A MINUS-GAMETE-SPECIFIC GENE IN FUSION-DEFECTIVE *CHLAMYDOMONAS* MUTANTS AND ANALYSIS OF BIOSYNTHETIC PATHWAYS UPREGULATED DURING GAMETOGENESIS

by

DMITRY Y. BROGUN

A dissertation submitted to the Graduate Faculty in Biology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2013

This manuscript has been read and accepted for the
Graduate Faculty in Biology in satisfaction of the
dissertation requirements for the degree of Doctor of Philosophy.

Dr. Charlene L. Forest, Brooklyn College

12-25-2012
Date

Chair of the Examining Committee

Dr. Laurel Eckhardt

12-25-2012
Date

Executive Officer

Dr. James T. Nishiura, Brooklyn College

Dr. Theodor R. Muth, Brooklyn College

Dr. Wegang Qiu, Hunter College

Dr. Andrew Singson, Rutgers University

Supervisory Committee

The City University of New York

ABSTRACT

A MINUS-GAMETE-SPECIFIC GENE IN FUSION-DEFECTIVE *CHLAMYDOMONAS* MUTANTS AND ANALYSIS OF BIOSYNTHETIC PATHWAYS UPREGULATED DURING GAMETOGENESIS
by

Dmitry Y. Brogun

Adviser: Charlene L. Forest Ph.D.

To gain insight into the mechanism of gamete fusion during fertilization, it is crucial to identify molecular, metabolic and genetic factors required for this process. Gamete fusion in *C. reinhardtii* cells proceeds via four genetically defined stages: 1) flagella recognition 2) signaling 3) mating structure adhesion and 4) fusion with subsequent zygote formation. During this study we used insertional and temperature sensitive conditional mutants that do not proceed to stage 4, but can agglutinate and adhere to each other via their mating structure. A homolog of the sex-restricted HAP2/GCS1 gene has been shown to prevent gamete fusion in *C. reinhardtii.* A SiteFinding-PCR search was conducted on the fusion-defective *Chlamydomonas* insertional mutants that could not be complemented with the wild type copy of the HAP2/GCS1 gene. We confirmed that mutant clone 5 had an insert in a *copia*-family retrotransposon on chromosome 13. We collaborated to discover that a gene, located 3′ to the retrotransposon is a minus-gamete specific gene (MGS). We hypothesized that MGS might be a second gene required for gamete fusion. Our main objective was to identify whether there is a defect in the DNA sequence in MGS in any of our fusion-defective mutants. We performed a chromosome walk on coding, promoter and 5′ upstream and 3′ downstream regulatory regions (UTR) of MGS via PCR. PCR

products then where sequenced and aligned. We used qRT-PCR to determine MGS expression levels in the control and fusion defective mutants. Analysis of the sequencing and expressional results showed no defect in the MGS gene.

For the systematics study we used comparative genomic and phylogenetic approaches enabling us to study metabolic pathways that are upregulated in gametes of *Chlamydomonas*. Congruent experimental results show that the nuclear-encoded and chloroplast localized MEP pathway leading to the biosynthesis of the isoprenoid precursor molecules is upregulated in *Chlamydomonas* gametes. It is expected, that the results from these studies will provide further insights into regulatory mechanisms occurring during gametogenesis, some of which might be necessary for gamete fusion in algae as well as in higher eukaryotic organisms.

ACKNOWLEDGEMENTS

First, I would like to acknowledge my family, especially my parents for always believing in my future goals and being there to provide love and comfort. I am grateful to my mentor and true friend Dr. Charlene L. Forest for her unconditional support, patience and guidance over my entire graduate career. From the beginning Dr. Forest has showed a great believe in my scientific thinking abilities and she has sharpened it to the point of allowing me to stay focused, both professionally and personally; therefore she has this sole touch of a truly eminent professor. I would like to thank Dr. James T. Nishiura for teaching me on how to think many steps ahead and be able to visualize my future academic career, as well as his continuous support on my behalf, and his valuable advice throughout my Ph.D. work. I would like to thank Dr. Theodore R. Muth for his support, valuable input and guidance as well. I am grateful to both of them for being in my committee. I was very honored and lucky to have Dr. Wegang Qiu as my committee adviser. As a student of molecular evolution, I am especially thankful to Dr. Qiu for his great advice on the phylogenetics part of my work. I thank Dr. Andrew Singson for joining the committee at the end of my dissertational work, and for his pioneering work that has advanced the gamete fusion field to the level that it is currently at, providing a solid foundation for future discoveries. I thank Kathryne Ray and Dr. Forests's former and current lab members with whom I have had a privilege to share my knowledge, for being friends and providing great support, and everyone else with whom I have had a privilege to interact both personally and professionally during my Ph.D. studies.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

CHAPTER 1
INTRODUCTION

1.1 Membrane Fusion Review

Membranes are physiological gatekeepers that control intra- and intercellular fusion. Membrane fusion mediates various vital biological processes ranging from fertilization and new tissue regeneration to syncytium formation (Chen and Olson, 2005). Membrane fusion is believed to be an energetically unfavorable and multi-protein-mediated process (White, 1990). Our search for the fusion proteins (fusogens) in *Chlamydomonas*, begins with forward genetics, allowing us to identify genes encoding them. Then, with the aid of a reverse genetics techniques, enabling us to identify mutation/s within them. This chapter describes the current knowledge of the genes and their products responsible for membrane fusion in model systems, with a particular emphasis and focus on the genes required for gamete fusion during fertilization.

1.1.1 Molecular Mechanism of Membrane Fusion

The molecular mechanism of membrane fusion is a spatially dependent phenomenon; therefore in order for membranes to fuse they must come into close proximity. During intercellular membrane fusion, i.e. fusion of gametes, gametes must overcome the electrostatic repulsion of the phospholipid bilayers (Jaffe and Cross, 1986). Charged fusion molecule/s in these membranes might be analogous to the viral fusion proteins (White et al., 1996). Integral and/or peripheral membrane proteins might play a vital role during the above-mentioned membrane fusion by facilitating conformational changes of the fusion-initiation sites of the membrane. Hydrophobic domains within fusogens tend to destabilize and bend participating membranes

toward one another, as it was shown in bilayer liposomes and in viral fusion proteins (White and Helenius, 1980; White et al., 2008; Wickner and Schekman, 2008). This promotes the sequential merger of the phospholipid monolayers, initially forming a reversible semi- hemifused intermediate, that is followed by complete fusion of the bilayers, as it was shown in enveloped viruses (Basanez, 2002; Chernomordik and Kozlov, 2005). In animal cells molecular composition and conformational structure of fusogens might very well be suited for the above-mentioned physiological function, since these fusogens accumulate at sites along the membrane enriched in molecules that favor membrane kinking (Basanez, 2002; Bentz and Mittal, 2000; Chernomordik and Kozlov, 2008; Corcoran et al., 2006; Lentz et al., 2000; Salsman et al., 2008; White et al., 2008; Wickner and Schekman, 2008) such as regions rich in cholesterol (Churchward et al., 2005).

1.1.2 Intracellular Vesicle Fusion

The canonical intracellular membrane fusion in all eukaryotes, which occurs during vesicle trafficking among organelles, is mediated by the conserved family of soluble N-ethyl-maleimide-sensitive factor attachment protein receptors (SNAREs) (Jahn and Scheller, 2006; Wickner and Schekman, 2008). Vesicle anchored (v-SNARE) and target anchored (t-SNARE) proteins (Sollner et al., 1993) interact to form a bundle of extended coiled-coil domains of α-helices bringing membranes together and facilitating fusion (Jahn and Scheller, 2006). SNAREs share the above-mentioned coiled-coil domains with enveloped viral fusogens, and these domains are believed to contribute the free energy necessary for membrane deformation (Basanez, 2002; Bentz, 2000; White et al., 2008; Harrison, 2008).

1.1.3 Intercellular Virus-Cell Fusion

Some enveloped viruses utilize an integral, pH dependent, fusogen, such as influenza hemagglutinin (HA), that belongs to the Class I viral fusion proteins (White, 1990). The HA fusion mechanism is mediated by the activated hydrophobic domain insertion into the host membrane and this is followed by extension of coiled-coil domains of two separated α-helices which fold upon each other to form a hairpin-like α-helical bundle that bend the membranes and promotes fusion (Jahn et al., 2003; Skehel and Wiley, 2000). The pH independent Class I viral fusion proteins activation is mediated by binding to a surface receptor, such as an envelope protein (Env) as in human immunodeficiency virus type 1 (HIV-1), causing the previously hidden fusion peptide to be exposed and inserted into the target membrane (McClure et al., 1988; Sinangil et al., 1988; Stein et al., 1987). This is followed by the similar membrane fusion mechanism as in HA described above. In contrast to the class I viral fusogens that utilize amphipatic α helices to mediate target fusion, members of the class II viral fusogens might be facilitated by the β barrels (Kuhn et al., 2002).

Non-enveloped viruses utilize the Fusion-Associated Small Transmembrane (FAST, ~10–14 kDa) proteins to mediate fusion of the infected and uninfected cells (Shmulevitz and Duncan, 2000). The proposed molecular mechanism of the FAST protein mediated fusion involves insertion of the N-proximal hydrophobic domain into the neighboring cells followed by the destabilization of the donor's cell plasma membrane caused by an intracellular, C-proximal, positively charged basic domain (Salsman et al., 2008; Shmulevitz and Duncan, 2000).

1.1.4 Myoblast Fusion

During intercellular fusion, myoblasts (muscle founder cells), utilize four-pass transmembrane proteins, tetraspanins, to cooperate with integrins allowing these cells to fuse, via an unknown molecular mechanism (Horsley and Pavlath, 2004; Tachibana and Hemler, 1999). Tetraspanins and integrins both belong to large transmembrane protein superfamilies, and each superfamily participates in cell–cell fusion between different tissues (Horsley and Pavlath, 2004; Tachibana and Hemler, 1999), in addition to their alternative functions in many cellular processes, such as: cytoskeleton scaffolding and extracellular matrix adhesion (Berditchevski, 2001).

1.1.5 Fertilization in *Drosophila*

During fertilization, *Drosophila*'s sperm penetrates the egg membrane intact without plasma membrane fusion (Evans, 2012). When the sperm enters the cytoplasm of the egg its membrane disintegrates with the aid of the sperm tetraspanin-like protein named Sneaky. This protein shows a similar multi domain organization as the dendritic cell–specific transmembrane protein (DC-STAMP) domain and *spe-42* found in *C.elegans*, described below (Wilson et al., 2006).

1.1.6 Fertilization and Development in *C.elegans*

To this date seven *spe*-fusogen (*spe*-sperm defective) candidates were isolated during forward genetic studies, in one of the best-studied model organisms, the hermaphrodite-nematode *Caenorbabditis elegans* (Singson et al., 2008). Mutants contain physiologically normal sperm that fail to fuse with egg plasma membranes in the spermatheca (Singson et al., 2008). The first cloned *spe*-class fusogen, *spe-9*, encodes a single pass sperm integral protein containing ten EGF-like repeats and is essential during nematode fertilization (Singson et al., 1998). Other

cloned *spe*-class fusogens include *spe*-38 that encodes a four transmembrane domain protein not homologous to mammalian CD9 tetraspanin (Chatterjee et al., 2005). *spe*-41 is a calcium channel that is involved in sperm-egg interaction (Xu and Sternberg, 2003) and *spe*-42 encodes a protein with a dendritic cell–specific transmembrane protein (DC-STAMP) motif (Kroft et al., 2005; Wilson et al., 2006), as was found in *Drosophila* Sneaky. Further genetic screens allowed isolation of *egg*-1 and *egg*-2, integral proteins with extracellular domains containing eight low-density lipoprotein (LDL) receptor repeats (Kanandale et al., 2005), and both are candidates to mediate sperm-oocyte fusion. Infertile hermaphrodites that lack both *egg*-1 and *egg*-2 show classical synthetic infertility, since hermaphrodites lacking functional *egg*-1 or *egg*-2 are only subfertile (Evans, 2012). The tyrosine phosphatase–like proteins, *egg*-3, *egg*-4, and *egg*-5, appear to mediate fertilization-triggered responses that are part of the oocyte-to-embryo transition (Cheng et al., 2009; Maruyama et al., 2007; Parry et al., 2009; Stitzel et al., 2007).

Similar genetic screens in the *C.elegans* have identified additional fusogens that are important during its development. A gene named epithelial fusion failure 1 (*eff*-1) that blocks epithelial cell membrane fusion, which is essential during the *C.elegans's* organ development (Mohler et al., 2002), contains an extracellular hydrophobic domain that is involved in localization of it to the membrane but not in fusion-pore formation (Kontani et al., 2005). Endogenous expression of *eff*-1 in normally non-fusing *C.elegans* cells (Shemer et al., 2004) and in Sf9 insect cells (Podbilewicz et al., 2006) promotes cell fusion. In contrast, *aff*-1 (anchor fusion failure 1) mediates fusion of the anchor cells and its endogenous expression results in fusion of cells that normally do not fuse in *C. elegans* (Sapir et al., 2007).

1.1.7 Fusion of Sea Urchin gametes

In the most widely used model organism to study gamete fusion, sea urchin *Strongylocentrotus purpuratus*, only bindin (Glaser et al., 1999; Ulrich et al., 1998) has been shown to mediate the adhesion and fusion of the plasma membranes of the sperm and egg. Bindin is the egg binding protein, a 30.5-kDa hydrophobic protein contained within the acrosome of the sea urchin's sperm (Glabe and Vacquier, 1977; Trimmer et al., 1986; Vacquier and Moy, 1977; Wassarman, 1987a). Upon the exocytosis of the sperm's acrosomal granule, bindin co-localizes at the acrosomal process and it mediates binding of the sperm to the vitelline layer of the egg (Glabe and Lennarz, 1979; Glabe and Vacquier, 1977). Bindin can be considered a hemagglutinin-like molecule whose activity can also be inhibited by fucoidin or glycoconjugates released from the egg surface by pronase digestion. Moreover, it also recognizes and binds to specific sequences of sugar residues as a lectin (Vacquier and Moy, 1977; Gao et al., 1997; Gao et al., 1986; Glabe et al., 1982; Lopo et al., 1982). In vitro studies have show that bindin associates specifically with gel-phase phospholipid vesicles (Glabe, 1985a), and it can induce fusion of mixed-phase phospholipid vesicles (Glabe, 1985b). Therefore, these observations suggest that bindin might play a dual role in fertilization: first, it mediates the attachment of sperm to egg and second, it facilitates fusion of the sperm and egg plasma membranes (Trimmer and Vacquier, 1986).

1.1.8 Mammalian Fertilization

Mammalian fertilization is a multi-step process that involves receptor–ligand interactions, ion–channel modulations, proteolysis cellular signaling and recognition between interacting gametes (Wassarman, 2002; Wassarman et al., 2005). *In vivo* fertilization in *viviparous* organisms takes

place in the ampulla of the oviduct, and under homeostasis results in the formation of one or several zygotes. This process is initiated by maturation of both gametes, an egg (oogenesis) and the sperm (first spermatogenesis, which is followed by spermiogenesis) (Gilbert, 2003). Upon ejaculation from the male's reproductive organ, mature sperm, enters the vaginal vestibule. Once sperm enter this reproductive tract, capacitation occurs in which the adherent epididymal/seminal plasma glycoproteins are removed from the sperm (Austin, 1951, 1952; Austin and Bishop, 1958; Chang and Pincus, 1951). Thus "hyperactivated" (Yanagimachi, 1970) sperm continue their journey to the site of fertilization, the ampulla of the oviduct (Wassarman, 1987a). During capacitation, the intracellular concentration of divalent Ca^{2+} cation increases, leading to elevated level of cAMP and a decrease in intracellular pH, thus preparing sperm to undergo the acrosome reaction (Harada et al., 2011). Subsequent studies using pharmacological inhibitors and monoclonal antibodies in algae (Snell et al., 1982) and in sea urchins (Trimmer et al., 1986) showed the importance of calcium ions during egg activation and fertilization.

To form a zygote, capacitated sperm must first penetrate layers of cumulus cells surrounding each ovulated egg, and then bind to and penetrate the egg extracellular coat (zona pellucida; ZP) (Wassarman, 1987b; Wassarman et al., 2004). The zona pellucida is the outer layer of egg cell, which consists of three glycoproteins ZP1, ZP2 and ZP3 (Wassarman et al., 2005) whith ZP2 and ZP3 present as long filaments which are crosslinked by ZP1 (Jovine et al., 2002; Wassarman and Mortillo, 1991). Acrosomal enzymes are released in a process called the acrosome from a lysosome-like vesicle overlying the sperm nucleus, that is derived from the Golgi during spermatogenesis (Eddy and O'Brien, 1994; Wassarman et al., 2005). Following the acrosome reaction, the released acrosomal enzymes digest the glycoproteins in zona pellucida (Buffone et

al., 2008; Zimmerman et al., 2011). After the sperm penetrates the cumulus layer and travels through the zona pellucida, the sperm gains access to the perivitelline space between the egg plasma membrane and the ZP (Wassarman and Litscher, 2009). Then, the equatorial segment of the sperm plasma membrane (Florman and Ducibella, 2006; Huang and Yanagimachi, 1985; Yanagimachi, 1994, 1998) and microvillar domain, the region of the egg ooplasm located away from the metaphase II spindle, interact, leading to the sperm-egg membrane fusion by an unknown mechanism (Evans, 2012).

After the sperm-egg plasma membrane fusion event, which is distinguished by the establishment of cytoplasmic continuity between the two cells, the egg completes its meiosis and the sperm-oocyte nuclei migrate towards one another to unite (karyogamy) and form a zygote (Clark and Dell, 2006). The egg undergoes the zona reaction, also known as a slow block, in which cortical granules release protease enzymes that diffuse into the zona pellucida via exocytosis leading to alternations in glycoprotein structure in the zona pellucida, thus preventing a polyspermy from other sperm that are still attached to the fertilized egg (Wassarman, 1987a).

1.1.9 Sperm and Egg Fusogen Candidates

Currently, many proteins have been identified that play a role during the fertilization process, however not all of them are bona fide fusogen molecules, that mediate fusion between sperm and egg plasma membranes, except sea urchin's bindin (see section 1.1.7) and Sp18 from abalone (Kresge et al., 2001). The novel sperm protein equatorin, that is co-localized to the equatorial segment of the sperm plasma membrane, is possibly involved in sperm-egg membrane interaction (Yamatoya et al., 2009). The monoclonal antibody MN-9 to equatorin reduced levels of sperm-egg fusion (Toshimori et al., 1998). Sperm equatorial segment protein 1 (SPESP1) does

not have an obvious hydrophobic transmembrane domain but it can bind to membranes (Wolkowicz et al., 2003) and thus it might be a peripheral membrane protein. Knockout Spesp1$^{-/-}$ male mice produce sperm with the reduced ability to undergo sperm-egg fusion (Fujihara et al., 2010). An immunoglobulin superfamily member, IZUMO1 on the mouse and mammalian sperm, has been shown to be an essential protein for sperm-oocyte fusion (Inoue et al., 2007). IZUMO1 has an immunoglobulin-like domain (Ig) as well as the Izumo domain (IZ) consisting of approximately 150-amino-acids. Currently, the molecular mechanism of function of IZUMO1 is not yet clear, although it might function by interacting with a mediator protein such as angiotensin-converting enzyme 3 (ACE3), since male mice with the Ace3 knockout shows a slight abnormality in the localization of IZUMO1 (Inoue et al., 2010). The tetraspanin CD9 is among the major fusogens identified thus far on the mouse egg (Kaji et al., 2000; Le Naour et al., 2000; Miyado et al., 2000) and is likely to function in corroboration with another tetraspanin, CD81, as double knockout Cd9$^{-/-}$/Cd81$^{-/-}$ female mice are completely infertile (Rubinstein et al., 2006). The family of the proteins on a sperm called a disintegrin and a metalloproteases (ADAMs) have been implicated in sperm-egg interaction, by interacting with heterodimeric membrane integrin proteins via their integrin ligand-like disintegrin domain (D) (Bigler et al., 2000; Eto et al., 2002; Takahashi et al., 2001; Zhu et al., 2000). In mice an Adam2$^{-/-}$ (fertilin β) knockout shows defect in sperm-egg membrane interactions (Cho et al., 1998; Han et al., 2009; Nishimura et al., 2001).

Apparently, we still do not know the full molecular mechanism/s in sperm-egg fusion. Moreover, it is still unknown what protein on the human sperm plasma membrane interacts with what protein on the plasma membrane of an egg to induce gamete fusion, although work to date provides some insight into this process.

1.2 Life Cycle of *Chlamydomonas*

The unicellular, green alga *Chlamydomonas reinhardtii* is a sexual, photosynthetic eukaryote, which is mostly found in soil and fresh water (Harris et al., 1989). Cells of *C. reinhardtii* contain a single cup-shaped chloroplast that occupies one third to almost one half of the cell volume (Bourque et al., 1971; Schotz et al., 1972). The two anteriorly situated flagella projecting from the basal bodies are used for motility and as sensory structures during the sexual mating cycle. The protein composition as well as the length of the flagella varies depending on the stage of the life cycle of *Chlamydomonas*. Gametes appear to have the longest flagella that contain large amount of the integral membrane glycoproteins, agglutinins, which allow the initial recognition of opposite mating type gametes (Goodenough et al., 1978). The eyespot apparatus is composed of two highly ordered layers of tetraterpenoid-rich lipid globules inside the *Chlamydomonas* chloroplast and is used to respond to photoshock as well as to perform phototaxis (Schmidt et al., 2006).

The alga switches between asexual and sexual reproductive cycles based on nitrogen availability (Sager and Granick, 1954) and it can be maintained indefinitely in pure cultures on defined growth media (Hudock and Rosen, 1976). Cells of *C. reinhardtii* mainly exist as a vegetative haploids that divide by mitosis. Morphologically, vegetative cells of *C. reinhardtii* are approximately 10 μm long and the presence of the hydroxyproline-rich glycoprotein cell wall gives them their oval shape. When vegetative cells of *C. reinhardtii* are induced to differentiate into mating competent gametes under nitrogen starvation many biochemical and morphological changes occur. During this differentiation, *C. reinhardtii* gametes show disintegration of chloroplast membranes as well as increased levels of accumulating secondary metabolites, such as: starch (Martin and Goodenough, 1975), neutral lipids TAG's (Moellering and Benning, 2009; Wang et al., 2009) and isoprenoid molecules. Of specific interest for secondary metabolite formation are pathways

involved in biosynthesis of isoprenoid molecules, (discussed in section 1.3). When the nitrogen concentration falls below threshold in the defined growth media, this stimulus induces sexual differentiation in *Chlamydomonas* (Sager and Granick, 1954). Cells of *C.reinhardtii* show isogamous heterothallic behavior resulting in the two mating types: mating type plus (mt^+) and mating type minus (mt^-) (Hudock and Rosen, 1976). The mating type locus, is approximately 1 megabase of DNA in length, has been mapped to location 30 cM from the centromere of the chromosome 6. Because, it is under recombinational repression, the *mt*-linked alleles can segregate together (Goodenough et al., 1995). There are two mt^+ specific and two mt^- specific genes found on each mating type locus. The mt^+ specific genes are *MTA*1 (MT locus, region a) and *FUS*1, found in region *a* and *c* in the mt^+ locus respectively, and they are only expressed in *plus* gametes (Goodenough et al., 1995). *FUS*1 mediates mt^+ adhesion to the mt^- structure during gametic cell interaction (Ferris et al., 1996), however the function of *MTA*1, which encodes a small polypeptide, currently is unknown (Ferris et al., 2002). *MTD*1 (MT locus, region d) and *MID* (minus dominance) genes are found in region d and f respectively on the mt^- locus; both are mt^- specific genes, only expressed in *minus* gametes, and both genes mediate *minus* gametic differentiation (Ferris et al., 2002; Lin and Goodenough, 2007). Under ammonium starvation on petri dishes (Martin and Goodenough, 1975; Sager and Granick, 1954) and continuous illumination, vegetative cells undergo their last two mitotic divisions and become mating competent gametes (Kates et al., 1968). They exit the cell cycle and remain in the state of quiescence or G_0 until gametes of opposite mating type fuse to form a diploid zygote that will eventually undergo meiosis (Goodenough et al., 2007). When young mating-competent gametes of opposite mating types (i.e. mt^+ and mt^-) come into an intimate contact, their initial recognition is mediated by the flagellar-exposed agglutinins, thus marking the initiation of mating (Bergman et

al., 1975; Forest and Togasaki, 1975; Sager and Granick, 1954). Agglutinin delivery to the gamete flagellar membrane surface (Adair et al., 1983) is promoted by the kinesin/dynein-mediated intraflagellar transport system, from the cell body (Kinoshita et al., 1992; Snell et al., 2004; Wang et al., 2006). Agglutinins are members of the hydroxyproline-rich glycoprotein (HRGP) family (Cassab, 1998), consisting of homologous glycoproteins of high molecular weight, rich in hydroxyproline (~ 12%) and serine (~ 10%) residues (Cooper et al., 1983). Plus agglutinin is encoded by the autosomal gene *SAG*1 (sexual agglutination), mapped to the chromosome 7 (Goodenough et al., 1978); minus agglutinin is encoded by the *SAD*1 (sexual adhesion) gene found in the mt^- locus on chromosome 6 (Ferris et al., 2002; Ferris et al., 2005; Goodenough et al., 2007). In mt^-, *SAD*1 is under the expressional control of the *MID* (minus dominance) gene which is absent from mt^+; even though *SAD*1 is present in mt^+ as well, it is transcriptionally silenced (Ferris et al., 2002). The above-mentioned initial recognition is followed by agglutination, that is mediated by adhesive interactions that later lead to flagellar tipping (Bergman et al., 1975; Forest et al., 1978). Agglutination also promotes conformational change in the structure of the tips of the flagella, thus allowing flagella to elongate and accumulate a dense fibrous tip material (FTM) in a specific region between the nine singlet A microtubules and the terminal membrane (Mesland et al., 1980). The above-mentioned flagellar interactions induce activation of the protein-kinase dependent cAMP signaling pathway (Goodenough, 1989; Pasquale and Goodenough, 1987). Molecular sexual signal amplification and transmission is mediated by calcium fluxes across the flagellar membrane (Snell et al., 1982). Subsequent Ca^{2+} binding to the intracellular membrane localized gamete-specific adenylate cyclase mediates a rapid elevation of the intracellular cAMP that is function as the secondary messenger in the above-mentioned sexual signaling pathway (Goodenough, 1989; Pasquale and Goodenough, 1987). As agglutination progresses, more

agglutinins are delivered to the flagella surface from the cell body via intraflagellar trafficking (Wang et al., 2006). In response to signaling events, cell release from the cell wall is aided by the enzyme autolysin. Initially, both mating type gametes secret a serine protease that is necessary to convert a periplasmic prometalloprotease to an active matrix degrading metalloprotease (autolysin) that is stored in vesicles at periplasmic space (Buchanan et al., 1989; Snell et al., 1989). The autolysin digests glycoproteins of the cell wall, that are rich in proline or hydroxyproline polypeptides (Waffenschmidt and Jaenicke, 1987), thus producing an opening in the cell wall at the flagellar proximity of the anterior region of both mating type gametes (Claes, 1971; Matsuda et al., 1985). Following the escape of the protoplasts through this opening, a "ghost-like" empty cell walls left behind. In addition, wall degradation leads to the rapid increase of carbohydrates in the culture medium (Solter and Gibor, 1977).

The above mentioned protein-kinase dependent cAMP signaling pathway that leads to elevated levels of intracellular cAMP that triggers flagellar tipping, and conversion and release of an active form of the autolysin enzyme, also leads to the activation of mating structures which are located on the cell membrane adjacent to the basal-body complex (Weiss et al., 1977) of one of the flagella (Friedmann et al., 1968; Sager and Granick, 1954) in both mating type gametes (Bergman et al., 1975; Claes, 1971). The activation of mating structures in mt^{+} results in a long, membrane-enclosed "fertilization tubule" similar to microvillus, containing polymerized actin filaments between the membrane and double-layered structure the "doublet zone" (Detmers et al., 1983; Goodenough et al., 1982; Goodenough et al., 1978; Goodenough and Weiss, 1975; Wilson et al., 1997), forming an acrosome-like (or microvillus like) extension (Goodenough and Weiss, 1975) with an apical glycoprotein layer called the "fringe" (Goodenough et al., 1982; Detmers et al., 1983). Activating the mating structure in mt^{-} gametes results in a slight elevation of the plasma

membrane producing "a dome like structure" (Goodenough and Weiss, 1975), mobilizing membrane particles to the center of the mating structure and produce a short-lived tubule with no microfilaments (Forest, 1983; Goodenough et al., 1982; Weiss et al., 1977). When the activated mating structures of opposite mating types come into intimate contact, the tip of the mt^+ mating structure recognizes and adheres to the activated mt^- mating structure (stage 3 in Forest, 1983). The apical region of the mt^+ gamete mating structure, the "fringe" (consisting of Fus1 protein), mediates adhesion to the membrane of the mt^- mating structure (Ferris et al., 1996; Misamore et al., 2003). One of the fusogens on the mt^- structure, is an integral protein coded for by the HAP2/GCS1 homolog that has been identified in a forward genetics screen (Liu et al., 2008b). Initial findings of sex-restricted HAP2/GCS1 have come from the genetic screens in *Arabidopsis thaliana* HAP2-HAPLESS 2 (Johnson et al., 2004) and *Lilium longiflorum* HAP2/GCS1-Generative Cell Specific 1 (Mori et al., 2006) *Arabidopsis hap*2/*gsc*1 mutants show reduced sperm targeting to the ovule due to the abnormal pollen tube growth, and in addition, the *hap*2 sperm that do reach the ovules fail to initiate fertilization (Johnson et al., 2004; Mori, 2006; von Besser et al., 2006). HAP2/GCS1 homologues have been found in genomes of all major eukaryotic taxa except fungi (Wong and Johnson, 2010). In *Chlamydomonas* its expression is in low levels in vegetative cells of both mating types, but increased in mt^- gametes (Mori et al., 2006).

Once the "fringe" mediated adhesion and fusion of the mt^+ "fertilization tubule" and the mt^- mating structure occur, the gametes are initially conjoined via a "cytoplasmic bridge" with the diameter of the microvillus (Goodenough et al., 2007), which then rapidly shortens, aligning the apical ends of the cells into opposition (Friedmann et al., 1968). The above-mentioned adhesion and subsequent gamete fusion lead to the production of the diploid quadriflagellated cell (QFC), or zygote (Forest, 1983, 1987; Harris et al., 1989; Sager and Granick, 1954) after a few hours, the zygote resorbs its

flagella and forms a thick zygospore wall. After maturation in the dark for about seven days, the mature zygote subsequently is re-introduced to light; the diploid zygote undergoes meiosis and releases four or eight haploid progeny, (see figure 1).

In vivo study of the molecular mechanisms of gamete membrane fusion, ie fertilization in higher eukaryotes, is an imperative yet a very difficult task. Therefore, the simplicity found during the mating processes in *Chlamydomonas* might serve as a useful model system allowing us to gain insight into the molecular mechanisms governing the gamete membrane fusion, which is found in more complex organisms. Moreover, it would allow us to find novel molecules ie. fusogens that mediate the above-mentioned gamete membrane fusion.

1.2.1 *Chlamydomonas* Forward Genetics

Our primary body of knowledge, regarding fusogens participating in gametic membrane fusion, comes from forward genetics. Therefore, mutants were produced using ultraviolet (UV) light. A streptomycin resistant strain, *sr-u*-2-23 mt^- cc 275 (*sru-2*) was used to allow for selection of mutants unable to fuse (Forest and Togasaki, 1975). The goal was to identify genes that are involved in gamete fusion in mt^- *Chlamydomonas* cells. Fusion defective mutants: *gam*1, *gam*10, *gam*11 were produced (Forest et al., 1978; Forest and Togasaki, 1975). Next, cells of the streptomycin resistant strain, *sru*-2-23 mt^- were randomly mutagenized by DNA insertion of pSP124S plasmid, producing 7 fusion defective mutants during 3 independent transformations. All of the mutants were able to recognize mt^+ cells, transmit signals, remove cell walls and adhere to mt^+ partners, suggesting that these mutants are blocked at the last stage of mating, ie fusion (Aksoy, PhD Thesis, 2008; Forest et al., 1978; La, Master Thesis, 1987; Lam, Master Thesis, 1991).

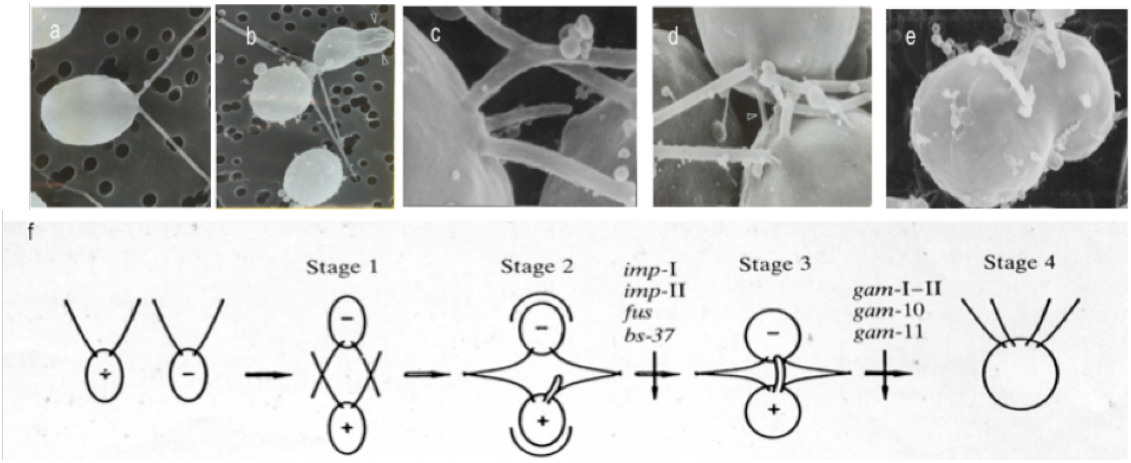


Figure 1. Mating stages in *Chlamydomonas* a-f. (Modified with permission after Forest 1978; 1983). Gametic plasma membrane interactions in *Chlamydomonas reinhardtii* can be separated into genetically definable stages (Forest, 1987). The first stage involves flagellar recognition, which is followed by signaling (Forest et al., 1978) activation of flagellar kinases (Wang et al., 2006; Wang and Snell, 2003) and cAMP production (Pasquale and Goodenough, 1987). These signals induce events shown in Stage 2, such as: flagellar tipping, a preference for interaction at the tips (Bergman et al., 1975), release of the enzyme autolysin (Claes, 1971) which digests the glycoprotein cell walls and activation of mating structures. Mating type + (mt^+) activation results in a long, membrane-enclosed mating tubule, containing polymerized actin filaments (Detmers et al., 1983). Mt^- cells mobilize membrane particles to the center of the mating structure (Weiss et al., 1977) and produce a short-lived tubule with no microfilaments (Goodenough et al., 1982). When the activated mating structures of opposite mating type cells come into contact, they adhere as shown in stage 3 (Forest, 1983) and finally, they fuse (stage 4). Stages 3 and 4, which involve interactions of the plasma membranes of the gametes, have been defined by mutations that block adhesion (stage 3), and fusion (stage 4) of the mating structures (Forest, 1987). Mutants defining adhesion stage 3 (*imp*-1, *imp*-11 and *fus*+) all are defective in or missing the *fus* gene, located in the mating type plus locus. Mutants defining gamete fusion stage 4, has been defined by *gam*-1-II, *gam*-10 and *gam*-11, all sex-limited mutants, expressed only in mt^-, and not linked to the mating type locus.

1.3 Isoprenoid Biosynthesis in *Chlamydomonas* Gametes

Isoprenoids are among the most ancient and abundant natural molecules found in all living organisms. Biosynthesis of isoprenoid molecules initiate from the two (C5) common precursors, isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). Both precursor molecules are found as the end products of two independent pathways. One pathway is the cytosolic mevalonate (MVA) pathway (Bach, 1995; Bloch, 1992; Bochar et al., 1999; Lichtenthaler, 1999; Lichtenthaler et al., 1997; Spurgeon and Porter, 1981), and the other is the methylerythritol phosphate (MEP) pathway, which occurs in the chloroplast (Rohmer et al., 1993; Schwender et al., 2001). Genes for enzymes of both pathways are encoded in the nucleus. Enzymes of the MVA pathway are ancestral to the archae origin (Lange et al., 2000) and MEP pathway enzymes in eukaryotes are ancestral to plastid bearing *Cyanobacteria* (Lange et al., 2000; Rohmer et al., 1993). Algae species might have either or both pathways. Under nitrogen starvation, vegetative cells of *Chlamydomonas* are induced to differentiate into mating competent gametes (see section 1.2). During this differentiation, many biochemical as well as morphological changes take place (Martin and Goodenough, 1975). Of particular interest is that *Chlamydomonas* mating competent gametes show elevated levels of secondary metabolites, such as - isoprenoid molecules.

1.3.1 MVA and MEP Biosynthetic Pathways

In all living organisms, isoprenoids are synthesized from isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP) (Rohmer, 1999). In higher plants both isomers are obtained via two independent pathways, either from a cytosolic mevalonate (MVA) pathway

and/or from a plastidial 2-*C*-methyl-D-erythritol 4-phosphate (MEP) pathway, which is also known as the non-mevalonate or 1-deoxy-D-xylulose-5-phosphate (DXOP) pathway (Schwender et al., 1996). Figure 2 depicts both pathways leading to the biosynthesis of the isoprenoid precursor molecules in higher plants. In contrast to the cytosolic MVA pathway, which requires three molecules of acetyl-coenzyme A (acetyl-CoA) for IPP generation, the MEP pathway uses pyruvate and glyceraldehyde-3-phosphate as substrates and involves eight plastid-localized nuclear-encoded enzymes (Schwender et al., 1996). Studies in higher plants show that seven enzymes of the MEP pathway are under light inducible control. They are:1-deoxy-D-xylulose-5-phosphate synthase (DXS), 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR), 2-C-methyl-D-erythritol-4-phosphate cytidyl transferase (MCT), 4-(cytidine 50-diphospho)-2-C-methyl-D-erythritol kinase (CMK), 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthase (MCS), 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate synthase (HDS), IPP-DMAPP isomerase (IPPI) methyl-2-(E)-butenyl-4-diphosphate synthase (HDS), IPP-DMAPP isomerase (IPPI) (Carretero-Paulet et al., 2002; Hsieh and Goodman, 2005; Ramos et al., 2008).

In higher plants the first enzyme in the MEP pathway is DXS which is believed to be the rate – controlling step for tetraprenoid biosynthesis (Matthews and Wurtzel, 2000). However, the MEP pathway could also be regulated at DXR and HDR enzymatic levels, with DXR considered to be a rate limiting step and HDR appearing to be the bottle-neck of the MEP pathway.

