hanayama, R. & Kawane, K. Autoimmunity and the Clearance of Dead Cells. *Cell* **140,** 619–630 (2010).

272. Engeland, M. van, Nieland, L. J. W., Ramaekers, F. C. S., Schutte, B. & Reutelingsperger, C. P. M. Annexin V-Affinity assay: A review on an apoptosis detection system based on phosphatidylserine exposure. *Cytometry* **31,** 1–9

273. Sivagnanam, U., Palanirajan, S. K. & Gummadi, S. N. The role of human phospholipid scramblases in apoptosis: An overview. *Biochim. Biophys. Acta* **1864,** 2261–2271 (2017).

274. Jabs, T. Reactive oxygen intermediates as mediators of programmed cell death in plants and animals. *Biochem. Pharmacol.* **57,** 231–245 (1999).

275. Van Acker, H. *et al.* The Role of Reactive Oxygen Species in Antibiotic-Induced Cell Death in Burkholderia cepacia Complex Bacteria. *PLoS ONE* **11,** (2016).

276. Bergamini, C. M., Gambetti, S., Dondi, A. & Cervellati, C. Oxygen, reactive oxygen species and tissue damage. *Curr. Pharm. Des.* **10,** 1611–1626 (2004).

277. Cadet, J. & Wagner, J. R. DNA Base Damage by Reactive Oxygen Species, Oxidizing Agents, and UV Radiation. *Cold Spring Harb. Perspect. Biol.* **5,** a012559 (2013).

278. Toyokuni, S. Reactive oxygen species-induced molecular damage and its application in pathology. *Pathol. Int.* **49,** 91–102 (1999).

279. Birben, E., Sahiner, U. M., Sackesen, C., Erzurum, S. & Kalayci, O. Oxidative Stress and Antioxidant Defense. *World Allergy Organ. J.* **5,** 9–19 (2012).

280. Circu, M. L. & Aw, T. Y. Reactive oxygen species, cellular redox systems, and apoptosis. *Free Radic. Biol. Med.* **48,** 749–762 (2010).

281. Duennwald, M. L. Growth Assays to Assess Polyglutamine Toxicity in Yeast. *J. Vis. Exp. JoVE* (2012). doi:10.3791/3461

282. Madeo, F. *et al.* A caspase-related protease regulates apoptosis in yeast. *Mol. Cell* **9,** 911–917 (2002).

283. Carmona-Gutiérrez, D. *et al.* The propeptide of yeast cathepsin D inhibits programmed necrosis. *Cell Death Dis.* **2,** e161 (2011).

284. Giorgini, F., Guidetti, P., Nguyen, Q., Bennett, S. C. & Muchowski, P. J. A genomic screen in yeast implicates kynurenine 3-monooxygenase as a therapeutic target for Huntington disease. *Nat. Genet.* **37,** 526–531 (2005).

285. Teng, X. & Hardwick, J. M. Quantification of genetically controlled cell death in budding yeast. *Methods Mol. Biol. Clifton NJ* **1004,** 161–170 (2013).

286. Kawane, K., Motani, K. & Nagata, S. DNA Degradation and Its Defects. *Cold Spring Harb. Perspect. Biol.* **6,** a016394 (2014).

287. Wyllie, A. H. Glucocorticoid-induced thymocyte apoptosis is associated with endogenous endonuclease activation. *Nature* **284,** 555–556 (1980).

288. Nagata, S., Nagase, H., Kawane, K., Mukae, N. & Fukuyama, H. Degradation of chromosomal DNA during apoptosis. *Cell Death Differ.* **10,** 108–116 (2003).

289. Holm, S. A Simple Sequentially Rejective Multiple Test Procedure. *Scand. J. Stat.* **6,** 65–70 (1979).

290. Holm's Sequential Bonferroni Procedure. in *Encyclopedia of Research Design* (SAGE Publications, Inc., 2010). doi:10.4135/9781412961288.n178

291. Boyd, V., Cholewa, O. M. & Papas, K. K. Limitations in the Use of Fluorescein Diacetate/Propidium Iodide (FDA/PI) and Cell Permeable Nucleic Acid Stains for

Viability Measurements of Isolated Islets of Langerhans. *Curr. Trends Biotechnol. Pharm.* **2,** 66–84 (2008).

292. Thakur, S., Cattoni, D. I. & Nöllmann, M. The fluorescence properties and binding mechanism of SYTOX green, a bright, low photo-damage DNA intercalating agent. *Eur. Biophys. J. EBJ* **44,** 337–348 (2015).

293. Roth, B. L., Poot, M., Yue, S. T. & Millard, P. J. Bacterial viability and antibiotic susceptibility testing with SYTOX green nucleic acid stain. *Appl. Environ. Microbiol.* **63,** 2421–2431 (1997).

294. Ramirez, C. N., Antczak, C. & Djaballah, H. Cell Viability Assessment: Toward Content-Rich Platforms. *Expert Opin. Drug Discov.* **5,** 223–233 (2010).

295. Kaplan, R., Jaye, M., Burgess, W. H., Schlaepfer, D. D. & Haigler, H. T. Cloning and expression of cDNA for human endonexin II, a Ca2+ and phospholipid binding protein. *J. Biol. Chem.* **263,** 8037–8043 (1988).

296. Sofiadis, A. *et al.* Proteomic profiling of follicular and papillary thyroid tumors. *Eur. J. Endocrinol.* **166,** 657–667 (2012).

297. Oparka, M. *et al.* Quantifying ROS levels using CM-H2DCFDA and HyPer. *Methods San Diego Calif* **109,** 3–11 (2016).

298. Gomes, A., Fernandes, E. & Lima, J. L. F. C. Fluorescence probes used for detection of reactive oxygen species. *J. Biochem. Biophys. Methods* **65,** 45–80 (2005).

299. Wardman, P. Fluorescent and luminescent probes for measurement of oxidative and nitrosative species in cells and tissues: progress, pitfalls, and prospects. *Free Radic. Biol. Med.* **43,** 995–1022 (2007).

300. Forkink, M., Smeitink, J. A. M., Brock, R., Willems, P. H. G. M. & Koopman, W. J. H. Detection and manipulation of mitochondrial reactive oxygen species in mammalian cells. *Biochim. Biophys. Acta* **1797,** 1034–1044 (2010).

301. UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* **43,** D204-212 (2015).

302. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45,** D158–D169 (2017).

303. Finn, R. D. *et al.* HMMER web server: 2015 update. *Nucleic Acids Res.* **43,** W30–W38 (2015).

304. Prakash, A., Jeffryes, M., Bateman, A. & Finn, R. D. The HMMER Web Server for Protein Sequence Similarity Search. *Curr. Protoc. Bioinforma.* **60,** 3.15.1-3.15.23 (2017).

305. Sinha, S. & Lynn, A. M. HMM-ModE: implementation, benchmarking and validation with HMMER3. *BMC Res. Notes* **7,** 483 (2014).

306. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).

307. Szklarczyk, D. *et al.* The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* **45,** D362–D368 (2017).

308. Skulachev, V. P. Cytochrome c in the apoptotic and antioxidant cascades. *FEBS Lett.* **423,** 275–280

309. Li, K. *et al.* Cytochrome c Deficiency Causes Embryonic Lethality and Attenuates Stress-Induced Apoptosis. *Cell* **101,** 389–399 (2000).

310. Teng, X. *et al.* Genome-wide Consequences of Deleting Any Single Gene. *Mol. Cell* **52,** 485–494 (2013).

311. McFarland, C. D. *et al.* The Damaging Effect of Passenger Mutations on Cancer Progression. *Cancer Res.* **77,** 4763–4772 (2017).

312. Pon, J. R. & Marra, M. A. Driver and Passenger Mutations in Cancer. *Annu. Rev. Pathol. Mech. Dis.* **10,** 25–50 (2015).

313. Conrad, T. M., Lewis, N. E. & Palsson, B. Ø. Microbial laboratory evolution in the era of genome-scale science. *Mol. Syst. Biol.* **7,** 509 (2011).

314. Flowers, J. M. *et al.* Whole-Genome Resequencing Reveals Extensive Natural Variation in the Model Green Alga Chlamydomonas reinhardtii. *Plant Cell* **27,** 2353–2369 (2015).

315. Gallaher, S. D., Fitz-Gibbon, S. T., Glaesener, A. G., Pellegrini, M. & Merchant, S. S. Chlamydomonas Genome Resource for Laboratory Strains Reveals a Mosaic of Sequence Variation, Identifies True Strain Histories, and Enables Strain-Specific Studies. *Plant Cell* **27,** 2335–2352 (2015).

316. Lamberg, A., Nieminen, S., Qiao, M. & Savilahti, H. Efficient Insertion Mutagenesis Strategy for Bacterial Genomes Involving Electroporation of In Vitro-Assembled DNA Transposition Complexes of Bacteriophage Mu. *Appl Env. Microbiol* **68,** 705–712 (2002).

317. Glenn, G. & Andreou, L.-V. Analysis of DNA by Southern blotting. *Methods Enzymol.* **529,** 47–63 (2013).

318. Chapter 5 - The Sexual Cycle. in *The Chlamydomonas Sourcebook (Second Edition)* (eds. Harris, E. H., Stern, D. B. & Witman, G. B.) 119–157 (Academic Press, 2009). doi:10.1016/B978-0-12-370873-1.00005-8

319. Chapter 8 - Chlamydomonas in the Laboratory. in *The Chlamydomonas Sourcebook (Second Edition)* (eds. Harris, E. H., Stern, D. B. & Witman, G. B.) 241–302 (Academic Press, 2009). doi:10.1016/B978-0-12-370873-1.00008-3

320. Xu, W., Coll, J. L. & Adamson, E. D. Rescue of the mutant phenotype by reexpression of full-length vinculin in null F9 cells; effects on cell locomotion by domain deleted vinculin. *J. Cell Sci.* **111 ( Pt 11),** 1535–1544 (1998).

321. de Vries, J. & Wackernagel, W. Detection of nptII (kanamycin resistance) genes in genomes of transgenic plants by marker-rescue transformation. *Mol. Gen. Genet. MGG* **257,** 606–613 (1998).

322. Transformation Rescue. in *Encyclopedia of Genetics, Genomics, Proteomics and Informatics* (ed. Rédei, G. P.) 2009–2009 (Springer Netherlands, 2008). doi:10.1007/978-1-4020-6754-9_17270

323. Hammill, A. K., Uhr, J. W. & Scheuermann, R. H. Annexin V staining due to loss of membrane asymmetry can be reversible and precede commitment to apoptotic death. *Exp. Cell Res.* **251,** 16–21 (1999).

324. Stuart, M. C. A., Reutelingsperger, C. P. M. & Frederik, P. M. Binding of annexin V to bilayers with various phospholipid compositions using glass beads in a flow cytometer. *Cytometry* **33,** 414–419

325. Rockenfeller, P. *et al.* Phosphatidylethanolamine positively regulates autophagy and longevity. *Cell Death Differ.* **22,** 499–508 (2015).

326. Kaur, J. & Debnath, J. Autophagy at the crossroads of catabolism and anabolism. *Nat. Rev. Mol. Cell Biol.* **16,** 461–472 (2015).

327. Galluzzi, L., Bravo-San Pedro, J. M. & Kroemer, G. Preface. in *Methods in Enzymology* (eds. Galluzzi, L., Bravo-San Pedro, J. M. & Kroemer, G.) **587,** xxiii–xxix (Academic Press, 2017).

**APPENDIX A**

**Computational Methods Used to Retrieve and Analyze Subject Sequences for BLASTp**

All computations were performed using a mid 2015 MacBook Pro Retina running macOS Sierra (version 10.12.5) and equipped with a 2.2 GHz Intel Core i7 processor and 16GB 1600 MHz DDR3 memory.

## KEY:

Any clarifying remarks that were "commented out" of the code are denoted by red text.

Commands within a blue box were carried out in the bash shell (Terminal).

"$" at the beginning of a line designates a new Terminal command:
"mattbreuer$" denotes a command from my local system
"ubuntu$" denotes a command from the AWS instance

Commands within a tan box were carried out using the R language within R Studio.

Text within a grey box describe non-code methods that were carried out.

**Retrieve Sequences for Database**

Three independent keyword searches were carried out within the NCBI Entrez Webpage.

The keywords used were
 "apoptosis"
 "necrosis"
"programmed cell death"

The results from each search were saved locally as three separate multi-FASTA files.