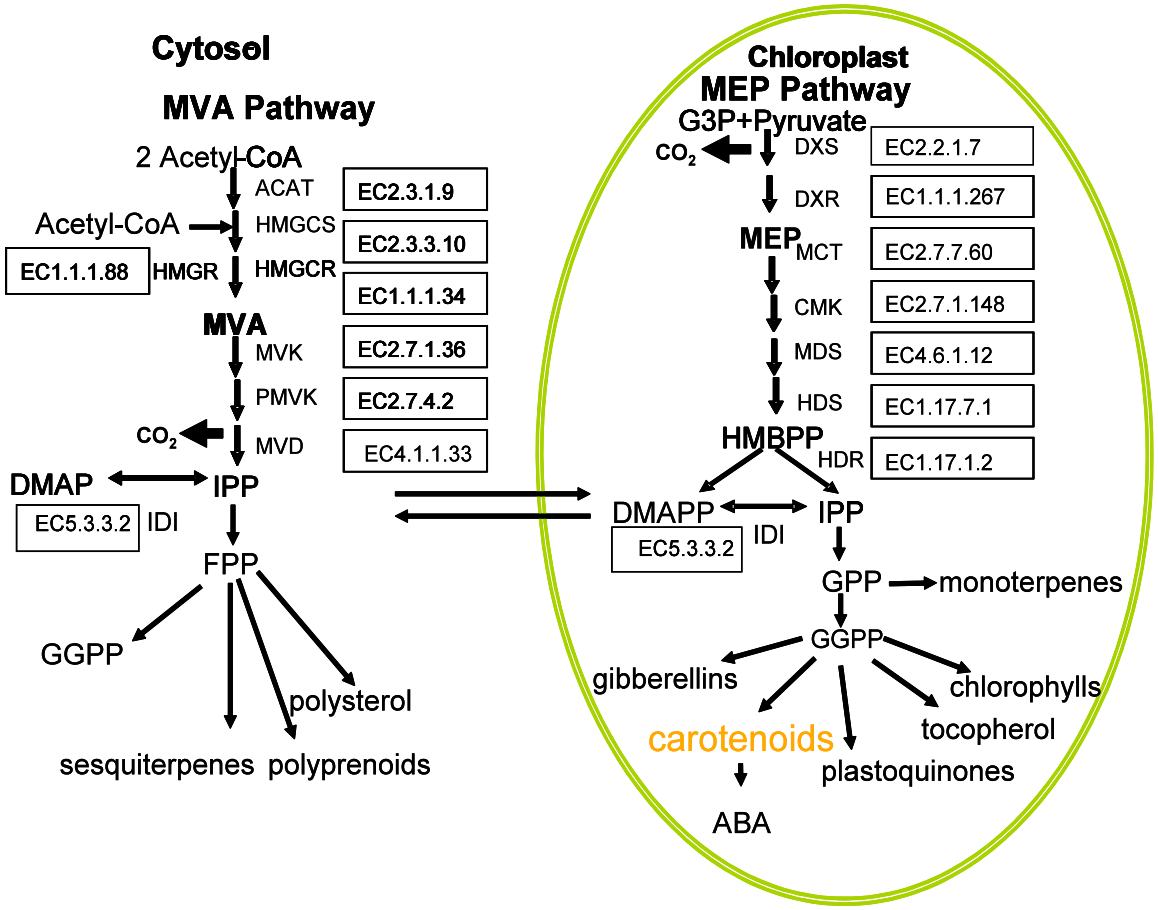


Figure 2. Spatial segregation of IPP biosynthesis in higher plants via MVA and MEP pathways. HMG-CoA, 3-Hydroxy-3-methylglutaryl CoA; MVA, mevalonic acid; MVAP, mevalonic acid 5-phosphate; MVAPP, mevalonic acid 5-diphosphate; IPP, isopentenyl diphosphate; DMAPP, dimethylallyl diphosphate; FPP, farnesyl diphosphate; Mt, mitochondrion; UQ, ubiquinone; GA-3-P, glyceraldehyde 3-phosphate; DOXP, 1-deoxy-D-xylulose-5-phosphate; MEP, 2-C-methyl-D-erythritol 4-phosphate; CDP-ME, 4-diphosphocytidyl- 2-C-methyl-D-erythritol; CDP-ME2P, 4-diphosphocytidyl- 2-C-methyl-D-erythritol 2-phosphate; ME-2,4cPP, 2-C-methyl-D-erythritol2,4-cyclodiphosphate; HMBPP, 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate; GGPP, geranylgeranyl diphosphate; GA, gibberellic acid; PQ, plastoquinone; ABA, abscisic acid. Enzymes of theMVA pathway: HMGS, HMG-CoA synthase; HMGR, HMG-CoA reductase; MVK, MVA kinase; PMK, MVAP kinase; MDD, MVAPP decarboxylase. Enzymes of the MEP pathway: DXS, MEP synthase; DXR,MEP reductoisomerase; MCT, CDP-ME synthase; CMK, CDP-ME kinase; MDS, ME-2,4cPP synthase; HDS, HMBPP synthase; HDR, HMBPP reductase. Names of enzymes and corresponding EC numbers written on the right site of the arrow indicating the alternating steps of the MEP and MVA pathways. (Lange and Ghassemian, 2003; Lange et al., 2000; Laule et al., 2003; Lichtenthaler, 1999; Rodriguez-Concepcion and Boronat, 2002).

1.4 Hypotheses and Specific Aims

Our long-term goal is to understand the mechanism of fusion used by gametes. Our hypothesis, supported by our genetic and our collaborators RNAseq data (Umen, personal communications), is that there are multiple genes controlling gamete fusion in *Chlamydomonas*. Our first goal is to determine whether *MGS* is impaired in any of our fusion defective mutants. *MGS* is located on chromosome 13 upstream from the *copia*-family LTR retrotransposon that is bearing the *ble* construct. Our second goal is to reconstruct the evolution of the MEP and MVA pathways, leading to the biosynthesis of the C5 isoprenoid precursor molecules, in algal species. It is expected, that the results gathered from this study will provide further insights into the mechanism of gamete fusion in algae.

CHAPTER 2

General Materials and Methods

2.1 General Reagents

2.1.1 Analytical Grade Reagents

We obtained our analytical grade general laboratory chemicals from either Sigma or Fisher Scientific (USA). Molecular biology grade agarose was purchased from invitrogen. We used agar for all of the algae cultures from Bacto Agar. Restriction enzymes and modifying enzymes were purchased from either Epicentre or New England Biolabs. Wide range DNA molecular weight marker Hi-Lo was obtained from Bionexsus.

2.1.2 General Solutions and Media

Agarose gel loading buffer (5×): 0.25% bromophenol blue, 50% TE, 50% glycerol

EDTA: ethylenediamine tetra-acetic acid, 0.5 M pH8.0

LB: 1% (w/v) bacto-tryptone, 0.5% (w/v) bacto-yeast extract, 170 mM sodium chloride

LB agar: LB solidified with 1.5% bacto-agar

TE Buffer: 10 mM Tris-HCl, pH8.0, 1 mM EDTA

TAE buffer (1×): 10 mM Tris-acetate, 0.5 mM EDTA pH 7.8

TAP growth medium Tris Acetate Phosphate (TAP) Growth Medium

Tris (hydroxymethyl aminomethane)	2.42 g
Beijerinck's solution	50 ml
Phosphate buffer	1.0 ml
Trace elements	1.0 ml

Glacial acetic acid	1.0 ml
dH_2O to 1000 ml	

SEM-N Nitrogen-Free Medium (TAP)

Nitrogen-free Beijerinck's solution	50 ml
Phosphate buffer for SEM-N	75 ml
Trace elements solution	1.0 ml
dH_2O to 1000ml	

Stock solutions used to make media:

Beijerinck's solution

NH_4Cl	8.00 g
$CaCl_2 \cdot 2H_2O$	1.00 g
$MgSO_4 \cdot 7H_2O$	2.00 g
dH_2O to 1000 ml	

Phosphate buffer for TAP (pH = 7.0)

K_2HPO4	174.18 g
KH_2PO4	136.09 g
dH_2O to 1000 ml	

Trace elements solution

KOH	20 g
EDTA (disodium salt)	50.0 g
$ZnSO_4 \cdot 7H_2O$	22.0 g
H_3BO_3	11.4 g

$MnCl_2 \cdot 4H_2O$	5.06 g
$FeSO_4 \cdot 7H_2O$	4.99 g
$CoCl_2 \cdot 6H_2O$	1.61 g
$CuSO_4 \cdot 5H_2O$	1.57 g
$(NH_4)_6Mo_7O_{24} \cdot 4H_2O$	1.10 g
dH_2O to 1000 ml	

Beijerinck's solution (for SEM-N)

$CaCl_2 \cdot 2H_2O$	1.0 g
$MgSO_4 \cdot 7H_2O$	2.0 g
dH_2O to 1000 ml	

Phosphate buffer (for SEM-N)

K_2HPO4	14.34 g
KH_2PO4	7.26 g
dH_2O to 1000 ml	

2.2 Algal Cell Culture

2.2.1 *Chlamydomonas* Cell Cultures and Gowth Conditions.

The wild type strains of *Chlamydomonas reinhardtii* selected as controls during this study, because of their high mating efficiency, were; CC-620 R3 NM mt^+ (mating type +) which we referred to as the R^+ and CC-621 NO mt^- (mating type -) referred as the NO^-. Strains CC-620 and CC-621 are subclones of the Smith 137c wild-type strain. The streptomycin resistant strain CC-275 (*sr-u-2-23,* mt^-) referred, as sru^- is the parental strain for all of our fusion defective mutants listed below:

a) UV mutants - *gam1, gam10 and gam11*.

b) Insertional mutants - 2-29, cl5, 2-8, 2-25 is a sub-clone of cl5.

Cultures of *C. reinhardtii* were stored on 90 mm Petri dishes containing TAP medium solidified with 1.5% agar. Cells were streaked on the above plates with a sterile wooden stick and cultured under a continuous fluorescent light source at 10 watts/m^2 at 25°C, unless otherwise specified. To induce gametogenesis log phase growth *C. reinhardtii* cell were grown on Petri plates on TAP growth media containing 1.5% agar for 5-7 day and then inoculated into the SEM-N media overnight to further promote induction of gametes. Cells were periodically checked under the microscope for mating competence.

2.3 Genomic DNA Isolation

Genomic DNA from *wt*, insertional and conditional mutant cells of *Chlamydomonas* was isolated with the Qiagen DNeasy Plant Mini Kit (cat# 69104) according to the manufactures protocol and/or by phenol-chloroform extraction. Prior to DNA purification, all the cell cultures were subjected to double clean up (swim-up) through a cotton plug utilizing the ability of *Chlamydomonas* to perform phototaxis for 2-3 hours in N-free liquid SEM-N media. Cells were collected from the top and spread onto Petri dishes with TAP growth media containing 1.5% agar. This procedure was repeated at least twice, ensuring that the cell cultures used for genomic DNA purification, as well as for other downstream applications listed below, were axenic.

2.4 Trizol DNA Purification

Prior to the Trizol total DNA purification procedure cells were flash-frozen in liquid nitrogen $2x10^7$ cells were then thawed in 1000 µl Trizol (Invitrogen) and incubated at 30°C for 5 minutes. To induce phase separation, 200 µl of chloroform was added and the resulting mixture was briefly mixed. The resulting homogenate was incubated at 30°C for 15 minutes and then centrifuged at 12,000 RPM for 15 minutes at 2-8°C. The DNA was precipitated with 300 µl 100% Ethanol (Sigma) from the organic phase and tubes with the DNA samples were incubated at 25°C for 2-3 minutes. Then tubes with DNA samples were centrifuged at 12,000 RPM for 5 minutes at room temperature. After removal of the phenol-ethanol supernatant, the DNA pellet was washed twice with 0.1 M sodium citrate in 10% ethanol. It was then stored in the washing solution for 30 minutes at room temperature with periodic mixing followed by centrifugation at 2,000 g for 5 minutes at 25ºC. The DNA pellet was resuspended in 75% ethanol and incubated for 15 min at room temperature with periodic mixing and then centrifuged at 2,000 g for 5 minutes at 25º C. After removal of the ethanol, the DNA pellet was briefly air-dried at room temperature. The DNA pellet was dissolved in 0.4 ml 0.008 N NaOH and then centrifuged at 12,000 g for 10 minutes. The resulted supernatant containing the dissolved DNA was stored at -30 º C freezer. DNA quality was assayed by 0.9% agarose gel-electrophoresis.

2.5 Quantification of DNA

DNA concentration was determined by the Ultrospec 1000 spectrophotometer (Pharmacia Biotech) at 260 nm (A260) using the formula below. DNA concentration (µg/µl) = (OD 260) x (dilution factor) x (50 µg DNA/ml)/1000 µl. The dilution factor was 500 since we used 2µl of the DNA and 998 µl of dH2O.

(Concentration of DNA in µg/µl) x (volume of DNA in µl) = DNA in µg. Expected yield: 5-7 ug DNA/10^6 cells.

2.6 Site Finding PCR

To map the physical location and to determine the genomic region flanking the plasmid pSP124S, which was used for DNA insertional mutagenesis of the host strain sru^- (cc 275) we performed Site Finding Polymerase Chain reaction (SF-PCR) (Tan et al., 2005). SF-PCR required us to utilize a sequential combination of the degenerate and plasmid pSP124S specific primers. See figures 3 and 4 (section 3.2.1 and 3.2.2 respectively), depicting a schematic representation of Site-Finding PCR modified after Tan et al., 2005.

2.6.1 Primary SF-PCR

500 to 800ng of undigested genomic DNA was used as the template for the primary SF-PCR reaction, and 1 µM of each degenerate SFP1 primer along with the plasmid specific Ble1 or Ble2 primer were used. The product of the primary PCR was diluted 1:10 and subjected to the secondary SF-PCR reaction.

2.6.2 Secondary SF-PCR

Diluted product from primary SF-PCR along with 1 µM degenerate SFP2 and RD223, RD225 or Ble2 insert specific primers were used during secondary SF-PCR. The SF-PCR reaction conditions and list of primers are listed below in the Table 1 and Table 2 respectively.

Table 1. SiteFinding PCR Thermal Cycling Conditions

Reaction	**Cycles**	**Thermal conditions**
SiteFinding PCR	1	92°C, 2 min 95°C, 1 min 25°C, 1 min 68°C, 10 min 68°C, 30 s
Primary	1 30 1	94°C, 1 min 95°C, 10 sec 60°C, 30 s 68°C, 6 min 72°C, 5 min
Secondary	1 30 1	94°C, 1 min 95°C, 10 sec 66°C, 30 s 72°C, 6 min 72°C, 5 min

Table 2. List of Primers used in SiteFinding PCR

Primer Name	**Sequence** 5' to 3'	**Tm**
SiteFinder Primers	(CACGACACGCTACTCAACACACCACCTCGC	
SF1	ACAGCGTCCTCAAGCGGCCGCNNNNNNGCC)	82.92
SFP1	CACGACACGCTACTCAACAC	62.45
SFP2	ACTCAACACACCACCTCGCACAGC	67.98
Gene Specific Primers		
Ble1	TGTTGTCCGGCACCACCTGGTC	68.26
Ble2	CTGATGAACAGGGTCACGTC	62.45
BleC	AGATGTTGAGTGACTTCTCTT	56.71
RD223	TTGGCTGCGCTCCTTCTGGCATTTAAATC	67.45
RD225	GATAAGCTTGATATCGAATTCC	57.08

2.6.3 Cycling Conditions Used for SiteFinding PCR

The setting of thermal cycling conditions was based on the study from the Tan et al., 2005 paper. In the Site Finding reaction, only minor changes were made in the thermal conditions, such as; the annealing temperature was changed according to the T_m of the gene specific primer used in each reaction.

2.6.4 Purification of PCR Product

SF-PCR reactions were performed on the DNA isolated from the cells of the *sru*$^-$ and three insertional fusion-defective mutants; 2-29, cl5 and 2-8. All bands from the gel in figure B were cut out, gel purified and sent out for sequencing to GENEWIZ.

2.7 General Methods in Nucleic Acid Amplification, Cloning and Sequencing

2.7.1 Oligonucleotide Synthesis

Oligonucleotide primers synthesized by Operon and Invitrogen were diluted to 100 μM with sterile nuclease free dH_2O.

Table 3. List of Primers used during chromosome walk

	Sequence			Name	Tm C°
5' -	AGGCGGAACTTTTGGGAGGGC	-	3'	fCAR4b	66.5
5' -	CGGGATAGGCTGCGAGTGCG	-	3'	rCAR4b	68.6
5' -	AGCAGGATGTGGCGGCAGGAAA	-	3'	fCAR3b	66.4
5' -	ACCCACACGCGCAACCCATC	-	3'	rCAR3b	66.6
5' -	GGAGATGGAGGAACATCGGGGGC	-	3'	fCAR2	69.9
5' -	TCCGCCTCCGCGACAAATGC	-	3'	rCAR2	66.6
5' -	TGTGCCCTACCTCCGACATGC	-	3'	fCAR1	66.5
5' -	CTGTGACGGCCCCCGATGTTC	-	3'	rCAR1	68.4
5' -	GAACAGCCCCTGAATGTACG	-	3'	fGTPase_1000	62.4
5' -	CGTGCTGTACGATCCAAGC	-	3'	rGTPase_1000	62.3
5' -	GACGAGCTGGTGCGGGCATT	-	3'	fCAR_2553	66.6
5' -	CAGCGGCACGGCATGACTGA	-	3'	rCAR_2553	66.6
5' -	ACGCCAACTCGGCAGCAAAC	-	3'	R3F47	68
5' -	CAAGTCCTGCCCTGCTTTGCC	-	3'	R3r48	68
5' -	ACACGCCGCCAAATCAACAAC	-	3'	R3r31	67
5' -	CGCGAAGTTGGAGCTTGCCTGT	-	3'	R1F16	60.24
5' -	GCCGGCGCAGTAGCCCAAAA	-	3'	R1r18	60.59
5' -	ACGAGGGGCAGGCACCAGAA	-	3'	fCAR18_6N	60.11
5' -	CGTAGGGCAGCGGGGACTCT	-	3'	rCAR18_1N	60.04
5' -	CGGCGGGTGCTCCTGTCAAA	-	3'	F34	71
5' -	AAGTGCCCGGCTGAACGCAA	-	3'	r31	70.38
5' -	TTCTGGTGCCTGCCCCTCGT	-	3'	r19	69.65
5' -	GGGCAGAGGCGGATGATGGC	-	3'	r11	69
5' -	CGTGGCTGTGACGACGGTAGC	-	3'	e15_F124	68.4
5' -	AGGCCCACAGGTCCGCCAAA	-	3'	e15_r130	66.6
5' -	CCACAGGTCCGCCAAACGCA	-	3'	e15_r24	71.86
5' -	AGGGTGGGCTTGCTCCGGTC	-	3'	e14_F16	70.1
5' -	GCGCTGACGGTTTTGGCAGG	-	3'	e13_F38	70.38
5' -	TGGCGGCAGCAATGTCGGAG	-	3'	e13_F6	66.6
5' -	GCCGCATCGAGACCTGCTGT	-	3'	e14_r58	66.6
5' -	GCGCGCATCCCACAAGCC	-	3'	e16_F46	66.7
5' -	CCAACGGGGACATGTTCTAC	-	3'	e12_F1	60.23
5' -	CATGCCCGTCACACTGTTC	-	3'	e12_r1	61.19
5' -	CGCCTGATGGTGATGGTG	-	3'	e11_F1	62
5' -	GAAGCCGCACTGCAAGGT	-	3'	e11_r1	62
5' -	GATGGTGGGCGGGTTGGGA	-	3'	e11_F18	70.74
5' -	ATCCAGCACGCAACCCGAAGC	-	3'	e11_r20	70.81
5' -	CGAGATCCAGCACGCAACCCG	-	3'	e11_r35	72.39
5' -	GCGCGTATGTCGGATTTTAC	-	3'	e10.5_F1	60.48
5' -	CGCCATTCTATGCATGTGTT	-	3'	e10_r2	59.57
5' -	CGGGGCGCAATCACCTACCAT	-	3'	e10_r6	70.05
5' -	CGCAATCACCTACCATAGCC	-	3'	e10_r3	60.49

Table 3. Continued.

	Sequence			Name	Tm C°
5' -	CGGTGGCGCGTATGTCGGATTTTA	-	3'	e10_F21	72
5' -	TGGCGCGTATGTCGGATTTTACG	-	3'	e10_F38	69.65
5' -	CAGGCGTATTGTCTCGTTGA	-	3'	e9_F1	59.86
5' -	CCACAGTGCGAGGTAGAACA	-	3'	e9_r1	59.9
5' -	TTCTGCTCACAGGCGTATTG	-	3'	e9_F2	60.01
5' -	TCTGCTCACAGGCGTATTGT	-	3'	e9_F24	59.47
5' -	CCACAGTGCGAGGTAGAACA	-	3'	e9_r24	59.9
5' -	GTGTTCTGCTCACAGGCGTA	-	3'	e9_F4	60.06
5' -	CGCAATCACCTACCATAGCC	-	3'	e9_r22	60.49
5' -	CGCCATTCTATGCATGTGTT	-	3'	e9_r17	59.17
5' -	GGTAGAACATGTCCCCGTTG	-	3'	e9r_13	60.23
5' -	TGGCCTTGTCCACAGTGATA	-	3'	e8_r1	60.1
5' -	TCCCACAGTGCGAGGTAGA	-	3'	e9_r100	60.41
5' -	TGTAGCTTCCACATGCCAAC	-	3'	e2_r1	59.72
5' -	GCACTCCATAAAACCGTGGA	-	3'	e2_r21	60.89
5' -	GTGCCATAGCAACCACCAT	-	3'	e2_r33	59.39
5' -	TTGGTACCTTCTCGCCTGAT	-	3'	e2_F33	59.69
5' -	CGCAGCCTTGATGTCTGTCT	-	3'	e1_F34	61.57
5' -	GCGCCTAAGCACTCCATAAA	-	3'	e2_r34	60.36
5' -	TCCACGGTTTTATGGAGTGC	-	3'	e2_F9N	60.98
5' -	GCCTTGACCTCCATCCAGT	-	3'	e9r_r7	60.06
5' -	AAGTCGAATCTACCGCAAGG	-	3'	e8_F1	59.34
5' -	AGTGATACCTGGGTGCAAGG	-	3'	e8_r2	59.99
5' -	TGACGCACTGCGGTAAGC	-	3'	e14_F1	62
5' -	TGTAGAGCCCGAGCCTTTTG	-	3'	e14_r1	62
5' -	TTGGTGGCAGCCATTAGG	-	3'	e14_F35	60.62
5' -	GTGGCAGCCATTAGGCAGAT	-	3'	e14F19A	62
5' -	GATGCCAGGCGTAGGAGAT	-	3'	e14r01A	59.9
5' -	TGCCGGTAAGCCGGGGGAAT	-	3'	e14_F7	71.58
5' -	GGGCTATTGGTGGACCTTTT	-	3'	e13F1C	60.19
5' -	GAGATGGGAGTGTCGGAAGT	-	3'	e13F60C	59.1
5' -	ACTGCGCTCACCACTAGCA	-	3'	e13r3G	60.78
5' -	CACTGGCCAACCCTCAAC	-	3'	e14r79E	60.09
5' -	CTGTCCACGCCATGTTTGT	-	3'	e13r60C	60.59
5' -	GCCTTCTTCATTCCTCGACA	-	3'	e13F01D	60.34
5' -	ACAATGCTACCGTCGTCACA	-	3'	e14r5	60.18
5' -	CGGTGAGCAAAAGGCATAG	-	3'	e14F47	59.43

Table 3. Continued.

	Sequence		Name	Tm C°
5' -	GGGCTATTGGTGGACCTTTT	- 3'	e13F1C	60.19
5' -	GAGATGGGAGTGTCGGAAGT	- 3'	e13F60C	59.1
5' -	ACTGCGCTCACCACTAGCA	- 3'	e13r3G	60.78
5' -	CACTGGCCAACCCTCAAC	- 3'	e14r79E	60.09
5' -	CTGTCCACGCCATGTTTGT	- 3'	e13r60C	60.59
5' -	GCCTTCTTCATTCCTCGACA	- 3'	e13F01D	60.34
5' -	ACAATGCTACCGTCGTCACA	- 3'	e14r5	60.18
5' -	CGGTGAGCAAAAGGCATAG	- 3'	e14F47	59.43
5' -	CTCAGCTCGCAGCACCACTA	- 3'	E14r5	62.84
5' -	TGATCGGCTAGGCAACAAG	- 3'	e14F3D	59.96
5' -	CCTGCTATGGCCTCCTGAT	- 3'	e14F15E	60.19
5' -	GCAGTTGTGACCCAGCTTCT	- 3'	e14r2	60.45
5' -	CGATGAGGAATGCAACACAA	- 3'	e14r1	60.66
5' -	GCCTTCTTCATTCCTCGACA	- 3'	e13F1H	60.34
5' -	CACACAGCCGTCATCAACTC	- 3'	e13r14H	60.32
5' -	GCTGGATGAAGCTGCAAAAC	- 3'	e14F1k	60.92
5' -	GACAGCAGTGGCAATCAGAA	- 3'	e14F45L	60
5' -	TGTAGAGCGCAAACAGACCA	- 3'	e14r21	61
5' -	TGCTTGGCAGGACAAGGTGGC	- 3'	e12_F26	70.1
5' -	AGCCCTCCTCCACGTCCTGC	- 3'	e12_r68	60.25
5' -	GCGCCTAAGCACTCCATAAA	- 3'	e2_r34	60.36
5' -	GCTGCCTATGCTTGCGGGTTGT	- 3'	e12F64	60.3
5' -	AGCGCAGAGCCTGTGAAGCA	- 3'	e12r139	59
5' -	CCGGCTGCTGTGTGCTGACC	- 3'	e10F16	69.93
5' -	GACGGCTGTGTGGCAGGACG	- 3'	e10F39	60
5' -	CGTAGGGCAGCGGGGACTCT	- 3'	e10r16	60
5' -	TGCTTGGCAGGACAAGGTGGC	- 3'	e12F26	59.84
5' -	GCTGGGTGCGTTGGGTGAAG	- 3'	e15F12	58
5' -	GCCGTCTGCCGTGCCCTTTA	- 3'	e5F3	59.7
5' -	CCTAGCCCCCGCTGACCTGA	- 3'	e4F5	59.5
5' -	CCGCATCCTCACCGCCACAA	- 3'	e4r87	71.83
5' -	ACCCGAAGCCGCACTGCAAG	- 3'	e4r71	60.25
5' -	CAGCGTGGCCGTGAAGGTGG	- 3'	e4r192	60.1
5' -	CCCCTAGGCTCCGCCCTTCC	- 3'	e4r135	60.3
5' -	GGTGCGATCGGCTTGCTTGG	- 3'	e4F4	58
5' -	AGGCGCCGAAGGAGCTGAGA	- 3'	e4r13	59.97
5' -	TGACGCAACATGGGCGGTTCT	- 3'	e4F22	58.1
5' -	GCCGGGGATGAACAATGTGGGT	- 3'	e4F112	58.54
5' -	CTGTGGCGGTGGCGCGTA	- 3'	e4F4A	58.67
5' -	GCTTCCTGCGCGACAGTGA	- 3'	e1F133	57
5' -	AGGCGCCGAAGGAGCTGAGA	- 3'	e1r11	69.9
5' -	GTCGTGCTACGTCGTGAATCCCCT	- 3'	e1F0	69.49
5' -	CGTAGCTTGACTAACGGACCCTAGC	- 3'	F0_60	65.4

2.7.2 PCR Amplification

PCR products were generated using a Techne Thermal Cycler-312.

PCRs (Polymerase Chain Reactions) were performed using reagents from

FailSafe PCR System Mix E (Epicentre) with 10 ng of genomic DNA in a 10-µl reaction mixture (10 nM of each primer, 50 µM dNTPs, 3 mM $MgCl_2$), under the following conditions: 2 min at 95°, 30 cycles of amplification [denaturation (30 sec at 94°)/annealing (.3-1 min at appropriate temperature depending on the primers)/polymerization (appropriate time at 72° depending on the primer sets)], and 10 min at 72°. To confirm the integrity of each sequence we analyzed at least two independently derived clones.

2.7.2.1 PCR Thermal Cycling Conditions

PCR thermal cycling conditions employed during a chromosome walk:

Reaction	Step	Thermal Settings	No. of Cycles
	1	95°C, 5 min	1
	2	94 °C, 30 s; Tm°C primer spec.; 72°C template spec.;	30
	3	72°C 5 min	1

2.7.2.2 Assembly of PCR Reactions

Assembly of a PCR reactions for 1 Reaction Tube in 10 μl Total Volume:

5 μl Mix E

10 μM Forward Primer

10 μM Forward Primer

0.1 μl Enzyme Mix (1.25 units)

1 μg Genomic DNA Template

0.9 μl dH_2O

2.7.2.3 Amplification of Targets with High Percent GC

The amplification of targets with high percent GC was accomplished by using Hot Start CleanAmp™ 7-deaza-dGTP (CleanAmp™ 7-deaza-dGTP Mix, TriLink BioTechnologies Inc., Cat. no. N-9504), 1x PCR buffer (supplemented with $MgCl_2$ to 2.5 mM; Invitrogen), 1.25 U *Taq* DNA polymerase (Invitrogen), 0.2 μM forward and reverse primer pairs.

2.8 Agarose Gel Electrophoresis

After PCR reactions where completed, amplified DNA samples where subjected to agarose gel electrophoresis. In order to obtain the DNA bands of specific interest 0.9% agarose gels were cast in 1X Tris-Acetate-EDTA (TAE, Invitrogen) buffer containing 1 μg/mL ethidium bromide. We loaded 100 ng of amplified DNA along with the 1 μl of the Agarose Gel loading DYE-5X (Fisher) in the total of 10 μl reactions into the each well. Gels were run in Mini Gel Migration

Tank, GEL XL or GEL XL UTRA-V2 (Labnet) at 100V for approximately 20 min, then visualized and photographed using the FloroChem™ 8900 (Alpha Innotech) imaging system.

2.8.1 PCR Products Gel Purification

Clean up of the conventional PCR and SF PCR products were performed using the Wizard SV Gel and PCR Clean-Up System (Promega) according to the manufacturer's protocol. Briefly, 6 µl from the PCR as well as SF PCR products were loaded on a 0.6 or 0.9% agarose gel and run at 50 volt for 45 minutes in 1X TAE Buffer. After the completion of the run each DNA band of specific interest was excised from the gel and on column purification was done by centrifugation as it stated in the manufactures protocol mentioned above. The purified DNA was eluted with 25 µl of Nuclease-Free Water (Quagen) and then stored at 4°C. DNA concentration was determined as described in section

2.9 Sequencing

Purified DNA samples along with the respective primers were sent for sequencing to Genewiz (South Plainfiled, NJ). Gel-purified DNA samples were sent in 10 µl reactions of ~ 40 ng/µl DNA concentration along with 5 pmol of the specific forward and reverse primer pair used to amply each fragment. The DNA concentration was determined as described in section 2.3.

2.10 Total RNA Purification from Algal Cells

Total RNA was purified using TRIzol reagent (Invitrogen) according to the manufacturer's protocol. All centrifuge and collection tubes were pretreated with RNaseZap (Ambion). Prior to the phenol-chloroform total RNA purification procedure mid-log-phase vegetative and gamete

wt as well as all fusion defective mutants cells were harvested at 5,000 rpm in a Superspeed RC2-B Automatic Refrigerated Centrifuge (Sorval) for 5 min. Aliquots of $2x10^7$ cells from each sample were flash-frozen in liquid nitrogen and were thawed in 1000 µl TRIzol reagent followed by incubation at 30°C for 5 min. To induce phase separation 200 µl of chloroform was added to each sample following with brief mixing. The homogenate was incubated at 30°C for 15 min and followed by centrifugation at 12,000 x g/15 min in an Eppendorf 5418 centrifuge at 2-8°C thereafter. The aqueous upper phase containing RNA was aspirated and transferred to a new microcentrifuge tube and 500 µl of iso-propanol (Sigma) added to each sample to precipitate the RNA. Samples were incubated for 3 min at 25°C and tubes were centrifuged at 12,000 x g/10 min at 2-8°C. The RNA pellets were washed in 75% ethanol and resuspended in 20-35 µl of RNase free water.

2.10.1 DNase Treatment of RNA

Any residual gDNA was removed by treatment with 2 Units/µl Turbo RNase-free DNase (Ambion) in a 50 µl reaction incubated at 37°C for 30 min.

2.10.2 RNA Re-purifucation

Resulting RNA was on column re-purified and extracted with RNeasy MinElute Cleanup Kit (QIAGEN). Briefly, samples were adjusted to 100 µl with dH_2O and 350 µl of RLT buffer was added and mixed with 250 µl of 100% ethanol by pipetting. Resulting mixture was transferred into the RNeasy MinElute spin column and centrifuged at 2-8°C at 8000 x g/15 sec in an Eppendorf 5418 centrifuge thereafter. The supernatant was discarded and 500 µl of RPE buffer

was added to each spin column that was centrifuged at 8000 x g/15 sec. This step was followed by adding 500 µl of 80% ethanol and centrifugation 8000 x g/2min, and latter 5 min centrifugation of the RNeasy MinElute column with open lid at 14,000 x g. The RNA was eluted with 14 µl RNase free water via 1 min centrifugation at maximum speed. All centrifugations were done at room temperature.

2.10.3 RNA Concentration Determination

RNA concentration was quantified on Qubit® 2.0 Fluorometer using RNA BR (Broad-Range) Assay Kit (Invitrogen) according to the manufactures protocol. Also, quality of re-purified RNA was measured by OD260 and OD280 of a 1:50 dilution (see section 2.5) and checked by gel electrophoresis described below.

2.10.4 Formaldehyde gel electrophoresis of RNA

All RNA equipment including gel tanks, gel casting trays and combs were subjected to the RNase denaturation treatment (Ambion) and then thoroughly rinsed with nuclease free dH2O. 1% formaldehyde denaturing gels were prepared by heating 0.5 g of agarose in 45 mL dH2O. After cooling to approximately 50ºC 5 mL of 10X denaturing gel buffer (Ambion) and 5 µL ethidium bromide (10 mg/mL) were added and gels were cast and left on the bench to polymerize for 1 hour. Samples of 5 µL RNA were prepared by the addition of 15 µL formaldehyde loading dye briefly spun and denatured by heating to 55ºC for 5 min and then loaded onto the gel. Each gel was run at 55V for 2 h in 1 × MOPS buffer.

2.11 qRT-PCR

2.11.1 qRT-PCR Oligonucleotide Synthesis

Oligonucleotide primers synthesized by Invitrogen were diluted to 100 μM with sterile dH_2O.

2.11.2 RT Reaction

First-strand cDNA synthesis from 1 μg of total RNA was done with SuperScript III Reverse Transcriptase (Invitrogen) for 50 min at 37°C. Following heat inactivation for 15 min at 70°C, 2 μL of cDNAs were used per PCR reaction.

2.11.3 Assembly of qRT-PCR Reactions

Real-time PCR was performed on an MJ Research Opticon (BIO-RAD Hercules, CA) using the SYBR green PCR master mix with the equivalent input of 50 ng RNA and 1 μl each primer. Primers selected to amplify sections of the MGS spanning the 235-bp exon were Forward, 5′-AGGATAGCGGCTTCTCTGGATTCGG -3′; and Reverse, 5′-CCCACTGAACGCCCAACATCTTCAG -3′. The expected RT amplicon was 129 bp. The 18S rDNA gene was amplified for each sample to ensure control of cDNA amplification from respective RNA samples Forward, 5′ - ATCTGCGAAAGCATTTGCCA-3′; Reverse, 5′- CGGCATCGTTTATGGTTGAGAC -3′). The expected product was 104 bp (cDNA). Thirty five RT–PCR cycles were carried out with denaturation for 1 min at 94°C, annealing for 30 sec at 57°C, and extension for 1 min at 72°C. We analyzed fold-changes (average and standard deviation) of MGS expression using the average threshold cycle (Ct) value of vegetative and gamete cells by the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). We quantified serially diluted samples by standard curve method for

quantification of mRNA abundance (Larionov et al, 2005). All qPCR products were visualized on a 2% agarose gel after 35 min 100V electrophoresis and melting curves were performed after qPCR to confirm that the amplification product was unique as well.

2.12 Cloning of the qRT-PCR Product

After qPCR products were visualized on a 2% agarose gel bands corresponding for 18S and MGS cDNA were cut out and gel purified using Wizard SV Gel and PCR Clean-Up kit (Promega) as it was described in section 2.1.

2.12.1 TOPO TA Cloning

Cloning Reaction was performed utilizing TOPO® TA Cloning (Invitrogen).
Briefly, 2 µl of gel purified product were added to the mixture of 1 µl salt solution, 1 µl TOPO® vector and 2 µl of dH_2O to a final volume of 6 µl. The resulting mixture was mixed gently and incubated for 5 minutes at room temperature followed by incubation on ice. Transformation of the One Shot® Chemically Competent *E. coli* cells was done with 2 µl of the TOPO® Cloning reaction from above. Followed by gentle mixing and incubated of tubes on ice for 20 min. Tubes with cells were heat-shocked fro 30 seconds at 42°C without shaking and immediately transferred to ice. We added 250 µl S.O.C. medium to each tube followed by shaking the tubes horizontally (200 rpm) at 37°C for1 hour. After transformation 10-50 µl of content was spreaded on selective plates containing kanamycin and incubated overnight at 37°C. 10 colonies were picked from each transformation reaction for downstream applications.

2.12.2 Midi Prep

A single *E.coli* colony was incubated at 37°C in a starter culture of 5 ml LB medium containing 50 µg/ml kanamycin under shaking conditions (300 rpm) for 7 hours. After initial incubation, 200 µl of started *E.coli* culture were added into a flask containing 100 ml of LB medium containing the same concentration of antibiotic mentioned above and incubated under shaking condition at 37°C for overnight. Plasmid extraction was performed using plasmid Midi Kit (QIAGEN) according to the manufactures protocol. Quantification of resulting plasmid DNA was done on a 0.8% agarose gel. Then samples were sent for sequencing to Genewiz (South Plainfield, NJ). Samples and sequencing primers were sent dissolved in water at room temperature as described in section 2.7.

2.13 Thin Layer Chromatography (TLC)

Total lipids, from all of the algae species examined in this study were extracted via chloroform:acetic acid (5:1v/v) mixture following the protocol after Folch et al., (1957) with minor modifications. All operations were carried out at 4°C. First mobile phase solvents: methylacetate:isopropanol:chloroform:methanol:KCl (v/v 25:25:25:10:4). Second mobile phase solvents; hexane:ethyl ether:acetic acid (v/v 70:30:2). 5µL from each total lipid extract was spotted on the TLC plate, allowed to dry and transferred in to the TLC chamber containing the elution mixtures specified above and allowed to run both phases for 50 min.

2.14 MEP and MVA DNA and Protein Sequences Searches

Search for MEP and MVA pathway DNA and protein sequences for *Chlamydomonas reinhardtii* and *Arabidopsis thaliana* was performed after (Frommolt et al., 2008). Gathered *C. reinhardtii*

and *A.thaliana* MEP and MVA pathway DNA and protein sequences were verified and used to perform BLAST with the e-value cut-off at 1.0e-10 (Altschul et al., 1997) and TBLASTN with the e-value cut-off at 1.0e-20 (Gertz et al., 2006) searches against the following publicly accessible data bases: 1) US DOE JGI (Joint Genome Institute) via Eukaryotic Genomics Web site (http://genome.jgi-psf.org), 2) NCBI (National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov), 3) TBestDB (Taxonomically Broad EST Database of the pan-Canadian Protist EST Program, (http://amoebidia.bcm.umontreal.ca/pepdb/searches/welcome.php). Among the investigated algae genomes were species representing three groups that are included in the green algae (Chlorophyta). The species of *Chlamydomonas reinhardtii* and *Volvox carteri* belong to the class of the *Chlorophyceae*. The class of the *Trebouxiophyceae* was represented by the unicellular alga *Chlorella* st.NC64A (Blanc et al., 2010) as well as *Chlorella vulgaris* (previously misidentified and now recognized as the *Coccomyxa* sp.169). The class of the *Prasinophyceae* was represented by the *Ostreococcus* st.RCC299*, Ostreococcus tauri* (Derelle et al., 2006), and *Ostreococcus lucimarinus* (Palenik et al., 2007); and two strains of *Micromonas*, *Micromonas pusilla CCMP1545 and Micromonas strain RCC299.* Eukaryotic Genomics browsers of mosses *Physcomitrella patens subsp. Patens,* the heteroconta (starmenophiels) *Thalassiosira psuedonana* (Armbrust et al., 2004), *Aureococcus anophagefferens, Phaeodactylum tricornutum*, *E.huxleyi* and *F.cylindrus* as well (US DOE JGI). Complete sequences of *Cyanophaseae* represented by *Synechosystis PCC6803, Anabaena sp. PCC7120 and Prochlorococcus marinus NATL2A* (US DOE JGI). Red algae was represented by protein sequences from *Cyanidioschyzon merolae* extracted from genome browser (http://merolae.biol.s.u-tokyo.ac.jp) at the University of Tokyo/Japan (Matsuzaki et al., 2004) and protein sequences extracted from genome database

(http://genomics.msu.edu/galdieria/references.html) at Michigan State University/United States (Barbier et al., 2005). Additional searches of the ORF (open reading frame) annotations and genome contings of fully sequenced algae genomes as well as the EST (expressed sequence tag) projects including Cryptophyte *Guillardia theta* CCMP2712 and Chlorarachniophyte *Bigelowiella natans* CCMP2755 (Curtis et al., 2012) were preformed as well.

2.15 Phylogenetic Analysis

Sets of orthologous proteins representing members of MVA and MEP pathways encoded in all genomes under this study were aligned in individual files with MUSCLE 3.6 (Edgar, 2004) or CLUSTALW 1.83 (Chenna et al., 2003) and written into NEXUS and PHYLIP format. Phylogenetic trees were inferred via the maximum likelihood (ML) and neighbor joining (NJ) methods using the programs 'Neighbor' and 'ProML' from the PHYLIP package, version 3.67 (Felsenstein, 1989). Branch support values were estimated with 100 bootstrap replicates. Additionally, phylogenetic trees were constructed in MrBayes (using the same alignment), version 3.1.2 (Huelsenbeck and Ronquist, 2001), running MC^3 for one million generations under the Jones amino acid substitution matrix with a fixed rate among sites. We sampled every 100^{th} generation and estimated posterior probabilities using the final 5,000 trees. Trees produced from all three methods are similar in their topologies. Branch lengths are proportional to the number of amino acid substitutions per site (see scale bar). Branch support values shown in the figures are derived from MrBayes.

CHAPTER 3

RESULTS

The goal of this lab is to identify the genes that are involved in *mt*$^-$ gamete fusion in *Chlamydomonas reinhardtii.* In this study, we are looking at a gene (Cre13.g605900.t1) we had previously identified (Aksoy, PhD Thesis, 2008).

3.1 Confirming the Flanking Genomic Region of the Insertions

Previous forward genetic experiments that utilized DNA insertional mutagenesis with the pSP124S plasmid, resulted in random insertions into the genome of the host *Chlamydomonas* strain *sru-2* (La, Masters thesis, 1987; Lam, Masters thesis, 1991; Aksoy, PhD Thesis, 2008). The above-mentioned random insertions into the genomic DNA could have disrupted gene coding sequences and/or regulatory regions that control gene expression. In order to confirm the physical location of the plasmid pSP124S in the fusion defective mutant cl5 [previously mapped by TAIL-PCR (Aksoy, PhD Thesis, 2008) and SF-PCR (Lai, Master Thesis, 2011; Shrestha, Master Thesis, 2011)] we ran additional SF-PCR experiments (Site-Finding Polymerase Chain Reaction, Tan et al., 2005). SF-PCR required us to utilize a sequential combination of degenerate and plasmid pSP124S specific primers. These experiments allowed us to walk into the sequences adjacent to the inserted plasmid in a collection of insertional fusion-defective mutants that had not been complemented with the wild type copy of the HAP2/GCS1 gene (Section 1.5).

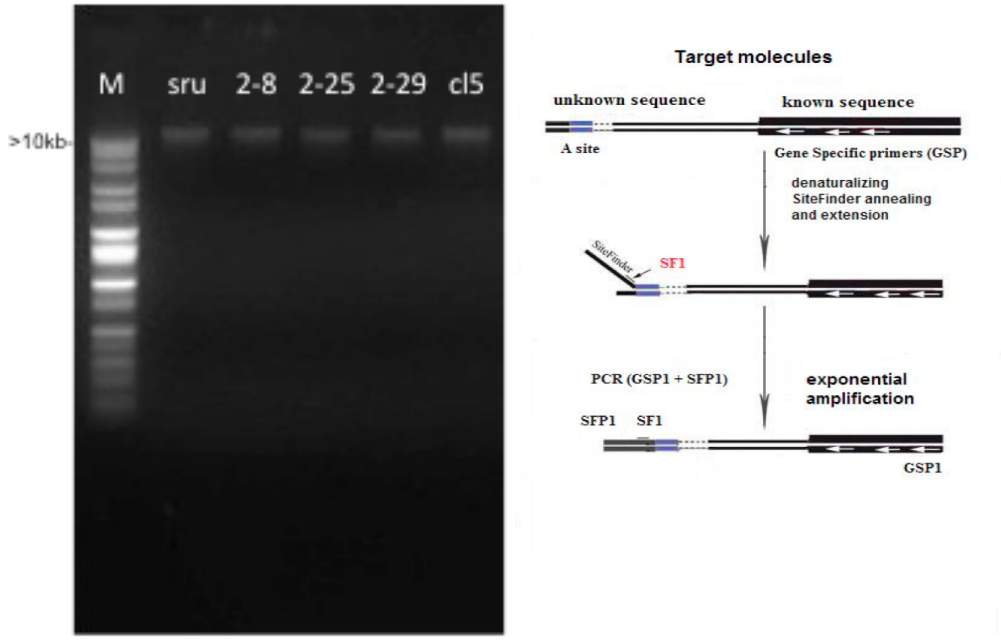
3.2 Verification of the Insert Location in Mutant cl-5 by SF-PCR

According to previously reported Southern Blot results, the mutant cl5 showed three bands, suggesting that there are three insertions in its gDNA (Aksoy, PhD Thesis, 2008). Utilizing SF-PCR to walk into the flanking regions from the insertion/s should allow us to verify the gene(s) or gene regulatory region/s that are disrupted by an insertion, one of which might be the cause of the fusion defective phenotype.

3.2.1 Pre-Primary and Primary SF-PCR Hybridization

The genomic DNA of the control strain *sru-2* as well as several insertional fusion-defective mutants (including cl5) utilized in SiteFinding-PCR are shown in figure 3A. During the pre-primary SF-PCR reaction we used degenerate Site Finder primer (SF1). The pre-primary step occurs at low temperature and produces annealing sites for the SF primers (figure 3B), therefore leading to the initiation of the Site Finding reaction. During the primary SF-PCR reaction the degenerate Site Finder primer (SFP1) along with the insert specific primer Ble1 were utilized, figure 3C. We did not expected to see any specific product from the control *sru-2*, host strain used to create the insertional fusion-defective mutants, except for the smearing effect usually seen after primary SF-PCR reactions (figure 4A1). As for the insertional fusion-defective mutant cl5, the smearing effect as well as the presence of non-specific DNA bands after the primary SF-PCR reaction indicated that the primary SF-PCR had amplified variable length DNA products (figure 4A1). The negative control labeled –C, where no DNA was added to the SF-PCR reaction, had no product or smear (figure 4A1), thus we could continue to the secondary SF-PCR.

(A) **(B)**



(C)

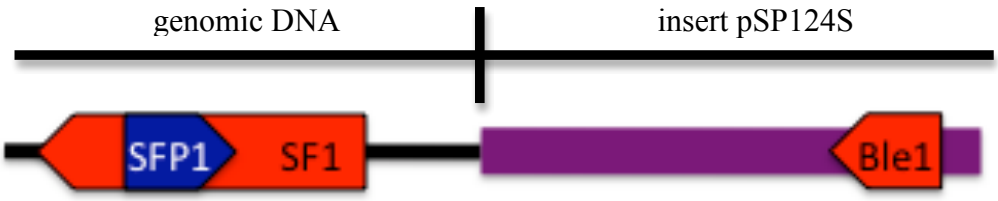


Figure 3. Schematic Diagram of Primary Site Finding PCR. A. 0.9% agarose gel image of the genomic DNA utilized in SiteFinding-PCR. **B.** Schematic outline of primary SF-PCR (Tan et al. 2005). **C.** Diagram indicating where the gene specific primer Ble1 binds to the insert and the relative position of SFP1 to SiteFinder 1 (SF1).

3.2.2 Secondary SF-PCR

The secondary SF-PCR for cl5 was performed using the primary SF-PCR products as templates. A nested gene specific primer, Ble2 that binds to the second exon of *ble* upstream of the primer used for primary PCR, as well as primers RD223 and RD225 (Table 2B) that bind upstream of

the primer Ble2, were paired with the same degenerate primer, SFP2 in three separate reactions, figures 4A2 & 4B. As expected, specific products were formed in the secondary nested SF-PCR for cl5, shown on the left of the gel (figure 4A2). The size of the three amplified DNA bands using the Ble2 primer were, ~ 950 bp (top band), ~ 450bp (middle band) and ~ 325 bp (lowest band). Three DNA bands were produced using the RD223 primer, ~ 700 bp (top band), ~ 400bp (middle band) and ~ 250 bp (lowest band). Two DNA bands were seen using the RD225 primer, a top DNA band at ~ 650 bp and a lower one at ~250 bp. Again the negative control reaction shown in –C/2° demonstrated that no contamination was present.

(A1)

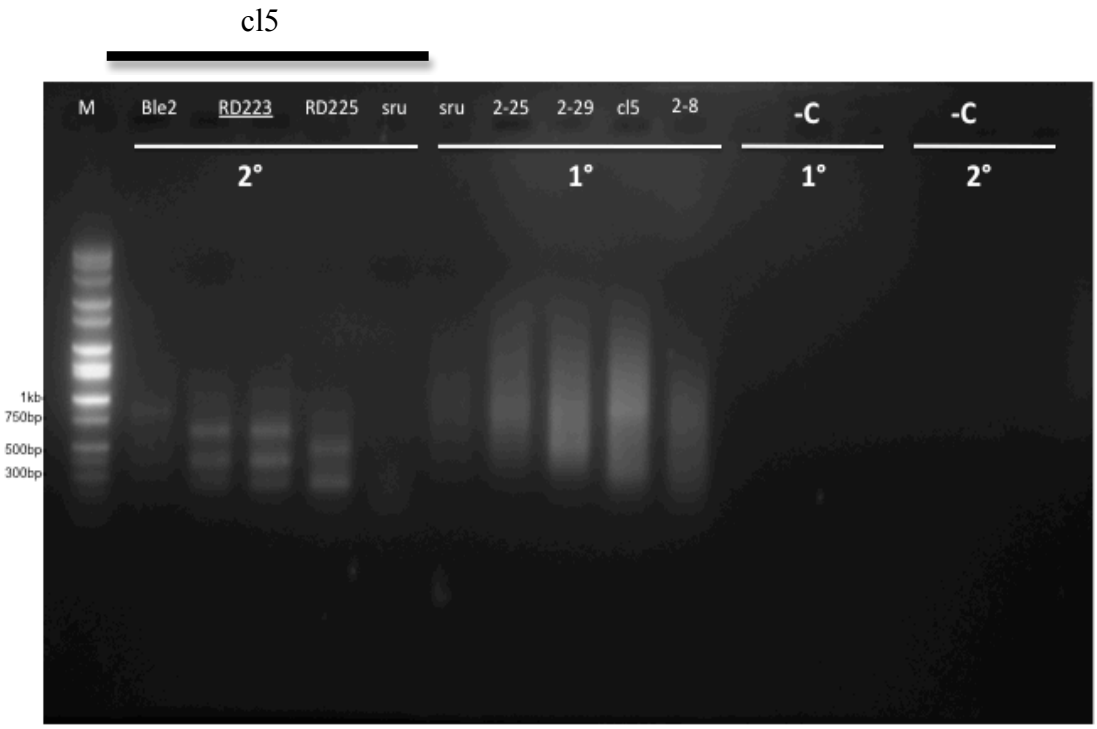
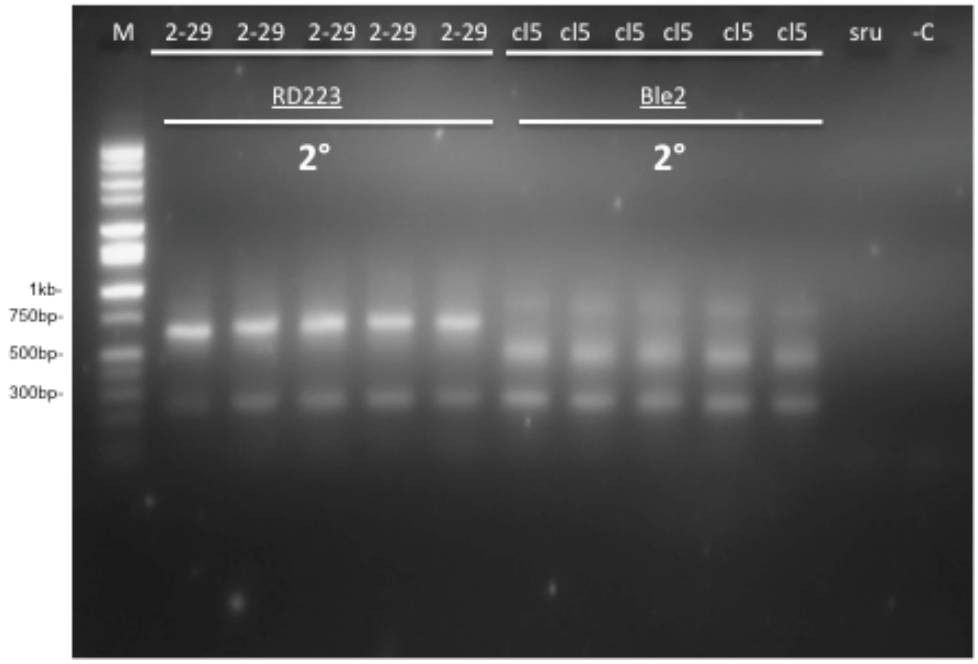
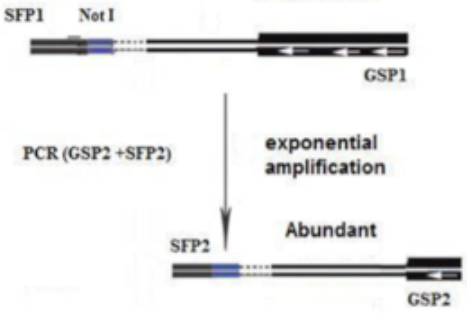


Figure 4 continued next page.

(A2)



(B)



(C)

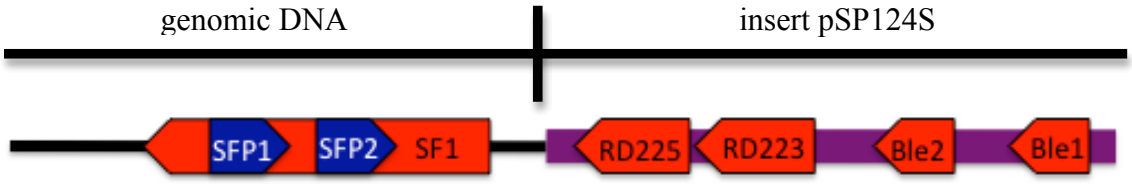


Figure 4. Results for Primary and Secondary SF-PCR. **A.** 0.9% agarose gel image for primary and secondary SF-PCR products. Primary SF-PCR reactions for multiple mutants including cl5 are shown in the middle of the gel A1. Secondary SF-PCR reactions for cl5 are shown on the left of the gel (A1) and on the right of the gel (A2). The negative controls, labeled –C/1° & –C/2°, where no DNA was added to the SF-PCR reactions showed no product (results from the other mutants shown will be discussed later) **B.** Schematic outline of the secondary SF-PCR (Tan et al., 2005). **C.** Diagram indicating where the gene specific primers Ble1, Ble2, RD223 and RD225 bind to the insert and their relative position with respect to SFP1, SFP2 and SF1 primers.

The DNA for each band was cut out; gel purified and sent for sequencing at Genewiz.

3.3 Sequencing Results for Mutant cl5

After the gel purified DNA samples from mutant cl5 were sequenced, we blasted the sequences we received from Genewiz into the Phytosome 8, JGI 4.3 *Chlamydomonas* database. According to the sequencing results, only the top size bands for each of the primer sets used during the secondary SF-PCR allowed us to verify the regions flanking the insert of plasmid pSP124S, which was used for DNA insertional mutagenesis (the other sequences showed only the plasmid). The sequence corresponding to the Ble2 DNA band at ~950 bp (figure 5A) produced three hits matching genomic sequences in the *Chlamydomonas* database. They are located on chromosomes 2, 7 and 13. The longest matching genomic sequence was located between 6297731 and 6297919 on chromosome 13, shown in figure 5B. The sequences corresponding to the other two top DNA bands ~700 bp using RD223 and ~650 bp using RD225 primer (figure 5A) produced six hits matching the genomic sequences in the *Chlamydomonas* database. They are located on chromosomes 1, 2, 5, 7, 12 and 13. The longest matching genomic sequence for both the RD223 and RD225 primers was located between 6297731 and 6297919 on chromosome 13, which is the same as the result obtained with the Ble2 primer, shown in figure 5B. The above-mentioned genomic sequence on chromosome 13 corresponds to the 3´ UTR of the Cre13.g606000.t1.1 gene that codes for an elongation factor-like protein (Ef-GTPase) as well as the partial coding sequence and 5´UTR of the Cre13.g6059550.t1.1 gene coding for a *copia*-family retrotransposon (RT), shown in figure 5C. The multiple hits returned after we blasted the base sequences of the SF-PCR purified DNA fragments into the Phytosome 8, JGI 4.3 Chlamy

database is due to a fact that the *Chlamydomonas* genome contains at least seven *copia*-family retrotransposons (results not shown).

(A)

Ble2

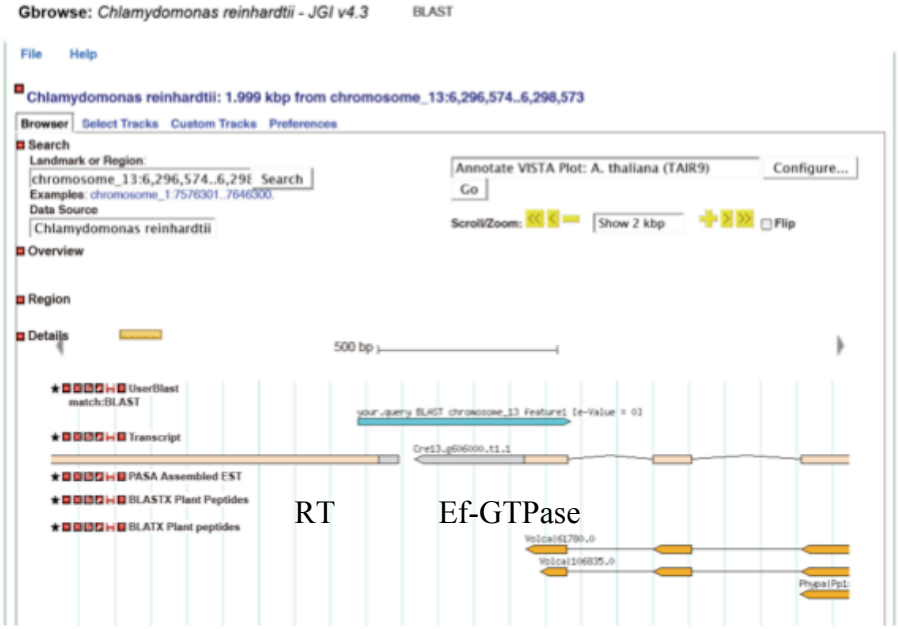
NNNNNNNNNNNGNNNGNNNNNCGTCNCCNCGAGTCCCGGGAGAACCCGAGCCGGTCGGTCCAGAACTCGACCGCTCCGGCGAC
NNNNCGCGCGGTGAGCACCGGAACGGCGCTGGTCAGCTTGGCCATCCTGCAAATGGAAACGGCGACGCAGGGTTAGATGCTGCT
TGATACGGCGACAGAGGAGCCAAAAGTGTTCNGCGACATGNGCGNCCCNGTTCNTCAGATNGGCANNACTTGACCNGACCCGCC
GACGAGTGNTCGATCCTCCTAGCTAATGATAAGAAGGCTTGCCAATTGCGAGCGAATTAACGACAAGCGGTGCTCATAACCTGTT
GCGAAGTTGGTNCGGCAAGCCTGAATTGCTAGCGGGTGCGTTACACACGATGCAGCTACGCTGNNTGNNACCTTCGNTAGTCTGG
ATATATAATGCGCATGTNTAGTTGACAGCAACACGCTTGGNCGGCGCNNNNNACCTCCCGACGGTCCGTGAACCGACACAATGCC
GTACTGGCAGTCGGGNAGGCGCTACAGATGGNCAGACAAGGAGGCACTTCTGCCTGATCAGGGCAGGCGCCTCGCTAACTTCCC
AGGNCCCCAAGTTCGGACCGCAGCTACCACCTCACCGGCAGTGCTGNCCGATAACCCCAAATTACCAACAGCTGGNANCATCTAC
TGCTGCGCCTGTTGCNNCCNNTTGNCCTTGNNNGCCGCGCCCCGCGTCCAGGACCTCNNTGCGCAGCCGCANCGCCGCAGGGGTC
ANCTCTATGAGCTCGTCCTGGCCAACATACCTGCAGNCGGACCGGGGNCGGCGGGTTTATGTTGCAGTNNGTAAGCACGGGNCCN
TCNTCTNNNNTNNNNNNNNNNNNNNNNNNNNNN

RD223
NNNNNNNNNNNNNNNNNNNANNNNTTCGCNNTATAGNNAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGCAGCA
TCATCTTGAAGCTGCTTGAACTTGACCAGTCCCGAGGACGAGTCATCCATCCTACCTAGCTAGCGATAATAAGGCTTGCCAACCGC
GAGCGAATTAACGACAAGCGGGGCTCATAACCTGTTGCGAAGTTGGTTCGGCAAGCCTGAATTGCTAGCGGGTGCGTTACACACG
ATGCAGCTACGCTGCCTGGCACCTTCGGTAGTCTGGGTATATAATGCGCATGTGTAGTTGACAGCAACACGCTTGGTCGGGGCTTA
AGACCTCCCGACGGTCCGTGAACCGACACAATGCCGTACTGGCAGTCGGGGAGGCGCTACAGATGGGCAGACAAGGAGGCACTT
CTGCCTGATCAGGGCAGGCGCCTCGCTAACTTCCCAGGGCCCCAAGTTCGGACCGCAGCTACCACCTCACCGGCAGTGCTGCCCG
ATAACCCAAATTACCAACAGCTGGCACCATCTACTGCTGCGCCTGTTGCGCCCGCTTGGCCTTGGTCCGCCGCGCCCCCGCGTCCA
GGACCTCCTTGCGCAGCCGCACCGCCGCAGGGGTCACCTCTATGAGCTCGTCCTGGGCAACATGCCTGCAGGCNNCCNCGCGGCC
GCTTGAGGACGCTGTGCGAGGNGNNNGNTTTTNANNNNCNTGNNCGNGANTTAAATGCCANAAGGAGCNCAANCNNANNNTTNT
NTNNTNTNNTTNNGNNNNTNNNGNNNANANAN

RD225
NNNNNNNNNNNNNNNNGNANNNNTTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGCAG
CATCATCTTGAAGCTGCTTGAACTTGACCAGTCCCGAGGACGAGTCATCCATCCTACCTAGCTAGCGATAATAAGGCTTG
CCAACCGCGAGCGAATTAACGACAAGCGGGGCTCATAACCTGTTGCGAAGTTGGTTCGGCAAGCCTGAATTGCTAGCGGG
TGCGTTACACACGATGCAGCTACGCTGCCTGGCACCTTCGGTAGTCTGGGTATATAATGCGCATGTGTAGTTGACAGCAA
CACGCTTGGTCGGGGCTTAAGACCTCCCGACGGTCCGTGAACCGACACAATGCCGTACTGGCAGTCGGGGAGGCGCTACA
GATGGGCAGACAAGGAGGCACTTCTGCCTGATCAGGGCAGGCGCCTCGCTAACTTCCCAGGGCCCCAAGTTCGGACCGCA
GCTACCACCTCACCGGCAGTGCTGCCCGATAACCCAAATTACCAACAGCTGGCACCATCTACTGCTGCGCCTGTTGCGCC
CGCTTGGCCTTGGTCCGCCGCGCCCCCGCGTCCAGGACCTCCTTGCGCAGCCGCACCGCCGCAGGGGTCACCTCTATGAG
CTCGTCCTGNGCAACATGCCTGCAGGCCCCCCNGCGGCCGCTTGANGACGCTGTGCGAGGTGNGNTGTTNANTN

Figure 5 continued next page.

(B)



(C)

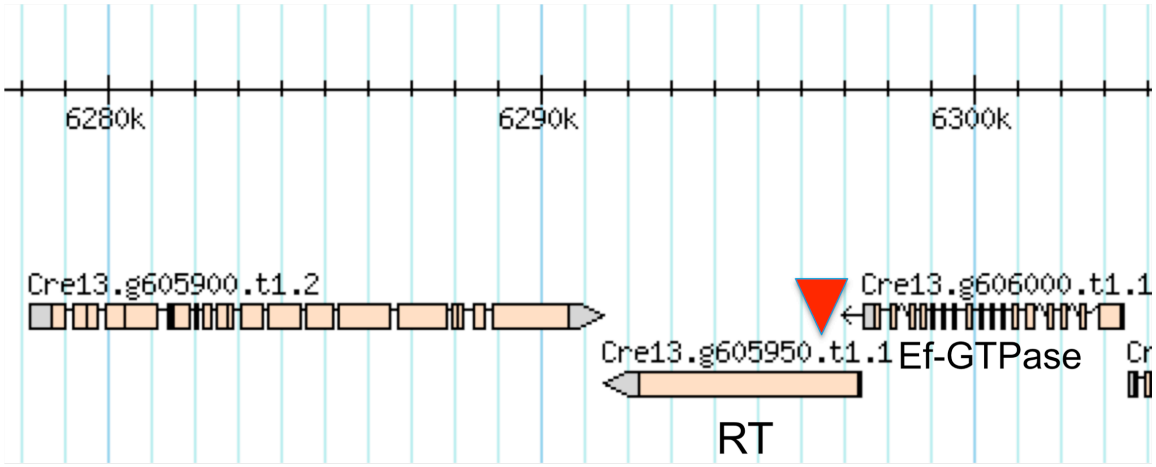


Figure 5. Sequencing result for mutant cl5. A. Sequences obtained from the secondary SF-PCR products. **B.** Phytosome 8, JGI 4.3 Chlamy DB blast results. **C.** Insert mapping, *copia*-family retrotransposon Cre13.g605950.t1.1 (RT) along with the point of insertion (inverted triangle) shown in the middle, Cre13.g605900.t1.2 shown on the left, the Cre13.g606000.t1.1 coding for Ef-GTPase is on the right.

Although Southern blots showed 3 bands for this mutant, we were only able to demonstrate one interesting SF-PCR.

3.4 Analysis of the DNA Regions Flanking the Insert

According to the analysis of the results that will be discussed in the sections 3.5 and 3.6 we decided to analyze regions on chromosome 13 that are proximal and distal to the *copia*-family retrotransposon.

3.4.1 Chromosome Walking Downstream From the Insertion

To attempt to demonstrate the molecular basis for the fusion incompetent phenotype in this mutant, we performed a chromosome walk through the regions flanking the insertion in the *copia*-family retrotransposon. To determine the DNA sequence, PCR reactions were performed on the genomic regions around the insertion site in cl5, (figure 6).

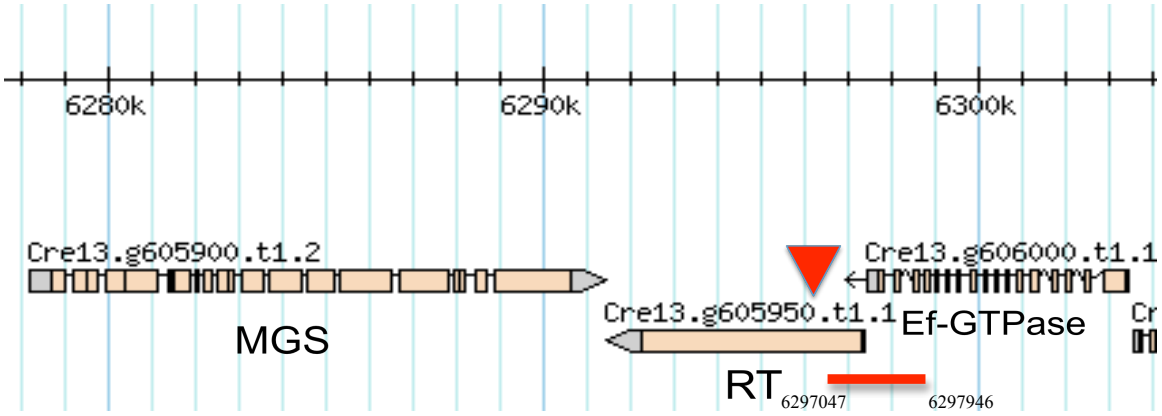


Figure 6. Schematic representation of the amplified DNA region flanking the insertion in cl5. Genome browser image of the region of the insertion in chromosome 13. MGS to the left of the RT and Ef-GTPase to the right of the RT. The inverted triangle shows the insertion site in the RT. The bar under the RT and Ef-GTPase indicates the downstream region of interest.

First genome specific primers were designed that were specific to the downstream site adjacent to the insertion on chromosome 13. In addition to mutant cl5, other insertional and conditional mutants as well as the *sru2* control strain and *wt* strains were also tested for polymorphisms in this region of chromosome 13. Primers fGTPase and rGTPase should produce a 899 bp product, amplifying the DNA region between 6297047 and 6297946 bp on chromosome 13. Figure 7 shows this expected PCR product from the control strain *sru2,* from wild types strains R^+ and NO^-, as well as from cl5 and the other insertional and conditional mutants tested.

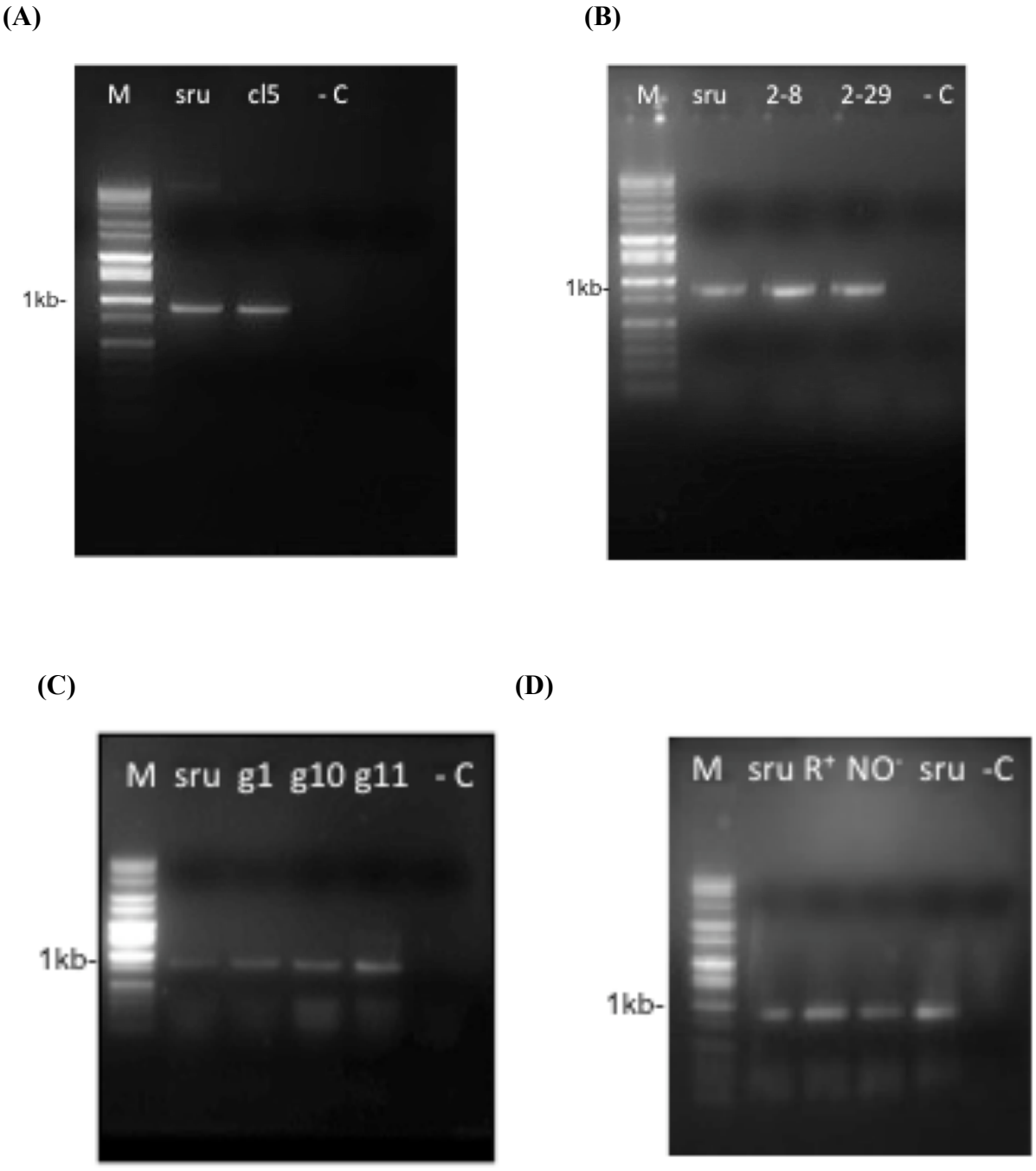


Figure 7. Chromosome walking from the region downstream from the insertion. A,B,C,D. All four agarose gels represent PCR with fGTPase and rGTPase that produced a 899 bp product. The *sru2*- control strain; cl5, 2-8, 2-29 insertional mutants; g1 (*gam*1), g10 (*gam*10) and g11 (*gam*11) conditional mutants; R^+ *wt mt*+, NO^- *wt mt*- (see Materials & Methods for strain ids). The last lanes are negative controls (–C) with no gDNA added. All bands from the PCR were gel extracted, gel purified and sent for sequencing at Genewiz. (M DNA marker).

The initial search for aberrant DNA sequences (such as deletion(s) or insertion(s)) in the mutant strains within the analyzed DNA region was based on screening for fragments that had altered mobility on a 0.9% agarose gel or by a failure to amplify a fragment. According to figure 7, all PCR amplification reactions produced the expected 899 bp product. All visible DNA fragments were gel purified and subjected to Sanger sequencing at Genewiz, enabling us to deduce their sequences and further check for any coding defect.

3.4.2 Sequencing of the Downstream Region Flanking the Insert (RT/Ef-GTPase Walk)

All bands from the PCR reactions were gel extracted, gel purified and sent for sequencing. Primer extension sequencing was performed by Genewiz, using Applied Biosystems BigDye version 3.1. The reactions were then run on Applied Biosystem's 3730xl DNA Analyzer. The sequencing results we received were blasted into the *Chlamydomonas* genome and aligned via the multiple sequence alignment algorithm ClustalW. The sequence alignment results showed that at position 6297404 on chromosome 13, corresponding to the intergenic region separating a *copia*-family retrotransposon and the Ef-GTPase, there is a base pair difference (C to T) in all of the strains we sequenced when compared to the CC-503 cw92 mt^+ strain that was used for genomic sequencing at JGI, (figure 8).

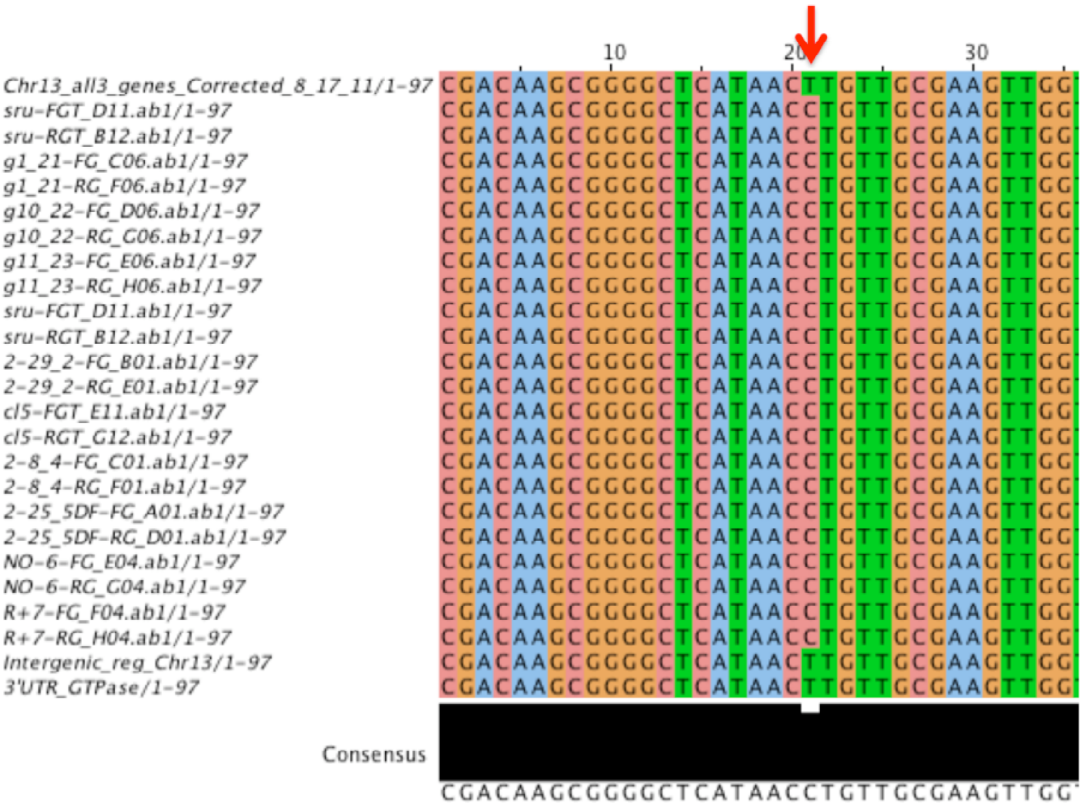


Figure 8. Sequence alignments for the region downstream from the insertion. Sequences for that downstream region were aligned via ClustalW. The red arrow indicates the position of a nucleotide base pair difference.

3.4.3 Chromosome Walking Upstream From The Insertion

Genome specific primers were designed for a site upstream from the insertion on chromosome 13. Primers fCAR and rCAR should produce a 2553 bp product, amplifying the DNA region spanning 6290574 and 6293127 bp on chromosome 13, (figure 9).

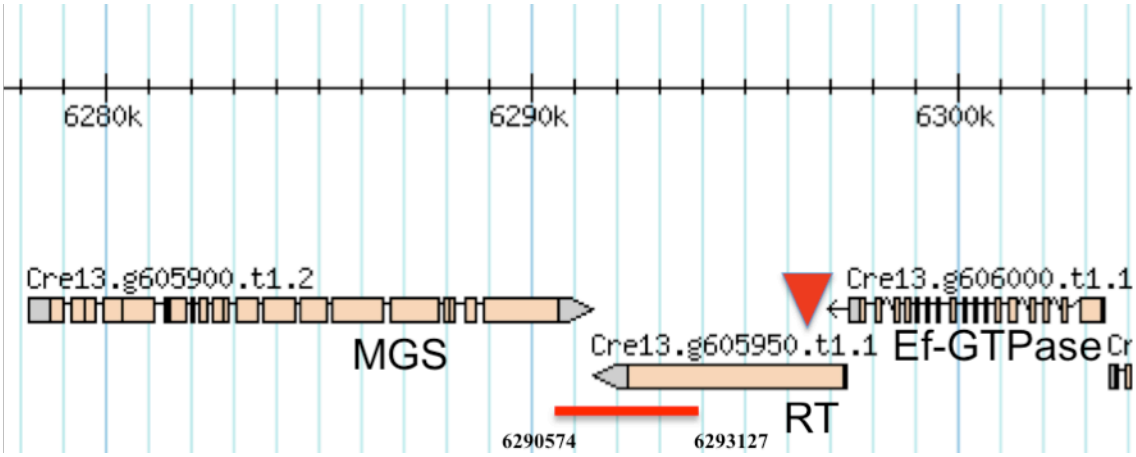


Figure 9. Schematic representation of the amplified DNA region flanking the insertion in cl5. Genome browser image of the region of the insertion in chromosome 13. MGS left of the RT and Ef-GTPase to the right of the RT. The inverted triangle shows the insertion site in the RT. The bar under RT and Ef-GTPase indicates the upstream region of interest.

In addition to the mutant cl5, other insertional and conditional mutants as well as the *sru2* control were tested. Figure 10 shows the expected PCR product from the control *sru2* strain, and the insertional and conditional mutants tested, including cl5.

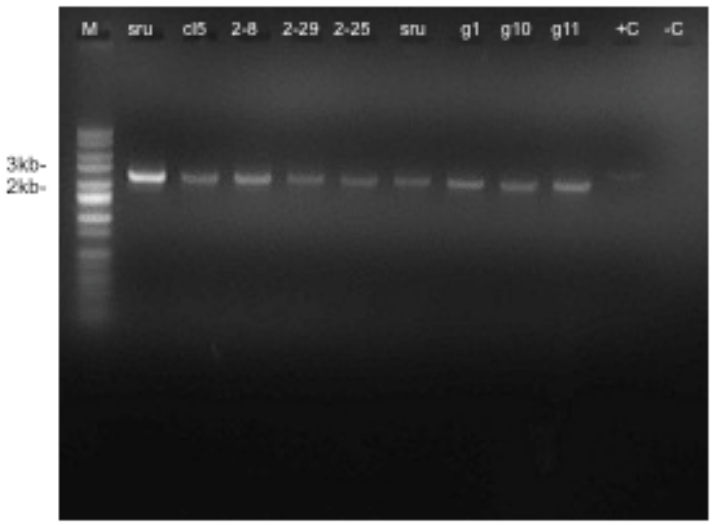


Figure 10. Chromosome walking in the region upstream from the insertion. PCR with fCAR and rCAR gave a 2553 bp product. (+C) is a positive control. The last lane is a negative control (–C) with no gDNA added. (M is a DNA marker).

The initial search for aberrant DNA sequences bearing mutations such as deletions or insertions within the above mentioned upstream region was based on screening for fragments that had altered mobility in the 0.9% agarose gel. Amplified DNA from all mutants produced the same size product (2553 bp); thus we confirmed that the upstream region adjacent to the insertion site did not contain any significant deletions or insertions as well. All bands from the PCR reactions were then gel extracted, gel purified and sent for sequencing at Genewiz.

3.4.4 Sequencing Results for the Upstream Region Flanking the Insert (MGS/RT Walk)

The sequencing results we received were blasted into the *Chlamydomonas* genome and aligned via the multiple sequence alignment algorithm ClustalW, (figure 11).