**Parse accession numbers from multi-FASTA files**

```
$ sed -n '/>[a-zA-Z0-9]/p'
/Volumes/Transcend/Research_for_Dr_Gaillard/Thesis/bioinformatics_files/1-
sequence_retrieval/1_searchresults_apoptosis.faa| awk -v col=1 '{print $col}' | sed 's/>//g' >
3_searchresults_apoptosis_accessions.txt

$ sed -n '/>[a-zA-Z0-9]/p'
/Volumes/Transcend/Research_for_Dr_Gaillard/Thesis/bioinformatics_files/1-
sequence_retrieval/1_searchresults_necrosis.faa | awk -v col=1 '{print $col}' | sed 's/>//g' >
3_searchresults_necrosis_accessions.txt
```

```
$ sed -n '/>[a-zA-Z0-9]/p'
/Volumes/Transcend/Research_for_Dr_Gaillard/Thesis/bioinformatics_files/1-
sequence_retrieval/1_searchresults_pcd.faa | awk -v col=1 '{print $col}' | sed 's/>//g' >
3_searchresults_pcd_accessions.txt
```

**Combine the three files of search results (accessions) into a single file**

```
$ cat 3_searchresults_apoptosis_accessions.txt 3_searchresults_necrosis_accessions.txt
1_searchresults_pcd.faa > 4_searchresults_apop_necro_pcd_accessions_raw.txt
```

**Determine the proteins which appeared in more than one keyword search**

```
# make a list of accession numbers

listA <- read.csv("/Volumes/Transcend/Thesis/bioinformatics_files/1-
sequence_retrieval/searchresults_apoptosis_accessions.txt",header=FALSE)
A <- listA$V1
listB <- read.csv("/Volumes/Transcend/Thesis/bioinformatics_files/1-
sequence_retrieval/searchresults_necrosis_accessions.txt",header=FALSE)
B <- listB$V1
listC <- read.csv("/Volumes/Transcend/Thesis/bioinformatics_files/1-
sequence_retrieval/searchresults_pcd_accessions.txt",header=FALSE)
C <- listC$V1

# determine the number of results from each search

length(A)
length(B)
length(C)

# identify the overlapping entries between the search results and assign the
output of each file to a variable

abc <- Reduce(intersect, list(A,B,C))
ab <- Reduce(intersect, list(A,B))
ac <- Reduce(intersect, list(A,C))
bc <- Reduce(intersect, list(B,C))

# export the lists of overlapping entries as text files

write(abc, file = "/Volumes/Transcend/Thesis/bioinformatics_files/1-
sequence_retrieval/overlapping_results_apop_necro_pcd.txt", ncolumns = 1, sep =
"")
write(ab, file = "/Volumes/Transcend/Thesis/bioinformatics_files/1-
sequence_retrieval/overlapping_results_apop_necro.txt", ncolumns = 1, sep = "")
write(ac, file = "/Volumes/Transcend/Thesis/bioinformatics_files/1-
sequence_retrieval/overlapping_results_apop_pcd.txt", ncolumns = 1, sep = "")
```

```
write(bc, file = "/Volumes/Transcend/Thesis/bioinformatics_files/1-
sequence_retrieval/overlapping_results_necro_pcd.txt", ncolumns = 1, sep = "")

# load package VennDiagram
library(VennDiagram)

# list assigns each list a label which will be added to the final image. fill
designates a color for each circle. cex is the size of the text. filename gives
the filepath in which the image will be saved
venn.diagram(list("Apoptosis"=A, "Necrosis"=B, "PCD"=C), fill =
c("turquoise","darkorchid1","seagreen3"), cex =
1.5,filename="/Volumes/Transcend/Thesis/bioinformatics_files/1-
sequence_retrieval/venn_uniq+overlap.png")
```

## Remove duplicate accession numbers from the file of combined search results

```
# count the number of lines in the combined list of search results
mattbreuer$ wc 4_searchresults_apop_necro_pcd_accessions_raw.txt
91687

# remove duplicate accessions from the combined list
mattbreuer$ sort 4_searchresults_apop_necro_pcd_accessions_raw.txt -n | uniq >
5_searchresults_apop_necro_pcd_accessions_no_dups.txt

# count the number of lines in the combined list of accessions after removal of
duplicate entries
mattbreuer$ wc 5_searchresults_apop_necro_pcd_accessions_no_dups.txt
85656
```

## Confirm the appropriate number of entries were removed from the combined file

Number of lines in the raw file - number of lines in the no_dups file: 91687 -
85656 = 6,031 lines removed.

Sum of the repeated values from venn_search_results.png: 866 + 5112 + 9
+ 22(2) = 6,031 total duplicate entries.

## Retrieve the non-redundant search results as sequences in FASTA format

```
# install E-Direct locally
$ perl -MNet::FTP -e \
>   '$ftp = new Net::FTP("ftp.ncbi.nlm.nih.gov", Passive => 1);
>     $ftp->login; $ftp->binary;
>     $ftp->get("/entrez/entrezdirect/edirect.tar.gz");'
$ gunzip -c edirect.tar.gz | tar xf -
$ rm edirect.tar.gz
$ export PATH=$PATH:$HOME/edirect
$ ./edirect/setup.sh
```

# script (6_get_fasta_script.sh) which uses E-Direct to fetch the sequence(s), in FASTA format, given an accession number(s) as the input:

```
#!/usr/bin/bash
while read id
do
efetch -db protein -id "$id [uid]" -format fasta < /dev/null
done
```

# first, take the unique accession numbers and strip the ".version" from the end of each accession. then, fetch each of the sequences in fasta format by executing 6_get_fasta_script.sh from the command line.
```
$ sed 's/\.[[:digit:]]//g' /Volumes/Transcend/Thesis/bioinformatics_files/1-
sequence_retrieval/5_searchresults_apop_necro_pcd_accessions_no_dups.txt | bash
6_get_fasta_script.sh > 7_searchresults_apop_necro_pcd_fasta_no_dups.faa
```

# count the number of sequences that are retrieved to confirm that all 85656 sequences are accounted for.
```
$ grep ">" 7_searchresults_apop_necro_pcd_fasta_no_dups.faa -c
85656
```

## Retrieve taxonomic classification of the source organism for all unique search results

# script (9_get_taxonomy_script.sh) which uses e-direct to fetch taxonomic classification of the source organism for each of the accession numbers

```
#!/usr/bin/bash
while read id
do
efetch -db protein -id "$id [uid]" -format gpc < /dev/null |
 xtract -pattern INSDSeq -element INSDSeq_accession-version     INSDSeq_taxonomy
INSDSeq_organism
done
```

# first, take the unique accession numbers from 5_searchresults_apop_necro_pcd_accessions_no_dups.txt and strip the ".version" from the end of each accession. then, fetch the taxonomic classification of the source organism using 9_get_taxonomy_script.sh.
```
$ sed 's/\.[[:digit:]]//g' /Volumes/Transcend/Thesis/bioinformatics_files/1-
sequence_retrieval/5_searchresults_apop_necro_pcd_accessions_no_dups.txt | bash
10_get_taxonomy_script.sh > 11_search_results_taxonomy.txt
```

# confirm that all 85656 entries are in the output file
```
$ wc 11_search_results_taxonomy.txt
85656
```

## Reformat the taxonomic classifications

# first, replace the separators in the second column with tabs. then, parse the "kingdom" column. then, count the occurrence of each column and remove the duplicate entries

```
$ sed -E $'s/;[[:space:]]/\t/g' 11_search_results_taxonomy.txt | awk -F '\t' '{print $3}' | sort | uniq -
c | sort -nr > 12_kingdom_distribution_of_search_results+counts.txt
```

**Determination of phylogenetic distribution**

Went through this file and manually assigned the kingdoms into the one of the
following taxonomic groups:

> animals
> plants
> protists
> fungi

for each of the four groups, I summed the counts of their constituent
kindgoms. Lastly, I summed the counts of each kingdom to ensure that all
85656 unique search results were accounted for.

**Create a pie chart showing the taxonomic distribution of the search results**

```
data <- read.csv(file="/Volumes/Transcend/Thesis/bioinformatics_files/1-
sequence_retrieval/13_group_distribution_of_search_results+counts.csv", sep=",",
header=FALSE)
slices <- data$V1
slicelabels <- data$V2
pct <- round(slices/sum(slices)*100)
slicelabels <- paste(slicelabels, pct)

# add percent symbols to labels

slicelabels <- paste(slicelabels,"%",sep="") # add % to labels
# plot the data as a piechart

library(plotrix)
pie3D(slices,edges=NA,radius=0.98,height=0.1,theta=pi/3,start=0.3,border=par("fg")
,  col=c("tomato","green3","slateblue","skyblue"),labels=slicelabels,labelpos=NULL,
labelcol=par("fg"),labelcex=1,sector.order=NULL,explode=0.0,shade=0.8,
    mar=c(3.5,3.5,3.5,3.5),pty="s")
title(main="Phylogenetic Distribution of Search Results")
```

**Appendix B**

**BLASTp of *C. reinhardtii* Proteome against the Combined Search Results**

**Setup of AWS Instance**

Launched a new AWS EC2 instance (r3.2xlarge) from the AWS marketplace webpage. ncbi-blast-2.6.0+ software came pre-installed on the instance.

Created and downloaded a unique keypair.pem.txt file to grant remote access to the instance.

**Connect to AWS Instance**

```
# granted permissions to keypair.pem.txt
mattbreuer$ chmod 400 keypair.pem.txt

# connected to aws instance using a secure shell
mattbreuer$ ssh -i keypair.pem.txt ubuntu@54.193.124.117
```

**Perform BLASTp in AWS instance, using the *C. reinhardtii* proteome as the queries, and the unique search results as the subjects**

```
# in new terminal window, uploaded the unique search results to the aws instance in fasta format

mattbreuer$ scp -i keypair.pem.txt ~/Volumes/Transcend/Thesis/bioinformatics_files/2-blast/
1_searchresults_apop_necro_pcd_fasta_no_dups.faa ubuntu@54.193.124.117:

# uploaded the C. reinhardtii proteome to the aws instance
mattbreuer$ scp -i keypair.pem.txt ~/Volumes/Transcend/Thesis/bioinformatics_files/2-blast/
1_Creinhardtii_v5.5.proteome.fa ubuntu@54.193.124.117:

# in the first terminal window, confirmed that the files had been uploaded to the aws instance
ubuntu$ ls –l

# created a blastable database from the unique search results
ubuntu$ makeblastdb -in 1_searchresults_apop_necro_pcd_fasta_no_dups.faa -dbtype 'prot' -out
2_db_searchresults_apop_necro_pcd

# performed blastp on aws instance
ubuntu$ blastp -query 1_Creinhardtii_v5.5.proteome.fa -db
2_searchresults_apop_necro_pcd_accessions_no_dups_blastdb -evalue 1 -max_target_seqs 500
-outfmt 11 -out 3_blastp_crproteome_against_searchresults.asn
```

**Download and reformat BLAST results**

```
# converted the blast results to tabular format
ubuntu$ blast_formatter -archive 3_blastp_crproteome_against_searchresults.asn -outfmt "7
qseqid bitscore evalue qcovs pident sacc" -out 4_blastp_crproteome_against_searchresults.txt
```

```
# downloaded the blast databases to my computer
mattbreuer$ scp -i keypair.pem.txt ubuntu@54.193.124.117:/
2_searchresults_apop_necro_pcd_accessions_no_dups_blastdb.phr .

mattbreuer$ scp -i keypair.pem.txt ubuntu@54.193.124.117:/
2_searchresults_apop_necro_pcd_accessions_no_dups_blastdb.psq .

mattbreuer$ scp -i keypair.pem.txt ubuntu@54.193.124.117:/
2_searchresults_apop_necro_pcd_accessions_no_dups_blastdb.pin .

# downloaded the blast results to my computer
mattbreuer$ scp -i keypair.pem.txt ubuntu@54.193.124.117:/
3_blastp_crproteome_against_searchresults.asn .
mattbreuer$ scp -i keypair.pem.txt ubuntu@54.193.124.117:/
4_blastp_crproteome_against_searchresults.txt .

# remove all lines containing hash marks in the file containing the tabular format of the
blast results
mattbreuer$ sed '/#/d' 4_blastp_crproteome_against_searchresults.txt
```

## Filter BLAST results

```
# create a new file of filtered blast results, keeping only the lines in which the percent
identity is greater than or equal to 25 and the bitscore is greater than or equal to 100
mattbreuer$ awk '($2 >= 100 ) && ($5 >= 25)' 4_blastp_crproteome_against_searchresults.txt >
5_blastp_crproteome_against_searchresults_filtered.txt
```

## Count the number of results in the filtered BLAST results

```
mattbreuer$ awk '($5 >= 25) && (29.94 >= $5) { ++count } END{ print count }'
/Volumes/Transcend/Thesis/bioinformatics_files/2-
blasts/5_blastp_crproteome_against_searchresults_filtered.txt

mattbreuer$ awk '($5 >= 30) && (39.94 >= $5) { ++count } END{ print count }'
/Volumes/Transcend/Thesis/bioinformatics_files/2-
blasts/5_blastp_crproteome_against_searchresults_filtered.txt
mattbreuer$ awk '($5 >= 40) && (49.94 >= $5) { ++count } END{ print count }'
/Volumes/Transcend/Thesis/bioinformatics_files/2-
blasts/5_blastp_crproteome_against_searchresults_filtered.txt

mattbreuer$ awk '($5 >= 50) && (59.94 >= $5) { ++count } END{ print count }'
/Volumes/Transcend/Thesis/bioinformatics_files/2-
blasts/5_blastp_crproteome_against_searchresults_filtered.txt

mattbreuer$ awk '($5 >= 60) && (69.94 >= $5) { ++count } END{ print count }'
/Volumes/Transcend/Thesis/bioinformatics_files/2-
blasts/5_blastp_crproteome_against_searchresults_filtered.txt

mattbreuer$ awk '($5 >= 70) && (79.94 >= $5) { ++count } END{ print count }'
/Volumes/Transcend/Thesis/bioinformatics_files/2-
blasts/5_blastp_crproteome_against_searchresults_filtered.txt
```

```
mattbreuer$ awk '($5 >= 80) && (89.94 >= $5) { ++count } END{ print count }'
/Volumes/Transcend/Thesis/bioinformatics_files/2-
blasts/5_blastp_crproteome_against_searchresults_filtered.txt

mattbreuer$ awk '($5 >= 90) && (100 >= $5) { ++count } END{ print count }'
/Volumes/Transcend/Thesis/bioinformatics_files/2-
blasts/5_blastp_crproteome_against_searchresults_filtered.txt
```