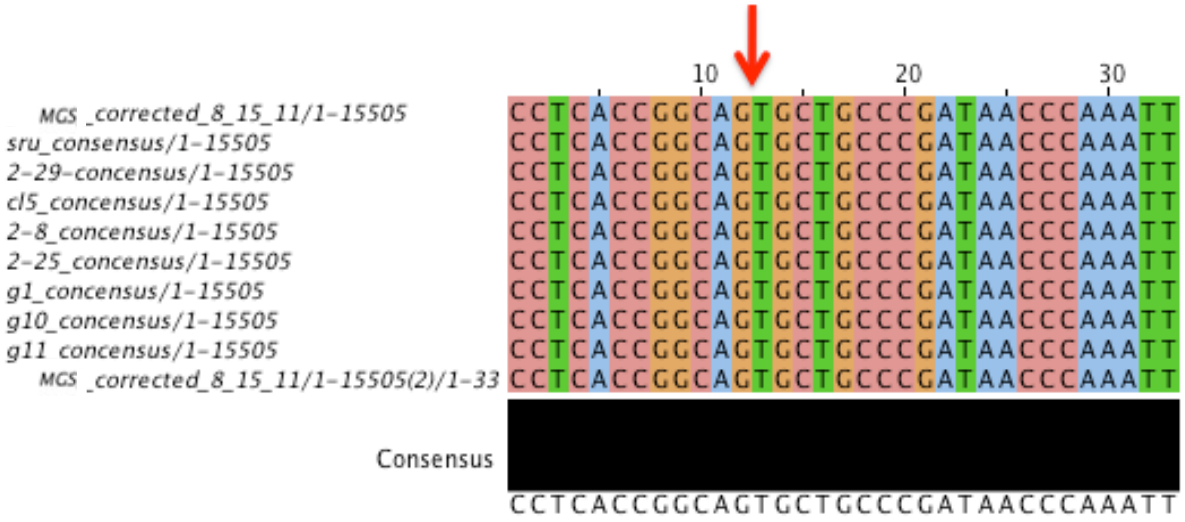


Figure 11. Sequence alignments for the region upstream the insertion. Sequenced MGS gene regions were aligned via the ClustalW algorithm. The red arrow indicates the border between the 3′ UTR of a *copia*-family retrotransposon (to the right of an arrow) and the adjacent intergenic region (to the left of the arrow).

The sequence alignment results showed that the DNA upstream region of *wt* and mutants spanning bp 6290574 and 6293127, covering part of the coding region of the *copia*-family retrotransposon and its entire 3′ UTR region and the upstream intergenic region between the RT and MGS as well as the part of the 3′ UTR region of the MGS, does not show any mutations.

3.5 Expressional Analysis of MGS

3.5.1 EST Data

We confirmed that the plasmid insertion in cl5 occurred in chromosome 13; as concluded from the sequencing result of SF-PCR amplification product, the exact base position is 6297731 (section 3.3). Next we analyzed Expressed Sequence Tags (EST's) for the ~ 100 kb region radiating from the point of insertion on chromosome 13. Only one gene, (Cre13.g605900.t1) located upstream from the *copia*-family retrotransposon RT (Cre13.g605950.t1) showed an unusual expression pattern, with all of the EST's coming from the gametic library. The EST data from this region of chromosome 13, from the updated JGI ChlamyDB 4.3 Phytosome 8 genome site, is shown in figure 12, including the retrotransposon and the gene models flanking it. To the right of RT is Cre13.g606000.t1 annotated as having homologies to both- an elongation factor and GTP binding protein (Ef-GTPase). To the left of RT is Cre13.g605900.t1 with a carboxypeptidase motif.

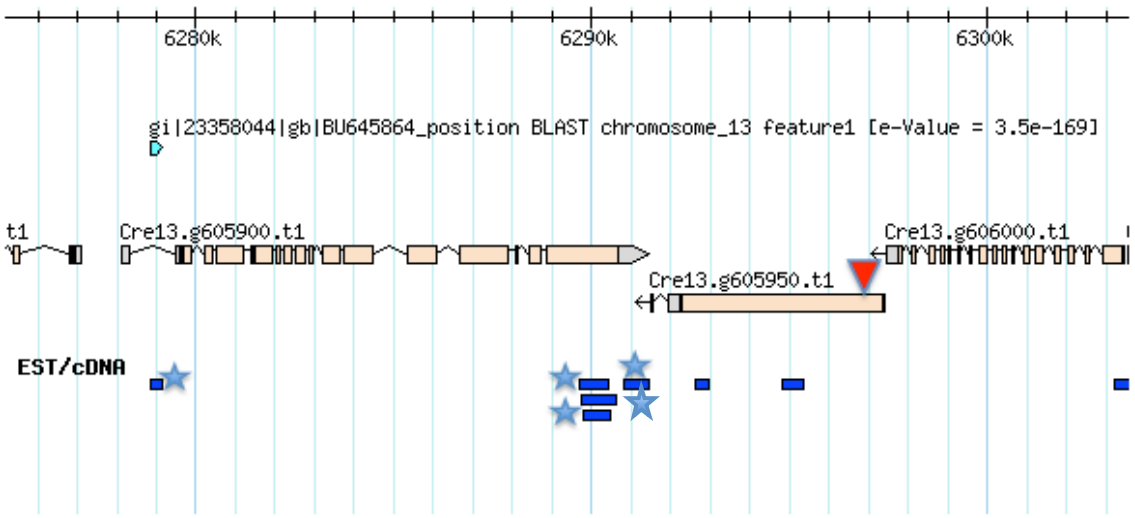


Figure 12. Genomic Analysis: EST data. EST's for the *copia*-family retrotransposon (Cre13.g605950.t1) along with a point of insertion (inverted triangle) is shown in the middle, Cre13.g605900.t1 with a carboxypeptidase motif is shown on the left, the Cre13.g606000.t1 coding for an Ef-GTPase is on the right. Gametic EST's with stars next to them are shown below each gene model.

3.5.2 454 Data

Results of the 454 transcriptome analysis (http://genomes.mcdb.ucla.edu) of the whole genome sequencing data for vegetative and gametic cells indicate an elevated level of Cre13.g6059500.t1 expression during gametic stage, figure 13B. Note the gametic data is more compressed due to the high level of expression of this gene in gametes.

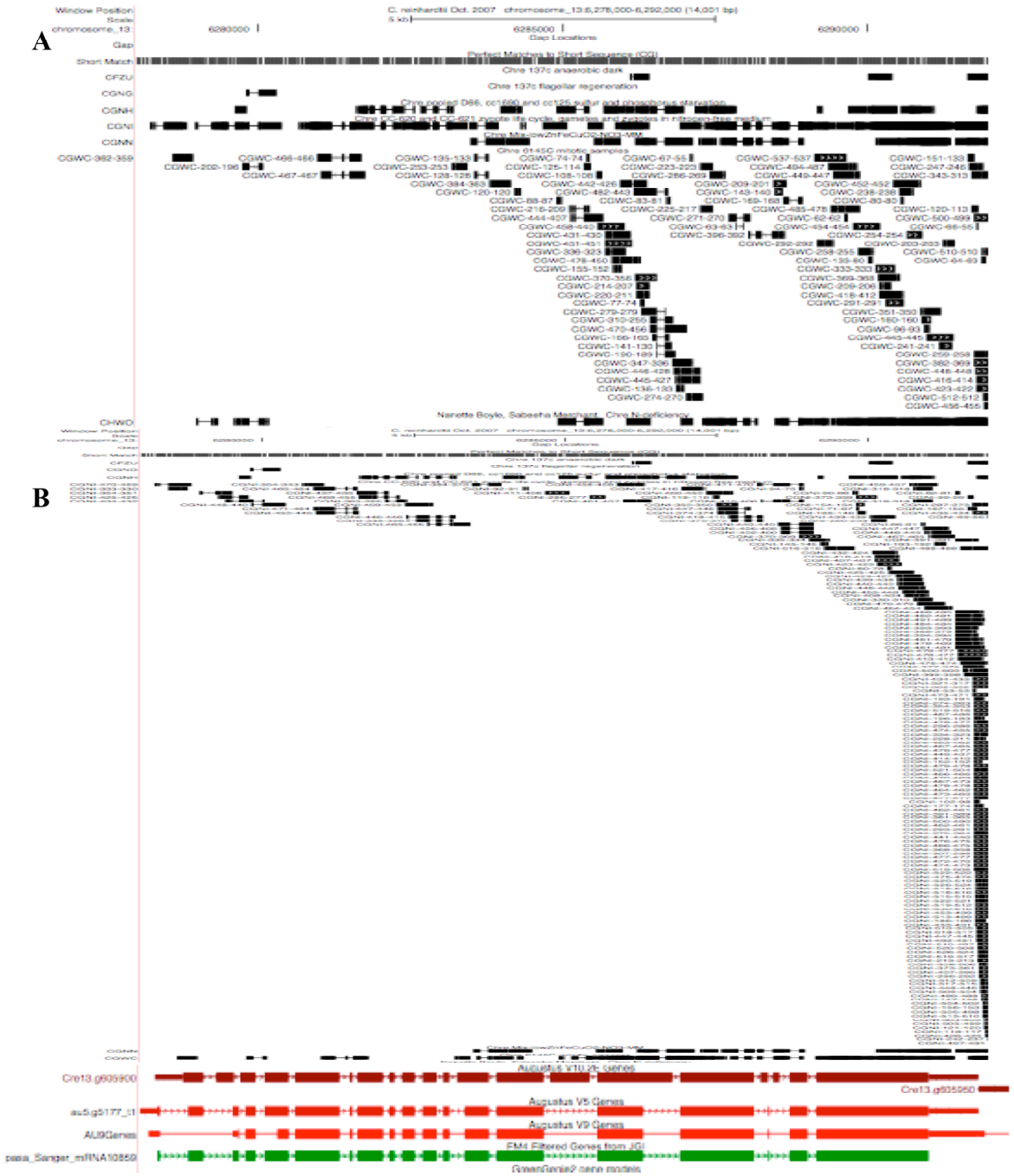


Figure 13. Genomic Analysis: **454 Data.** The 454 whole transcriptome expression data for Cre13.g605900.t1 during (A) vegetative and (B) gametic stages.

3.5.3 RNAseq Data

Although we knew gene Cre13.g605900.t1 was upregulated in the *Chlamydomonas* gametes, we did not know if this upregulation was mating type specific. Since the fusion defective phenotype was expressed only in *mt*$^-$ mutants we hypothesized that if this gene is involved in gamete fusion, it should also be specifically upregulated only in minus gametes. The available EST data cannot distinguish between mating types (section 3.5.1). In collaboration with Dr. James Umen, we gained access to RNAseq (high-throughput transcriptome analysis) data for the genomic region covering the chromosome 13 loci from 6240493 to 6356688. Dr. Umen's analysis was specifically designed to compare expression of the two mating types. We analyzed the expression of Cre13.g605900.t1 and genes in the region and, comparing their expression in vegetative cells and gametes of both mating types. The results of the RNAseq analysis (data not shown) supported our initial hypothesis that the Cre13.g605900.t1 gene is a **minus gamete specific gene** we now called MGS.

3.5.4 qRT-PCR of the MGS *Chlamydomonas* Wild Type

We continued our collaboration with Dr. James Umen and Dr. Peter DeHoff. They confirmed their RNAseq (high-throughput transcriptome data) analysis for the MGS via qRT-PCR for wild type mt$^+$ and mt$^-$. They utilized 18S as an internal control during qRT-PCR. Analysis of their results in the apparent fold changes in MGS expression by the standard curve method (see Materials and Methods) indicated that changes in its expression based on total RNA input to the reverse transcriptase reaction were 2.5 to 5 fold higher in the wild-type mating type minus gametes compared to the mating type plus gametes. When normalized to 18S, MGS mRNA showed a gradual decrease in abundance 10 min, 30 min, 2h and 3h after zygote formation in

wild-type *Chlamydomonas reinhardtii* strains CC-124 mt^- and CC-125 mt^+, (figure 14). The results of qRT-PCR analysis support the RNAseq data indicating that the MGS is a **minus gamete specific gene**.

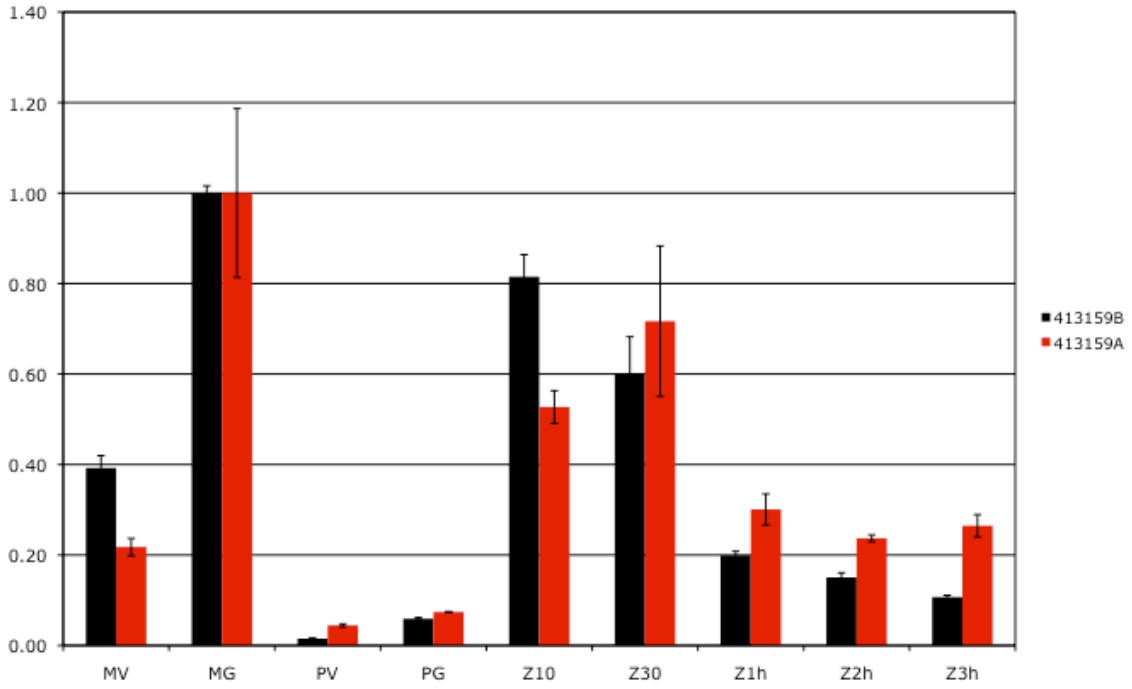


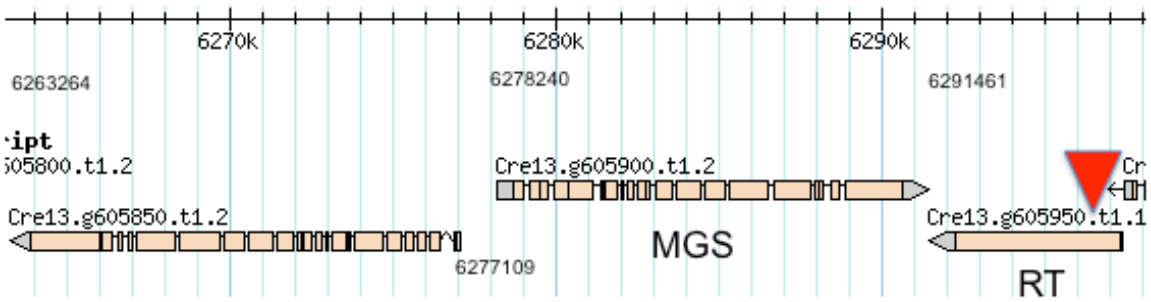
Figure 14. qRT-PCR of MGS in *Chlamydomonas* wild type. Relative abundance, expressed as a fold change, of MGS mRNA from *Chlamydomonas* wild-type cells of mating types minus (M) strain CC-124 and plus (P) strain CC-125 was normalized to 18S and measured by qRT-PCR, during vegetative (V), gametic (G) and zygotic (Z) and post-zygotic stages (10 min, 30 min, 2h and 3h after fusion). The standard deviations of the two repeats (413159B & 413159A), each done in triplicate are shown.

3.6 Molecular Analysis of MGS Gene

The SF-PCR, EST, RNAseq and qRT-PCR data suggested that the gene we call MGS shows all of the characteristics of a gene that might be involved in gamete fusion. Therefore, our rationale was that analysis of the gene in our fusion-defective mutants might allow us to find the site of the

mutations. MGS is found on chromosome 13 between 6278240 - 6291461 on the latest (8) version of Phytosome, JGI 4.3 (figure 15A). The coding sequence is on the positive DNA strand, spanning a 13222 bp. region, with an over 64% GC rich sequence. The cDNA transcript length is 10882 bp. and the CDS sequence is 9639 bp. MGS is flanked with prominent 5′ and 3′ UTR regions and its coding regions consist of 20 exons that are intervened by 19 introns, (figure 15B).

(A)



(B)

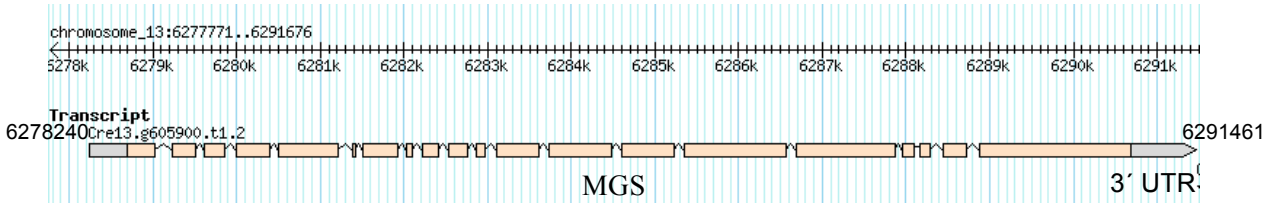


Figure 15. Genomic analysis of MGS. . (A) Genome browser image of the region of the MGS gene on chromosome 13. The inverted triangle shows the insertion site on RT. MGS is to the left of RT and to the right of Cr13.g605850.t1.2 (B) Schematic representation of MGS, including the 5′ and the 3′ UTR regions.

The protein product is expected to be 281 kDa and have 3212 aas., with multiple hydrophobic domains predicted by different algorithms. According to a standard one such as Kyte and Doolittle, there are 16 hydrophobic segments predicted, suggesting the presence of

transmembrane domains, (figure 16). Although others, such as TMM do not predict these transmembrane domains (results not shown).

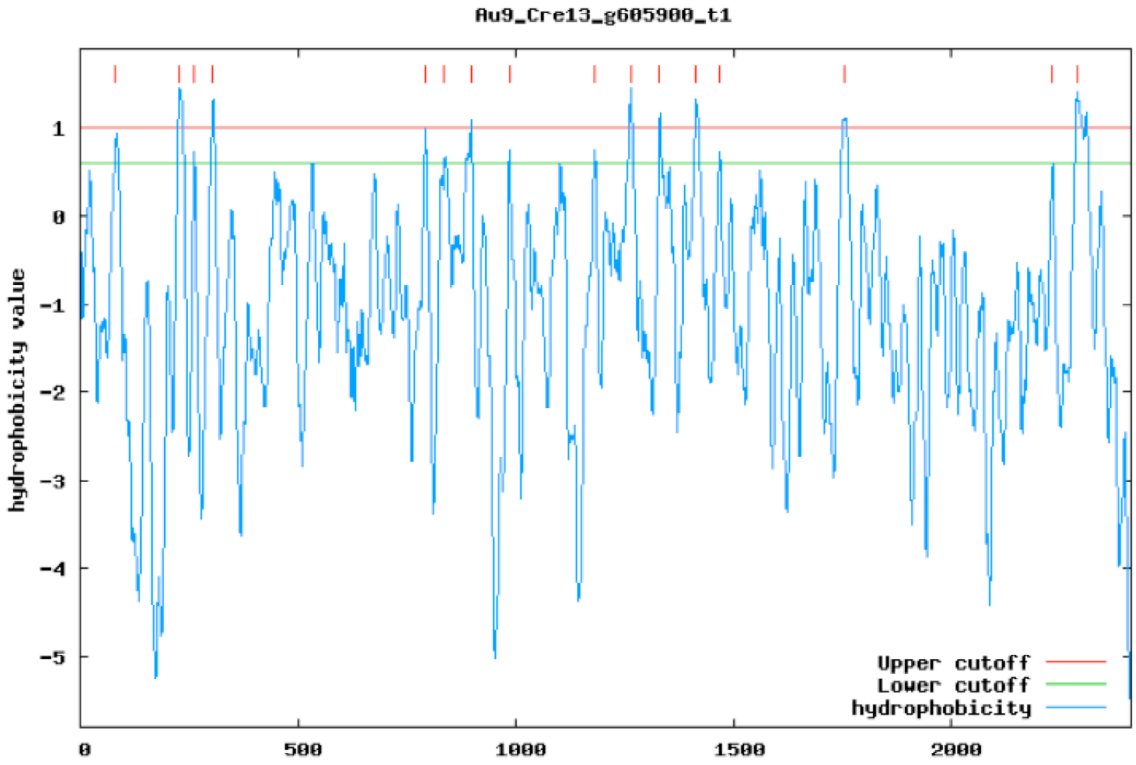


Figure 16. Hydropathy plot. The panel above represents a hydropathy profile of the MGS protein according to the algorithm of Kyte and Doolittle, averaged over a window of seven residues (hydrophobic above the red line).

3.7 Chromosome Walk on MGS

To allow for the molecular analysis of the changes in the MGS gene, its DNA sequence including the promoter, 5′ and the 3′ UTR regions were divided into 18 regions of ~ 1 kb (figure 17). These DNA regions were further subjected to an upstream (away from the point of insertion) chromosome walk via PCR. Region specific forward and reverse primers were designed to amplify them (Table 2). The searches for aberrant DNA sequences bearing mutations

causing the fusion-defective phenotype were conducted within the above-mentioned 18 regions; such mutants would be identified based on the sequencing reactions.

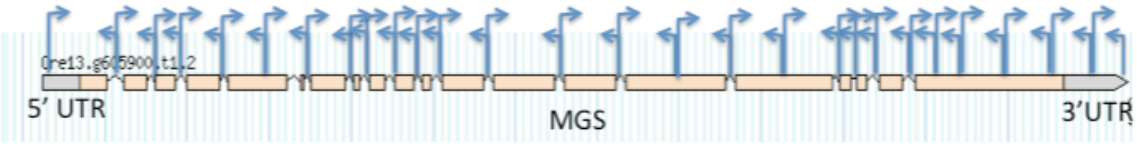


Figure 17. Molecular Analysis of MGS. Schematic representation of MGS. The DNA sequence including the 5´ and the 3´ UTR were divided into the 18 regions of ~ 1 kb.

In addition to the control strain *sru*2 and the insertional fusion-defective mutant cl5, the chromosome walk was also conducted on all of our *Chlamydomonas* fusion-defective mutants that were unable to be complemented with the wild type (*wt*) copy of the HAP2/GCS1 gene (sections 1.5 & 3.1).

3.7.1 PCR Analysis of the 3′ UTR and Exons 19 and 20 of the MGS Gene

The chromosome walk on the 3´ UTR region was 797 bp. in length and the sequence, spanning from 6290679 to 6291475 bp, on chromosome 13 had a 56.5% GC content. The adjacent last exon walk covering from 6288879 to 6290678 bp on chromosome 13 was 1799 bp and had a 68% GC content. In addition to mutant cl5, other insertional as well as conditional mutants were also tested. The intergenic region, last exon and the 3` UTR were divided into four regions of ~ 1 kb and PCR primers were designed to amplify them. Region IV covered the intergenic region and the proximal part of the 3` UTR. Region III encompassed the distal part of the 3` UTR and spanned the border of the last exon (20) and the 3` UTR; region I and II covered the remaining part of the last exon (20), the adjacent intron and exon 19 (figure 18).

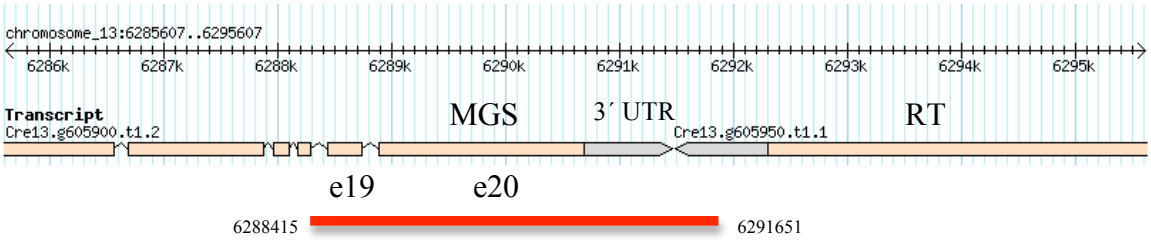


Figure 18. Schematic representation of the amplified 3′ UTR region and adjacent exons 20 and 19. Genome browser image of the MGS region on chromosome 13. MGS is to the left of the RT. The bar under MGS indicates the amplified region of interest.

A PCR reaction with primers fCAR4and rCAR4, covering the intergenic region and the distal part of the 3′ UTR region IV gave the expected 880 bp product (figure 19). The reaction with primers R3F47 and R3r31, covering part of the last exon region III and part of the proximal part of the 3′ UTR region IV gave the expected product of 599 bp. A PCR reaction with primer R3F47 along with R3r48 gave the expected 746 bp product for the distal part of the MGS 3′ UTR regions III and IV as well (figure 20). A PCR reaction with primer pair fCar1 and rCar1 gave the 1008 bp expected product and PCR with primer fCar2 and rCar2 gave the 1217 bp expected product as well (figure 21). These results indicated that all tested mutants including cl5 did not contain any visible insertion or deletion in the 3′ UTR region or the last exon of MGS. All visible DNA fragments from those four regions were gel purified and further subjected to Sanger sequencing at Genewiz, enabling us to deduce their DNA coding sequences and further check for their integrity.

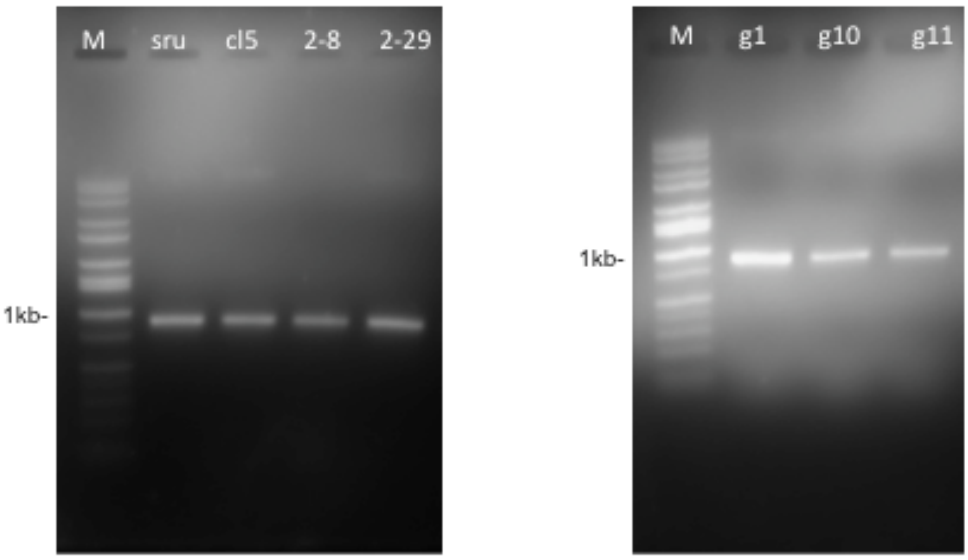


Figure 19. PCR of MGS region IV. PCR of the MGS gene region IV covering the intergenic region and the distal part of the 3´ UTR. PCR with primers fCAR4 and rCAR4 gave the expected 880 bp (M is a DNA marker).

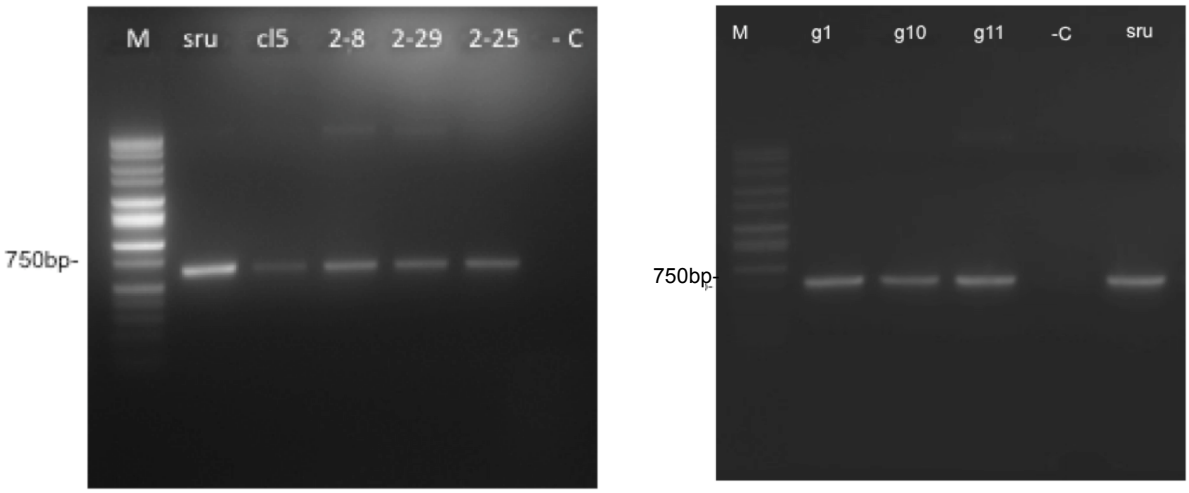


Figure 20. PCR of MGS regions III and IV. PCR of the MGS gene region III and IV covering the distal part of the 3´ UTR and part of the last exon. PCR with primers R3F47 and R3r48 gave the expected 746 bp band. Lane (–C) is a negative control with no gDNA added. (M is a DNA marker).

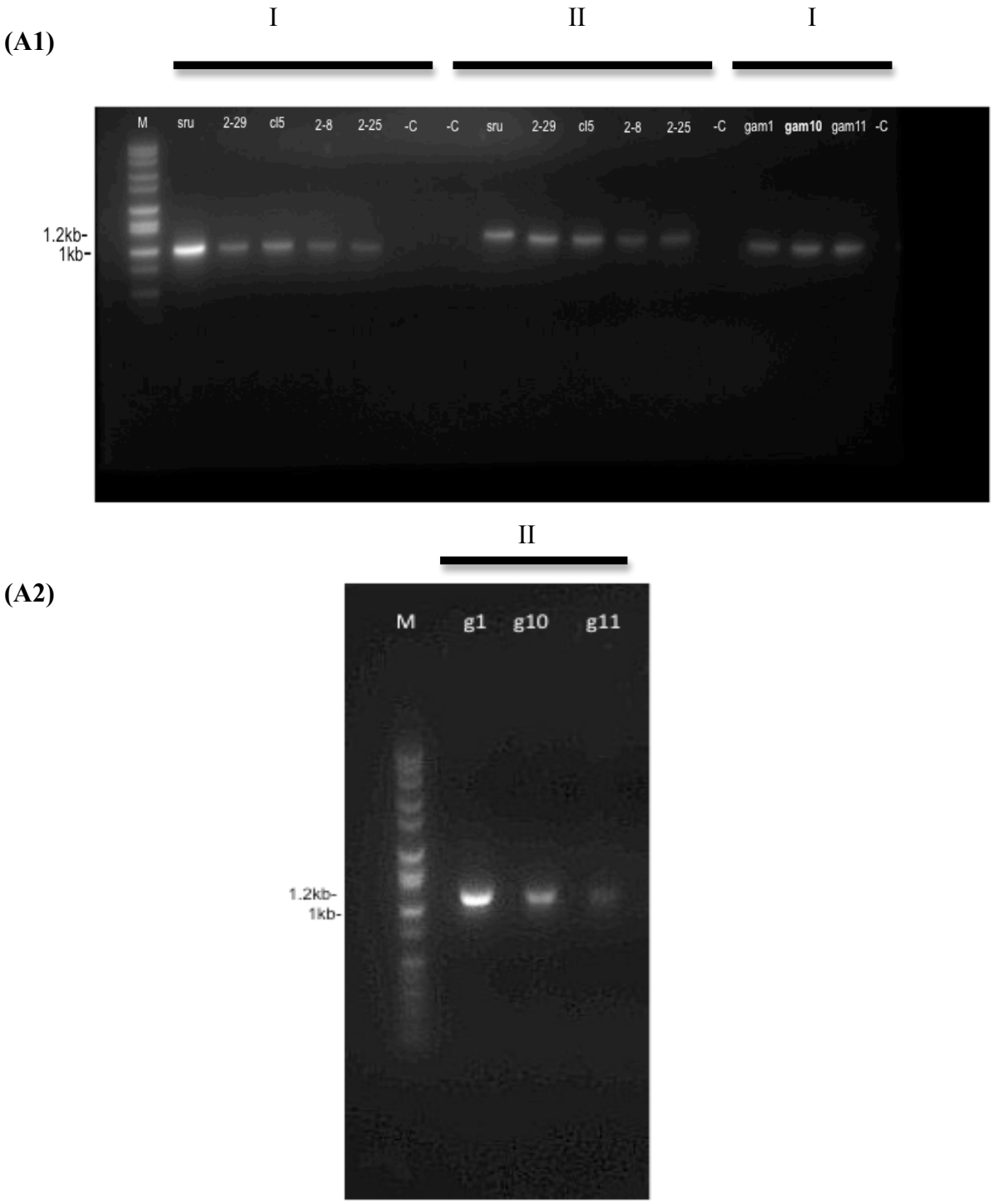


Figure 21. PCR of MGS regions I and II. PCR of the exons 19 and 20 (regions I and II) of the control strain *sru2* as well as the insertional and conditional fusion-defective mutants (including cl5). PCR with fCar1 and rCar1 (region I) gave the 1008 bp expected product (left and right gel (A1)) and PCR with fCar2 and rCar2 (region II) gave the 1217 bp expected product as well (middle of the gel (A1) and gel (A2)). (–C) is a negative control with no gDNA added. (M is a DNA marker).

3.7.1.1 Sequencing of the 3′ UTR and Exons 19 and 20 of the MGS Gene

The sequencing results we received from the Genewiz were blasted into the *Chlamydomonas* genome and aligned via the multiple sequence alignment algorithm ClustalW, (figure 22). Consensus sequences were produced for each alignment by aligning forward and reverse reads received from sequencing facility.

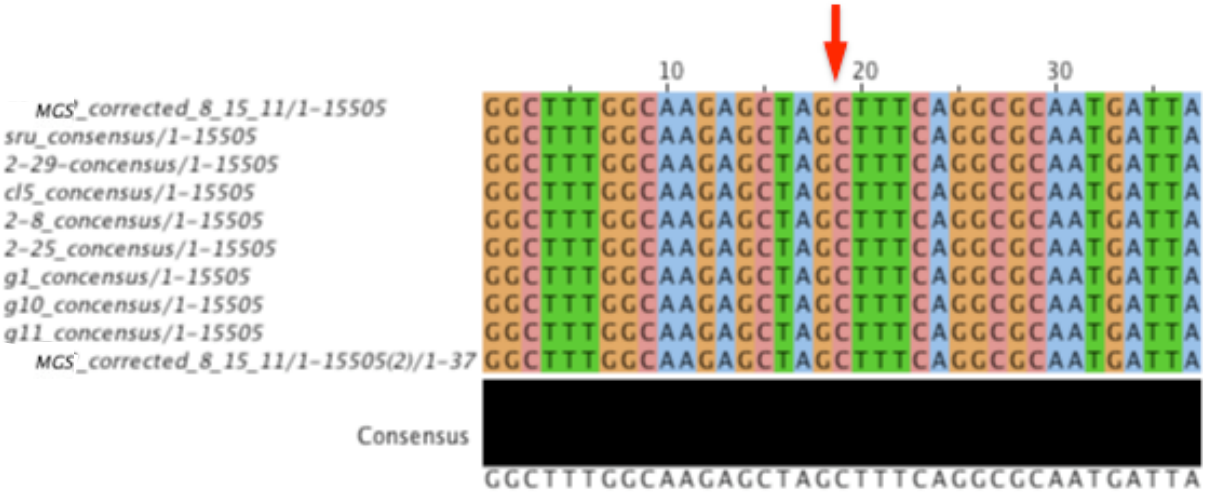


Figure 22. A portion of the sequence alignments for the 3′ UTR region and adjacent exon 20. The sequenced MGS gene regions were aligned via ClustalW algorithm. The red arrow indicates the border between the last exon (20, (to the left of the arrow) and the 3′ UTR region of MGS (to the right of the arrow).

The sequence alignment results from all of the mutants showed that the MGS region spanning 6288879 and 6290679, that cover the entire 3′ UTR region and last exon, is not mutated in any of the mutants analyzed including cl5.

3.7.2 PCR Analysis of Exons 18, 17 and Part of Exon 16 of the MGS Gene

The chromosome walk on exons 18, 17 and part of exon 16 was 2357 bp in length and the sequence, from 6287273 to 6289629 bp., on chromosome 13 had a 64.4% GC content (figure 23).

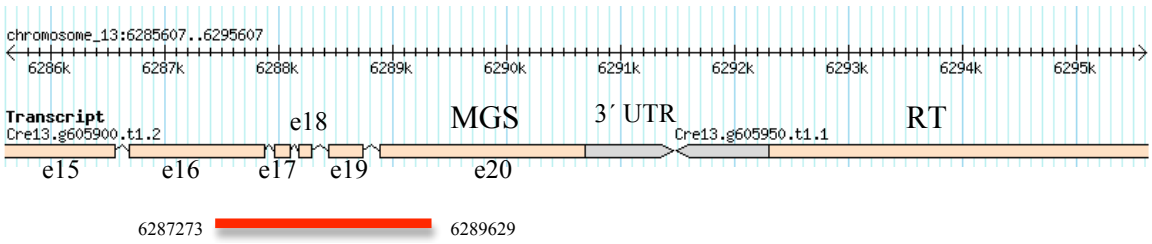


Figure 23. Schematic representation of amplified exons 18, 17 and part of the exon 16. Genome browser image of this MGS region on chromosome 13. MGS is to the left of the RT. The bar under MGS indicates the amplified region of interest.

A PCR reaction with primers fCAR18_6N and R1r18, covering the exons 18, 17 and part of the exon 16 gave the expected 2357 bp product (figure 24).

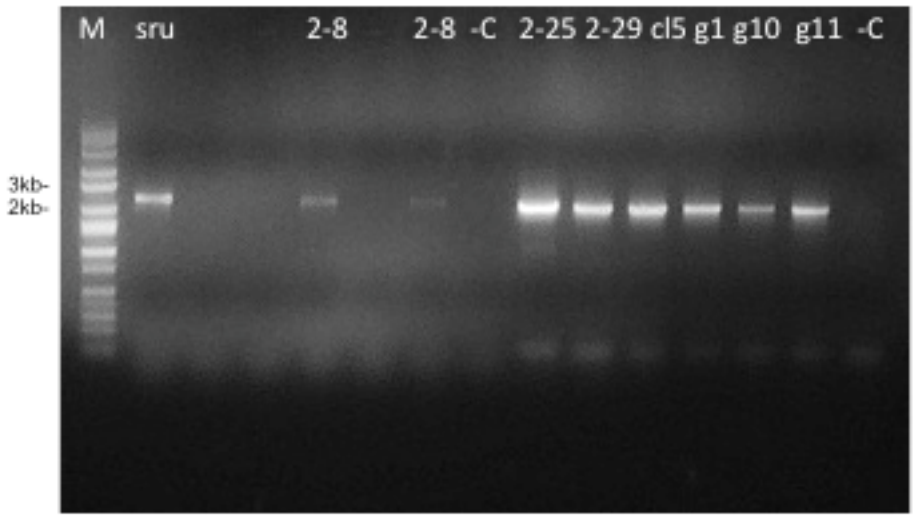


Figure 24. Chromosome walking on MGS Exons 18, 17 and part of 16 via PCR. PCR of the exons 18, 17 and part of exon 16 for the control strain *sru2* as well as the insertional and conditional fusion-defective mutants (including cl5). PCR with primers fCAR18_6N and R1r18 gave a 2357 bp expected product. (–C) is a negative control with no gDNA added. (M is a DNA marker).

3.7.2.1 Sequencing of Exons 18, 17 and Part of Exon 16 of the MGS Gene

The sequencing results we received from the Genewiz were blasted into the *Chlamydomonas* genome and aligned via the multiple sequence alignment algorithm ClustalW, (figure 25).