**Plot the distribution of percent identities between C. reinhardtii and the subject sequences**

```
#Vertical Bar Graph in RStudio

percent_identity_counts <- c(33967, 33967, 8836, 1719, 783, 238, 69, 114)
percent_identity_ranges <- (c("25-29", "30-39", "40-49", "50-59", "60-69", "70-79", "80-89", "90-100"))
percent_identity_ranges <- paste(percent_identity_ranges,"%",sep="") # add % to labels
#shift the parameters of the graph so that it will fit onto the output page
par(mar=c(5, 6, 3, 1),mgp=c(3,1,0))

#load the RColorBrewer package
library(RColorBrewer)

#take the stock color palette "Spectral" and turn it into a ramped palette. This is so that I can use it on datasets containing more than 11 data points (the Spectral color palette has only 11 colors in it)

SpectralRamp<-colorRampPalette(brewer.pal(11,"Spectral"))

#plot

barplot(percent_identity_counts,horiz=FALSE,inside=TRUE,
      axes=TRUE,xlab="Percent Identity of Query Sequence to Subject
Sequence",ylab="Number of Subject
Sequences",beside=FALSE,space=0.2,xpd=FALSE,ylim=c(0,35000),axisnames=TRUE,
names.arg=percent_identity_ranges,las=1,cex=1,cex.names=1,cex.axis=0.85,col=Sp
ectralRamp(length(percent_identity_counts)))
```

**Retrieve the *C. reinhardtii* proteins which aligned strongly with a protein from the search results**

```
# from the list of filtered results, parse the query identifiers, remove the duplicate entries, then
write to a new file
```

```
mattbreuer$ awk -v col=1 '{print $col}' 5  blastp  crproteome  against  searchresults  filtered.txt |
sort -n | uniq >
6_blastp_crproteome_against_searchresults_filtered_query_accessions_no_dups.txt
```

# then, uploaded the file into the phyotomine search tool on the phytozome webpage
and created a new list. then, downloaded the sequences of each identifier from that
list in fasta format
(7_blastp_crproteome_against_searchresults_filtered_query_fastas_no_dups.fa).

# APPENDIX C

## Analysis of CrPPPs with BLAST2GO

Setup of AWS instance

```
#uploaded the query sequences to the previously-launched aws instance
mattbreuer$ scp -i keypair.pem.txt /Volumes/Transcend/Thesis/bioinformatics_files/2-
blasts/7_blastp_crproteome_against_searchresults_filtered_query_f astas_no_dups.fa :

# connected to the aws instance
mattbreuer$ ssh -i keypair.pem.txt ubuntu@54.193.124.117

# confirmed that the file had been uploaded to the aws instance
ubuntu@ip-172-31-6-69:~$ ls
```

Performed BLASTp of potential *C. reinhardtii* PCD proteins against the RefSeq protein
database

```
# performed blastp in the aws instance
blastp -query 7_blastp_crproteome_against_searchresults_filtered_query_fastas_ no_dups.fa -db
refseq_protein -max_target_seqs 1000 -evalue 0.1 - num_threads 8 -outfmt 5 -out
8_blastp_potential_creinhardtii_pcd_proteins_against_refseq_prote in.xml
```

Formatted the BLAST2GO output

```
# Used BLAST2GO to assign Gene Ontology terms to the potential C. reinhardtii PCD
proteins

# parse columns 1 and 2 from the dataset. for lines (c. reinhardtii proteins) containing
multiple go terms, replace the seperators of the go terms with a newline character and
a tab character. this places the additional go terms onto a new line and moves them
into the second column. for all empty cells in column one, replace with the value in the
cell directly above.

awk -F'\t' '{ OFS="\t"; print $1 "\t" $2 }' 2_b2g_results_seqid+goid+goname.txt | sed -E
$'s/;[[:space:]]/\ \t/g' | awk -F\\t -v COL=1 ' BEGIN{OFS="\t";}
$COL=="" {$COL = saved} {saved = $COL; print}
' > 3 b2g seq ids+go ids.txt

#parse columns 1 and 3 from the dataset. for lines containing multiple go terms,
replace the separators of the go terms with a newline character and a tab character.
this places the additional go terms onto their own new line and moves them into the
second column. for all empty cells in column one, replace with the value in the cell
directly above
awk -F'\t' '{ OFS="\t"; print $1 "\t" $3 }' 2_b2g_results_seqid+goid+goname.txt | sed -E
$'s/;[[:space:]]/\ \t/g' | awk -F\\t -v COL=1 ' BEGIN{OFS="\t";}
$COL=="" {$COL = saved}
{saved = $COL; print}
' > 3 b2g_seq_ids+go_names.txt

# combined the correlated go ids and go names columns
```

```
paste 3_b2g_seq_ids+go_ids.txt 3_seq_ids+go_names.txt | awk -F\t 'BEGIN{OFS="\t";} {print
$1,$2,$4}' > 4_seq_ids+go_ids+go_names.txt
```

Analyzed the BLAST2GO data

```
# counted the total number of biological processes, cellular components, and molecular
functions that were assigned to the c. reinhardtii blast results by blast2go
$grep -c 'P:' 4_b2g_seq_ids+go_ids+go_names.txt $grep -c 'C:'
4_b2g_seq_ids+go_ids+go_names.txt $grep -c 'F:' 4_b2g_seq_ids+go_ids+go_names.txt

# counted the number of occurrences of each go id/name pair which was assigned to
the c. reinhardtii blast dataset. then, removed the "f:", "c:", and "p:" from the go terms to
make them prettier when graphed
awk -F\t 'BEGIN{OFS="\t";} {print $2,$3}' 4_b2g_seq_ids+go_ids+go_names.txt | grep 'C:' |
sort | uniq -c | sort -nr | sed -E $"s/^ +//" | sed -e "s/ /$(printf '\t')/" | sed -e 's/[[:alpha:]]:://g' >
5_b2g_cellular_component_counts.txt
awk -F\t 'BEGIN{OFS="\t";} {print $2,$3}' 4_b2g_seq_ids+go_ids+go_names.txt | grep 'F:' |
sort | uniq -c | sort -nr | sed -E $"s/^ +//" | sed -e "s/ /$(printf '\t')/" | sed -e 's/[[:alpha:]]:://g' >
5_b2g_molecular_function_counts.txt
awk -F\t 'BEGIN{OFS="\t";} {print $2,$3}' 4_b2g_seq_ids+go_ids+go_names.txt | grep 'P:' |
sort | uniq -c | sort -nr | sed -E $"s/^ +//" | sed -e "s/ /$(printf '\t')/" | sed -e 's/[[:alpha:]]:://g' >
5_b2g_biological_process_counts.txt

# counted the number of unique go terms assigned for biological processes, cellular
components, and molecular functions
wc -l 5_b2g_biological_process_counts.txt

wc -l 5_b2g_cellular_component_counts.txt

wc -l 5_b2g_molecular_function_counts.txt

# took the count files and opened them in excel. retained the first 15 lines. then added up the
counts in column one of lines 16-end, placed the sum in line 16, and titled it "other".
```

Created bar graphs in RStudio to represent the BLAST2GO results

```
#horizontal bar graph to depict biological processes assigned by blast2go

#read input file containing data

biological_processes <-
read.csv(file="/Volumes/Transcend/Thesis/bioinformatics_files/4-
blast2go/6_b2g_biological_process_counts_for_R.csv", sep=",", header=FALSE)

#take all column 1 values from input file and assign them to the term
"frequencies"

frequencies <- biological_processes$V1
```

```
#take all column 3 values from input file and assign them to the term
"processes"
processes <- biological_processes$V3

#shift the parameters of the graph so that it will fit onto the output page
par(mar=c(3, 13, 1, 0),mgp=c(2,0.5,0))

#create a color palette
palette1 <- c("darkgreen", "forestgreen", "lightgreen", "lightcyan",
"lightsteelblue1", "pink")

#create a ramped palette using the above color palette
palette1Ramp <- colorRampPalette(palette1)

#plot
barplot(frequencies,horiz=TRUE,inside=TRUE, axes=TRUE, xlab="Number of
Results",beside=FALSE,space=0.2,xpd=FALSE,xlim=c(0,1600),ax
isnames=TRUE,names.arg=processes,
las=1, cex=0.8,cex.names=0.8,col=palette1Ramp(length(processes))) #add labels
for bar heights text(x=frequencies,y=barpos,labels=frequencies,cex=0.8,pos=4)

#horizontal bar graph to depict cellular components assigned by blast2go

#read input file containing data
cellular_components <-
read.csv(file="/Volumes/Transcend/Thesis/bioinformatics_files/4-
blast2go/6_b2g_cellular_component_counts_for_R.csv", sep=",", header=FALSE)

#take all column 1 values from input file and assign them to the term
"frequencies"
frequencies <- cellular_components$V1

#take all column 3 values from input file and assign them to the term
"processes"
processes <- cellular_components$V3

#shift the parameters of the graph so that it will fit onto the output page
par(mar=c(3, 12, 1, 1),mgp=c(2,0.5,0))

#load the RColorBrewer package
library(RColorBrewer)
```

```
#take the stock color palette "Spectral" and turn it into a ramped palette.
This is so that I can use it on datasets containing more than 11 data points
(the Spectral color palette has only 11 colors in it) RdYlBuRamp<-
colorRampPalette(brewer.pal(11,"RdYlBu"))

#plot
barplot(frequencies,horiz=TRUE,
axes=TRUE,xlab="Number of
Results",beside=FALSE,space=0.2,xpd=FALSE,xlim=c(0,500),axi
snames=TRUE,names.arg=processes,
las=1, cex=0.8,cex.names=0.8,col=RdYlBuRamp(length(processes)),densi
ty=1000,angle=270)

#add labels for bar heights
text(x=frequencies,y=barpos,labels=frequencies,cex=0.8,pos=4)

#horizontal bar graph to depict molecular functions assigned by blast2go

#read input file containing data
molecular_functions <-
read.csv(file="/Volumes/Transcend/Thesis/bioinformatics_files/4-
blast2go/6_b2g_molecular_function_counts_for_R.csv", sep=",", header=FALSE)

#take all column 1 values from input file and assign them to the term
"frequencies"
frequencies <- molecular_functions$V1

#take all column 3 values from input file and assign them to the term
"processes"
processes <- molecular_functions$V3

#shift the parameters of the graph so that it will fit onto the output page
par(mar=c(3, 12, 1, 1),mgp=c(2,0.5,0))

#create a color palette

palette2 <- c("darkslategrey", "darkcyan", "cadetblue4", "cadetblue2",
"cadetblue1", "aquamarine", "aquamarine1", "darkseagreen1", "snow2",
"thistle1", "plum1", "plum2", "plum3")
```

```
#create a ramped palette using the above color palette
palette2Rmp <- colorRampPalette(palette2)

#plot
barpos <- barplot(frequencies,horiz=TRUE, axes=TRUE,xlab="Number of
Results",beside=FALSE,space=0.2,xpd=FALSE,xlim=c(0,2000),ax
isnames=TRUE,names.arg=processes,
las=1, cex=0.8,cex.names=0.8,col=palette2Rmp(length(processes)),densit
y=1000,angle=45)

#add labels for bar heights
text(x=frequencies,y=barpos,labels=frequencies,cex=0.8,pos=4)
```

## APPENDIX D

**Image Processing for Determination of the Percentage of Fluorescent Cells**

**ImageJ macro used to subtract all of the "hot pixels" present in the fluorescent images**

```
waitForUser("Select the directories in the following order:\n\n1. Input
directory\n\n2. Output Directory\n\n3. Directory to move images already
analyzed")


dir1 = getDirectory("Choose a Directory")
dir2 = getDirectory("Choose a Directory")
dir3 = getDirectory("Choose a Directory")
list = getFileList(dir1)

for (i=0
i<list.length
i++) { showProgress(i+1, list.length)
open(dir1+list[i])

run("Brightness/Contrast...")

selectWindow(list[i])
run("Maximize")

waitForUser("Adjust Brightness/Contrast")

selectWindow(list[i])

saveAs("Tiff", dir2 + list[i] + "_no_bg")
run("Close")

selectWindow("B&C");
run("Close");
File.rename(dir1 + list[i], dir3 + list[i])


}
```

**ImageJ macro used to crop the brightfield and fluorescent images of the hemacytometer**

```
waitForUser("Select the directories in the following order:\n\n1. Brightfield
Directory\n\n2. Fluorescence Directory \n\n3. Brightfield Output
Directory\n\n4. Fluorescence Output Directory")
```

```
dir1 = getDirectory("Choose a Directory")
dir2 = getDirectory("Choose a Directory")
dir3 = getDirectory("Choose a Directory")
dir4 = getDirectory("Choose a Directory")

list1 = getFileList(dir1)

for (i=0
i<list1.length
i++) { showProgress(i+1, list1.length)
open(dir1+list1[i])

pc_orig_filename = File.name;
unwanted_part = "_pc.tif"
len1=lengthOf(pc_orig_filename)
len2=lengthOf(unwanted_part)
len3=indexOf(pc_orig_filename, unwanted_part)
string1=substring(pc_orig_filename, len2+len3,len1)
string2=substring(pc_orig_filename, 0,len3)
shortened_name=string2+string1

selectWindow(list1[i])

open(dir2 + shortened_name + "_fl_clean.tif_no_bg.tif")
fl_orig_filename = File.nameWithoutExtension

run("Images to Stack", "name=Stack title=[] use");
run("Scale to Fit");

waitForUser("Select Area of Interest", "Select the perimeter of the
hemacytometer")
run("Crop")

run("Stack to Images")

saveAs("tiff", dir4 + fl_orig_filename+"_cropped")
close();

saveAs("tiff", dir3 + pc_orig_filename+"_cropped")
close()
```

**ImageJ macro used to automatically score the number of cells in the cropped brightfield image of the hemacytometer grid**

```
waitForUser("Select the directories in the following order:\n\n1.Input
directory\n\n2. Output Directory\n\n3. Directory to move images already
analyzed")


dir1 = getDirectory("Choose a Directory")
dir2 = getDirectory("Choose a Directory")
dir3 = getDirectory("Choose a Directory")
list = getFileList(dir1)

setBatchMode(true)

for (i=0
i<list.length
i++) { showProgress(i+1, list.length)
open(dir1+list[i])

run("Scale to Fit")

run("Bandpass Filter...", "filter_large=40 filter_small=3 suppress=None
tolerance=5 autoscale saturate")
setOption("BlackBackground", false);
run("Smooth")

run("Make Binary")
run("Watershed")

run("Despeckle")
run("Despeckle")
run("Despeckle")
run("Erode")
run("Watershed")
run("Dilate")
run("Watershed")

run("Analyze Particles...", "size=25-Infinity circularity=0.30-1.00 summarize
add")

selectWindow(list[i])
run("Close")
```

```
open(dir1 + list[i]);
name = File.nameWithoutExtension

roiManager("Measure");
saveAs("Measurements", dir2 + "measurements_" + name + ".csv")

roiManager("Show All without labels")
roiManager("Set Color", "red")
roiManager("Set Line Width", 0)
run("Flatten")

saveAs("Tiff", dir2 + name + "_counted_overlay")
roiManager("reset")

run("Close")

File.rename(dir1 + list[i], dir3 + list[i])

}


selectWindow("Summary")


saveAs("Text", dir2+"Cell Counts ")
```

**ImageJ macro used to manually verify the counts produced from the automated counts of the cropped brightfield image. Note that you must select the "multi-point" tool from the ImageJ bar for this to work.**

```
waitForUser("Select the directories in the following order:\n\n1.Input
directory\n\n2. Output Directory\n\n3. Directory to move images already
analyzed")

dir1 = getDirectory("Choose a Directory")
dir2 = getDirectory("Choose a Directory")
dir3 = getDirectory("Choose a Directory")

list = getFileList(dir1)

for (i=0
i<list.length
i++) { showProgress(i+1, list.length)
```

```
open(dir1+list[i])

setBatchMode(false)
run("Maximize");
run("In [+]");
run("In [+]")
setLocation(0,0,1425,875)
run("Point Tool...")
run("Properties... ", "show")
setBatchMode(true)

selectWindow("Counts_" + list[i]);
saveAs("Text", dir2 + "Manual Counts - " + list [i])
run("Close")

selectWindow(list[i]);
saveAs("Tiff", dir2 + "Manual Counts - " + list[i])
run("Close")

File.rename(dir1 + list[i], dir3 + list[i])
```

**ImageJ macro used to automatically score the number of cells in the cropped fluorescent image of the hemacytometer grid.**

```
waitForUser("Select the directories in the following order:\n\n1.Input
directory\n\n2. Output Directory\n\n3. Directory to move analyzed images")

dir1 = getDirectory("Choose a Directory")
dir2 = getDirectory("Choose a Directory")
dir3 = getDirectory("Choose a Directory")
list = getFileList(dir1)

setBatchMode(true)

for (i=0
i<list.length
i++) { showProgress(i+1, list.length)
open(dir1+list[i])

setMinAndMax(0, 50)
run("Convert to Mask")
```

```
run("Watershed");
run("Analyze Particles...", "size=5-Infinity add summarize")
roiManager("Show All without labels")


selectWindow(list[i])
run("Close")


open(dir1 + list[i]);
name = File.nameWithoutExtension;
File.rename(dir1 + list[i], dir3 + list[i])
roiManager("Measure");
saveAs("Measurements", dir2 + "measurements_" + name + ".csv")


roiManager("Show All without labels")
run("Flatten");
saveAs("Tiff", dir2 + name + "_counted_overlay")
roiManager("reset")


selectWindow("Results")
run("Close");
}
selectWindow("Summary");
saveAs("Text", dir2+"Fluorescent Cell Counts")


run("Close")
```

**ImageJ macro used to manually verify the counts produced from the automated counts of the cropped fluorescent image. Note that the "multi-point" tool from the ImageJ bar must be selected for this to work.**

```
waitForUser("Select the directories in the following order:\n\n1. Input
fluorescence directory\n\n2. Location of Brightfield Images to Overlay \n\n3.
Output Directory\n\n4. Location to move analyzed images")


dir1 = getDirectory("Choose a Directory")
dir2 = getDirectory("Choose a Directory")
dir3 = getDirectory("Choose a Directory")
dir4 = getDirectory("Choose a Directory")


list1 = getFileList(dir1)


for (i=0
i<list1.length
```

```
i++) { showProgress(i+1, list1.length)
open(dir1+list1[i])

setBatchMode(true)

name = File.name

unwanted_part = "_fl_clean.tif_no_bg_cropped_counted_overlay.tif"
len1=lengthOf(name)
len2=lengthOf(unwanted_part)
len3=indexOf(name,unwanted_part)
string1=substring(name, len2+len3,len1)
string2=substring(name, 0,len3)
shortened_name=string2+string1

setMinAndMax(0,40)

setBatchMode(false)
selectWindow(list1[i])

open(dir2 + shortened_name + "_pc.tif_cropped.tif")
bf_image = File.nameWithoutExtension

run("Add Image...", "image=bf_image x=0 y=0 opacity=40")
run("Flatten")

selectWindow(list1[i]);
close;
selectWindow(bf_image + ".tif")

close

selectWindow(bf_image + "-1.tif")
run("Maximize");
run("In [+]");
run("In [+]")
setLocation(0,0,1425,875)

run("Point Tool...")

run("Properties... ", "show")
```

```
setBatchMode(true)

selectWindow("Counts_" + bf_image + "-1.tif")
saveAs("Text", dir3 + "Manual Counts - " + list1[i])
run("Close")

selectWindow(bf_image + "-1.tif");
saveAs("Tiff", dir3 + "Manual Counts - " + list1[i])
run("Close")

File.rename(dir1 + list1[i], dir4 + list1[i])
```

# APPENDIX E

## Analysis and Visualization of Image Data

**Analysis of Growth Curve Data**

```
library(dplyr)
library(ggplot2)
library(reshape2)
library(multcomp)
library(nlme)

raw_data <- read.csv(file="/Users/mattbreuer/Dropbox/Thesis/II - Mutant
Experiments/2. Growth Curve
Analysis/raw_data.csv",sep=",",header=TRUE)

summary(raw_data)

raw_data <- mutate(raw_data, strain = relevel(strain, ref = "wt"))

raw_data$density <- raw_data$cells_counted*20000

levels(raw_data$strain)

summ_data <- raw_data %>%
  group_by(time,strain) %>%
  summarise(mean_density = mean(density),
        stderror_density = sd(density)/sqrt(n()))

#make the plot first
plot <- ggplot(summ_data, aes(x=time,y=mean_density,colour=strain))

#make the plot theme
complete_plot <- plot + theme(panel.border = element_rect(colour = "black",
fill = NA),
                panel.background = element_rect(fill = "white", colour =
"grey"),
                panel.grid.major.x = element_blank(),
```

```
                        panel.grid.major.y = element_line(colour = "grey",
linetype = "solid", size=0.1),
                        axis.text = element_text(size=9),
                        axis.title = element_text(size=12),
                        axis.line.x = element_line(colour = "black", linetype =
"solid"),
                        axis.line.y = element_line(colour = "black", linetype =
"solid"),
                        axis.ticks.x = element_line(size = 0.5),
                        axis.ticks.y = element_blank(),
                        legend.box.background = element_rect(),
                        legend.key = element_rect(fill = "white", colour = "white"),
                        legend.text = element_text(face="italic")) +


  #then change the axes values and labels
  scale_x_continuous(breaks = c(1,2,3,4,5,6), expand = c(0,0.1)) +
  scale_y_continuous(limits = c(0, 12000000), breaks =
c(0,1000000,2000000,3000000,4000000,5000000,6000000,7000000,80000
00,9000000,10000000,11000000,12000000), expand = c(0,0)) +
  labs(x="Time (Days)", y = "Cell Density (cells/mL") +


  #then customize how the plots will look. error bars laid down first, then the
line,
  #then the black outline for each dot, then the colored dots on top of
everything.
  geom_errorbar(aes(ymin=mean_density - stderror_density,
              ymax=mean_density + stderror_density),
          width=0.1) +
  geom_line() +
  geom_point(colour = "black", size=1.5) +
  geom_point(size = 1.2) +
  ggsave("/Users/mattbreuer/Desktop/scatterplot.tiff", width = 8, height = 6,
dpi = 1500)



###statistical analysis
#first use the raw_data dataframe to create a new dataframe to use for stats
```

```
stats_df <- raw_data[,c(1,2,3,5)]

#relevel the dataframe to use the wt strain as the reference
stats_df <- mutate(stats_df, strain = relevel(strain, ref = "wt"))

#turn the time points into factors, since I used numerical values
stats_df$time <- factor(stats_df$time)

#carry out the anova
model <- aov(density ~ (strain*time) + Error(replicate/(strain*time)), data =
stats_df)
summary(model)

##post-hoc contrasts
stats_df$strain.time <- interaction(stats_df$strain,stats_df$time)
library(lme4)
library(lmerTest)

m <- lmer(density ~ strain.time + (1 | replicate:strain) + (1 | replicate:time),
stats_df)

summary(glht(m, linfct=
          c("strain.timewt.1 - strain.timeire1.1 = 0",
            "strain.timewt.1 - strain.timepig3.1 = 0",
            "strain.timewt.1 - strain.timebi1.1 = 0",
            "strain.timewt.2 - strain.timeire1.2 = 0",
            "strain.timewt.2 - strain.timepig3.2 = 0",
            "strain.timewt.2 - strain.timebi1.2 = 0",
            "strain.timewt.3 - strain.timeire1.3 = 0",
            "strain.timewt.3 - strain.timepig3.3 = 0",
            "strain.timewt.3 - strain.timebi1.3 = 0",
            "strain.timewt.4 - strain.timeire1.4 = 0",
            "strain.timewt.4 - strain.timepig3.4 = 0",
            "strain.timewt.4 - strain.timebi1.4 = 0",
            "strain.timewt.5 - strain.timeire1.5 = 0",
            "strain.timewt.5 - strain.timepig3.5 = 0",
```

```
        "strain.timewt.5 - strain.timebi1.5 = 0")),
    test = adjusted(type = "holm"))
```