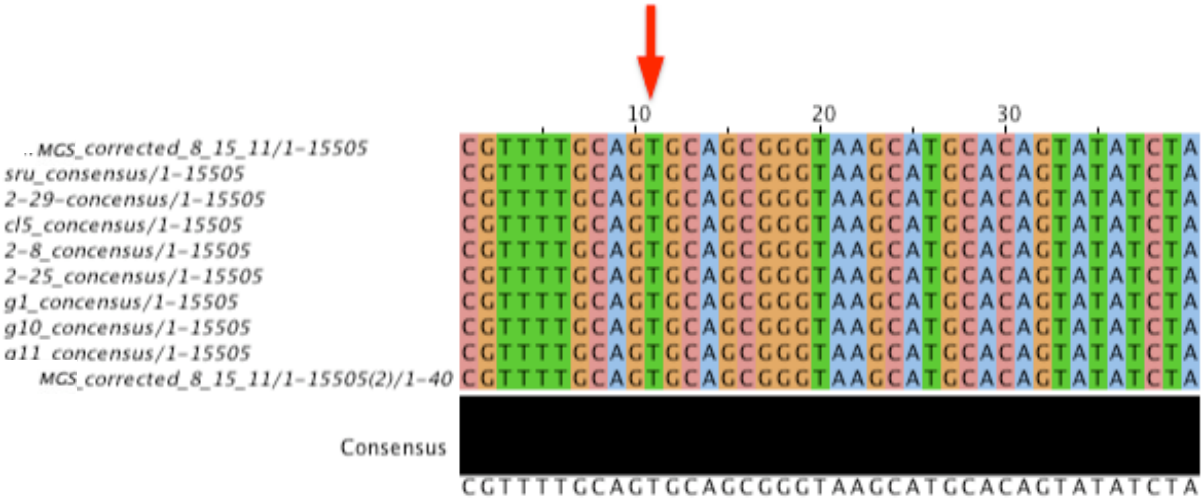


Figure 25. A portion of the sequence alignment for the exon 18. The sequenced MGS gene regions were aligned via ClustalW algorithm. The red arrow indicates border between the intron (to the left of the arrow) and the exon 18 region of MGS (to the right of an arrow).

The sequence alignment results from all of the mutants showed that the MGS region spanning between 6287273 and 6289629, that covers the entire 3´ UTR region and the last exon, is not mutated in any of the mutants analyzed, including cl5.

3.7.3 PCR Analysis of Exon 16 of the MGS gene

The chromosome walk on exon 16 was 2211 bp. in length and the sequence, spanning 6286124 to 6288334 bp. on chromosome 13 had a 64.9% GC content (figure 26).

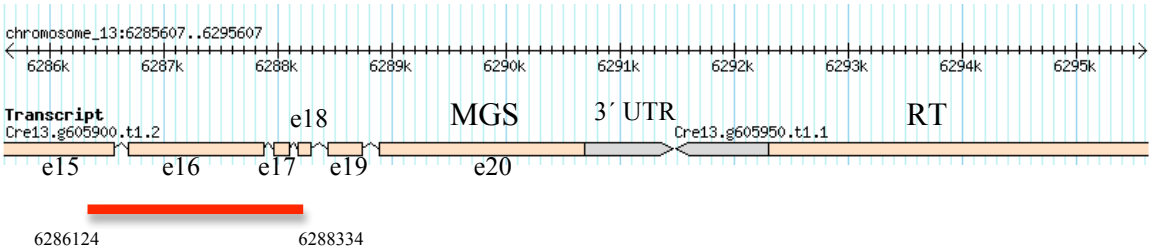


Figure 26. Schematic representation of amplified exon 16. Genome browser image of the MGS region on chromosome 13. MGS to the left of the RT. Bar under MGS indicates the amplified region of interest.

A PCR reaction with primers F34 and r31, covering exons 16 gave the expected product, 2211 bp (figure 27).

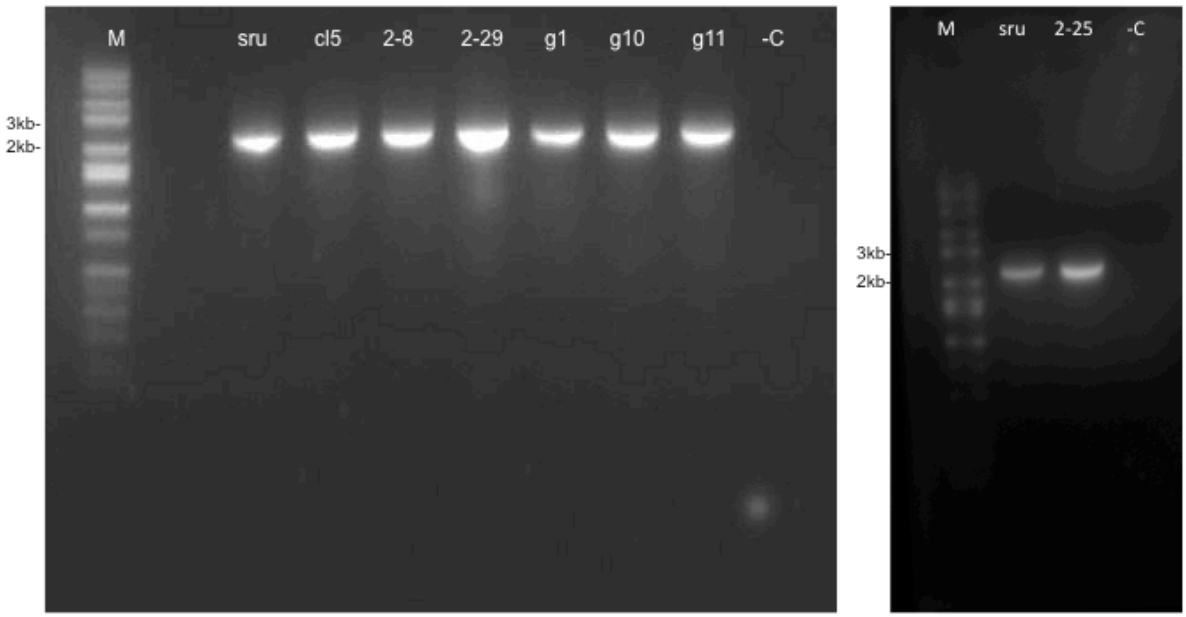


Figure 27. Chromosome walking on MGS Exon 16 via PCR. PCR of exon 16 of the control strain *sru2* as well as the insertional and conditional fusion-defective mutants (including cl5). PCR with F34 and r31 gave a 2211 bp expected product. (–C) is a negative control with no gDNA added. (M is a DNA marker).

3.7.3.1 Sequencing of Exon 16 of the MGS Gene

The sequencing results we received from the Genewiz were blasted into the *Chlamydomonas* genome and aligned via the multiple sequence alignment algorithm ClustalW, (figure 28).

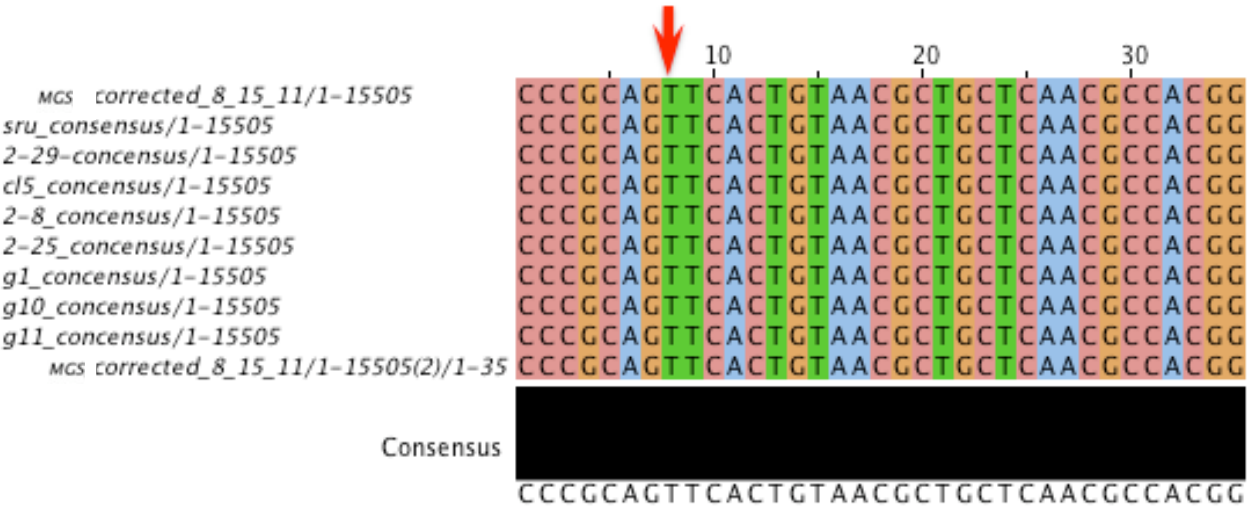


Figure 28. A portion of the sequence alignment for the exon 16. The sequenced MGS gene regions were aligned via ClustalW algorithm. The red arrow indicates the border between the intron (to the left of the arrow) and the exon 16 region of MGS (to the right of an arrow).

The sequence alignment results from all of the mutants showed that the MGS region spanning between 6286124 and 6288334, that covers the exon 16 region, is not mutated in any of the mutants analyzed, including cl5.

3.7.4 PCR Analysis of the Exons 15, 14, 13 and Part of Exon 12 of the MGS Gene

The chromosome walk on the exons 15, 14, 13 and part of exon 12 was 3864 bp in length and the sequence, spanning from 6283239 to 6287102 bp, on chromosome 13 had a 67.3% GC

content (figure 29).

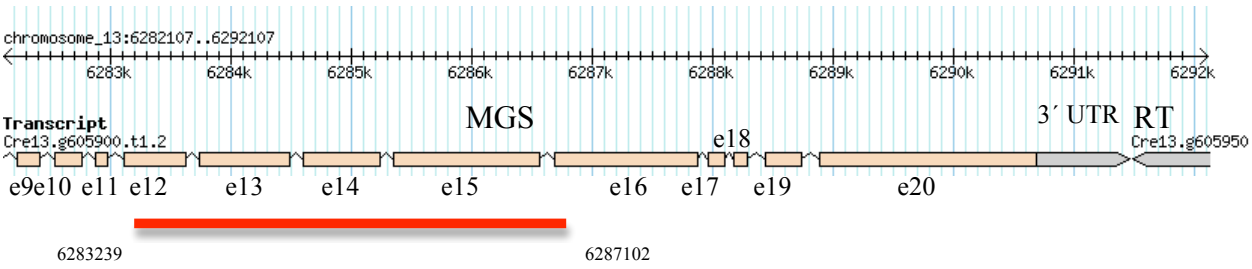


Figure 29. Schematic representation of the amplified exons 15, 14, 13 and part of 12. Genome browser image of the MGS region on chromosome 13. MGS to the left of the RT. The bar under MGS indicates the amplified region of interest.

A PCR reaction with primers e13F6 and r11, covering exons 15, 14, 13 and part of exon 12 gave the expected product of 3864 bp (figure 30).

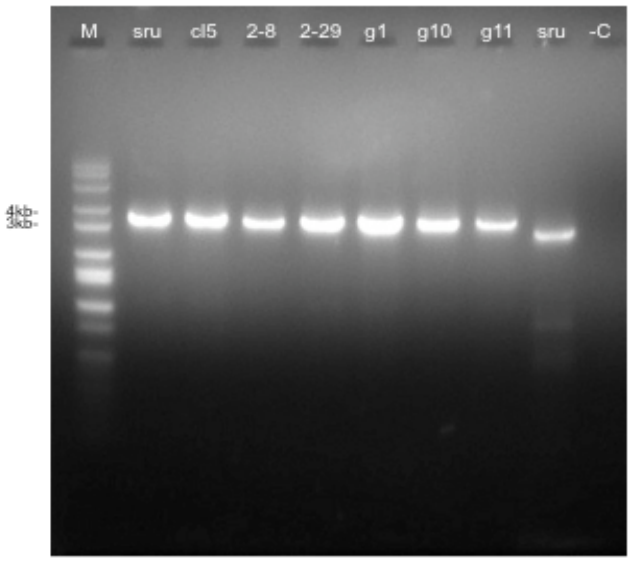


Figure 30. Chromosome walking on MGS exons 15, 14, 13 and part of 12 via PCR. PCR of exons 15, 14, 13 and part of 12 of the control strain *sru2* as well as the insertional and conditional fusion-defective mutants (including cl5). PCR with primers e13F6 and r11gave a 3864 bp expected product. (–C) is a negative control with no gDNA added. (M is a DNA marker).

3.7.4.1 Sequencing of the Exons 15, 14, 13 and Part of Exon 12 of the MGS Gene

The sequencing results we received from the Genewiz were blasted into the *Chlamydomonas* genome and aligned via the multiple sequence alignment algorithm ClustalW, (figure 31).

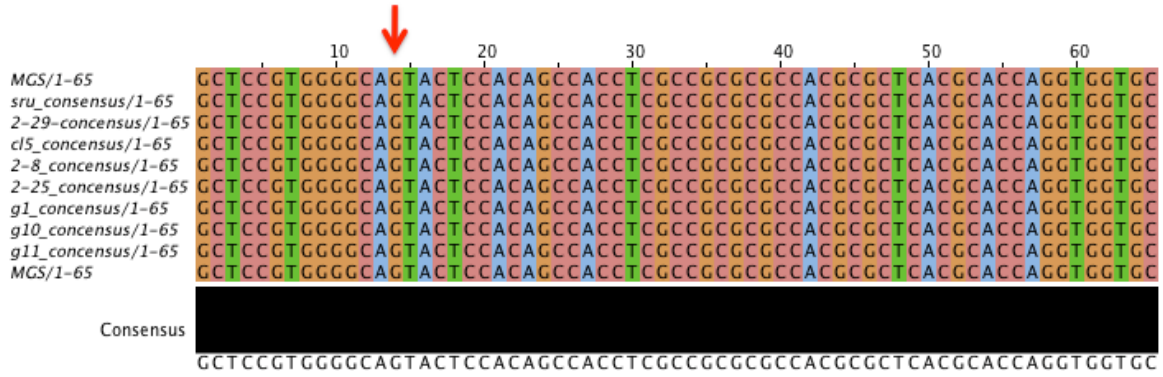


Figure 31. A portion of the sequence alignment covering exons 15, 14, 13 and part of the exon 12. The sequenced MGS gene regions were aligned via ClustalW algorithm. The red arrow indicates border between the intron (to the left of the arrow) and the exon 12 region of the MGS (to the right of an arrow).

The sequence alignment results from all of the mutants showed that the MGS region spanning 6283239 to 6287102 bp, that covers exons 15, 14, 13 and part of 12, is not mutated in any of the mutants analyzed, including cl5.

3.7.5 PCR Analysis of the Exons 5, 6, 7, 8, 9, 10, 11 and part of the exon 12 of the MGS

The chromosome walk on the exons 5, 6, 7, 8, 9, 10, 11 and part of the exon 12 was 2995 bp in length and the sequence, spanning 6280359 to 6283454 bp, on chromosome 13 had a 64.3% GC content.

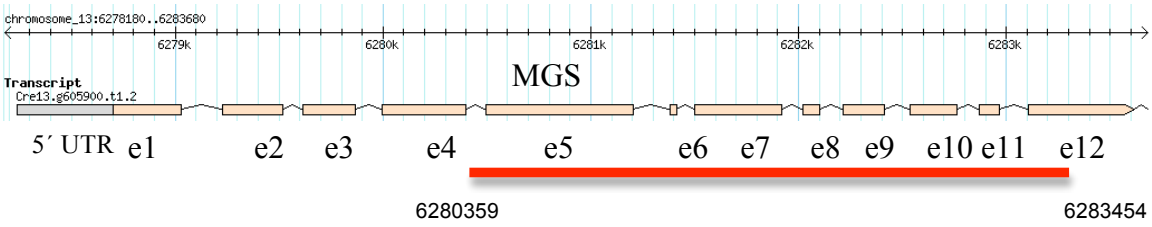


Figure 32. Schematic representation of the amplified exons 5, 6, 7, 8, 9, 10, 11 and part of 12. Genome browser image of the MGS region on chromosome 13. The bar under MGS indicates the amplified region of interest.

A PCR reaction with primers e9_F1 and e12_r1, covering the exons 5, 6, 7, 8, 9, 10, 11 and part of 12 gave the expected product of 2995 bp (figure 33).

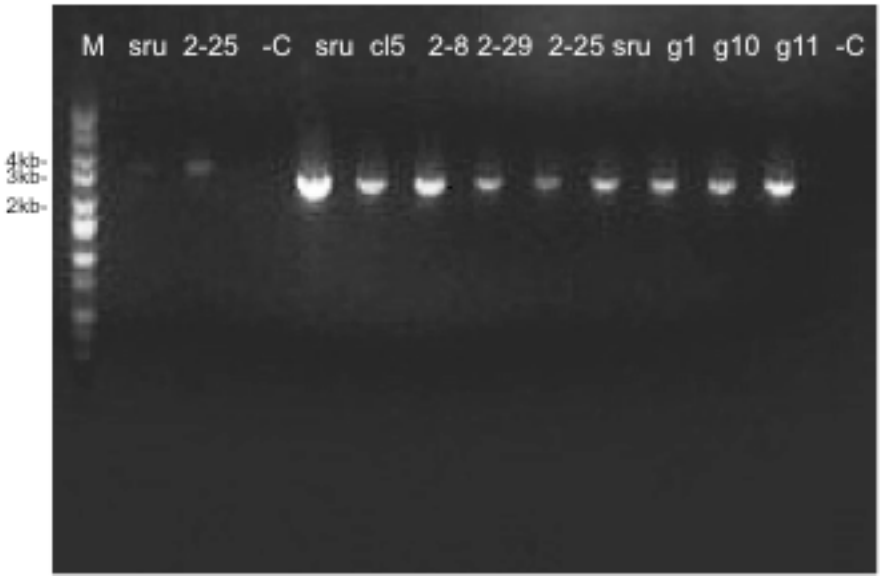


Figure 33. Chromosome walking on MGS exons 5, 6, 7, 8, 9, 10, 11 and part of 12 via PCR. PCR of exons 5, 6, 7, 8, 9, 10, 11 and part of 12 of the control strain *sru2* as well as the insertional and conditional fusion-defective mutants (including cl5). PCR with primers e9_F1 and e12_r1 gave a 2995 bp expected product. (–C) is a negative control with no gDNA added. (M is a DNA marker).

3.7.5.1 Sequencing of the Exons 5, 6, 7, 8, 9, 10, 11 and part of the exon 12 of the MGS

The sequencing results we received from the Genewiz were blasted into the *Chlamydomonas* genome and aligned via the multiple sequence alignment algorithm ClustalW, (figure 34).

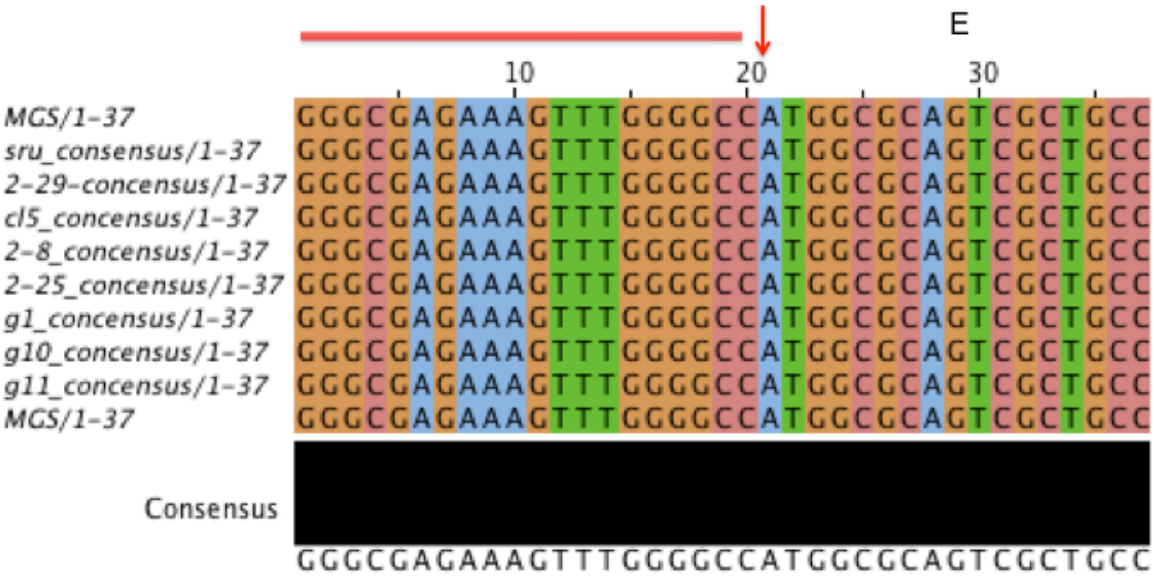


Figure 34. A portion of the sequence alignments for the exons 5, 6, 7, 8, 9, 10, 11 and part of 12. The sequenced MGS gene regions were aligned via ClustalW algorithm. The red arrow indicates border between the intron (to the left of the arrow) and the exon 7 region of the MGS (to the right of an arrow).

The sequence alignment results from all of the mutants showed that the MGS region spanning between 6280359 and 6283454, that cover the entire 3´ UTR region and last exon, is not mutated in any of the mutants analyzed including cl5.

3.7.6 PCR Analysis of the Promoter, 5´ UTR Region and Exons 1, 2, 3 & 4 of the MGS Gene

The chromosome walk on the promoter was 962 bp. and had a 55.8% GC content covering from 6277009 to 6277970 bp on chromosome 13. The adjacent 5´ UTR region was 323 bp in length and had a 56.1% GC content, spanning from 6277971 to 6278294 bp. position on chromosome 13. The walk on the adjacent exons 1, 2, 3, and 4 with their respective introns was 2246 bp from 6278295 to 6280540 bp on chromosome 13 had a 62.1% GC content (figure 35). In addition to mutant cl5, some other insertional as well as conditional mutants were also tested.

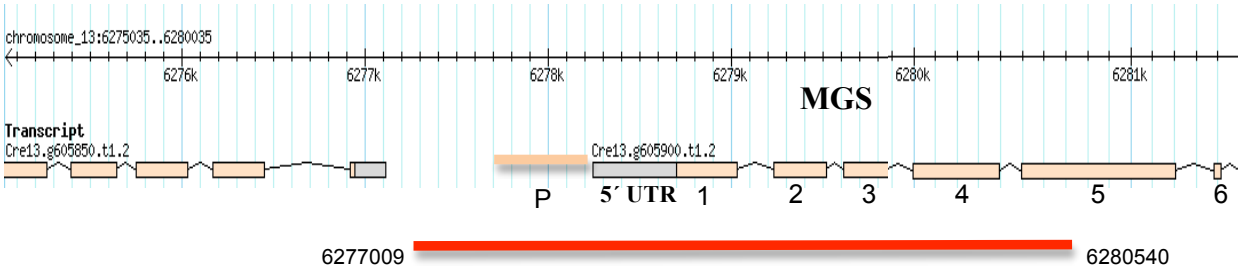


Figure 35. Schematic representation of the amplified promoter, 5′ UTR region and exons 1, 2, 3 and 4 via PCR. Genome browser image of the promoter (P) region, 5´ UTR, exons 1, 2, 3 and 4 along with the adjacent introns. The bar in front of the MGS represents promoter region, the bar below indicates the region of interest.

A PCR reaction with primers F0_60 and e4r13, covering promoter, the 5´ UTR region and exons 1, 2, 3 and 4 along with their respective introns gave the expected product of 3532 bp (figure 36). The analysis of these results indicated that all tested mutants including cl5 did not contain any visible insertion or deletion in the promoter, 5´ UTR region as well as exons 1, 2, 3 and 4 along with their respective introns. All visible DNA fragments from these regions were gel purified

and further subjected to Sanger sequencing at Genewiz, enabling us to deduce their DNA coding sequences and further check for their integrity.

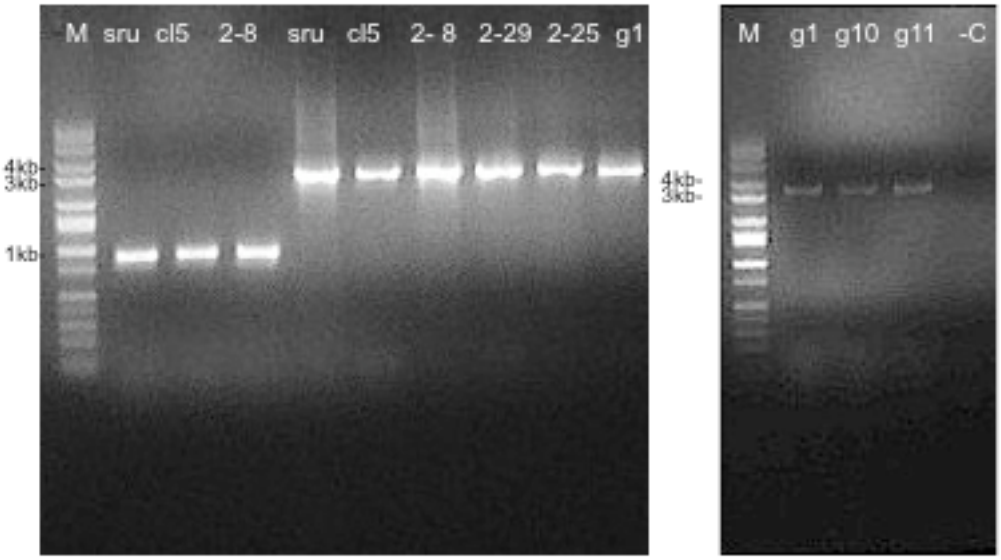


Figure 36. Chromosome walking on MGS via PCR. PCR of the promoter, 5´ UTR region, and exons 1, 2, 3 and 4 of the control strain *sru2-2* as well as the insertional fusion-defective mutants (including cl5). PCR with e1F0 and e2_r34 gave a 802 bp expected product (left) and PCR with F0_60 and e4r13 gave the 3532 bp expected product as well (middle). (–C) is a negative control with no gDNA added. (M is a DNA marker).

3.7.6.1 Sequencing of the Promoter, 5´ UTR Region and Exons 1, 2, 3 & 4 of the MGS Gene

The sequencing results we received from the Genewiz were blasted into the *Chlamydomonas* genome and aligned via the multiple sequence alignment algorithm ClustalW, (figure 37).

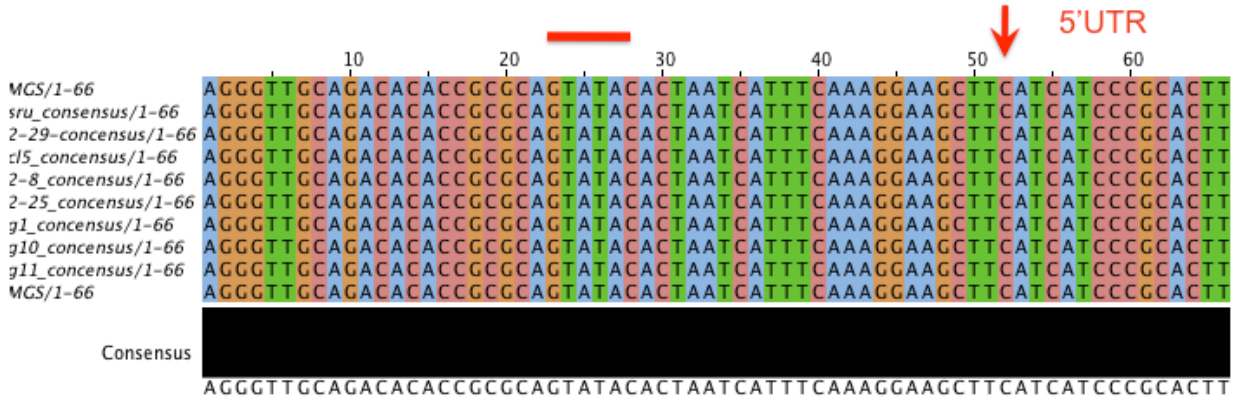


Figure 37. Partial sequence alignment for MGS promoter, 5´ UTR region and exons 1, 2, 3 & 4 . Sequences were aligned via ClustalW algorithm. The bar above indicates the TATA box, arrow indicates the border between promoter and 5´ UTR region.

The sequence alignment results from all of the mutants showed that the MGS region spanning between 6277009 and 6280540, that cover the entire promoter, 5´ UTR region, and exons 1, 2, 3 and 4, is not mutated in any of the mutants analyzed, including cl5.

3.8 qRT-PCR of MGS in Fusion-Defective Mutants

We performed qRT-PCR to determine whether the pSP124S vector insertion in the *copia*-family RT resulted in an increase or decrease in the level of MGS expression in the insertional fusion-defective mutant cl5. In addition to the control strain *sru*2 and mutant cl5, MGS expression was also tested in the *Chlamydomonas* fusion-defective mutants that were unable to be complemented with the wild type copy of the HAP2/GCS1 gene (see sections 1.5 & 3.1). Genome specific primers were designed that were specific to two target sites in exon 20. For

target 1 (T-1): Primers F_1 and r_1 should produce a 160 bp product, amplifying the cDNA region spanning 6289691and 6289837 bp, (64.8% GC) (figure 38). For target 2 (T-2): Primers F_2 and r_2 (see section 4.2) should produce a 239 bp product, amplifying the cDNA region spanning 6289392 and 6289629 bp, (71.5% GC) (figure 38). We utilized target 1 (T-1) to analyze the expression of MGS, since it was used during our collaborative analysis allowing us to identify MGS as a minus gamete specific gene (see section 3.5.4).

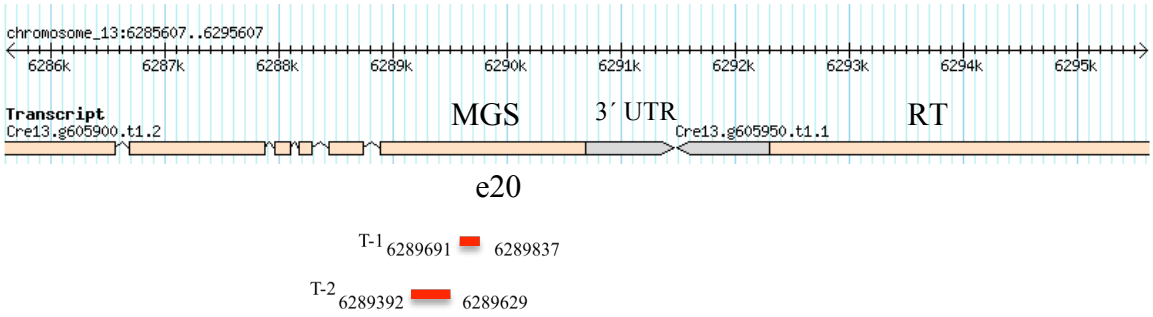


Figure 38. **Genome browser image of MGS showing qRT-PCR target regions**. Schematic representation of the two target regions (T-1) and (T-2) in exon 20. RT is to the right of MGS. The two bars under MGS indicate the two amplified regions of interest.

We utilized 18S as an internal control during qRT-PCR. Primers 18S_F and 18S_r (Section 4.2) should produce a 116 bp product with 45.6% GC.

3.8.1 qRT-PCR of MGS in the Insertional Fusion-Defective Mutant cl5

Processed RNA was checked for its integrity via quality control gels prior to reverse transcriptase reaction (figure 39).

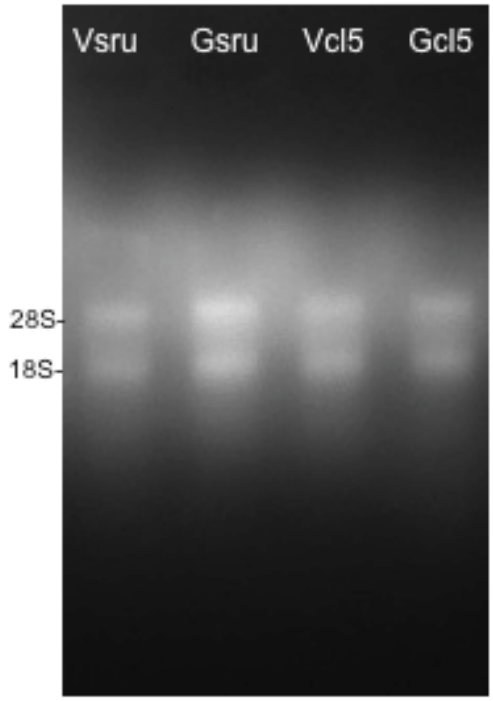


Figure 39. RNA Quality Control gel. RNA quality control gel indicating intactness of the RNA as the starting material for reverse transcriptase reaction. Vegetative (V) and gametic (G) RNA samples of the control *sru*2 and insertional fusion-defective mutant cl5. 28S and 18S are rRNA bands.

We analyzed the results in the apparent fold changes in MGS expression by the standard curve method; this analysis, based on total RNA input to the reverse transcriptase reaction, indicated no differences in MGS expression in the vegetative and gametic samples when comparing the control strain *sru* with all of the tested insertional fusion-defective mutants, including cl5 (figure 40).

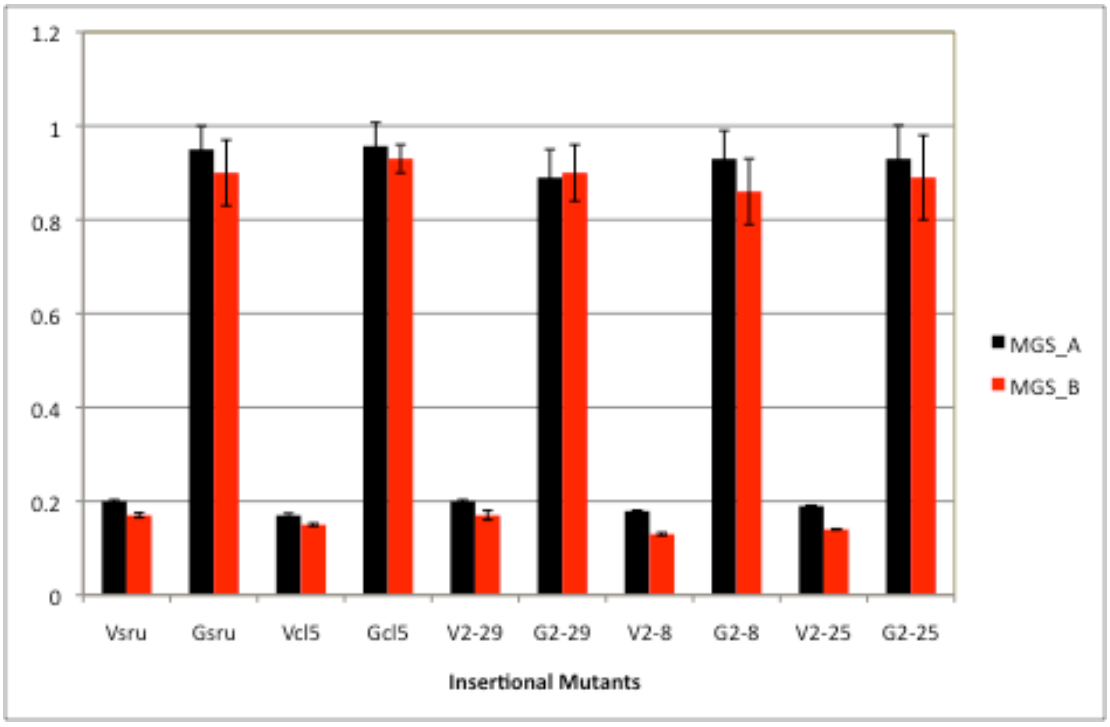


Figure 40. qRT-PCR of the MGS gene in *Chlamydomonas* Insertional Fusion-Defective Mutants. Relative abundance, expressed as a fold change, of MGS mRNA from *Chlamydomonas* control *sru*2 cells and insertional fusion-defective mutants including cl5 were normalized to 18S and measured by qRT-PCR, during vegetative (V) and gametic (G) stages. The standard deviations of the two repeats (MGS_A & MGS_B), each done in triplicate are shown. Primers F-1 and r_1 were used.

The analysis of these results indicate that the insertion in the *copia*-family RT does not affect the level of MGS expression in the insertional any of the fusion-defective mutants tested, including cl5. Also, the analysis of this result supports previous qRT-PCR and RNAseq data indicating that the MGS is a minus gamete specific gene (see section 3.5). Figure 41 shows the expected qRT-PCR products for the 18S rRNA, MGS targets T-1 and T-2 from the control *sru2* strain and

mutant cl5. All bands from the qRT-PCR reactions were gel extracted, gel purified, and sent for sequencing at Genewiz (see section 3.4.4). Again we can see the difference in expression of MGS in gametes and vegetative cells.

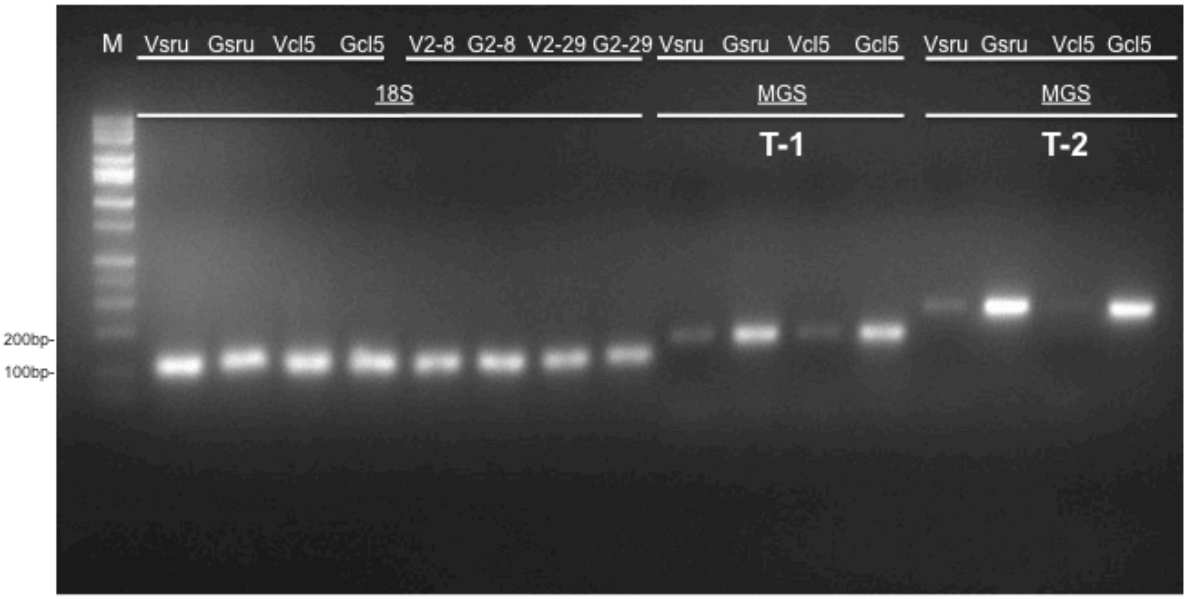


Figure 41. qRT-PCR of MGS Expression of MGS in *Chlamydomonas* Insertional Fusion-Defective Mutant cl5 . cDNA from the vegetative (V) and gametic (G) samples of the control *sru*2 and insertional fusion-defective mutants, including cl5, show the expected product for 18S 109 bp, target 1 (T-1) 160 bp and target 2 (T-2) 239 bp.

3.8.2 qRT-PCR of MGS in the Conditional Fusion-Defective Mutants

Processed RNA was checked for its integrity via quality control gels prior to reverse transcriptase experiments (figure 42).

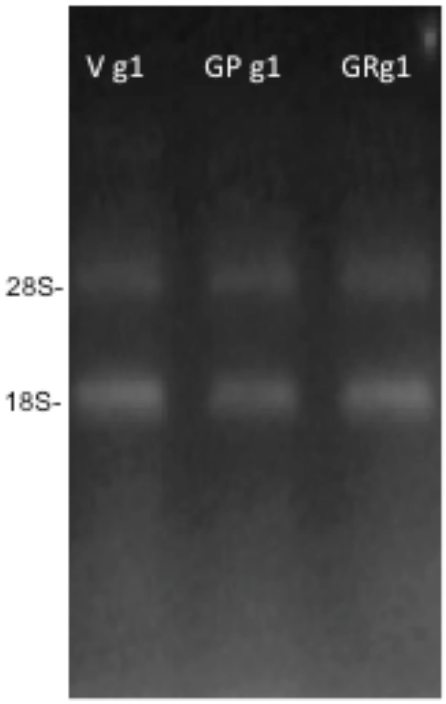


Figure 42. RNA Quality Control gel . RNA quality control gel indicating intactness of the RNA as the starting material for reverse transcriptase reactions. Vegetative (V) and gametic (G) RNA samples from conditional fusion-defective mutant *gam*1 (g1) at permissive (P) and restrictive (R) temperature conditions. 28S and 18S are rRNA bands.