**Analysis of Percent Fluorescent Cells with SYTOX Green**

```
library(dplyr)
library(ggplot2)
library(reshape2)
library(multcomp)
library(nlme)

raw_data <- read.csv(file="/Users/mattbreuer/Dropbox/Thesis/II - Mutant
Experiments/5. SYTOX Assay/raw_data.csv",sep=",",header=TRUE)

summary(raw_data)

raw_data$percent_fluorescence <-
(((raw_data$auto_counted_cells_fluorescence +
raw_data$uncounted_fluorescent_cells -
raw_data$non_cells_fluorescence_counted) /
(raw_data$auto_counted_cells_total + raw_data$uncounted_total_cells -
raw_data$miscounted_total_cells)*100))

raw_data <- mutate(raw_data, strain = relevel(strain, ref = "wt"))

levels(raw_data$strain)

summ_data <- raw_data %>%
  group_by(time,strain) %>%
  summarise(mean_fluorescence = mean(percent_fluorescence),
        stderror_fluorescence = sd(percent_fluorescence)/sqrt(n()))

#make the plot first
plot <- ggplot(summ_data, aes(x=time,y=mean_fluorescence,colour=strain))

#make the plot theme
complete_plot <- plot + theme(panel.border = element_rect(colour = "black",
fill = NA),
```

```
                        panel.background = element_rect(fill = "white", colour =
"grey"),

                        panel.grid.major.x = element_blank(),

                        panel.grid.major.y = element_line(colour = "grey",
linetype = "solid", size=0.1),

                        axis.text = element_text(size=9),

                        axis.title = element_text(size=12),

                        axis.line.x = element_line(colour = "black", linetype =
"solid"),

                        axis.line.y = element_line(colour = "black", linetype =
"solid"),

                        axis.ticks.x = element_line(size = 0.5),

                        axis.ticks.y = element_blank(),

                        legend.key = element_rect(fill = "white", colour = "white"),

                        legend.box.margin = margin(t = 2, r = 2, b = 2, l = 2, unit
= "pt"),

                        legend.box.background = element_rect(colour = "black"),

                        legend.text = element_text(face="italic", size =10)) +


  #then change the axes values and labels
  scale_x_continuous(breaks = c(0,30,60,90,120,150,180,210,240), expand
= c(0,3)) +

  scale_y_continuous(limits = c(0, 105), breaks =
c(0,10,20,30,40,50,60,70,80,90,100,110), expand = c(0,0)) +

  labs(x="Time (minutes)", y = "Percentage of Fluorescent Cells") +


  #then customize how the plots will look. error bars layed down first, then
the line,
  #then the black outline for each dot, then the colored dots on top of
everything.
  geom_vline(xintercept=120, linetype=3) +

  geom_label(aes(x=150, label="Removal from \nHeat Stress", y=95),
colour="black", angle=0, size=3.5) +

  geom_segment(aes(x = 135, y = 95, xend = 120, yend =95, colour =
"black"),

          arrow = arrow(length = unit(0.2, "cm")),colour = "black") +

  geom_errorbar(aes(ymin=mean_fluorescence - stderror_fluorescence,
            ymax=mean_fluorescence + stderror_fluorescence),
```

```
        width=2) +
  geom_line() +
  geom_point(colour = "black", size=1.5) +
  geom_point(size = 1.2) +
  ggsave("~/Desktop/sytox_scatterplot.tiff", width = 8, height = 6, dpi = 1500)

###statistical analysis
#first use the raw_data dataframe to create a new dataframe to use for stats
stats_df <- raw_data[,c(1,2,3,10)]

#relevel the dataframe to use the wt strain as the reference
stats_df <- mutate(stats_df, strain = relevel(strain, ref = "wt"))

#turn the time points into factors, since I used numerical values
stats_df$time <- factor(stats_df$time)

#carry out the anova
model <- aov(percent_fluorescence ~ (strain*time) +
Error(replicate/(strain*time)), data = stats_df)
summary(model)

##post-hoc contrasts
stats_df$strain.time <- interaction(stats_df$strain,stats_df$time)
library(lme4)
library(lmerTest)

m <- lmer(percent_fluorescence ~ strain.time + (1 | replicate:strain) + (1 |
replicate:time), stats_df)

summary(glht(m, linfct=
          c("strain.timewt.30 - strain.timeire1.30 = 0",
            "strain.timewt.30 - strain.timepig3.30 = 0",
            "strain.timewt.30 - strain.timebi1.30 = 0",
            "strain.timewt.60 - strain.timeire1.60 = 0",
            "strain.timewt.60 - strain.timepig3.60 = 0",
            "strain.timewt.60 - strain.timebi1.60 = 0",
```

```
                    "strain.timewt.90 - strain.timeire1.90 = 0",
                    "strain.timewt.90 - strain.timepig3.90 = 0",
                    "strain.timewt.90 - strain.timebi1.90 = 0",
                    "strain.timewt.120 - strain.timeire1.120 = 0",
                    "strain.timewt.120 - strain.timepig3.120 = 0",
                    "strain.timewt.120 - strain.timebi1.120 = 0",
                    "strain.timewt.150 - strain.timeire1.150 = 0",
                    "strain.timewt.150 - strain.timepig3.150 = 0",
                    "strain.timewt.150 - strain.timebi1.150 = 0",
                    "strain.timewt.180 - strain.timeire1.180 = 0",
                    "strain.timewt.180 - strain.timepig3.180 = 0",
                    "strain.timewt.180 - strain.timebi1.180 = 0")),
          test = adjusted(type = "holm"))
```

**Analysis of Percent Fluorescent Cells with FDA**

```
library(dplyr)
library(ggplot2)
library(reshape2)
library(multcomp)
library(nlme)


raw_data <- read.csv(file="/Users/mattbreuer/Dropbox/Thesis/II - Mutant
Experiments/3. Fluorescein Diacatate Hydrolysis
Assay/raw_data.csv",sep=",",header=TRUE)


summary(raw_data)


raw_data$percent_fluorescence <-
(((raw_data$auto_counted_cells_fluorescence +
raw_data$uncounted_fluorescent_cells -
raw_data$non_cells_fluorescence_counted) /
(raw_data$auto_counted_cells_total + raw_data$uncounted_total_cells -
raw_data$miscounted_total_cells)*100))


raw_data <- mutate(raw_data, strain = relevel(strain, ref = "wt"))


levels(raw_data$strain)
```

```
summ_data <- raw_data %>%
  group_by(time,strain) %>%
  summarise(mean_fluorescence = mean(percent_fluorescence),
          stderror_fluorescence = sd(percent_fluorescence)/sqrt(n()))

#make the plot first
plot <- ggplot(summ_data, aes(x=time,y=mean_fluorescence,colour=strain))

#make the plot theme
complete_plot <- plot + theme(panel.border = element_rect(colour = "black",
fill = NA),
                  panel.background = element_rect(fill = "white", colour =
"grey"),
                  panel.grid.major.x = element_blank(),
                  panel.grid.major.y = element_line(colour = "grey",
linetype = "solid", size=0.1),
                  axis.text = element_text(size=9),
                  axis.title = element_text(size=12),
                  axis.line.x = element_line(colour = "black", linetype =
"solid"),
                  axis.line.y = element_line(colour = "black", linetype =
"solid"),
                  axis.ticks.x = element_line(size = 0.5),
                  axis.ticks.y = element_blank(),
                  legend.key = element_rect(fill = "white", colour = "white"),
                  legend.box.margin = margin(t = 2, r = 2, b = 2, l = 2, unit
= "pt"),
                  legend.box.background = element_rect(colour = "black"),
                  legend.text = element_text(face="italic", size =10)) +

  #then change the axes values and labels
  scale_x_continuous(breaks = c(0,60,120,180,240), expand = c(0,3)) +
  scale_y_continuous(limits = c(0, 105), breaks =
c(0,10,20,30,40,50,60,70,80,90,100,110), expand = c(0,0)) +
  labs(x="Time (minutes)", y = "Percentage of Fluorescent Cells") +
```

```r
  #then customize how the plots will look. error bars layed down first, then the line,
  #then the black outline for each dot, then the colored dots on top of everything.
  geom_vline(xintercept=120, linetype=3) +
  geom_label(aes(x=156, label="Removal from \nHeat Stress", y=95), colour="black", angle=0, size=3.5) +
  geom_segment(aes(x = 135, y = 95, xend = 120, yend =95, colour = "black"),
            arrow = arrow(length = unit(0.2, "cm")),colour = "black") +
  geom_errorbar(aes(ymin=mean_fluorescence - stderror_fluorescence,
              ymax=mean_fluorescence + stderror_fluorescence),
          width=2) +
  geom_line() +
  geom_point(colour = "black", size=1.5) +
  geom_point(size = 1.2) +
  ggsave("~/Desktop/fda_scatterplot.tiff", width = 8, height = 6, dpi = 1500)

###statistical analysis
#first use the raw_data dataframe to create a new dataframe to use for stats
stats_df <- raw_data[,c(1,2,3,10)]

#relevel the dataframe to use the wt strain as the reference
stats_df <- mutate(stats_df, strain = relevel(strain, ref = "wt"))

#turn the time points into factors, since I used numerical values
stats_df$time <- factor(stats_df$time)

#carry out the anova
model <- aov(percent_fluorescence ~ (strain*time) + Error(replicate/(strain*time)), data = stats_df)
summary(model)

##post-hoc contrasts
stats_df$strain.time <- interaction(stats_df$strain,stats_df$time)
library(lme4)
library(lmerTest)
```

```
m <- lmer(percent_fluorescence ~ strain.time + (1 | replicate:strain) + (1 |
replicate:time), stats_df)

summary(glht(m, linfct=
        c("strain.timewt.30 - strain.timeire1.30 = 0",
          "strain.timewt.30 - strain.timepig3.30 = 0",
          "strain.timewt.30 - strain.timebi1.30 = 0",
          "strain.timewt.60 - strain.timeire1.60 = 0",
          "strain.timewt.60 - strain.timepig3.60 = 0",
          "strain.timewt.60 - strain.timebi1.60 = 0",
          "strain.timewt.90 - strain.timeire1.90 = 0",
          "strain.timewt.90 - strain.timepig3.90 = 0",
          "strain.timewt.90 - strain.timebi1.90 = 0",
          "strain.timewt.120 - strain.timeire1.120 = 0",
          "strain.timewt.120 - strain.timepig3.120 = 0",
          "strain.timewt.120 - strain.timebi1.120 = 0",
          "strain.timewt.150 - strain.timeire1.150 = 0",
          "strain.timewt.150 - strain.timepig3.150 = 0",
          "strain.timewt.150 - strain.timebi1.150 = 0",
          "strain.timewt.180 - strain.timeire1.180 = 0",
          "strain.timewt.180 - strain.timepig3.180 = 0",
          "strain.timewt.180 - strain.timebi1.180 = 0")),
    test = adjusted(type = "holm"))
```

**Analysis of Percent Fluorescent Cells with CM-H$_2$DCFDA**

```
library(dplyr)
library(ggplot2)
library(reshape2)
library(multcomp)
library(nlme)


raw_data <- read.csv(file="/Users/mattbreuer/Dropbox/Thesis/II - Mutant
Experiments/4. ROS Accumulation
Assay/ros/combined/raw_data.csv",sep=",",header=TRUE)
```

```
summary(raw_data)

raw_data$percent_fluorescence <-
(((raw_data$auto_counted_cells_fluorescence +
raw_data$uncounted_fluorescent_cells -
raw_data$non_cells_fluorescence_counted) /
(raw_data$auto_counted_cells_total + raw_data$uncounted_total_cells -
raw_data$miscounted_total_cells)*100))

raw_data <- mutate(raw_data, strain = relevel(strain, ref = "wt"))

levels(raw_data$strain)

summ_data <- raw_data %>%
  group_by(time,strain) %>%
  summarise(mean_fluorescence = mean(percent_fluorescence),
       stderror_fluorescence = sd(percent_fluorescence)/sqrt(n()))

#make the plot first
plot <- ggplot(summ_data, aes(x=time,y=mean_fluorescence,colour=strain))

#make the plot theme
complete_plot <- plot + theme(panel.border = element_rect(colour = "black",
fill = NA),
                 panel.background = element_rect(fill = "white", colour =
"grey"),
                 panel.grid.major.x = element_blank(),
                 panel.grid.major.y = element_line(colour = "grey",
linetype = "solid", size=0.1),
                 axis.text = element_text(size=9),
                 axis.title = element_text(size=12),
                 axis.line.x = element_line(colour = "black", linetype =
"solid"),
                 axis.line.y = element_line(colour = "black", linetype =
"solid"),
                 axis.ticks.x = element_line(size = 0.5),
                 axis.ticks.y = element_blank(),
                 legend.key = element_rect(fill = "white", colour = "white"),
```

```
                    legend.box.margin = margin(t = 2, r = 2, b = 2, l = 2, unit
= "pt"),

                    legend.box.background = element_rect(colour = "black"),
                    legend.text = element_text(face="italic", size =10)) +


  #then change the axes values and labels
  scale_x_continuous(breaks = c(0,30,60,90,120), expand = c(0,3)) +
  scale_y_continuous(limits = c(0, 105), breaks =
c(0,10,20,30,40,50,60,70,80,90,100,110), expand = c(0,0)) +
  labs(x="Time (minutes)", y = "Percentage of Fluorescent Cells") +


  #then customize how the plots will look. error bars layed down first, then
the line,
  #then the black outline for each dot, then the colored dots on top of
everything.
  geom_vline(xintercept=120, linetype=3) +
  geom_label(aes(x=95, label="Removal from \nHeat Stress", y=95),
colour="black", angle=0, size=3.5) +
  geom_segment(aes(x = 105, y = 95, xend = 120, yend = 95, colour =
"black"),
          arrow = arrow(length = unit(0.2, "cm")),colour = "black") +
  geom_errorbar(aes(ymin=mean_fluorescence - stderror_fluorescence,
            ymax=mean_fluorescence + stderror_fluorescence),
          width=2) +
  geom_line() +
  geom_point(colour = "black", size=1.5) +
  geom_point(size = 1.2) +
  ggsave("~/Desktop/ros_scatterplot.tiff", width = 8, height = 6, dpi = 1500)

###statistical analysis
#first use the raw_data dataframe to create a new dataframe to use for stats
stats_df <- raw_data[,c(1,2,3,10)]

#relevel the dataframe to use the wt strain as the reference
stats_df <- mutate(stats_df, strain = relevel(strain, ref = "wt"))

#turn the time points into factors, since I used numerical values
```

```
stats_df$time <- factor(stats_df$time)

#carry out the anova
model <- aov(percent_fluorescence ~ (strain*time) +
Error(replicate/(strain*time)), data = stats_df)
summary(model)

##post-hoc contrasts
stats_df$strain.time <- interaction(stats_df$strain,stats_df$time)
library(lme4)
library(lmerTest)

m <- lmer(percent_fluorescence ~ strain.time + (1 | replicate:strain) + (1 |
replicate:time), stats_df)

summary(glht(m, linfct=
          c("strain.timewt.30 - strain.timeire1.30 = 0",
            "strain.timewt.30 - strain.timepig3.30 = 0",
            "strain.timewt.30 - strain.timebi1.30 = 0",
            "strain.timewt.60 - strain.timeire1.60 = 0",
            "strain.timewt.60 - strain.timepig3.60 = 0",
            "strain.timewt.60 - strain.timebi1.60 = 0",
            "strain.timewt.90 - strain.timeire1.90 = 0",
            "strain.timewt.90 - strain.timepig3.90 = 0",
            "strain.timewt.90 - strain.timebi1.90 = 0",
            "strain.timewt.120 - strain.timeire1.120 = 0",
            "strain.timewt.120 - strain.timepig3.120 = 0",
            "strain.timewt.120 - strain.timebi1.120 = 0")),
      test = adjusted(type = "holm"))
```

**Analysis of Percent Fluorescent Cells with Annexin V**

```
library(dplyr)
library(ggplot2)
library(reshape2)
library(multcomp)
library(nlme)
```

```r
raw_data <- read.csv(file="/Users/mattbreuer/Dropbox/Thesis/II - Mutant
Experiments/6. AnnexinV Assay/raw_data.csv",sep=",",header=TRUE)


summary(raw_data)


raw_data$percent_fluorescence <-
(((raw_data$auto_counted_cells_fluorescence +
raw_data$uncounted_fluorescent_cells -
raw_data$non_cells_fluorescence_counted) /
(raw_data$auto_counted_cells_total + raw_data$uncounted_total_cells -
raw_data$miscounted_total_cells)*100))


raw_data <- mutate(raw_data, strain = relevel(strain, ref = "wt"))


levels(raw_data$strain)


summ_data <- raw_data %>%
  group_by(time,strain) %>%
  summarise(mean_fluorescence = mean(percent_fluorescence),
        stderror_fluorescence = sd(percent_fluorescence)/sqrt(n()))


#make the plot first
plot <- ggplot(summ_data, aes(x=time,y=mean_fluorescence,colour=strain))


#make the plot theme
complete_plot <- plot + theme(panel.border = element_rect(colour = "black",
fill = NA),

                    panel.background = element_rect(fill = "white", colour =
"grey"),

                    panel.grid.major.x = element_blank(),

                    panel.grid.major.y = element_line(colour = "grey",
linetype = "solid", size=0.1),

                    axis.text = element_text(size=9),

                    axis.title = element_text(size=12),

                    axis.line.x = element_line(colour = "black", linetype =
"solid"),
```

```
                        axis.line.y = element_line(colour = "black", linetype =
"solid"),

                        axis.ticks.x = element_line(size = 0.5),
                        axis.ticks.y = element_blank(),
                        legend.key = element_rect(fill = "white", colour = "white"),
                        legend.box.margin = margin(t = 2, r = 2, b = 2, l = 2, unit
= "pt"),

                        legend.box.background = element_rect(colour = "black"),
                        legend.text = element_text(face="italic", size =10)) +


  #then change the axes values and labels
  scale_x_continuous(breaks = c(0,60,120,180,240), expand = c(0,3)) +
  scale_y_continuous(limits = c(0, 105), breaks =
c(0,10,20,30,40,50,60,70,80,90,100,110), expand = c(0,0)) +
  labs(x="Time (minutes)", y = "Percentage of Fluorescent Cells") +


  #then customize how the plots will look. error bars layed down first, then
the line,
  #then the black outline for each dot, then the colored dots on top of
everything.
  geom_vline(xintercept=120, linetype=3) +
  geom_label(aes(x=155, label="Removal from \nHeat Stress", y=95),
colour="black", angle=0, size=3.5) +
  geom_segment(aes(x = 135, y = 95, xend = 120, yend =95, colour =
"black"),
           arrow = arrow(length = unit(0.2, "cm")),colour = "black") +
  geom_errorbar(aes(ymin=mean_fluorescence - stderror_fluorescence,
              ymax=mean_fluorescence + stderror_fluorescence),
          width=2) +
  geom_line() +
  geom_point(colour = "black", size=1.5) +
  geom_point(size = 1.2) +
  ggsave("~/Desktop/annexinV_scatterplot.tiff", width = 8, height = 6, dpi =
1500)


###statistical analysis
#first use the raw_data dataframe to create a new dataframe to use for stats
```

```
stats_df <- raw_data[,c(1,2,3,10)]

#relevel the dataframe to use the wt strain as the reference
stats_df <- mutate(stats_df, strain = relevel(strain, ref = "wt"))

#turn the time points into factors, since I used numerical values
stats_df$time <- factor(stats_df$time)

#carry out the anova
model <- aov(percent_fluorescence ~ (strain*time) +
Error(replicate/(strain*time)), data = stats_df)
summary(model)

##post-hoc contrasts
stats_df$strain.time <- interaction(stats_df$strain,stats_df$time)
library(lme4)
library(lmerTest)

m <- lmer(percent_fluorescence ~ strain.time + (1 | replicate:strain) + (1 |
replicate:time), stats_df)

summary(glht(m, linfct=
          c("strain.timewt.60 - strain.timeire1.60 = 0",
            "strain.timewt.60 - strain.timepig3.60 = 0",
            "strain.timewt.60 - strain.timebi1.60 = 0",
            "strain.timewt.120 - strain.timeire1.120 = 0",
            "strain.timewt.120 - strain.timepig3.120 = 0",
            "strain.timewt.120 - strain.timebi1.120 = 0",
            "strain.timewt.180 - strain.timeire1.180 = 0",
            "strain.timewt.180 - strain.timepig3.180 = 0",
            "strain.timewt.180 - strain.timebi1.180 = 0",
            "strain.timewt.240 - strain.timeire1.240 = 0",
            "strain.timewt.240 - strain.timepig3.240 = 0",
            "strain.timewt.240 - strain.timebi1.240 = 0")),
      test = adjusted(type = "holm"))
```

**Analysis of Growth Area**

```
library(dplyr)
library(ggplot2)
library(reshape2)
library(multcomp)
library(nlme)

raw_data <- read.csv(file="/Users/mattbreuer/Dropbox/Thesis/II - Mutant
Experiments/7. Plating Assay/raw_data.csv",sep=",",header=TRUE)

summary(raw_data)

raw_data <- mutate(raw_data, strain = relevel(strain, ref = "wt"))

levels(raw_data$strain)

summ_data <- raw_data %>%
  group_by(time,strain) %>%
  summarise(mean_area = mean(total_area),
        stderror_total_area = sd(total_area)/sqrt(n()))

#first use the raw_data dataframe to create the dataframe to use for boxplot
boxplot_df <- raw_data[,c(1,2,3,5)]

#then add a column to the boxplot dataframe, specifying the combination of
the independent variables (strain and time) for each row
boxplot_df$strain.time <- interaction(boxplot_df$strain, boxplot_df$time)

#then create and save the boxplot
box_plot <- ggplot(boxplot_df, aes(x = strain.time, y = total_area, fill =
strain)) +
  geom_boxplot() +
  theme(panel.border = element_rect(colour = "black", fill = NA),
      panel.background = element_rect(fill = "white", colour = "grey"),
      panel.grid.major.x = element_blank(),
      panel.grid.major.y = element_line(colour = "grey", linetype = "solid",
size=0.1),
```

```
    axis.text = element_text(size=10),
    axis.text.x = element_text(face="italic"),
    axis.title = element_text(size=13),
    axis.line.x = element_line(colour = "black", linetype = "solid"),
    axis.line.y = element_line(colour = "black", linetype = "solid"),
    axis.ticks.x = element_line(size = 0.5),
    axis.ticks.y = element_blank(),
    legend.box.background = element_rect(),
    legend.key = element_rect(fill = "white", colour = "white"),
    legend.text = element_text(face="italic", size = 10)) +
coord_cartesian(xlim = NULL, ylim = c(0, 6.5)) +
scale_y_continuous(breaks = c(0,1,2,3,4,5,6,7), expand = c(0,0)) +
scale_x_discrete(labels = c("wt.0" = "wt",
                  "bi1.0" = "bi1",
                  "ire1.0" = "ire1",
                  "pig3.0" = "pig3",
                  "wt.30" = "wt",
                  "bi1.30" = "bi1",
                  "ire1.30" = "ire1",
                  "pig3.30" = "pig3",
                  "wt.60" = "wt",
                  "bi1.60" = "bi1",
                  "ire1.60" = "ire1",
                  "pig3.60" = "pig3",
                  "wt.90" = "wt",
                  "bi1.90" = "bi1",
                  "ire1.90" = "ire1",
                  "pig3.90" = "pig3",
                  "wt.120" = "wt",
                  "bi1.120" = "bi1",
                  "ire1.120" = "ire1",
                  "pig3.120" = "pig3",
                  "wt.150" = "wt",
                  "bi1.150" = "bi1",
                  "ire1.150" = "ire1",
                  "pig3.150" = "pig3",
```

```
                          "wt.180" = "wt",
                          "bi1.180" = "bi1",
                          "ire1.180" = "ire1",
                          "pig3.180" = "pig3"),
                 expand = c(0.025,0)) +
   labs(x="Strain", y = "Area (Relative)") +
   facet_grid(. ~ time, labeller = label_both, scales = "free") +
   theme(strip.text.x = element_text(size = 11, colour = "black", angle = 0)) +
   geom_label(aes(label="Removal from \nHeat Stress", y=6), colour="black",
angle=0, size=3.5) +
   ggsave("~/Desktop/boxplot_test.tiff", width = 12, height = 6, dpi = 1500)


###statistical analysis
#first use the raw_data dataframe to create a new dataframe to use for stats
stats_df <- raw_data[,c(1,2,3,5)]


#relevel the dataframe to use the wt strain as the reference
stats_df <- mutate(stats_df, strain = relevel(strain, ref = "wt"))


#turn the time points into factors, since I used numerical values
stats_df$time <- factor(stats_df$time)


#carry out the anova
model <- aov(total_area ~ (strain*time) + Error(replicate/(strain*time)), data
= stats_df)
summary(model)


##post-hoc contrasts
stats_df$strain.time <- interaction(stats_df$strain,stats_df$time)
library(lme4)
library(lmerTest)


m <- lmer(total_area ~ strain.time + (1 | replicate:strain) + (1 | replicate:time),
stats_df)
summary(glht(m, linfct=
          c("strain.timewt.30 - strain.timeire1.30 = 0",
```

```
        "strain.timewt.30 - strain.timepig3.30 = 0",
        "strain.timewt.30 - strain.timebi1.30 = 0",
        "strain.timewt.60 - strain.timeire1.60 = 0",
        "strain.timewt.60 - strain.timepig3.60 = 0",
        "strain.timewt.60 - strain.timebi1.60 = 0",
        "strain.timewt.90 - strain.timeire1.90 = 0",
        "strain.timewt.90 - strain.timepig3.90 = 0",
        "strain.timewt.90 - strain.timebi1.90 = 0",
        "strain.timewt.120 - strain.timeire1.120 = 0",
        "strain.timewt.120 - strain.timepig3.120 = 0",
        "strain.timewt.120 - strain.timebi1.120 = 0",
        "strain.timewt.180 - strain.timeire1.180 = 0",
        "strain.timewt.180 - strain.timepig3.180 = 0",
        "strain.timewt.180 - strain.timebi1.180 = 0")),
    test = adjusted(type = "holm"))
```

**APPENDIX F**
**Results of Statistical Analysis**

**Annexin V**
**Results of ANOVA**
Error: replicate

|  | Df | Sum_Sq | Mean_Sq | F_Value | Pr(>F) |
|---|---|---|---|---|---|
| Residuals | 1 | 136 | 136 | | |

Error: replicate:strain

|  | Df | Sum_Sq | Mean_Sq |
|---|---|---|---|
| strain | 3 | 6812 | 2270 |

Error: replicate:time

|  | Df | Sum_Sq | Mean_Sq |
|---|---|---|---|
| time | 4 | 49544 | 12386 |

Error: replicate:strain:time

|  | Df | Sum_Sq | Mean_Sq |
|---|---|---|---|
| strain:time | 12 | 5264 | 438.7 |

Error: Within

|  | Df | Sum_Sq | Mean_Sq | F-value | Pr(>F) | Significance |
|---|---|---|---|---|---|---|
| strain | 3 | 1287 | 428.9 | 10.93 | 0.000183 | *** |
| time | 4 | 9435 | 2358.7 | 60.08 | 7.33E-11 | *** |
| strain:time | 12 | 1630 | 135.8 | 3.46 | 0.006976 | ** |
| Residuals | 20 | 785 | 39.3 | | | |

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
1

**Results of Post-Hoc Contrasts**
Simultaneous Tests for General Linear Hypotheses

Linear Hypotheses:

| | Estimate | Std_Error | z-value | Pr(>\|z\|) | Significance |
|---|---|---|---|---|---|
| strain.timewt.60 - strain.timeire1.60 == 0 | -0.367 | 5.21E+00 | -7.00E-02 | 1 | |
| strain.timewt.60 - strain.timepig3.60 == 0 | -1.119 | 5.209 | -0.215 | 1 | |
| strain.timewt.60 - strain.timebi1.60 == 0 | 2.909 | 5.209 | 0.558 | 1 | |
| strain.timewt.120 - strain.timeire1.120 == 0 | -4.236 | 5.209 | -0.813 | 1 | |
| strain.timewt.120 - strain.timepig3.120 == 0 | 8.095 | 5.21E+00 | 1.55E+00 | 0.601 | |
| strain.timewt.120 - strain.timebi1.120 == 0 | 41.188 | 5.209 | 7.908 | 2.40E-14 | *** |
| strain.timewt.180 - strain.timeire1.180 == 0 | 26.017 | 5.209 | 4.995 | 3.53E-06 | *** |
| strain.timewt.180 - strain.timepig3.180 == 0 | 43.087 | 5.209 | 8.272 | 2.22E-15 | *** |
| strain.timewt.180 - strain.timebi1.180 == 0 | 55.24 | 5.21E+00 | 1.06E+01 | < 2e-16 | *** |
| strain.timewt.240 - strain.timeire1.240 == 0 | 30.381 | 5.209 | 5.833 | 3.81E-08 | *** |
| strain.timewt.240 - strain.timepig3.240 == 0 | 34.038 | 5.209 | 6.535 | 5.09E-10 | *** |
| strain.timewt.240 - strain.timebi1.240 == 0 | 59.365 | 5.209 | 11.397 | < 2e-16 | *** |