We analyzed the results in the apparent fold changes in MGS expression by the standard curve method; these analyses based on total RNA input to the reverse transcriptase reaction, indicated no differences in MGS expression in the vegetative and gametic samples when comparing the control strain *sru* with all of the tested conditional fusion-defective mutants under both restrictive and permissive temperature conditions (figure 43).

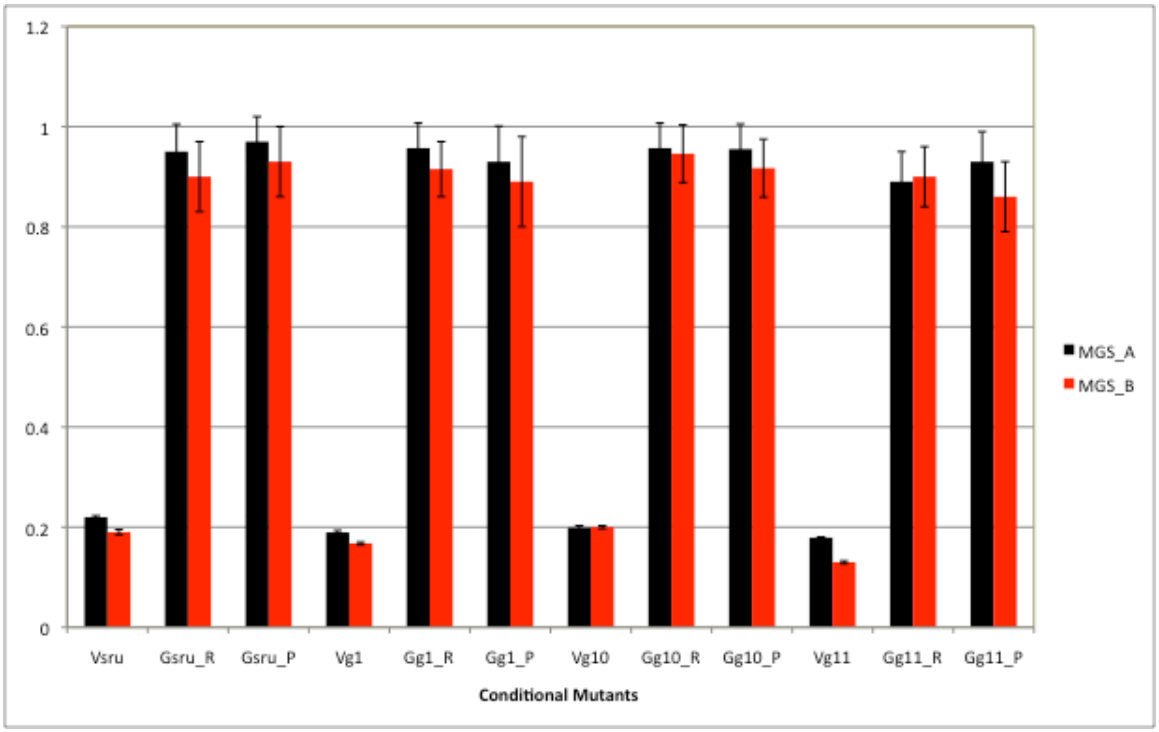


Figure 43. qRT-PCR of MGS in *Chlamydomonas* Conditional Fusion-Defective Mutants. Relative abundance, expressed as a fold change, of MGS mRNA from control *sru* (*sru*2) cells and conditional fusion-defective mutants g1 (*gam*1), g10 (*gam*10), g11 (*gam*11) was normalized to 18S and measured by qRT-PCR, during vegetative (V) and gametic (G) stages at restrictive 32 C° (R) and permissive 25 C° (P) temperatures. The standard deviations of the two repeats (MGS_A & MGS_B), each done in triplicate is shown.

The analysis of this result indicates that the level of MGS expression is not affected in any of the tested conditional fusion-defective mutants. Figure 44 shows the expected qRT-PCR products for the 18S rRNA, MGS targets T-1 and T-2 from the mutant *gam*1. All bands from the qRT-PCR reactions were gel extracted, gel purified, and sent for sequencing at Genewiz (see section 3.4.4).

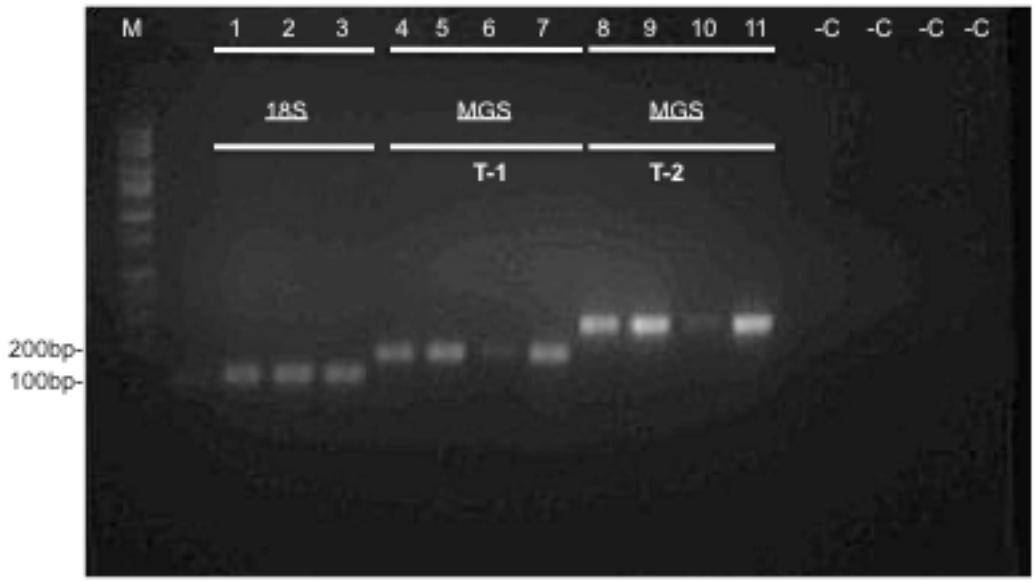


Figure 44. qRT-PCR of MGS Expression in Conditional Fusion-Defective Mutants. Lanes 1-3, primers 18S_F and 18S_r were used to amplify the message for 18S for cDNA from the vegetative (V) and gametic (G) samples of the conditional fusion-defective mutant *gam*1 under restrictive (Lane 2) and permissive (Lane 3) conditions. The expected products of 109 bp are seen for 18S. Lanes 4-7, primers F_1 and r_1 were used to amplify the message for the MGS gene target 1 (T-1) for cDNA from the mutant *gam*1 vegetative (V) (Lane 6) and gametic (G) samples under restrictive (Lanes 4 & 5) and permissive conditions (Lane 7). The expected products of 160 bp are seen for T-1. Lanes 8-11, primers F_2 and r_2 were used to amplify the message for the MGS gene target 2 (T-2) for cDNA from the vegetative (V) (Lane 10) and gametic (G) samples of the mutant *gam*1 under restrictive (Lanes 8 & 9) and permissive conditions (Lane 11). The expected products of 239 bp are seen for T-2. (-C is the negative control with no cDNA added).

We did not observe any signal amplification in the reactions originating from reverse transcriptase negative controls, therefore any gDNA contamination was excluded.

3.8.3 Sequencing Results for the qRT-PCR

All bands from the qRT-PCR reactions were gel extracted, gel purified, and sent for sequencing at Genewiz.

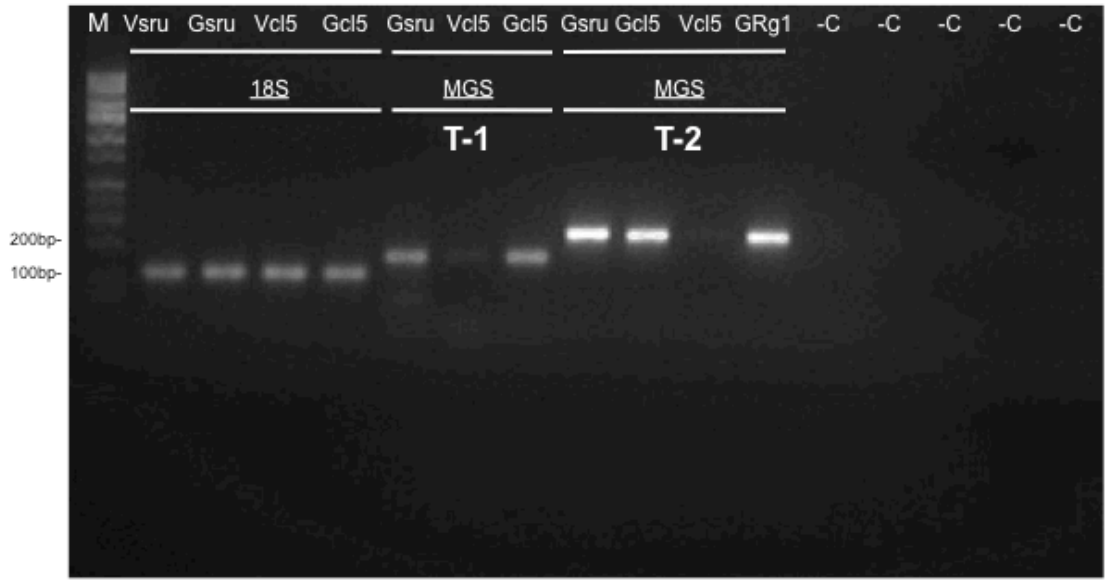
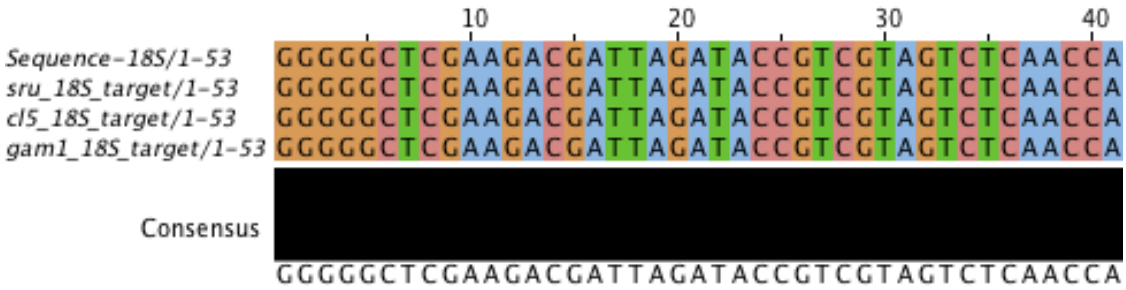


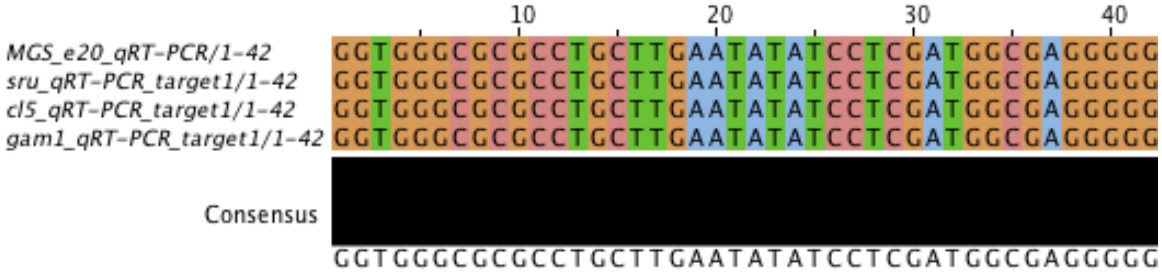
Figure 45. Sequencing of the qRT-PCR products. cDNA from the vegetative (V) and gametic (G) samples of the control *sru*2 and fusion-defective mutant cl5 show the expected size products for 18S, 109 bp. Target 1 (T-1) 160 bp. and target 2 (T-2) 239 bp (-C is the negative control with no cDNA added).

The sequencing results we received from Genewiz were aligned via the multiple sequence alignment algorithm ClustalW, figure 46.

(A)



(B)



(C)

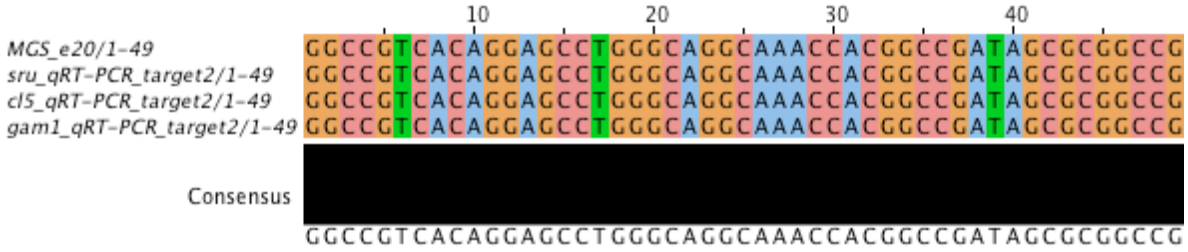


Figure 46. A portion of the sequence alignments for the qRT-PCR products. Sequences for the control strain *sru*-2 and fusion-defective mutants cl5 and *gam*1 were aligned via ClustalW. **A.** 18S rRNA **B.** MGS exon 20 Target 1 (T-1). **C**. MGS exon 20 target 2 (T-2).

3.9 Additional Cloning of the Flanking Genomic Region of the Insertions

The MGS chromosome walk (see section 3.7) and qRT-PCR data (see section 3.8) suggested that the insertional fusion-defective mutants 2-8 and 2-29 show no defects in the sequence or the expression of the MGS gene. Thus we are left with a series of mutants produced by DNA insertional mutagenesis with the pSP124S plasmid (La, Masters thesis, 1987; Lam, Masters thesis, 1991; Aksoy, PhD Thesis, 2008). These insertional fusion-defective mutants have not been complemented with the wild type copy of the HAP2/GCS1 gene (see section 1.5) and do not appear to be defective in MGS. Therefore, to map the physical location of the plasmid pSP124S, in both mutants, we applied reverse genetics and performed Site Finding Polymerase Chain Reaction (SF-PCR) (Tan et al., 2005).

3.9.1 Identification of the Insert Location in Mutant 2-8 by SF-PCR

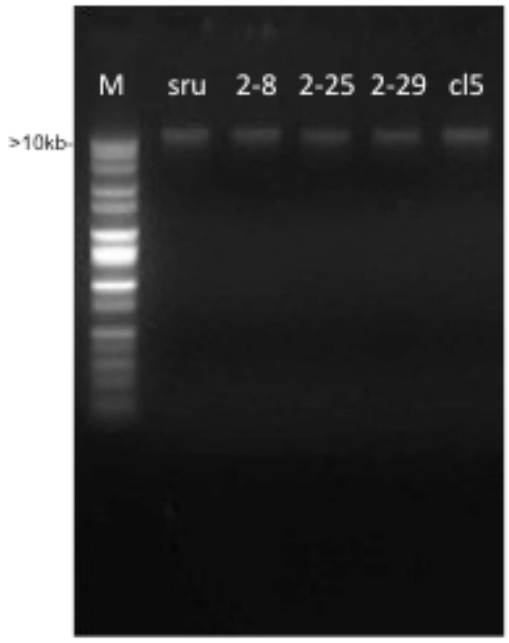
According to the Southern Blot results, mutant 2-8 had no band suggesting that there is a possibility of some structural change in the DNA of the insertion, making it undetectable by Southern blotting (Aksoy, PhD Thesis, 2008). Utilizing SF-PCR to walk into the regions flanking the insert should allow us to determine the gene/s or gene regulatory region/s that are disrupted by the insertion, which we hope to prove, caused the fusion defect.

3.9.2 Pre-Primary and Primary SF-PCR hybridization in Mutant 2-8

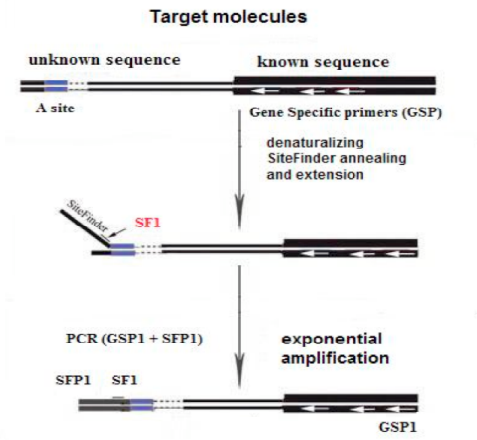
The genomic DNA of the control strain *sru-2* as well as several insertional fusion-defective mutants (including 2-8) utilized in SiteFinding-PCR are shown in figure 3A. During the pre-primary SF-PCR reaction we used degenerate Site Finder primer (SF1). The pre-primary step occurs at low temperature and produces annealing sites for the SF primers (figure 47B), allowing

for the initiation of the Site Finding reaction. During the primary SF-PCR reaction, the degenerate Site Finder primer (SFP1) along with the insert specific primer Ble1 were utilized, (figure 47C).

(A)



(B)



(C)

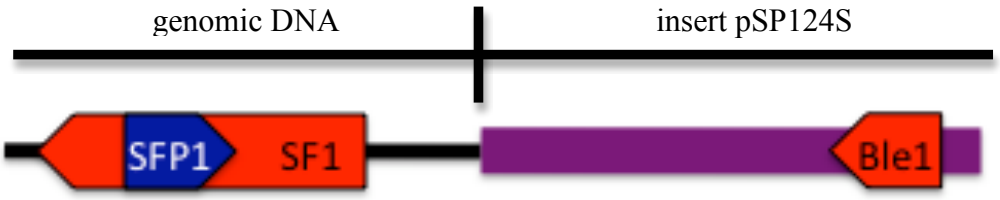


Figure 47. Genomic DNA and Schematic of the Pre-primary and Primary SF-PCR mutant 2-8. **A.** 0.9% agarose gel image of the genomic DNA utilized in SiteFinding-PCR. **B.** Schematic outline of primary SF-PCR (Tan et al. 2005). **C.** Diagram indicating where the gene specific primer Ble1 binds to the insert and the relative position of SFP1 to SiteFinder 1(SF1).

We did not expected to see any specific product from the control *sru-2*, host strain used to create the insertional fusion-defective mutants, except for the smearing effect usually seen after primary SF-PCR reactions (figure 48). For the insertional fusion-defective mutant 2-8, the smearing effect as well as the presence of non-specific DNA bands after the primary SF-PCR reaction indicated that the primary SF-PCR had amplified variable length DNA products (figure 48). The negative control labeled –C/1º, where no DNA was added to the SF-PCR reaction, had no product or smear (figure 48); thus we could continue to the secondary SF-PCR.

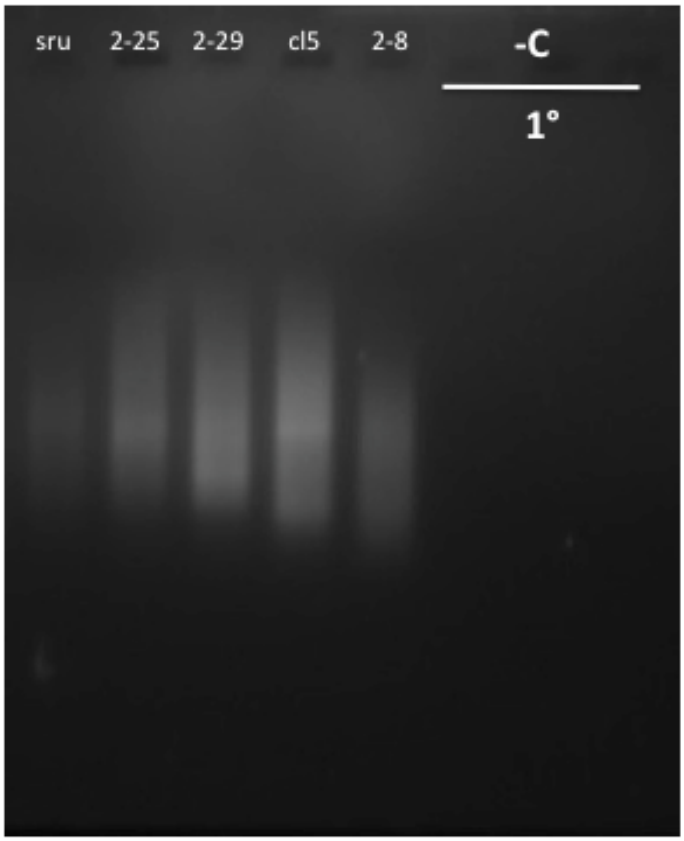
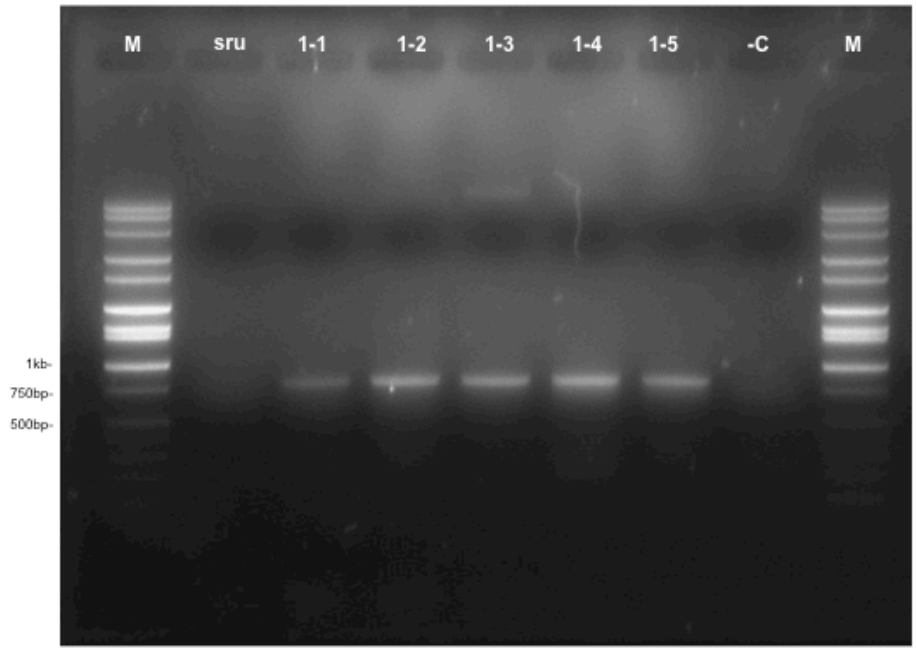


Figure 48. Results for Primary SF-PCR mutant 2-8. A. 0.9% agarose gel image for primary SF-PCR products. Primary SF-PCR reactions for multiple mutants including 2-8 are shown. Negative control, labeled –C/1°.

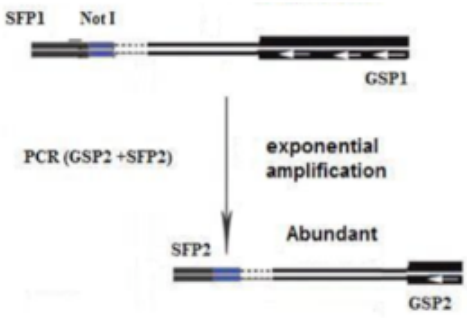
3.9.3 Secondary SF-PCR in Mutant 2-8

The secondary SF-PCR for 2-8 was performed using the primary SF-PCR products as templates. A nested gene specific primer, RD223 that binds to the second exon of *ble* upstream of the primer Ble2 (Table 2B), was paired with the same degenerate primer SFP2 (figures 49B & 49C). A specific product was formed in the secondary nested SF-PCR for 2-8, shown in figure 49A. The size of the one amplified DNA band using the RD223 primer was ~ 770 bp. Again the negative control reaction shown in –C demonstrated that no contamination was present.

(A)



(B)



(C)

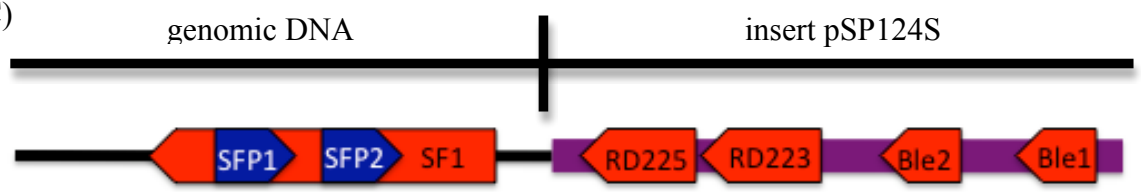


Figure 49. Secondary SF-PCR mutant 2-8. **A**. 0.9% agarose gel image for the secondary SF-PCR product for 2-8. Mutant 2-8 shows a distinct band at ~770 bp., lanes 1-1 through 1-5 show the products of primer RD223 and SFP2. The negative control, labeled –C, where no DNA was added to the secondary SF-PCR reactions showed no product. **B.** Schematic outline of secondary SF-PCR (Tan et al., 2005). **C.** Diagram indicating where the gene specific primers Ble1, Ble2, RD223 and RD225 bind to the insert and their relative position with respect to SFP1, SFP2 and SF1 primers.

The DNA for each band was cut out; gel purified and sent for sequencing at Genewiz.

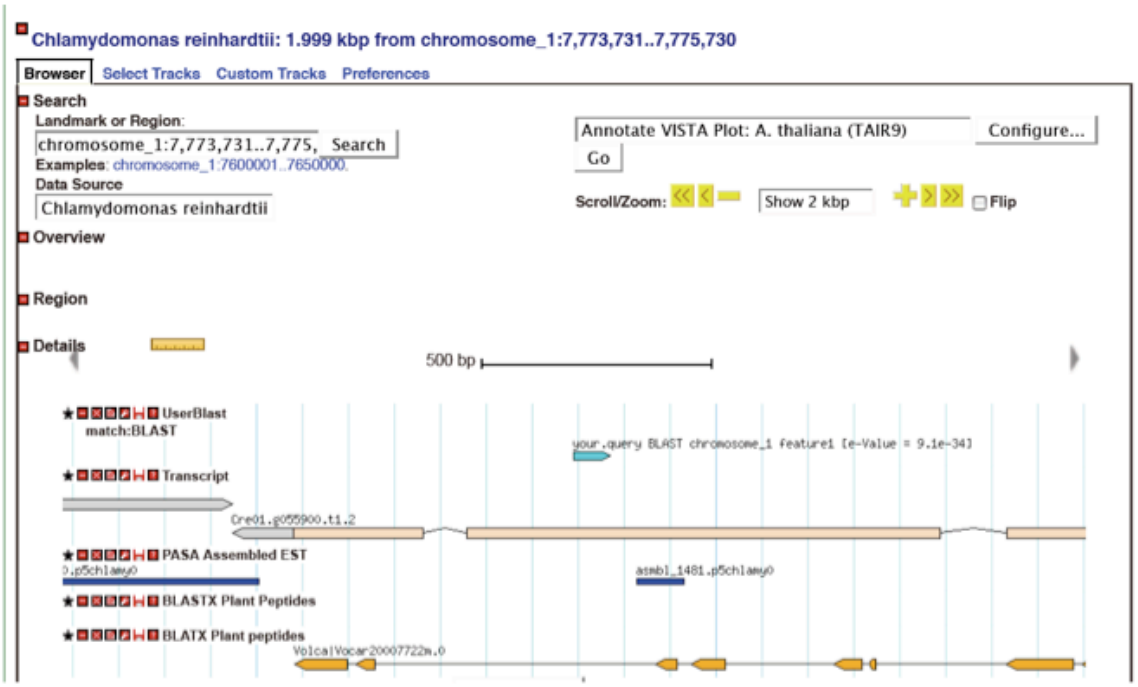
3.9.4 Sequencing Results for Mutant 2-8

When the gel purified DNA samples 1-2, 1-4 and 1-5 from mutant 2-8 were sequenced (figure 50A), we expected the sequences to contain the *rbcs2* promoter, followed by the pBluescript vector sequence, and then *Chlamydomonas* genomic sequences. We blasted the sequences we received from Genewiz into the Phytosome 8, JGI 4.3 *Chlamydomonas* database and the sequences produced a single hit matching the genomic sequence located between 7774691 and 7774770 bp (blue arrow) on Chromosome 1, as shown in figure 50B. The above-mentioned genomic sequence corresponds to the part of the 25th exon located upstream of the 3´UTR region of the Cre01.g055900.t1.2 gene coding for a flagellar/basal body protein, (figure 50C). Functional annotation of this locus indicated that the protein product of this gene might be involved in protein binding through three WD-40 domains and also it was found that this gene has homologs in higher eukaryotic organisms as well (discussed in section 5.1).

(A)

NNNNNNNNNNGNNNNNNNNNNNNNNNNNNNNGGNANNNNTTNNNNNNNNCNNNNNGCANNNNNNNNNNNNNNNNNNNNNNNNNNGNNNNNNNNNN
NNCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCA
GCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCANCCTGAATGGCGAAAGGAAATT
GTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCCCCCCT
GACCGGCGCTTCCGTATCAGCGCCTCCCCCAGCAGGACATGCTTCCCCTCTTTCGTCCTTGTCCTCGTCAAGGCCCCCCCGCGCCCGCTTGAGGA
CNCANTGCNAGGAGGGCNCNTTGAAAAAGNGNNGCCGGNGNTGGGGGGNTGNGGACTCTGCGTGTCCTAAACTTANGGGATGTTNAGGAAGGNCC
ANGGNCCCGGGAGCCANNGAGGGTTCCTGATGCTGCAGTCCCGGNAACGGCTGCAGGAGCNNNNNNNNNNNNNNNNNNNNNNNNNNNNACNNNNN
GNNNNNNGNNNNNNNNNNNNNNCNNNNGNNNNNNNNNNNNNNNNNNNTNNNNNNAGNNNNNNNNNNNNNNNN

(B)



(C)

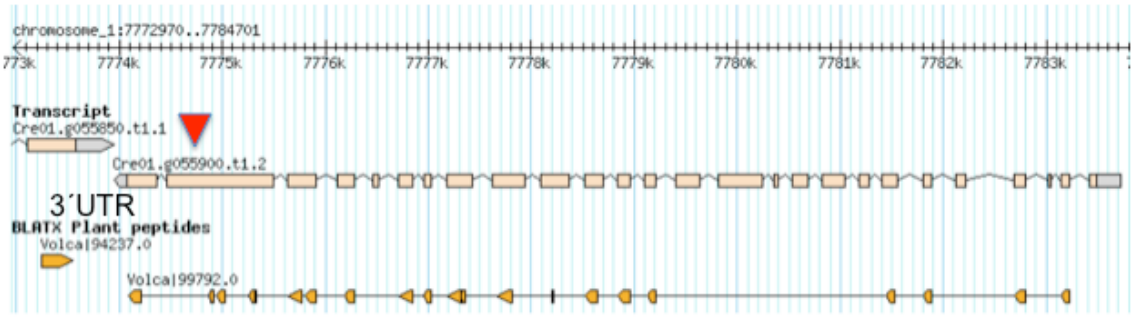


Figure 50. Sequencing result for mutant 2-8. A. Sequence of the secondary SF-PCR product. **B.** Phytosome 8, JGI 4.3 *Chlamydomonas* database blast result. **C**. Insert mapping Cre01.g055900.t1.2 gene coding for the flagellar/basal body protein along with the point of insertion (inverted triangle).

3.9.5 Identification of the Insert Location in Mutant 2-29 by SF-PCR

According to the Southern Blot results, mutant 2-29 had one band, suggesting that there is only one insertion (Aksoy, PhD Thesis, 2008). Utilizing SF-PCR to walk into the regions flanking the insert should allow us to determine the gene/s or gene regulatory region/s that are disrupted by the insertion, which we hope to prove, caused the fusion defect.

3.9.6 Pre-Primary and Primary SF-PCR hybridization for Mutant 2-29

The genomic DNA of the control *sru-2* strain as well as several insertional fusion-defective mutants (including 2-29) utilized in SiteFinding-PCR are shown in figure 47A. Experiments were carried out as described for mutant 2-8 (see section 3.9.1). We did not expect to see any specific product from the control *sru-2*, host strain used to create the insertional fusion-defective mutants, except for the smearing effect usually seen after primary SF-PCR reactions (figure 51). For the insertional fusion-defective mutant 2-29, the smearing effect as well as the presence of non-specific DNA bands after the primary SF-PCR reaction indicated that the primary SF-PCR had amplified variable length DNA products (figure 51). The negative control, labeled –C, where no DNA was added to the SF-PCR reaction, had no product or smear (figure 51), thus we could continue to the secondary SF-PCR.

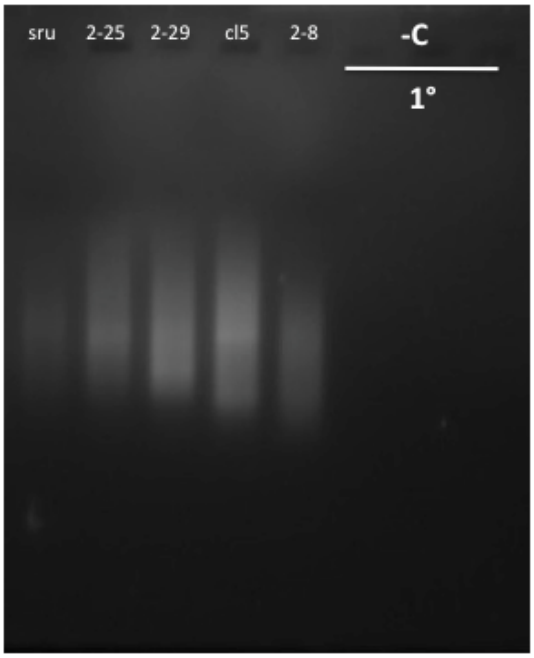


Figure 51. Primary SF-PCR mutant 2-29. A. 0.9% agarose gel image for primary SF-PCR products. Primary SF-PCR reactions for multiple mutants including 2-29 are shown. The negative control, labeled –C/1° where no DNA was added to the SF-PCR reactions showed no product.

3.9.7 Secondary SF-PCR in Mutant 2-29

The secondary SF-PCR for 2-29 was performed using the primary SF-PCR products as templates. A nested gene specific primer, RD223 that binds to the second exon of *ble* upstream of the primer Ble2 (Table 2B), was paired with the same degenerate primer SFP2 (figures 47B & 47C). The specific product formed in the secondary nested SF-PCR for 2-29 is shown in figure 52. Two amplified DNA bands were seen using the RD223 primer, a top band at ~ 690 bp and a lower one at ~250 bp. The negative control reaction, shown in –C, demonstrated that no contamination was present.

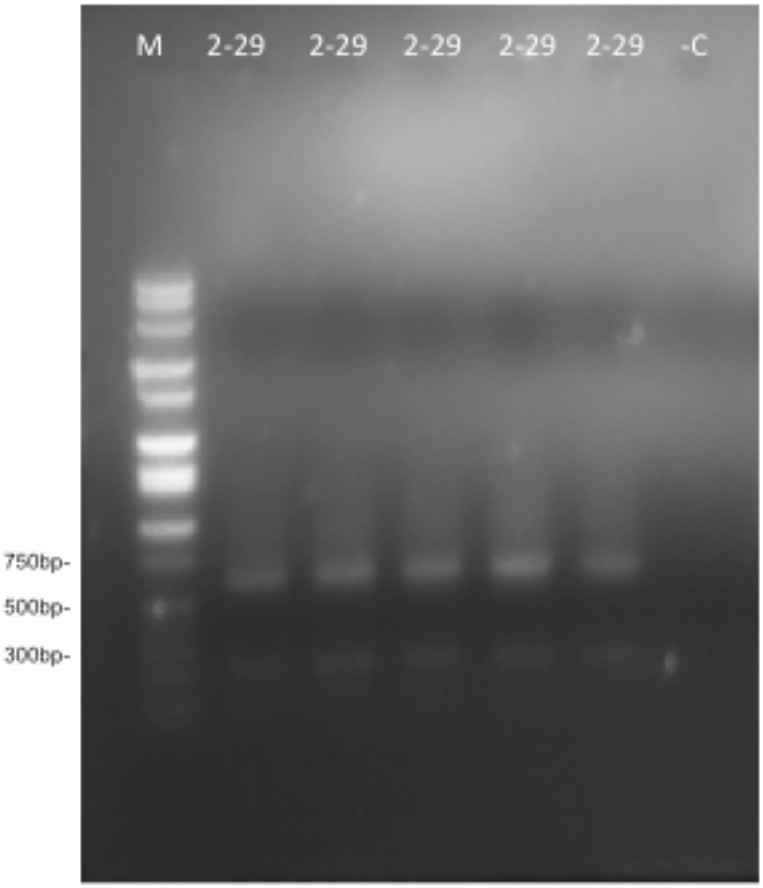


Figure 52. Secondary SF-PCR mutant 2-29. **A**. 0.9% agarose gel image for the secondary SF-PCR product for 2-29. The negative control, labeled –C, where no DNA was added to the secondary SF-PCR reaction showed no product.

The DNA for each band was cut out; gel purified and sent for sequencing at Genewiz.

3.9.8 Sequencing Results for Mutant 2-29

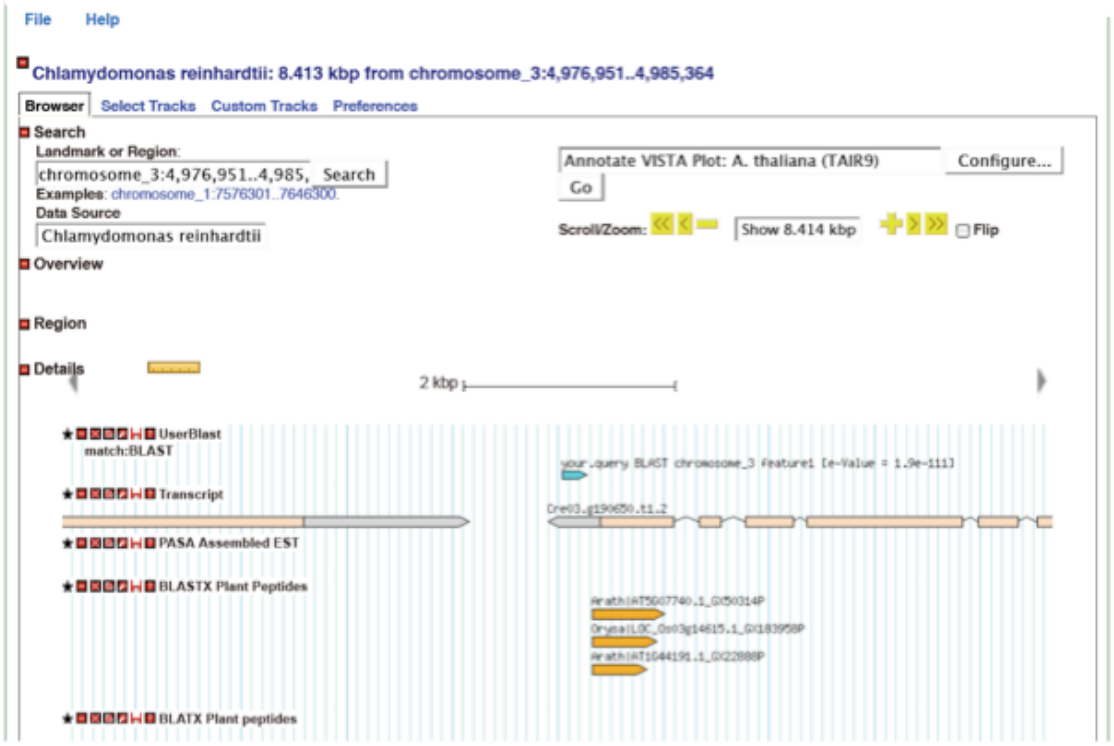
When the gel purified DNA samples from mutant 2-29 were sequenced (figure 53A), we blasted the sequences we received from Genewiz into the Phytosome 8, JGI 4.3 *Chlamydomonas* database. The sequence corresponding to the top DNA band (~690 bp) produced a single hit matching the genomic sequence located between 4981201 and 4981274 bp on Chromosome 3,

(figure 53B, see blue arrow). The above-mentioned genomic sequence corresponds to the 3´UTR of the Cre03.g190650.t1.2, a gene with no functional annotation and no putative conserved domains found (figure 53C). The sequence corresponding to the bottom band ~250 bp produced no hits in the Phytosome 8, JGI 4.3 *Chlamydomonas* database. Nonetheless, it matched the partial sequence of the *rbcs2* promoter and the pBluescript vector sequence upon its alignment with pSP124S insert vector sequence, data not shown.

(A)

```
NNNNNNNNNGNNNNNNNCGGTACCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTG
GGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGA
TCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGT
TAAATCAGCTCATTTTTTAACCAAACGCATAGCATGTGGGCCTAGCCGTGTGGCCGTGAGACAATCCTGTATGTATTTCTAATGAGA
ACAATCATGCAGTACGGTAGCAAGTATATATCTGGCGCTTCGTGGTGGCCCCAAGCCCCGCACCGCCTCTGTCTGTAAGGTTCTCNC
CCCGGCACCTGACATGCGCATTGACCTTGAGCTTTAAATAATAATGCCCACCCTGCACACCGTGCNNACGGACAGGAGCGTCTT
```

(B)



(C)

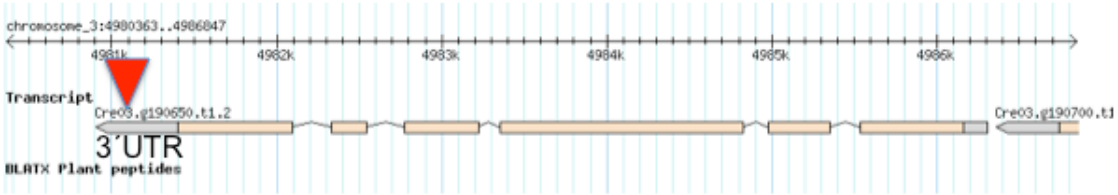


Figure 53. Sequencing result for mutant 2-29. A. Sequence of the secondary SF-PCR product. B. Phytosome 8, JGI 4.3 *Chlamydomonas* database blast result. C. Insert mapping lead to Cre03.g190650.t1.2 gene with no functional annotation and no putative conserved domains found along with the point of insertion (inverted triangle).

3.10 Thin Layer Chromatography (TLC) of the *Chlamydomonas* gametes

To determine whether *Chlamydomonas* gametes accumulate secondary metabolites, total cellular lipid extracts were prepared from vegetative and gametic cells (on an equal cell count basis) and separated by TLC (see section 2.7.9). This result provides further evidence that *Chlamydomonas* gametes show increased levels of accumulating secondary metabolites, such as: isoprenoid molecules, (figure 54).

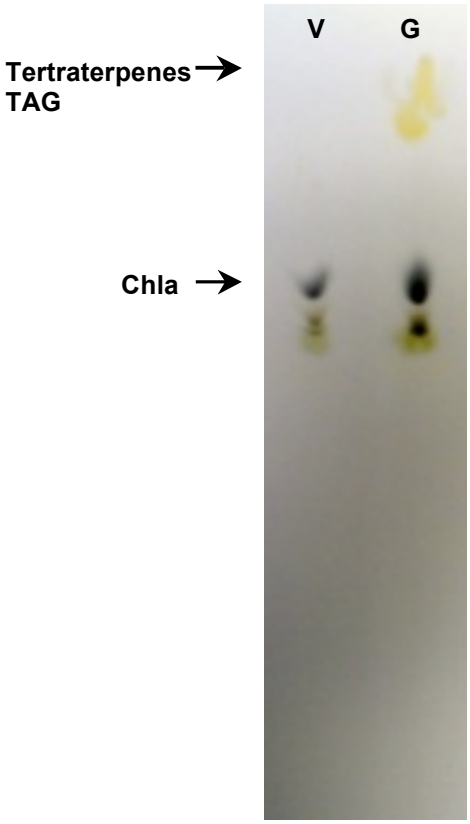


Figure 54. *Chlamydomonas* gametes accumulate secondary metabolites. 2-D TLC plate analysis of total cellular extracts from vegetative vs gametic *Chlamydomonas* wt cells *cw15* mt+. (V-vegetative, G-gametic, Chla-chlorophyll a, Tertraterpenes- isoprenoids, TAG-triacyl glycerides).

3.11 454 Data of the MEP Regulatory Enzymes

Results of the 454 transcriptome analysis (http://genomes.mcdb.ucla.edu) of the whole genome sequencing data indicate an elevated level of expression of the genes coding for the regulatory enzymes of the MEP pathway (DXS, DXR and HDS see section 1.3.1) in *Chlamydomonas* gametes, therefore indicating regulation of expression of these genes at the transcriptional level during gametic stage, figure 55.

(A)

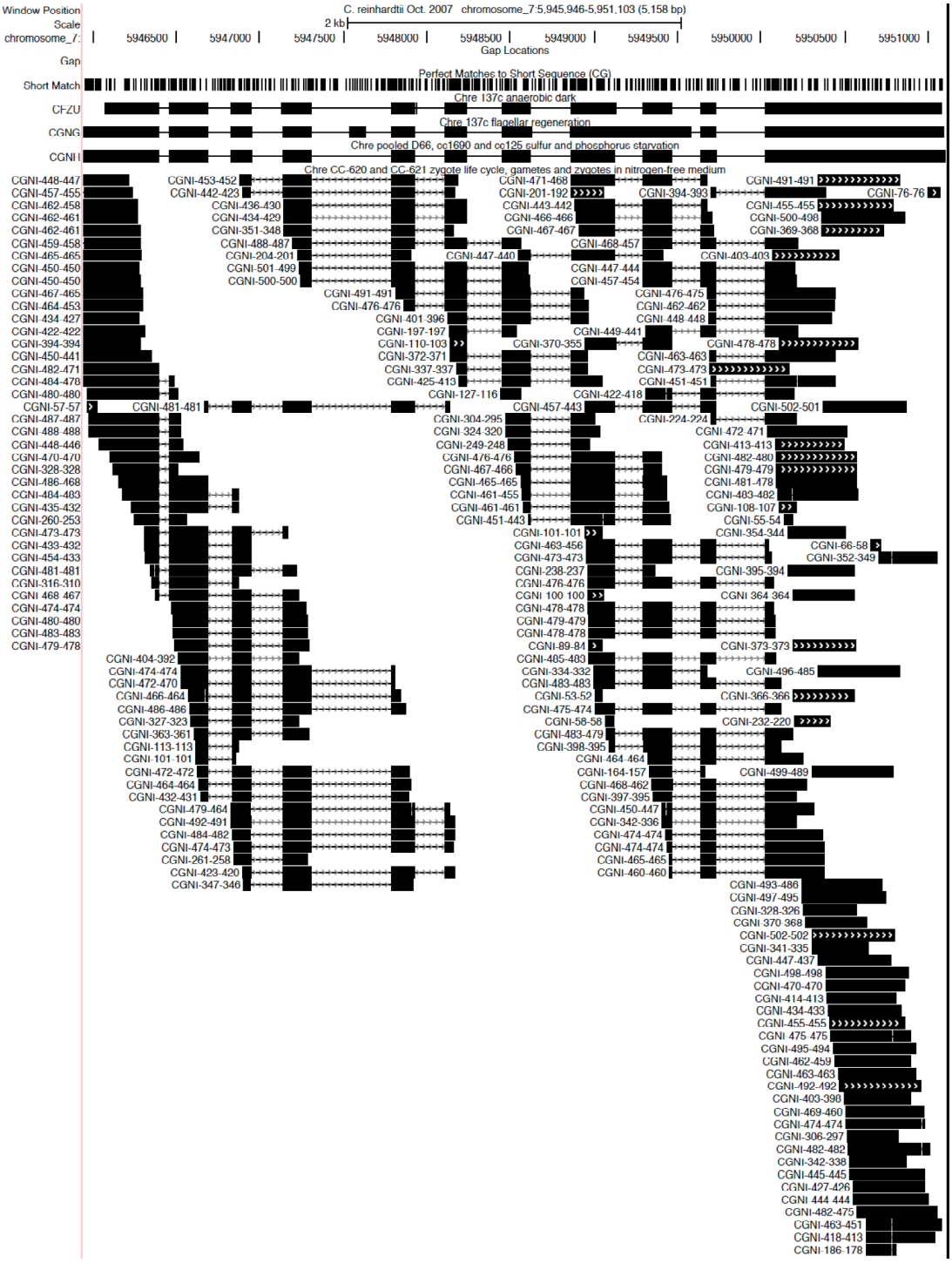


Figure 55 continued on the next page.

(B)

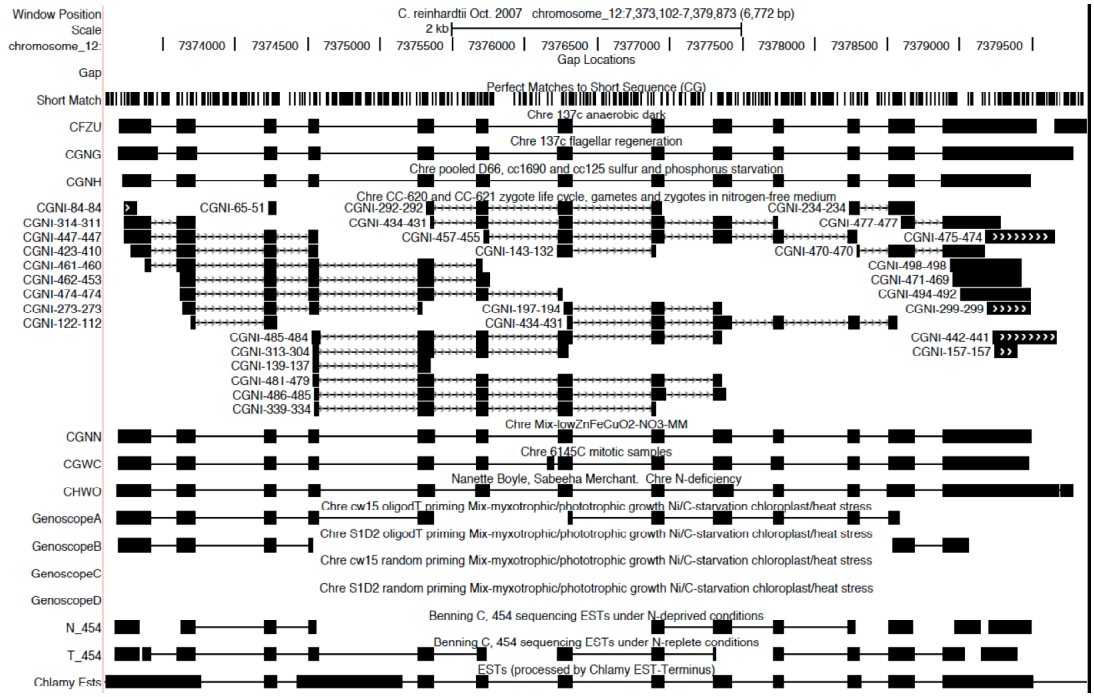


Figure 55 continued on the next page.

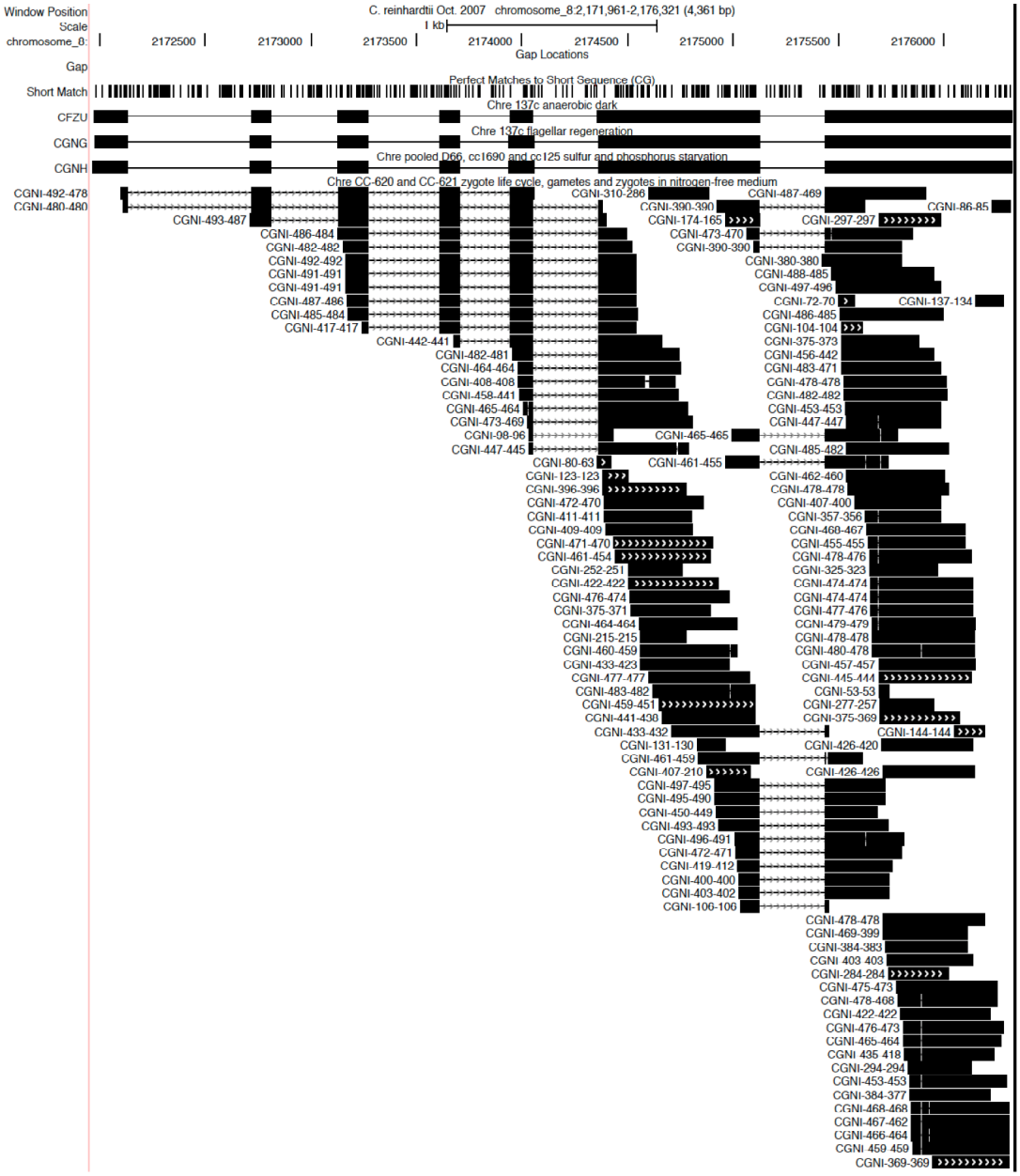
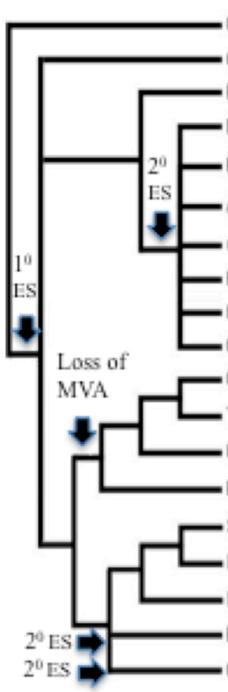


Figure 55. Genomic Analysis of the regulatory MEP pathway enzymes: **454 Data.** The 454 whole transcriptome expression data for (A) DXS (B) DXR (C) HDS during gametic stages.

3.12 Phylogenetic Distribution of MEP and MVA Pathway Genes

To investigate the presence of the MVA and the MEP pathway genes in algae, the available genomes of a variety of algae were investigated. During a cross-genomes search approach, genes for all the seven enzymes participating in the MEP pathway were identified in a variety of algal genomes using presence-absence analysis. We identified homologs in algal genomes using BLAST of DNA sequences from *Arabidopsis thaliana* and *Chamydomonas reinhardtii* with the e-value cut-off at 1.0e-10 and by using TBLASTN of protein sequences from the above-mentioned species with the e-value cut-off at 1.0e-20 (for the list of publically available fully sequenced as well as EST projects algal genomes see Material and Methods section 2.14). Further, to be able to dissect the gene gain or loss based on parsimony reconstruction, we reconstructed ancestral states of gene presence and absence on the internal nodes of the species tree (Table 4) using post-order traversal (Page and Holmes, 2005).

Results depicted in Table 4, confirms that green algae only contained genes for the MEP pathway and lacked the genes coding for enzymes of the mevalonate (MVA) pathway. This result also confirmed that green algae synthesize their isoprenoids exclusively through the MEP pathway (Frommolt et al., 2008; Schwender et al., 2001).



Sequenced organisms		MEP pathway								MVA pathway						
		DXS [EC:2.2.1.7]	DXR [EC:1.1.1.267]	MCT [EC:2.7.7.60]	CMK [EC:2.7.1.148]	MDS [EC:4.6.1.12]	HDS [EC:1.17.7.1]	HDR [EC:1.17.1.2]	IDI [EC:5.3.3.2]	AACT [EC:2.3.1.9]	HMGCS [EC:2.3.3.10]	HMGR [EC:1.1.1.88]	HMGCR [EC:1.1.1.34]	MVK [EC:2.7.1.36]	PMVK [EC:2.7.4.2]	MVD [EC:4.1.1.33]
Cyanophyta		+	+	+	+	+	+	+	+[b]	+	+	--	--	--	--	--
Glaucophyta*		+	--	--	--	--	+	+	nd	--	+	--	+	+	--	+
Rhodophyta		+	+	+	+	+	+	+	+[c]	+/--	+/--	+/--	+/--	+/--	+/--	+/--
Dinophyta	Alveolata	+	+	+	+	+	+	+	nd	+	+	+	+	+	+	+
Perkinsidae	Alveolata	+	+	+	+	+	+	+	nd	+	+	+	+	+	+	+
Apicomplexa	Alveolata	+	+	+	+	+	+	+	--	--	--	--	--	--	--	--
Ciliata	Alveolata	--	--	--	--	--	--	--	+[c]	+	+	+	+	+	+	+
Heterokontophyta	Chromalveolata	+	+	+	+	+	+	+	+[c]	+	+	+	+	+	+	+
Haptophyta	Chromalveolata	+	+	+	+	+	+	+	+[c]	+	+	+	+	+	+	+
Cryptophyta	Chromalveolata	+	--	--	--	+	+	+	+[c]	nd	nd	nd	nd	nd	nd	nd
Chlorophycea	Chlorophyta	+	+	+	+	+	+	+	+[c]	+	+	--	--	--	--	--
Trebouxiophycea	Chlorophyta	+	+	+	+	+	+	+	+[c]	+	+	--	--	--	--	--
Ulvophyceae	Chlorophyta	+	+	+	+	+	+	+	+[c]	+	+	--	--	--	--	--
Prasinophyceae	Chlorophyta	+	+	+	+	+	+	+	+[c]	+	+	--	--	--	--	--
Seed Plants	Streptophyta	+	+	+	+	+	+	+	+[c]	+	+	+	+	+	+	+
Mosses	Streptophyta	+	+	+	+	+	+	+	+[c]	+	+	+	+	+	+	+
Mesostigmales	Streptophyta	+	+	+	+	+	+	+	nd	--	+	--	+	--	--	+
Euglenophyta		+	--	+	--	--	--	+	+[c]	--	--	--	+	+	+	+
Chlorarachniophyta		+	--	--	--	--	+	--	nd	--	--	--	--	--	--	+

Table 4 continued on the next page.

Table 4. Schematic diagram of the MEP and MVA pathways gene distribution. Phylogeny of the species tree for the MEP and MVA pathways distribution is based on analysis of the rRNA and ITS sequences this study and after Delwiche 1999; Gould et al 2008. (*) indicate presence of EST's only; b-bacterial; e-eukaryotic; ND-not determined; Enzymes of the MEP pathway: DXS, MEP synthase; DXR,MEP reductoisomerase; MCT, CDP-ME synthase; CMK, CDP-ME kinase; MDS, ME-2,4cPP synthase; HDS, HMBPP synthase; HDR, HMBPP reductase. Enzymes of the MVA pathway: HMGS, HMG-CoA synthase; HMGR, HMG-CoA reductase; MVK, MVA kinase; PMK, MVAP kinase; MDD, MVAPP decarboxylase.

All algae species used in this study have genes coding for MEP pathway enzymes, but green algae lack genes coding for enzymes of the MVA pathway. The results of our parsimony analysis indicate that the Chlorophyta has lost the MVA pathway (see Supplementary Material) showing consistency with the previous phylogenetic analysis (Frommolt et al., 2008; Grauvogel and Petersen, 2007) as well as the biochemical analysis (Schwender et al., 2001). The significance of this discovery is that it confirms that green algae synthesize isoprenoids exclusively via the MEP pathway. The results from our analysis depicted in Table 4, reveal that green algae only contained genes for the MEP pathway and lacking genes coding for enzymes of the MVA pathway. Also our findings reveal that eukaryotic algae genomes used during this study contain nuclear encoded and latter chloroplast localized MEP pathway enzymes and each MEP pathway enzyme is encoded by a single gene. Also, our phylognenetic analysis supports the hypothesis postulating that MEP pathway genes were acquired from cyanobacterial ancestor of plastids via primary endosymbiotic event. The results of our analysis reveal that plastid bearing Cyanobacterial ancestor representative contain the MEP pathway enzymes and each enzyme is encoded by a single gene as well, (see supplementary materials).

3.12.1 Phylogenetic analysis of IDI

The Isopentienyl Diphosphate Isomerase (IDI) is the last enzyme participating in the formation of the C5 isoprene units and it is found in MEP and MVA pathways. There are two different types of IDI known, type-1 is found in *Chlamydomonas* and the rest of the Chloroplastida and type-2 is found in cyanobacteria (Bouvier et al., 2005). Therefore the IDI presence in *Chlamydomonas* is most likely due to its horizontal gene transfer (HGT) from the host cell that engulfed the *Cyanobacterial* plastid progenitor (Cunningham and Gantt, 2000). Figure 56 shows results of our phylogenetic analysis and Table 5 indicates the locations of the IDI gene in the fully sequenced algal genomes.

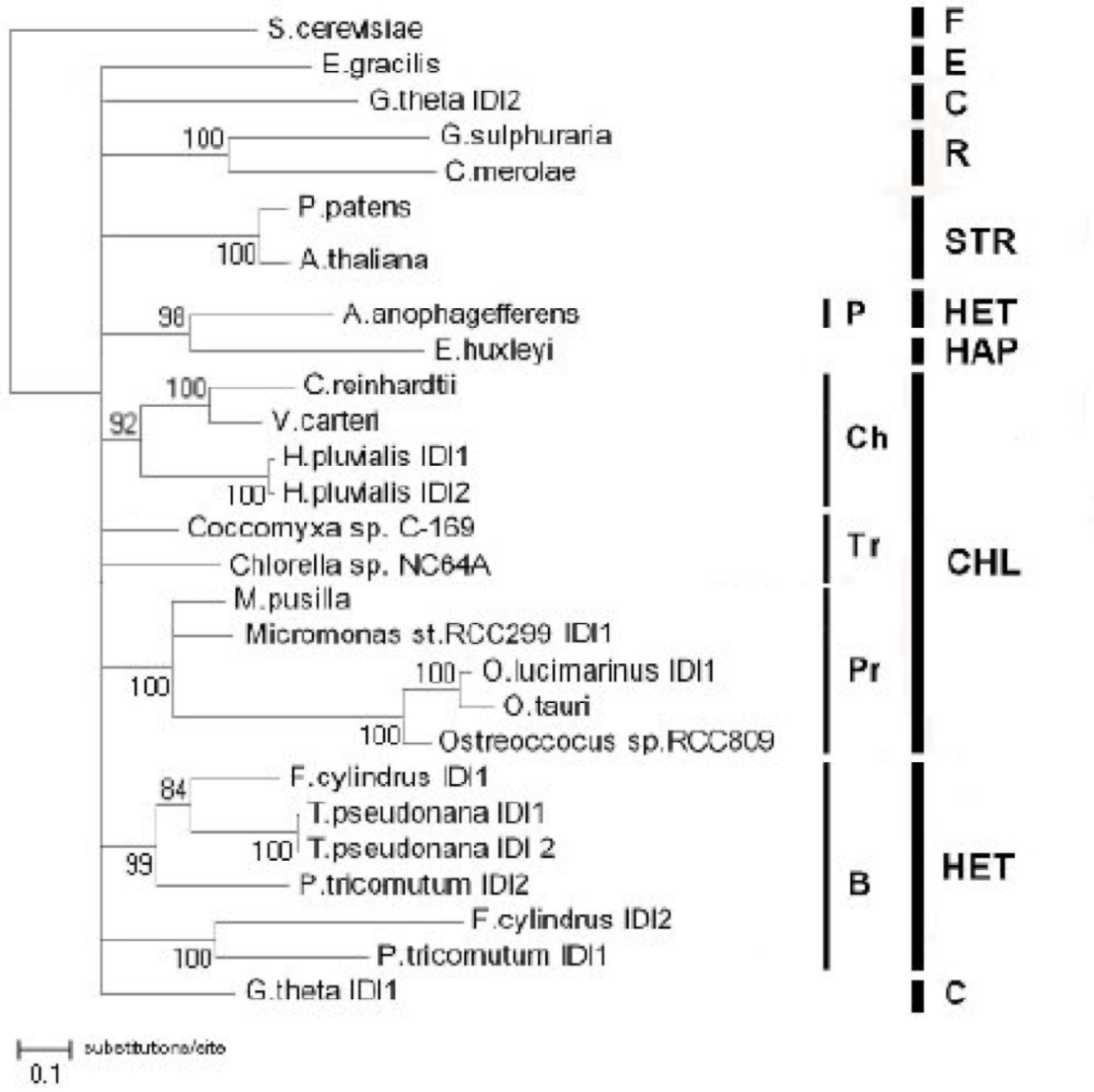


Figure 56. Phylogenetic tree for the IDI. Majority-rule consensus of 5,000 trees saved during Bayesian analyses of IDI protein sequences for 27 OTU's. Numbers above and below branches indicate clade frequency ≥80% (Bayesian posterior probability values). Abbreviations: A, Alveolata; B, Bacillariophyceae; Ch, Chlorophyceae; CHL, Chlorophyta; C, Chromista; CYA, Cyanobacteria; E, Euglenophyta; GL, Glaucophyta; HAP, Haptophyta; HET Hetrokontophyta; M, Metazoa, P, Pelagophyceae; Pr, Prasinophyceae; R, Rhodophyta; T, Trebouxiophyceae.

Species	Gene locations IDI_5.3.3.2
Aureococcus anophagefferens	Auran1/scaffold_11:1196450-1197226
Chlamydomonas reinhardtii	scaffold_44:27959-32220
Chlorella sp. NC64A	ChlNC64A_1/scaffold_16:646040-647998
Coccomyxa C-169	scaffold_2:3448181-3449385 scaffold_9:1254736-1255716
Emiliania huxleyi CCMP1516	Emihu1/scaffold_83:60073-63114
Micromonas pusilla CCMP1545	MicpuC2/scaffold_1:372804-373244 scaffold_15:258244-259225
Micromonas st. RCC299	Chr_13:510542-511536
Ostreococcus RCC809	OstRCC809_1/scaffold_14:509222-509998
Ostreococcus lucimarinus	Ost9901_3/Chr_12:405415-406065 Chr_20:517542-518312
Ostreococcus tauri	Ostta4/Chr_14.0001:484757-485168
Phaeodactylum tricornutum	Phatr2/chr_8:549830-550495 chr_10:89343-90249
Thalassiosira pseudonana	Thaps3/chr_6:1130041-1130820 chr_1:133217-134061
Volvox carteri	Volca1/scaffold_84:221836-222906
Fragilariopsis cylindrus	scaffold_8:1760140-1761259 scaffold_44:389875-390984 scaffold_6:1704484-1705281
Cyanidioschyzon merolae	c02f0006 158612 / c20f0011 222599>223252

Table 5. Location of the genes coding for the IDI enzyme.

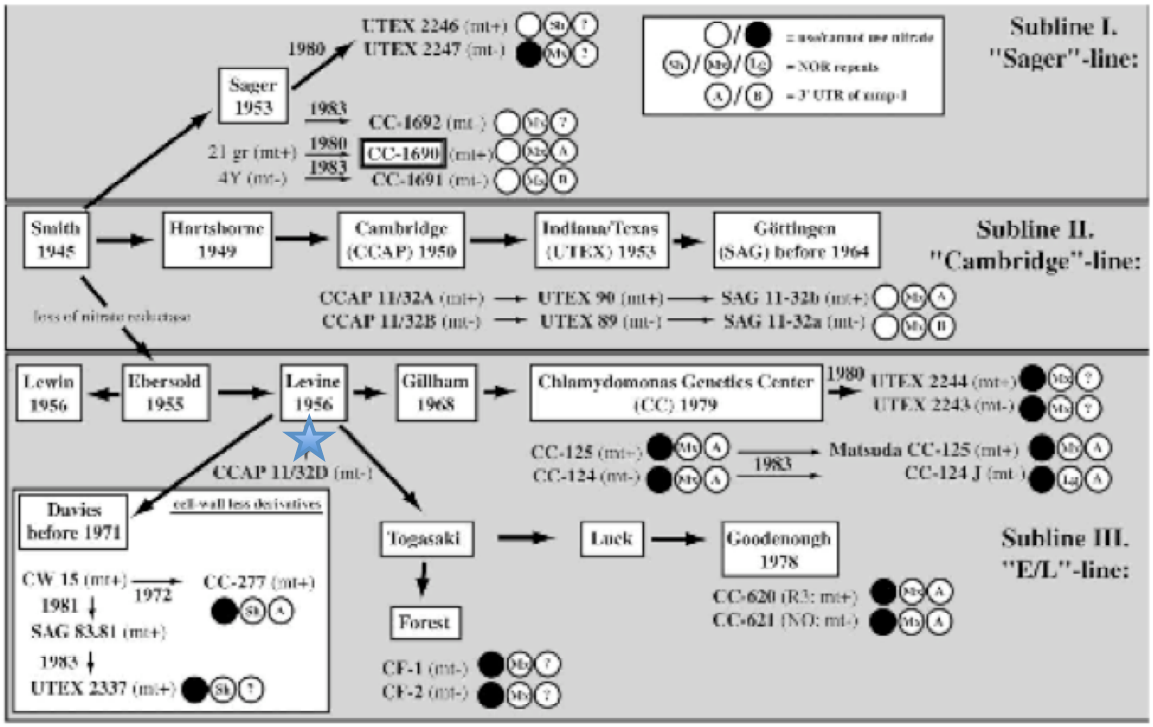
CHAPTER 4

DICUSSION

The membrane fusion of eukaryotic gametes is still an elusive phenomenon, since we do not know what fusogen/s mediate this process. Our initial body of knowledge, regarding fusogens participating in gametic membrane fusion in *Chlamydomonas*, has come from applied forward genetics (Forest, 1978; La, Masters thesis, 1987; Lam, Masters thesis, 1991; Aksoy, PhD thesis, 2008; Liu et al., 2008). Previous forward genetic experiments that utilized DNA insertional mutagenesis with the pSP124S plasmid resulted in random insertions into the genome of the host *Chlamydomonas* strain *sr-u*-2-23 mt^- cc 275, (*sru-2)* (La, Masters thesis, 1987; Lam, Masters thesis, 1991; Aksoy, PhD Thesis, 2008). This streptomycin resistant strain can be utilized to select for mutants unable to fuse (Forest and Togasaki, 1975). The above-mentioned mutagenesis by random DNA insertion of pSP124S plasmid produced 7 fusion defective mutants during 4 independent transformations. All of the mutants are able to recognize mt^+ cells, transmit signals, remove cell walls and adhere to mt^+ partners, suggesting that these mutants are blocked at the last stage of mating, i.e. fusion (Forest, 1978; La, Masters thesis, 1987; Lam, Masters thesis, 1991; Aksoy, PhD thesis, 1998). Liu et al used an insertional mutant created by Greg Pazour (University of Massachusetts) to show that the defect in the homolog of the sex-restricted HAP2/GCS1 gene prevents gamete fusion in *C. reinhardtii* (Liu et al., 2008a). Using complementation analysis, half of our mutants have been shown to be defective in this homolog of the sex-restricted HAP2/GCS1 gene. We continued our search for fusogen molecules in the fusion-defective mutants that could not be complemented with a wild type copy of the HAP2/GCS1 gene.

During this study we began by using SF-PCR to confirm the location of the pSP124 plasmid insertion in the fusion-defective mutant cl5. This insertion was mapped within the *copia*-family retrotransposon RT (Cre13.g605950.t1) in scaffold-44 at position 523923 (Aksoy, PhD thesis, 1998). According to the Phytosome 8, JGI 4.3 *Chlamydomonas* database, the updated insert position is in chromosome 13 with the base position being 6297731. Next we analyzed the upstream and downstream regions from the point of insertion on chromosome 13 using chromosome walking by PCR. We compared these regions for all of the fusion-defective mutants that could not be complemented with the wild type copy of the HAP2/GCS1 gene. The chromosome walk on the region downstream from the point of insertion revealed that at position 6297404 on chromosome 13, corresponding to the intergenic region separating the *copia*-family retrotransposon and the Ef-GTPase, there was a base pair difference (C to T) in all of the strains we sequenced when compared to the CC-503 cw92 mt^+ strain that was used for genomic sequencing at US DOE JGI. We believe that the CC-503 cw92 mt^+ strain might have acquired this random point mutation at some point after the transfer of the above-mentioned strain from Levine to Davies (figure 57 A &B).

(A)


Subline I.
"Sager"-line:
Sager 1953
1980
UTEX 2246 (mt+)
UTEX 2247 (mt-)
1983
CC-1692 (mt-)
21 gr (mt+)
4Y (mt-)
1980
1983
CC-1690 (mt+)
CC-1691 (mt-)
= NOR repeats
= 3' UTR of nmp-1
Subline II.
"Cambridge"-line:
Smith 1945
Hartshorne 1949
Cambridge (CCAP) 1950
Indiana/Texas (UTEX) 1953
Göttingen (SAG) before 1964
loss of nitrate reductase
CCAP 11/32A (mt+)
CCAP 11/32B (mt-)
UTEX 90 (mt+)
UTEX 89 (mt-)
SAG 11-32b (mt+)
SAG 11-32a (mt-)
Lewin 1956
Ebersold 1955
Levine 1956
Gillham 1968
Chlamydomonas Genetics Center (CC) 1979
1980
UTEX 2244 (mt+)
UTEX 2243 (mt-)
CC-125 (mt+)
CC-124 (mt-)
1983
Matsuda CC-125 (mt+)
CC-124 J (mt-)
CCAP 11/32D (mt-)
Subline III.
"E/L"-line:
Davies before 1971
cell-wall-less derivatives
CW 15 (mt+)
1972
CC-277 (mt+)
1981
SAG 83.81 (mt+)
1983
UTEX 2337 (mt+)
Togasaki
Luck
Goodenough 1978
CC-620 (R3; mt+)
CC-621 (NO; mt-)
Forest
CF-1 (mt-)
CF-2 (mt-)

(B)

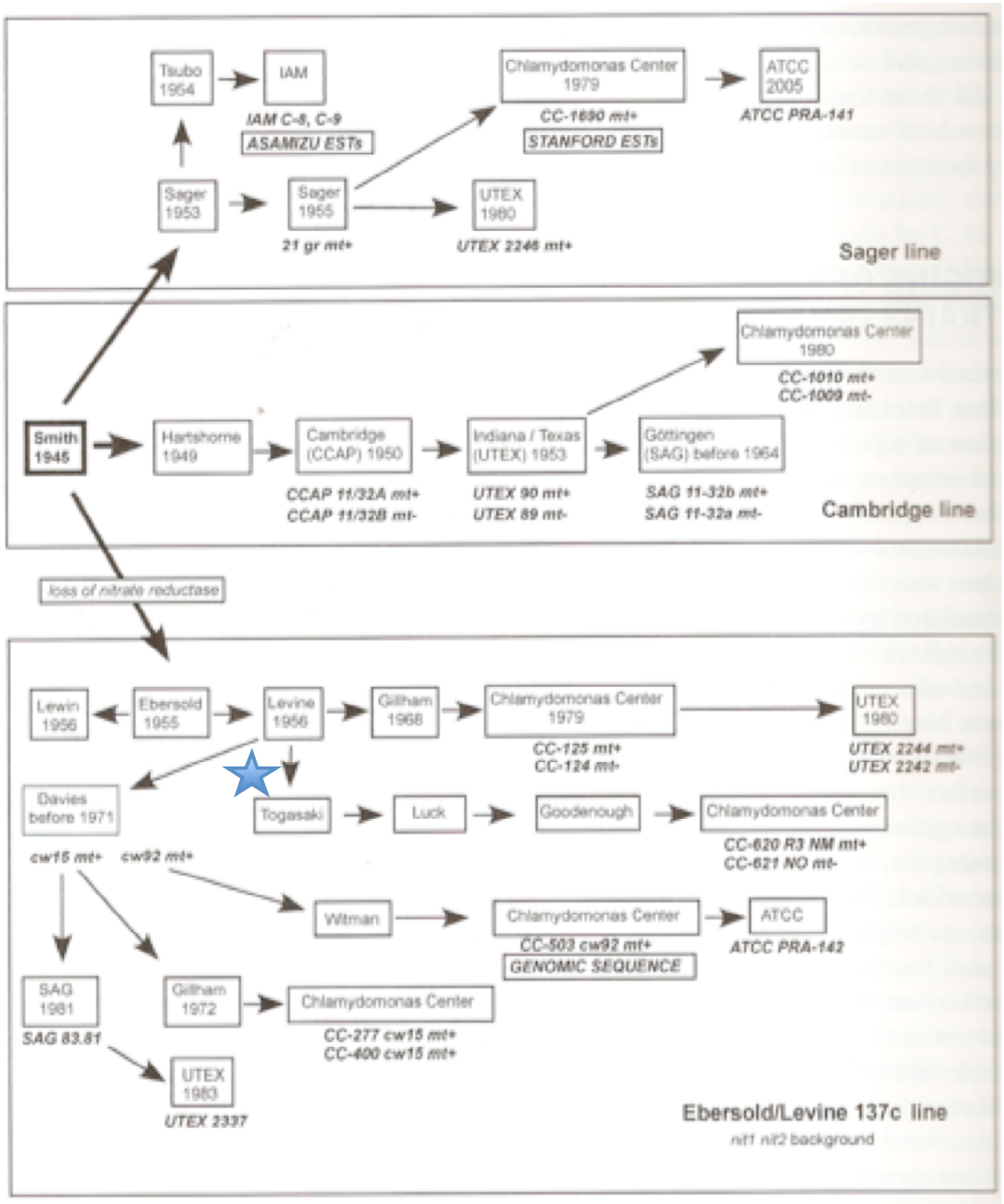


Figure 57. Schematic diagram of various *Chlamydomonas* strains. (*Chlamydomonas* Sourcebook v2). (Stars indicate the branch point).

The wild type strains that we used, CC-620 R3 mt^+ and CC-621 NO mt^-, came from Goodenough and were deposited into the *Chamydomonas* center in 1978 and our control/parental strain *sr-u-2-23* mt^- cc 275, (*sru-2)* came from Gillham and it was deposited into the *Chamydomonas* center in 1972.

To determine the presence of gamete-expressed gene/s in this region of chromosome 13, we performed genomic analysis of the ~100 kb region radiating from the point of the pSP124 plasmid insertion on chromosome 13. Analysis of the publicly available *Chlamydomonas* ESTs and the 454 database revealed that the Cre13.g605900.t1 gene, located downstream from the *copia*-family retrotransposon RT containing the plasmid insert in mutant cl5, showed only gametic EST's (Phytosome 8, JGI 4.3) and much higher gametic expression in the 454 database compared to vegetative expression (http://genomes.mcdb.ucla.edu). To determine whether Cre13.g605900.t1 is expressed in mt^+ or mt^- gametes or both, we collaborated with Dr. James Umen, and gained access to RNAseq (high-throughput transcriptome analysis) data for the genomic region spanning the chromosome 13 loci from 6240493 to 6356688. We found that Cre13.g605900.t1 (and no other gene in this region) showed **minus gamete specific expression** (Umen and DeHoff unpublished data) leading us to name the gene MGS. We then confirmed the RNAseq data with the use of qRT-PCR analysis again showing that MGS is a **minus gamete specific gene**.

MGS is located on chromosome 13 between 6278240 - 6291461 on the latest version of Phytosome, JGI 4.3 The coding sequence is on the positive DNA strand, spanning a 13222 bp. region, with an over 64% GC rich sequence. The cDNA transcript length is 10882 bp. and the CDS sequence is 9639 bp. The protein product is expected to be 281 kDa and have 3212 aas, with a low-complexity carboxypeptidase domain. Multiple hydrophobic domains are predicted

by some algorithms (KD and GET) while others (ProFamm) suggest the protein is found outside the cell.

To determine whether MGS shows any defect in either the coding DNA or the regulatory regions, we preformed a chromosome walk via PCR. We analyzed the entire sequence, including the promoter, 3′ and 5′ UTR regions and all of the intron and exon boundaries. We did not find defects in any of the fusion-defective mutants, including cl5. To determine whether the insertion of pSP124 in the *copia*-family retrotransposon RT had increased or decreased the expression of MGS in mutant cl5 or any of the other non-*gcs*1 mutants, we performed qRT-PCR. Again, we did not observe any abnormal expression. All mutants showed the same minus-gamete-specific expression seen in wild type. Despite the fact that the fusion-defective mutants showed no defect in MGS DNA or the gene's expression in all tested fusion-defective mutants including cl5, the significance of the finding that MGS is a minus-gamete specific gene provides an opportunity for its utilization as a genetic marker in future molecular and bioinformatic experiments. For future studies we propose to use reverse genetics analysis via amiRNA (artificial micro RNA interferance) to detrmine the function of MGS gene.