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- holm method)

**FDA**
**Results of ANOVA**

Error: replicate

| | Df | Sum_Sq | Mean_Sq | F-Value | Pr(>F) |
|---|---|---|---|---|---|
| Residuals | 1 | 34 | 34 | | |

Error: replicate:strain

| | Df | Sum_Sq | Mean_Sq |
|---|---|---|---|
| strain | 3 | 63.03 | 21.01 |

Error: replicate:time

| | Df | Sum_Sq | Mean_Sq |
|---|---|---|---|
| time | 6 | 106454 | 17742 |

Error: replicate:strain:time

| | Df | Sum_Sq | Mean_Sq |
|---|---|---|---|
| strain:time | 18 | 1241 | 68.94 |

Error: Within

| | Df | Sum_Sq | Mean_Sq | F-Value | Pr(>F) | Significance |
|---|---|---|---|---|---|---|
| strain | 3 | 24 | 8 | 1.128 | 0.3546 | |
| time | 6 | 19445 | 3241 | 453.736 | <2e-16 | *** |
| strain:time | 18 | 249 | 14 | 1.94 | 5.61E-02 | . |
| Residuals | 28 | 200 | 7 | | | |

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

**Results of Post-Hoc Contrasts**
Simultaneous Tests for General Linear Hypotheses

Linear Hypotheses:

| | Estimate | Std_Error | z-value | Pr(>\|z\|) | Significance |
|---|---|---|---|---|---|
| strain.timewt.30 - strain.timeire1.30 == 0 | -3.45347 | 2.01904 | -1.71 | 1 | |
| strain.timewt.30 - strain.timepig3.30 == 0 | -4.28793 | 2.02E+00 | -2.12E+00 | 0.505 | |
| strain.timewt.30 - strain.timebi1.30 == 0 | 12.42874 | 2.01904 | 6.156 | 1.27E-08 | *** |
| strain.timewt.60 - strain.timeire1.60 == 0 | 10.88579 | 2.01904 | 5.392 | 1.12E-06 | *** |
| strain.timewt.60 - strain.timepig3.60 == 0 | -3.37246 | 2.01904 | -1.67 | 1 | |
| strain.timewt.60 - strain.timebi1.60 == 0 | 0.87 | 2.02E+00 | 4.31E-01 | 1 | |
| strain.timewt.90 - strain.timeire1.90 == 0 | 1.96083 | 2.02E+00 | 9.71E-01 | 1 | |
| strain.timewt.90 - strain.timepig3.90 == 0 | -3.79664 | 2.02E+00 | -1.88E+00 | 0.841 | |
| strain.timewt.90 - strain.timebi1.90 == 0 | -14.80224 | 2.02E+00 | -7.33E+00 | 4.10E-12 | *** |
| strain.timewt.120 - strain.timeire1.120 == 0 | 0.02267 | 2.01904 | 0.011 | 1 | |
| strain.timewt.120 - strain.timepig3.120 == 0 | -0.73855 | 2.01904 | -0.366 | 1 | |
| strain.timewt.120 - strain.timebi1.120 == 0 | -2.93006 | 2.01904 | -1.451 | 1 | |
| strain.timewt.150 - strain.timeire1.150 == 0 | -0.36866 | 2.01904 | -0.183 | 1 | |
| strain.timewt.150 - strain.timepig3.150 == 0 | -0.12112 | 2.01904 | -0.06 | 1 | |
| strain.timewt.150 - strain.timebi1.150 == 0 | -0.8987 | 2.01904 | -0.445 | 1 | |
| strain.timewt.180 - strain.timeire1.180 == 0 | -0.11958 | 2.01904 | -0.059 | 1 | |
| strain.timewt.180 - strain.timepig3.180 == 0 | -0.4451 | 2.01904 | -0.22 | 1 | |

strain.timewt.180 -
strain.timebi1.180 == 0          -0.29238          2.01904          -0.145          1
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- holm method)

**Growth Curve**
**Results of ANOVA**

Error: replicate

| | Df | Sum_Sq | Mean_Sq | F-Value | Pr(>F) |
|---|---|---|---|---|---|
| Residuals | 1 | 2.58E+13 | 2.58E+13 | | |

Error: replicate:strain

| | Df | Sum_Sq | Mean_Sq |
|---|---|---|---|
| strain | 3 | 3.51E+12 | 1.17E+12 |

Error: replicate:time

| | Df | Sum_Sq | Mean_Sq |
|---|---|---|---|
| time | 6 | 1.13E+15 | 1.88E+14 |

Error: replicate:strain:time

| | Df | Sum_Sq | Mean_Sq |
|---|---|---|---|
| strain:time | 18 | 7.51E+12 | 4.17E+11 |

Error: Within

| | Df | Sum_Sq | Mean_Sq | F-Value | Pr(>F) | Significance |
|---|---|---|---|---|---|---|
| strain | 3 | 1.44E+11 | 4.81E+10 | 0.078 | 0.971 | |
| time | 6 | 2.34E+14 | 3.89E+13 | 62.95 | 5.84E-15 | *** |
| strain:time | 18 | 5.81E+12 | 3.23E+11 | 0.523 | 9.23E-01 | |
| Residuals | 28 | 1.73E+13 | 6.18E+11 | | | |

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

**Results of Post-Hoc Contrasts**
Simultaneous Tests for General Linear Hypotheses

Linear Hypotheses:

| | Estimate | Std_Error | z-value | Pr(>|z|) | Significance |
|---|---|---|---|---|---|
| strain.timewt.1 - strain.timeire1.1 == 0 | 120000 | 617626 | 0.194 | 1 | |
| strain.timewt.1 - strain.timepig3.1 == 0 | -33333 | 617626 | -0.054 | 1 | |
| strain.timewt.1 - strain.timebi1.1 == 0 | 146667 | 617626 | 0.237 | 1 | |
| strain.timewt.2 - strain.timeire1.2 == 0 | -1246667 | 617626 | -2.018 | 0.61 | |
| strain.timewt.2 - strain.timepig3.2 == 0 | -566667 | 617626 | -0.917 | 1 | |
| strain.timewt.2 - strain.timebi1.2 == 0 | -1080000 | 617626 | -1.749 | 1 | |
| strain.timewt.3 - strain.timeire1.3 == 0 | -800000 | 617626 | -1.295 | 1 | |
| strain.timewt.3 - strain.timepig3.3 == 0 | -826667 | 617626 | -1.338 | 1 | |
| strain.timewt.3 - strain.timebi1.3 == 0 | -1446667 | 617626 | -2.342 | 0.287 | |
| strain.timewt.4 - strain.timeire1.4 == 0 | -240000 | 617626 | -0.389 | 1 | |
| strain.timewt.4 - strain.timepig3.4 == 0 | -413333 | 617626 | -0.669 | 1 | |
| strain.timewt.4 - strain.timebi1.4 == 0 | -640000 | 617626 | -1.036 | 1 | |
| strain.timewt.5 - strain.timeire1.5 == 0 | 66667 | 617626 | 0.108 | 1 | |
| strain.timewt.5 - strain.timepig3.5 == 0 | -360000 | 617626 | -0.583 | 1 | |
| strain.timewt.5 - strain.timebi1.5 == 0 | -320000 | 617626 | -0.518 | 1 | |

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- holm method)

**Plating Assay**
**Results of ANOVA**

Error: replicate

|        | Df | Sum_Sq   | Mean_Sq  |
|--------|----|----------|----------|
| strain | 1  | 3.71E-02 | 3.71E-02 |

Error: replicate:strain

|        | Df | Sum_Sq   | Mean_Sq  |
|--------|----|----------|----------|
| strain | 3  | 2.15E+00 | 7.17E-01 |

Error: replicate:time

|        | Df | Sum_Sq   | Mean_Sq  |
|--------|----|----------|----------|
| strain | 1  | 1.69E+01 | 1.69E+01 |
| time   | 5  | 222.19   | 44.44    |

Error: replicate:strain:time

|             | Df | Sum_Sq | Mean_Sq |
|-------------|----|--------|---------|
| strain      | 1  | 0.407  | 0.4074  |
| strain:time | 17 | 10.485 | 0.6168  |

Error: Within

|             | Df | Sum_Sq   | Mean_Sq  | F-Value | Pr(>F)   | Significance |
|-------------|----|----------|----------|---------|----------|--------------|
| strain      | 3  | 3.00E-01 | 1.02E-01 | 1.569   | 2.19E-01 |              |
| time        | 6  | 5.28E+01 | 8.81E+00 | 135.935 | <2e-16   | ***          |
| strain:time | 18 | 5.51     | 0.306    | 4.727   | 0.000129 | ***          |
| Residuals   | 28 | 1.81     | 0.065    |         |          |              |

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

**Results of Post-Hoc Contrasts**
Simultaneous Tests for General Linear Hypotheses

Linear Hypotheses:

| | Estimate | Std_Error | z-value | Pr(>\|z\|) | Significance |
|---|---|---|---|---|---|
| strain.timewt.30 - strain.timeire1.30 == 0 | -0.60833 | 0.27979 | -2.174 | 0.386 | |
| strain.timewt.30 - strain.timepig3.30 == 0 | -2.56867 | 0.27979 | -9.181 | < 2e-16 | *** |
| strain.timewt.30 - strain.timebi1.30 == 0 | -1.52033 | 0.27979 | -5.434 | 7.72E-07 | *** |
| strain.timewt.60 - strain.timeire1.60 == 0 | 0.19967 | 0.27979 | 0.714 | 1 | |
| strain.timewt.60 - strain.timepig3.60 == 0 | -0.404 | 0.27979 | -1.444 | 1 | |
| strain.timewt.60 - strain.timebi1.60 == 0 | -0.126 | 0.27979 | -0.45 | 1 | |
| strain.timewt.90 - strain.timeire1.90 == 0 | -0.23933 | 0.27979 | -0.855 | 1 | |
| strain.timewt.90 - strain.timepig3.90 == 0 | -0.18867 | 0.27979 | -0.674 | 1 | |
| strain.timewt.90 - strain.timebi1.90 == 0 | -0.18667 | 0.27979 | -0.667 | 1 | |
| strain.timewt.120 - strain.timeire1.120 == 0 | -0.07633 | 0.27979 | -0.273 | 1 | |
| strain.timewt.120 - strain.timepig3.120 == 0 | -0.29133 | 0.27979 | -1.041 | 1 | |
| strain.timewt.120 - strain.timebi1.120 == 0 | 0.01367 | 0.27979 | 0.049 | 1 | |
| strain.timewt.180 - strain.timeire1.180 == 0 | -0.137 | 0.27979 | -0.49 | 1 | |
| strain.timewt.180 - strain.timepig3.180 == 0 | -0.15967 | 0.27979 | -0.571 | 1 | |
| strain.timewt.180 - strain.timebi1.180 == 0 | -0.08067 | 0.27979 | -0.288 | 1 | |

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- holm method)

**ROS**
**Results of ANOVA**
Error: replicate

|  | Df | Sum_Sq | Mean_Sq | F-Value | Pr(>F) |
|---|---|---|---|---|---|
| Residuals | 1 | 5.75E+01 | 5.75E+01 | | |

Error: replicate:strain

|  | Df | Sum_Sq | Mean_Sq |
|---|---|---|---|
| strain | 3 | 2.24E+02 | 7.47E+01 |

Error: replicate:time

|  | Df | Sum_Sq | Mean_Sq |
|---|---|---|---|
| time | 4 | 2.78E+04 | 6.94E+03 |

Error: replicate:strain:time

|  | Df | Sum_Sq | Mean_Sq |
|---|---|---|---|
| strain:time | 12 | 1.35E+04 | 1.13E+03 |

Error: Within

|  | Df | Sum_Sq | Mean_Sq | F-Value | Pr(>F) | Significance |
|---|---|---|---|---|---|---|
| strain | 3 | 4.60E+01 | 1.53E+01 | 0.384 | 0.7657 | |
| time | 4 | 5.75E+03 | 1.44E+03 | 36.041 | 7.05E-09 | *** |
| strain:time | 12 | 2.24E+03 | 1.87E+02 | 4.686 | 1.20E-03 | ** |
| Residuals | 20 | 7.98E+02 | 3.99E+01 | | | |

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

**Results of Post-Hoc Contrasts**

Simultaneous Tests for General Linear Hypotheses

Linear Hypotheses:

| | Estimate | Std_Error | z-value | Pr(>\|z\|) | Significance |
|---|---|---|---|---|---|
| strain.timewt.30 - strain.timeire1.30 == 0 | 63.643 | 3.985 | 15.97 | <2e-16 | *** |
| strain.timewt.30 - strain.timepig3.30 == 0 | 63.942 | 3.985 | 16.045 | <2e-16 | *** |
| strain.timewt.30 - strain.timebi1.30 == 0 | 37.396 | 3.985 | 9.384 | <2e-16 | *** |
| strain.timewt.60 - strain.timeire1.60 == 0 | -5.891 | 3.985 | -1.478 | 0.4181 | |
| strain.timewt.60 - strain.timepig3.60 == 0 | -7.257 | 3.985 | -1.821 | 0.3431 | |
| strain.timewt.60 - strain.timebi1.60 == 0 | -4.755 | 3.985 | -1.193 | 0.4656 | |
| strain.timewt.90 - strain.timeire1.90 == 0 | -64.299 | 3.985 | -16.134 | <2e-16 | *** |
| strain.timewt.90 - strain.timepig3.90 == 0 | -51.06 | 3.985 | -12.812 | <2e-16 | *** |
| strain.timewt.90 - strain.timebi1.90 == 0 | -53.849 | 3.985 | -13.512 | <2e-16 | *** |
| strain.timewt.120 - strain.timeire1.120 == 0 | -3.588 | 3.985 | -0.9 | 0.4656 | |
| strain.timewt.120 - strain.timepig3.120 == 0 | -10.308 | 3.985 | -2.586 | 0.0582 | . |
| strain.timewt.120 - strain.timebi1.120 == 0 | -6.897 | 3.985 | -1.731 | 0.3431 | |

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Adjusted p values reported -- holm method)

**SYTOX**
**Results of ANOVA**

| | Df | Sum_Sq | Mean_Sq | F-value | Pr(>F) | Significance |
|---|---|---|---|---|---|---|
| Residuals | 1 | 4.701 | 4.701 | | | |

| | Df | Sum_Sq | Mean_Sq |
|---|---|---|---|
| strain | 3 | 7288 | 2429 |

| | Df | Sum_Sq | Mean_Sq |
|---|---|---|---|
| time | 6 | 34661 | 5777 |

| | Df | Sum_Sq | Mean_Sq |
|---|---|---|---|
| strain:time | 18 | 8953 | 497.4 |

| | Df | Sum_Sq | Mean_Sq | F-value | Pr(>F) | |
|---|---|---|---|---|---|---|
| strain | 3 | 1593 | 530.8 | 93.54 | 1.06E-14 | *** |
| time | 6 | 5787 | 964.6 | 169.97 | <2.0E-16 | *** |
| strain:time | 18 | 2547 | 141.5 | 24.94 | 1.18E-12 | *** |
| Residuals | 28 | 159 | 5.7 | | | |

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

**Results of Post-Hoc Contrasts**

Simultaneous Tests for General Linear Hypotheses

Linear Hypotheses:

| | Estimate | Std. Error | z value | Pr(>\|z\|) | Significance |
|---|---|---|---|---|---|
| strain.timewt.30 - strain.timeire1.30 == 0 | 0.7359 | 2.5368 | 0.29 | 1 | |
| strain.timewt.30 - strain.timepig3.30 == 0 | 3.5045 | 2.5368 | 1.381 | 0.66855 | |
| strain.timewt.30 - strain.timebi1.30 == 0 | 0.5101 | 2.5368 | 0.201 | 1 | |
| strain.timewt.60 - strain.timeire1.60 == 0 | 4.1492 | 2.5368 | 1.636 | 0.5096 | |
| strain.timewt.60 - strain.timepig3.60 == 0 | 11.9458 | 2.5368 | 4.709 | 2.24E-05 | *** |
| strain.timewt.60 - strain.timebi1.60 == 0 | 6.4741 | 2.5368 | 2.552 | 0.07496 | . |
| strain.timewt.90 - strain.timeire1.90 == 0 | -2.8254 | 2.5368 | -1.114 | 0.79611 | |
| strain.timewt.90 - strain.timepig3.90 == 0 | 15.52 | 2.5368 | 6.118 | 9.48E-09 | *** |
| strain.timewt.90 - strain.timebi1.90 == 0 | 4.9373 | 2.5368 | 1.946 | 0.30973 | |
| strain.timewt.120 - strain.timeire1.120 == 0 | 9.8931 | 2.5368 | 3.9 | 0.00077 | *** |
| strain.timewt.120 - strain.timepig3.120 == 0 | 17.4135 | 2.5368 | 6.864 | 7.34E-11 | *** |
| strain.timewt.120 - strain.timebi1.120 == 0 | 44.8609 | 2.5368 | 17.684 | < 2e-16 | *** |
| strain.timewt.150 - strain.timeire1.150 == 0 | 18.9602 | 2.5368 | 7.474 | 1.01E-12 | *** |
| strain.timewt.150 - strain.timepig3.150 == 0 | 17.6055 | 2.5368 | 6.94 | 4.70E-11 | *** |
| strain.timewt.150 - strain.timebi1.150 == 0 | 63.8686 | 2.5368 | 25.177 | < 2e-16 | *** |
| strain.timewt.180 - strain.timeire1.180 == 0 | 25.5299 | 2.5368 | 10.064 | < 2e-16 | *** |
| strain.timewt.180 - strain.timepig3.180 == 0 | 28.7827 | 2.5368 | 11.346 | < 2e-16 | *** |
| strain.timewt.180 - strain.timebi1.180 == 0 | 76.627 | 2.5368 | 30.206 | < 2e-16 | *** |

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Adjusted p values reported -- holm method)

**VITA**

---

Curriculum Vitae – Fall 2018
Matthew R. Breuer

---

## Education

**M.S. Biology**                                                                December 2018
Sam Houston State University, Huntsville, Texas
    Primary Advisor: Anne R. Gaillard
    **Master's Thesis**: "Large Scale Prediction and Reverse Genetics Analysis of Programmed Cell Death Genes in *Chlamydomonas Reinhardtii*"
    Graduate Coursework: Experimental Physiology, Virology, Genomics and Bioinformatics, Model Organisms, Experimental Design in Biology, Cancer Biology, Advanced Molecular Genetics

**B.S. Biomedical Sciences, Minor in Philosophy**                                  May 2016
Sam Houston State University, Huntsville, Texas
    Primary Advisor: Anne R. Gaillard
    Undergraduate GPA: 3.3
    Elective Coursework: Biochemical Analysis of Proteins, Metabolism, Cell Biology, Molecular Biology

## Professional Experience

**Graduate Research Assistant**                                          Summer 2016-Current
Department of Biological Sciences, Sam Houston State University
    Primary Investigator: Anne R. Gaillard

**Head Teaching Assistant – General Microbiology**                        Fall 2016-Current
Department of Biological Sciences, Sam Houston State University

**Graduate Teaching Assistant**                                           Fall 2016-Current
Department of Biological Sciences, Sam Houston State University
    **Laboratory Courses Taught**: Molecular Biology, Cell Biology, General Microbiology.

**Undergraduate Research Assistant**                                  Spring 2014-Spring 2016
Department of Biological Sciences, Sam Houston State University
    Primary Investigator: Anne R. Gaillard

**Undergraduate Teaching Assistant**                                   Fall 2014-Spring 2016
Department of Biological Sciences, Sam Houston State University
    **Laboratory Courses Taught**: Cell Biology, Molecular Biology, General Microbiology), Introductory Applied Microbiology, and Introductory Cell Biology.

## Honors and Recognitions

**2ⁿᵈ Place Graduate Poster Presentation in Microbial Genetics**                     Fall 2016
    Fall meeting of the Texas Branch of the American Society for Microbiology

**Outstanding Undergraduate Research Award**                     Spring 2016
    SHSU Department of Biological Sciences

**2ⁿᵈ Place Undergraduate Poster Presentation in Molecular Microbiology**                     Fall 2015
    Fall meeting of the Texas Branch of the American Society for Microbiology

**1ˢᵗ Place Undergraduate Oral Presentation**                     Spring 2015
    Spring meeting of the Texas Branch of the American Society for Microbiology

## Professional Affiliations and Leadership Experience

**Beta Beta Beta Biological Honor Society, Delta Tau Chapter**
    President                     Spring 2015-Spring 2016
    Vice-President                     Fall 2014

**American Society for Microbiology, SHSU Chapter**
    Vice-President                     Fall 2014-Fall 2015

## Scholarly Presentations

**Breuer, M.R.** Gaillard, A.R. 2018. Large-Scale Prediction of Programmed Cell Death Genes and Reverse Genetic Analysis of Mutant Strains in *Chlamydomonas reinhardtii*. Poster presentation given at the 18ᵗʰ International Conference on the Cell and Molecular Biology of Chlamydomonas. Washington, D.C., Virginia.

**Breuer, M.R.** Gaillard, A.R. 2017. A Reverse Genetics Approach to Elucidating the Molecular Basis of Programmed Cell Death in *Chlamydomonas reinhardtii*. **Poster presentation** given at the fall meeting of the Texas Branch of the American Society for Microbiology. College Station, TX.

**Breuer, M.R.** Gaillard, A.R. 2017. Preliminary Analysis of Select Programmed Cell Death Genes in *Chlamydomonas reinhardtii*. **Poster presentation** given at the spring meeting of the Texas Branch of the American Society for Microbiology. New Braunfels, TX.

**Breuer, M.R.** Cho, H. Choudhary, M. and Gaillard, A.R. 2016. Potential Programmed Cell Death Genes in *Chlamydomonas reinhardtii*. **Poster presentation** given at the fall meeting of the Texas Branch of the American Society for Microbiology. Dallas, TX.

**Breuer, M.R.** Keathley, A.R. and Gaillard, A.R. 2016. Using two-dimensional gel electrophoresis to isolate a potential p53-like protein from *Chlamydomonas reinhardtii*. **Oral presentation** given at the Sam Houston State University Undergraduate Research Symposium. Huntsville, TX.

**Breuer, M.R.** Keathley, A.R. and Gaillard, A.R. 2016. Using two-dimensional gel electrophoresis to isolate a potential p53-like protein from *Chlamydomonas reinhardtii*. **Oral presentation** given at the Beta Beta Beta Annual Southwestern Conference. Cedar Hill, TX.

Gaillard, A.R. **Breuer, M.R.** and Keathley, A.R. 2015. Distinguishing between programmed cell death and necrosis in *Chlamydomonas*. **Poster presentation** given at the American Society for Cell Biology. Washington D.C.

**Breuer, M.R.** Keathley, A.R. and Gaillard, A.R. 2015. Population-level analysis of programmed cell death in *Chlamydomonas*. **Poster presentation** given at the fall meeting of the Texas Branch of the American Society for Microbiology. Huntsville, TX.

Keathley, A.R. **Breuer, M.R.** and Gaillard, A.R. 2015. Distinguishing between programmed cell death and necrosis in *Chlamydomonas*. **Poster presentation** given at the fall meeting of the Texas Branch of the American Society for Microbiology. Huntsville, TX.

Keathley, A.R. **Breuer, M.R.** and Gaillard, A.R. 2015. Is the amount of programmed cell death affected by genetic diversity in populations of *Chlamydomonas*? **Poster presentation** given at the fall meeting of the Texas Branch of the American Society for Microbiology. Huntsville, TX.

**Breuer, M.R.** Keathley, A.R. and Gaillard, A.R. 2015. Programmed cell death is distinct from necrosis in *Chlamydomonas*. **Oral presentation** given at the Sam Houston State University Undergraduate Research Symposium. Huntsville, TX.

**Breuer, M.R.** Keathley, A.R. and Gaillard, A.R. 2015. Programmed cell death is distinct from necrosis in *Chlamydomonas*. **Oral presentation** given at the spring meeting of the Texas Branch of the American Society for Microbiology. New Braunfels, TX.

**Breuer, M.R.** Keathley, A.R. and Gaillard, A.R. 2015. Programmed cell death is distinct from necrosis in *Chlamydomonas*. **Oral presentation** given at the Beta Beta Beta Annual Southwestern Conference. Kingston, OK.

**Breuer, M.R.** Keathley, A.R. and Gaillard, A.R. 2014. Investigating the role of programmed cell death in unicellular organisms. **Poster presentation** given at the fall meeting of the Texas Branch of the American Society for Microbiology.

## Technical Skills

### Laboratory Experience
### Protein Techniques

Protein extraction, purification, and quantification. Two-dimensional gel electrophoresis (isoelectric focusing and SDS-PAGE), gel staining (silver stain, coomassie). Western Blotting, imaging and analysis of polyacrylamide gels and Western blots.

**Nucleic Acid Techniques:** DNA extraction, purification, and quantification. Polymerase chain reaction, agarose gel electrophoresis, molecular cloning (plasmid ligation, transformation, etc.), Southern blotting, imaging and analysis of agarose gels.

**Cell Culturing:** Microbial (bacteria, *Chlamydomonas reinhardtii*, *Saccharomyces cerevisiae*), nematode (*Caenorhabditis elegans*), and mammalian (mouse fibroblast) culturing.

**Microscopy:** Brightfield, darkfield, phase-contrast, qualitative and quantitative fluorescence microscopy.

**Miscellaneous Techniques/Assays:** Conducting and analyzing data from growth curves, various colorimetric/fluorescent assays, scratch assay, analytical centrifugation (ultracentrifugation, discontinuous sucrose gradient centrifugation), hemacytometer cell scoring, DNA laddering assay.

**Computing Experience**

**Languages:** Used the Unix command line (Bash), R, and PERL to organize, format, and filter large sets of biological data.

**Software:** Automated batch analysis of microscopic image sets using custom macros in **ImageJ**, statistical analysis and visualization of data using the R language in **RStudio**, analysis of biological sequence data using **Geneious**, remote retrieval of NCBI datasets using **E-Utilities**, creation of custom databases and local BLASTp analysis using **BLAST+**, prediction of protein function by assignment of Gene Ontology terms to amino acid sequences in **BLAST2GO**, statistical analysis of biological data in **SAS**.