In addition to the MGS analysis, we also applied SF-PCR to identify as well as to clone the regions flanking the inserted plasmid in the remaining insertional fusion-defective mutants that showed no defect in the sequence or expression of the MGS gene and that were not complemented with the wild type copy of the HAP2/GCS1 gene. We were able to clone the flanking region of the insertion sites of mutants 2-8 and 2-29.

Since 2-8 had no band by Southern blotting and had lost Zeocin resistance (Aksoy, PhD Thesis, 2008), this suggested that there was a loss of some or all of the plasmid, possibly with some structural change in the DNA adjacent to or in part of the plasmid. Although this made it very

challenging, we were able to clone the flanking genomic region for mutant 2-8 which we showed to be in chromosome 1. The insertion occurred in position 7774691, corresponding to part of the 25th exon of the Cre01.g055900.t1.2 gene, which codes for a flagellar/basal body protein. This is annotated as a conserved protein found in ciliated organisms, upregulated during deflagellation with a protein binding function (Phytosome 8, JGI 4.3 *Chlamydomonas* database). Initially this protein was identified in the flagella and basal body proteome via comparative genomics (Li et al., 2004). According to the 454 transcriptome analysis it is expressed under different conditions listed in the order of its highest expression: 1) flagellar regeneration 2) mitotic 3) mix of low Zn, Fe, CuO_2, NO_3 4) sulfur and phosphorous starvation 5) gametic. We would expect to see EST expression of Cre01.g055900.t1.2 under the above-mentioned conditions since it initially was defined as a flagella-expressed gene (http://genomes.mcdb.ucla.edu). Functional annotation of the gene product indicates the presence of the three WD-40 repeats (also known as WD or beta-transducin repeats) that consist of short ~40 amino acid motifs, often terminating in a Trp-Asp (W-D) dipeptide. These repeated WD40 motifs have been shown to act as sites of transient protein-protein interactions, and platforms for the assembly of protein complexes or mediators of interplay among other proteins (http://pfam.sanger.ac.uk). We used prediction tools to analyze the protein sequence of Cre01.g055900.t1.2. Toppred prediction using GES and KD algorithms show that the putative protein might have eleven transmembrane domains, and TMM showed none. This gene disruption might be the cause of a fusion-defective phenotype in mutant 2-8; therefore it might be involved in the final stage of the *Chlamydomonas* gametes membrane fusion and it would be interesting to pursue further.

The mutant 2-29 had only one insertion that disrupted a 3´ UTR of the predicted gene (Cre03.g190650.t1.2) in chromosome 3. This gene shows no functional annotation and no

putative conserved domains, except seven gametic ESTs out of a total of 21 (http://genomes.mcdb.ucla.edu) and has no known functional homologues in any other organisms. Upon Blasting the protein product of this gene against the NCBI database the returned results produced matches only with a hypothetical protein from the *Chlamydomonas* genome and not any other organism. We used prediction tools to analyze the protein sequence of Cre03.g190650.t1.2. Toppred prediction using GES and KD algorithms show that the putative protein might have seven transmembrane domains and TMM showed none. Again, further analysis of this gene might reveal if it is indeed involved in gamete fusion.

The other focus of this work was to confirm that in *Chlamydomonas* gametes, secondary tetraprenoid biosythesis is dependent on up-regulation of genes involved in the MEP pathway by employing a comparative phylogenomics *in silico* approach. Over-accumulation of these secondary metabolites in *Chlamydomonas* gametes had been experimentally shown (Martin and Goodenough, 1975) and the significance of the tetraprenoid biosynthesis during developmental processes such as gameogenesis had been pointed out in the past as well (Grossman et al., 2004). With that in mind, we have examined the appearance and reconstructed the evolution of genes and enzymes evolved in isoprenoid C5 precursor molecule biosynthesis in both pathways (MEP and MVA) from prokaryotic to eukaryotic algal genomes. As the complete gene sequences for the MEP pathway of *C.reinhardtii* are publically available (Grossman et al., 2004), a comparative phylogenomics approach was taken to reconstruct the evolution of isoprenoid biosynthesis in different species of algae, including those whose genomes were recently sequenced. To be able to dissect the direction of the gene flow of the MEP and MVA pathways initially we employed a presence-absence analysis. Then we reconstructed the evolution of the MEP and MVA pathways for all of the fully sequenced algal genomes and confirmed that all

investigated algae have the MEP pathway. Also, we confirmed that diatoms have both isoprenoid pathways with MEP pathway enzymes that are closely related to the green (Chlorophyta) and red (Rhodophyta) MEP pathway enzymes. These results are consistent with the previous phylogenetic analyses that were shown by Grauvogel and Petersen 2007, and Frommolt et al., 2008. All green algae belonging to the division Chlorophyta and at least one of the red alga representatives belonging to Rhodophyta showed loss of MVA pathway, which is congruent with the previous biochemical (Schwender et al., 2001) and phylogenetic data (Grauvogel and Petersen, 2007; Frommolt et al., 2008). The individual phylogenies of the MEP pathway enzymes (see supplementary Figure 1) further suggest that eukaryotic algae with oxygenic photosynthesis have inherited these genes from prokaryotes. The horizontal gene transfer has been shown to play a significant role in distribution of the genes leading to biosynthesis of the C5 isoprene unit and explaining their mosaic relationship in both pathways (Lange, 2000). However, we have also found that we don't have enough resolving power to enable us to truly establish evolutionary relationship among different algal species. Therefore, more sequences from additional algal species are needed to determine more accurate evolutionary relationships for enzymes found in both isoprenoid pathways.

For future phylogenetic studies we propose to use syntheny analysis to confirm gene loss due to pseudogenization or deletion, or whether each gene loss took place independently or as a whole group. Moreover, in Chlorophyta missing MVA pathway genes could be verified using synteny analysis, including searching for orthologs of neighboring genes using BLAST, identifying whether the missing ORFs are due to pseudogenization or deletion or transcription silencing. Also, by using synteny data of Streptophyta as an outgroup and Chlorophyta as an ingroup, one would be able to estimate on which branch the deletion or pseudogenization had occurred.

Taken together, the results of our phylogenetic analysis indicating either the presence or loss of MVA and/or MEP pathways in algal species supports a single endosymbiotic event that resulted in the formation of primary photosynthetic eukaryotes. The presence and absence of both pathways in evolutionarily related organisms can shed light on plastid evolutionary origin as well as the genetic forces that acted on the organism under this study.

In conclusion, I have performed a comparative genomic analysis, SF-PCR to confirm the point of insertion, chromosome walk via PCR and analyzed expression of MGS via qRT-PCR in all fusion-defective mutants that could not be complemented with the *wt* copy of the HAP2/GCS1 gene. I have identified, cloned and sequenced the flanking region from the point of insertion in two of the mutants. Also, I have performed systematic and phylogenetic analysis of the MEP and MVA pathways.

Appendix

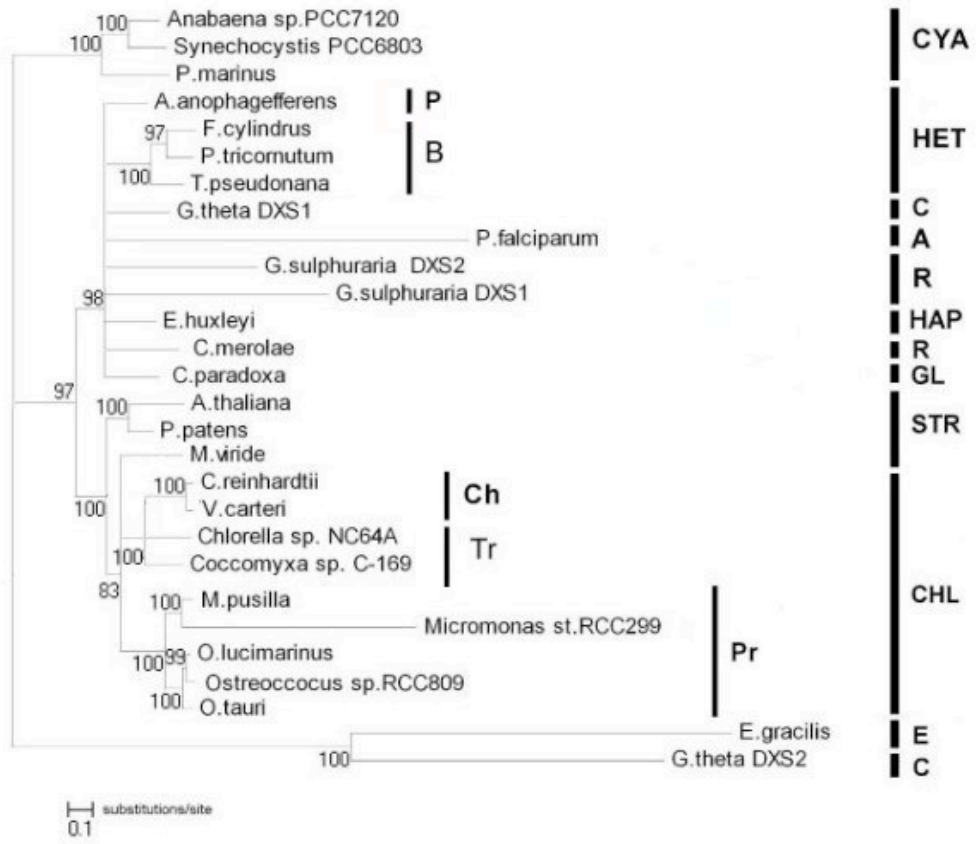
Supplementary Figure 1

Phylogenetic trees for the MEP and MVA proteins involved in isoprenoid biosynthesis. Shown are **(A)** DXS; **(B)** DXR; **(C)** MCT; **(D)** CMK; **(E)** MDS; **(F)** HDS; **(G)** HDR; **(H)** AACT; **(I)** HMGCS; **(J)** HMGCR; **(K)** MVK; **(L)** PMVK; **(M)** MVD.

The results of majority-rule consensus of 5,000 trees saved during Bayesian analyses of MEP and MVA pathway protein sequences are shown. Numbers above and below branches indicate clade frequency (Bayesian posterior probability values). Branch lenghts are proportional to number of substitutions per site (see scale bar). Abbreviations: A, Alveolata; B, Bacillariophyceae; Ch, Chlorophyceae; CHL, Chlorophyta; C, Chromista; CYA, Cyanobacteria; E, Euglenophyta; GL, Glaucophyta; HAP, Haptophyta; HET Hetrokontophyta; M, Metazoa, P, Pelagophyceae; Pr, Prasinophyceae; R, Rhodophyta; T, Trebouxiophyceae.

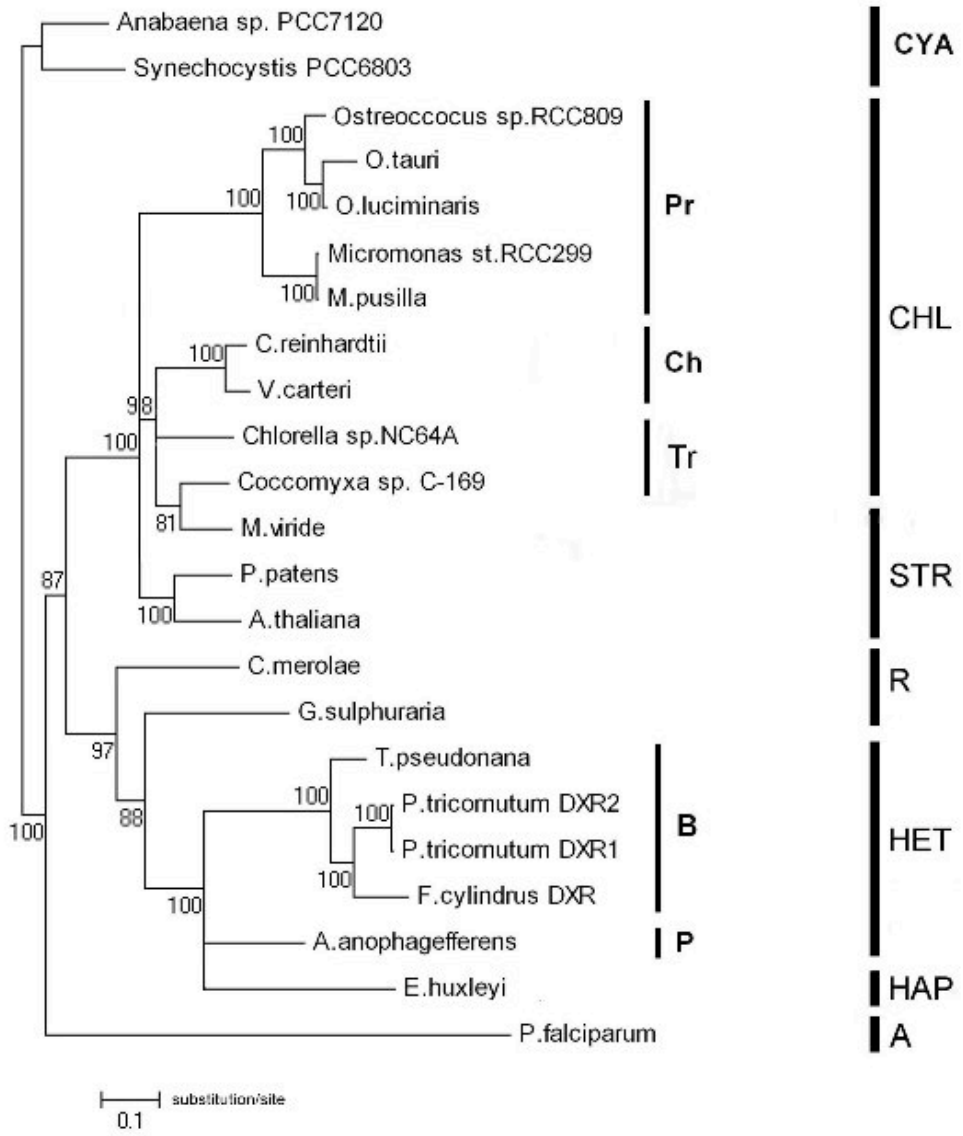
Supplementary Figure 1A

DXS



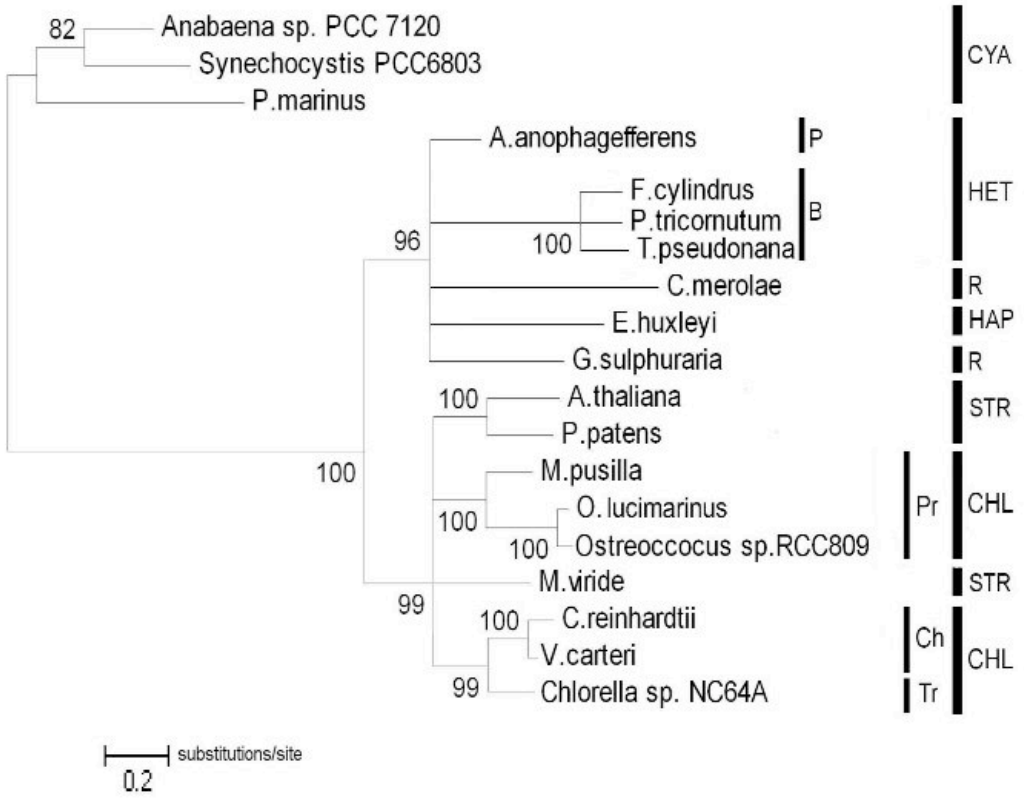
Supplementary Figure 1B

DXR



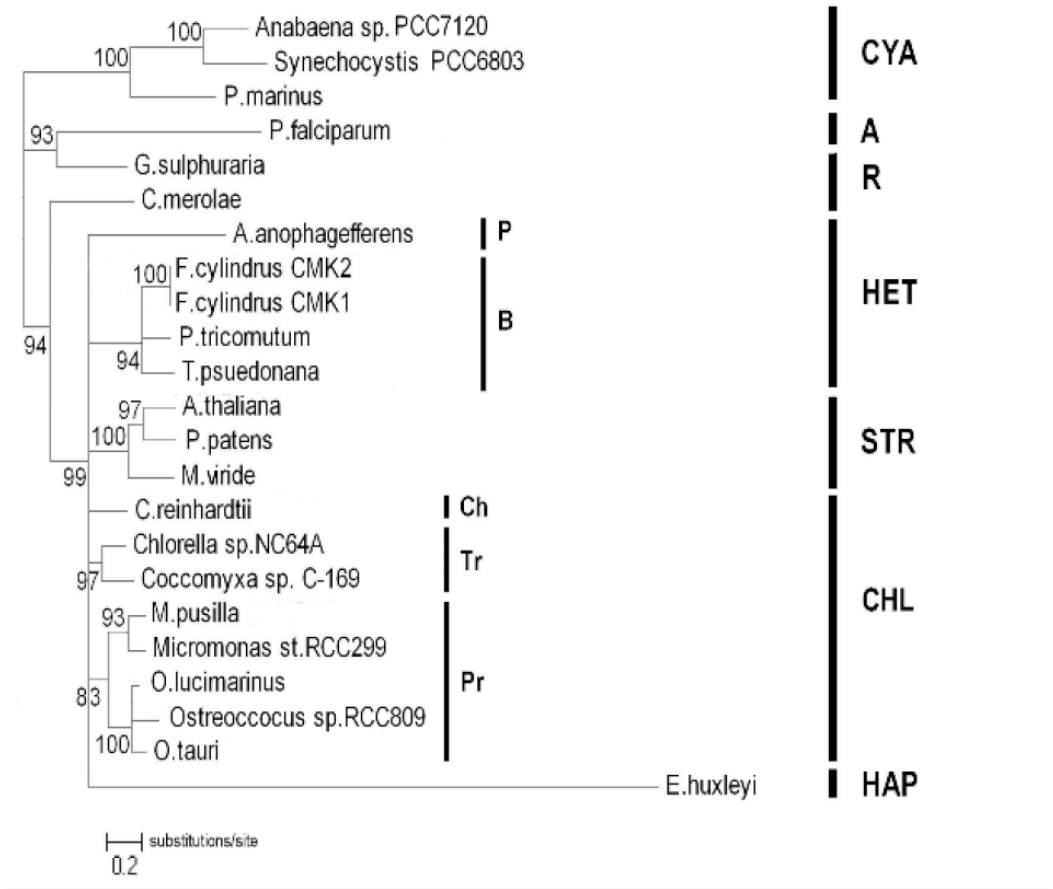
Supplementary Figure 1C

MCT



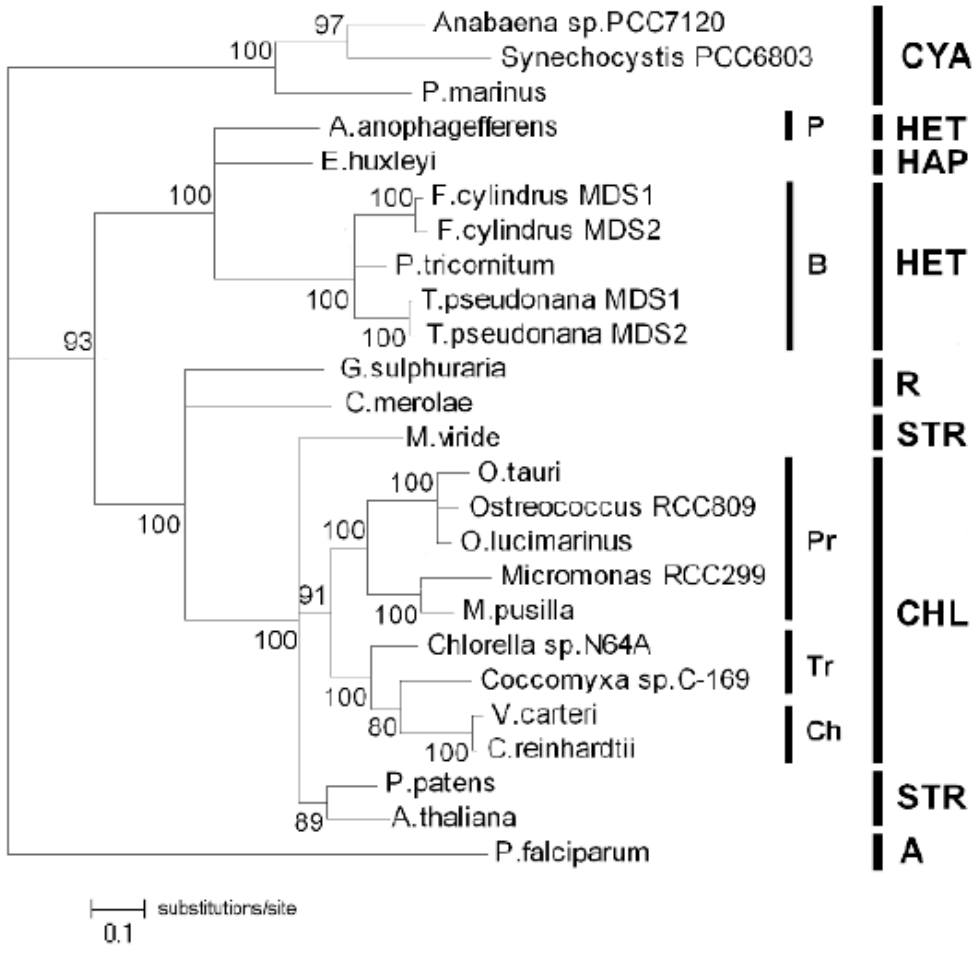
Supplementary Figure 1D

CMK



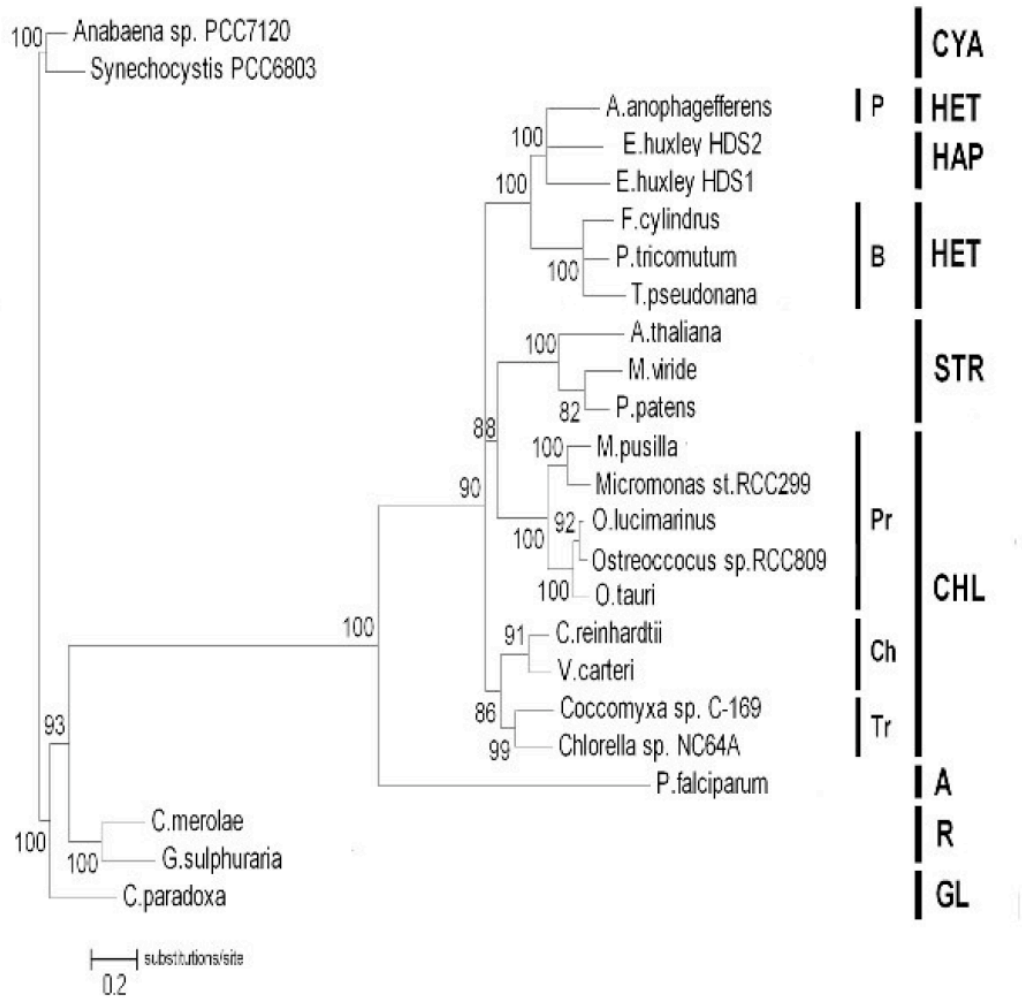
Supplementary Figure 1E

MDS



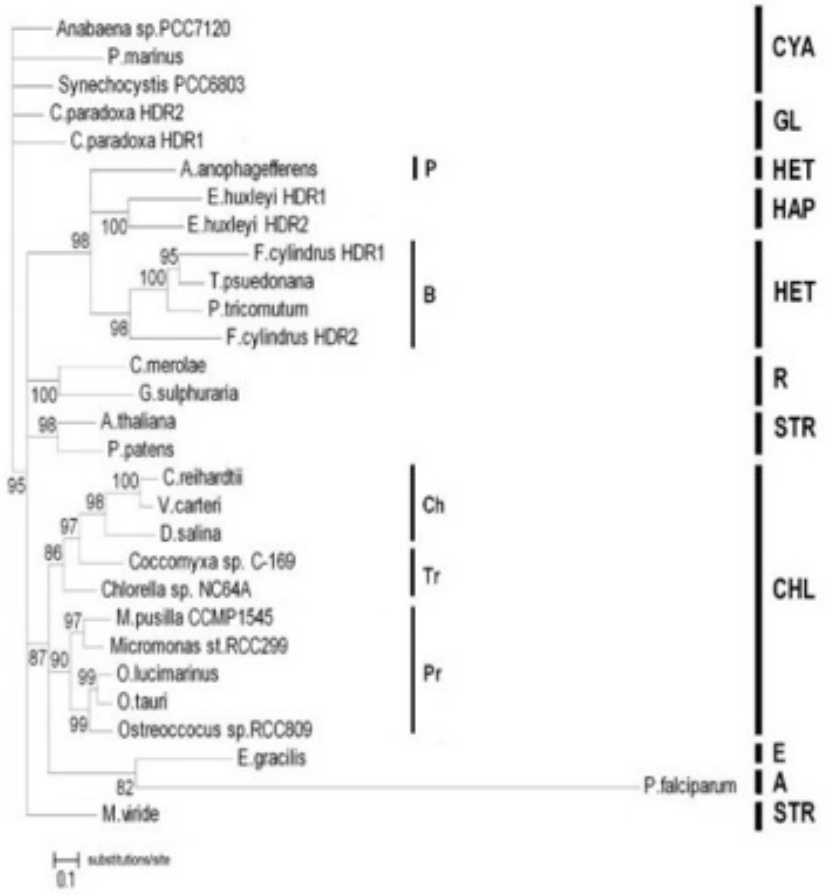
Supplementary Figure 1F

HDS



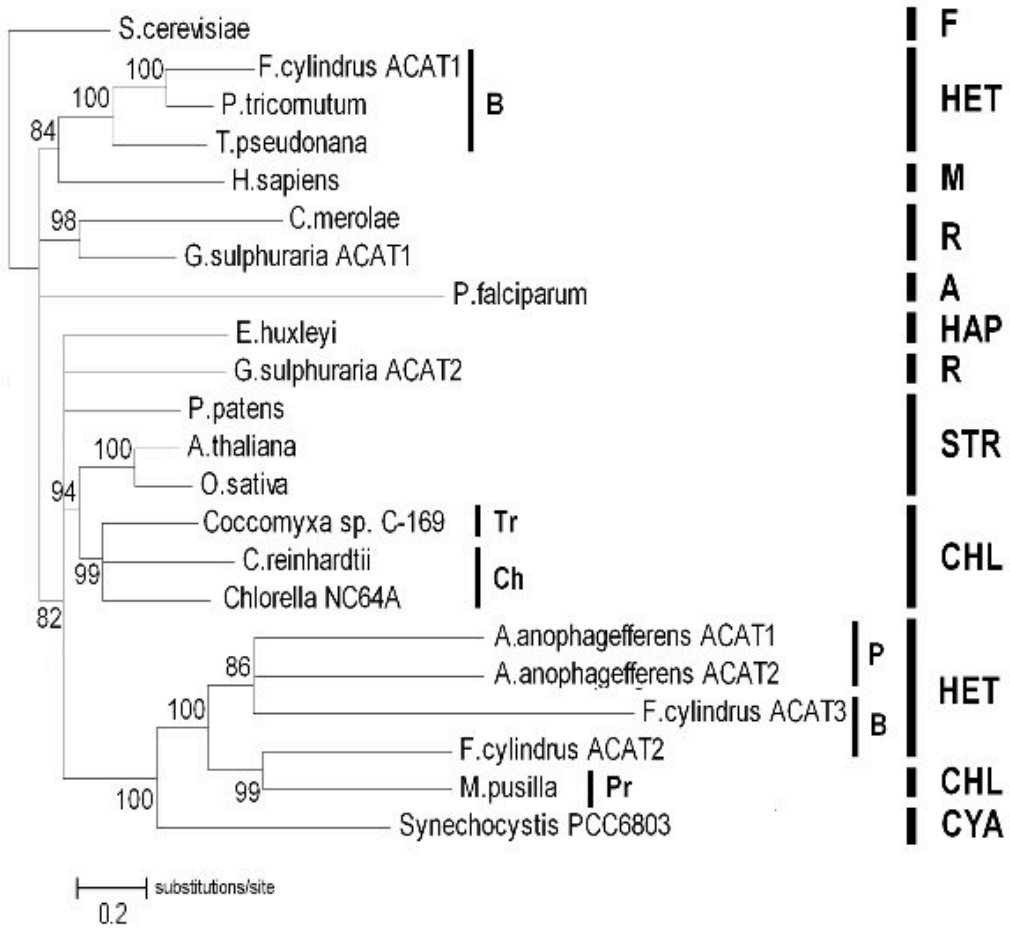
Supplementary Figure 1G

HDR



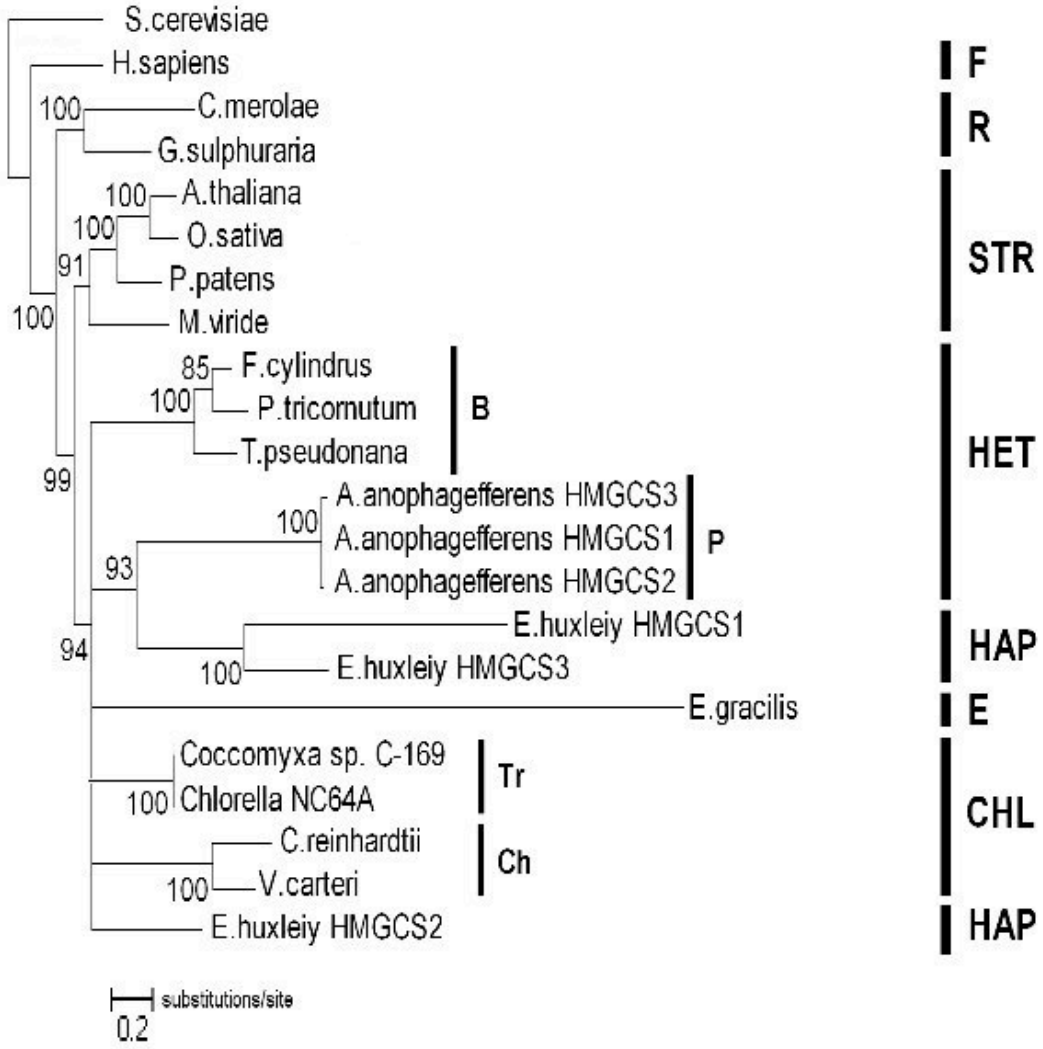
Supplementary Figure 1G

ACAT



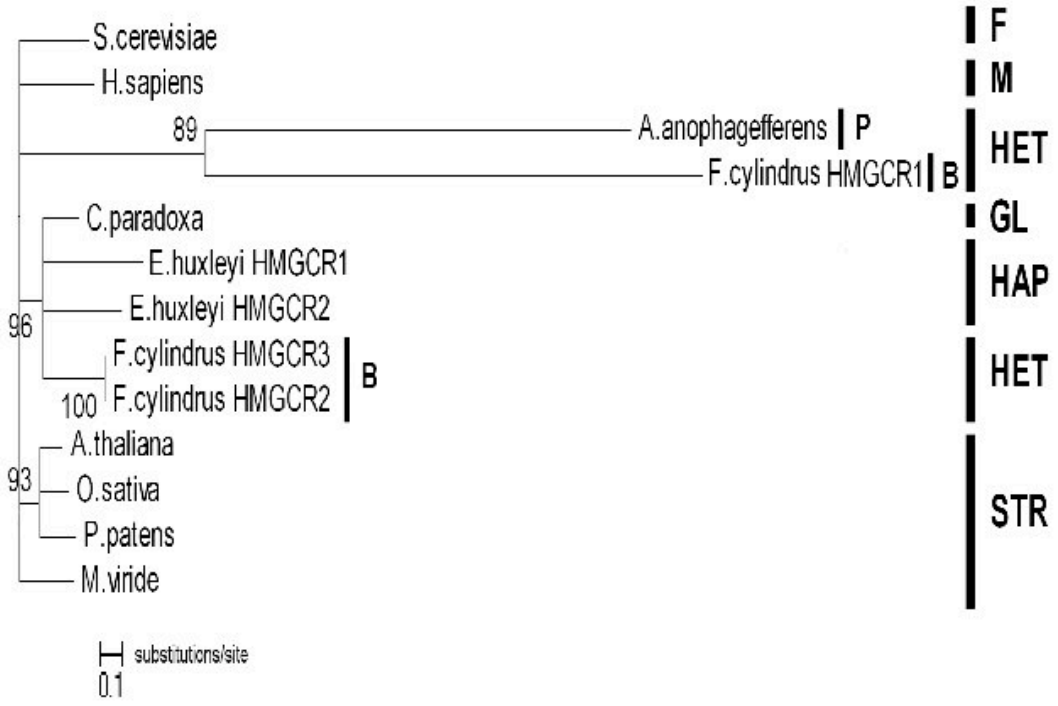
Supplementary Figure 1I

HMGCS



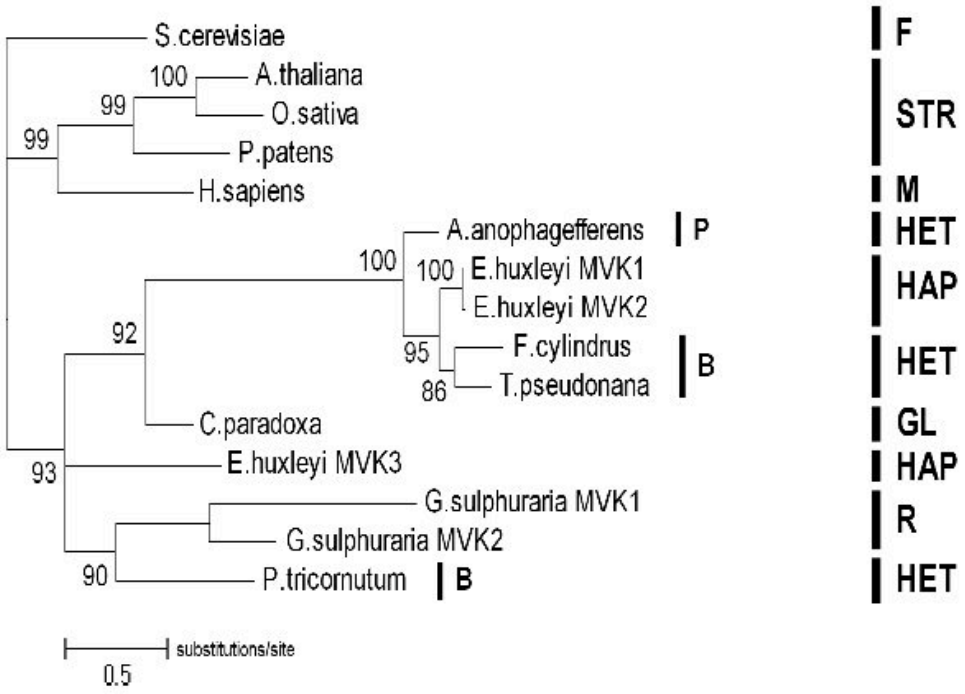
Supplementary Figure J

HMGCR



Supplementary Figure 1K

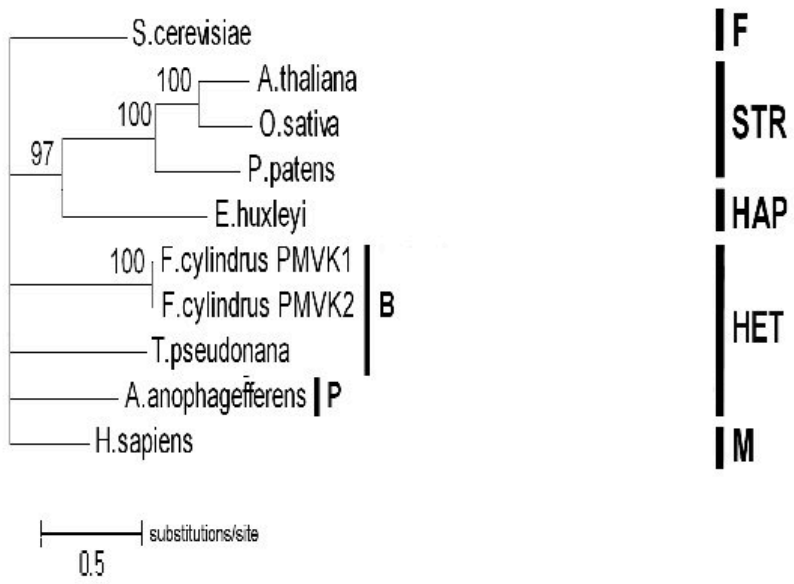
MVK



.

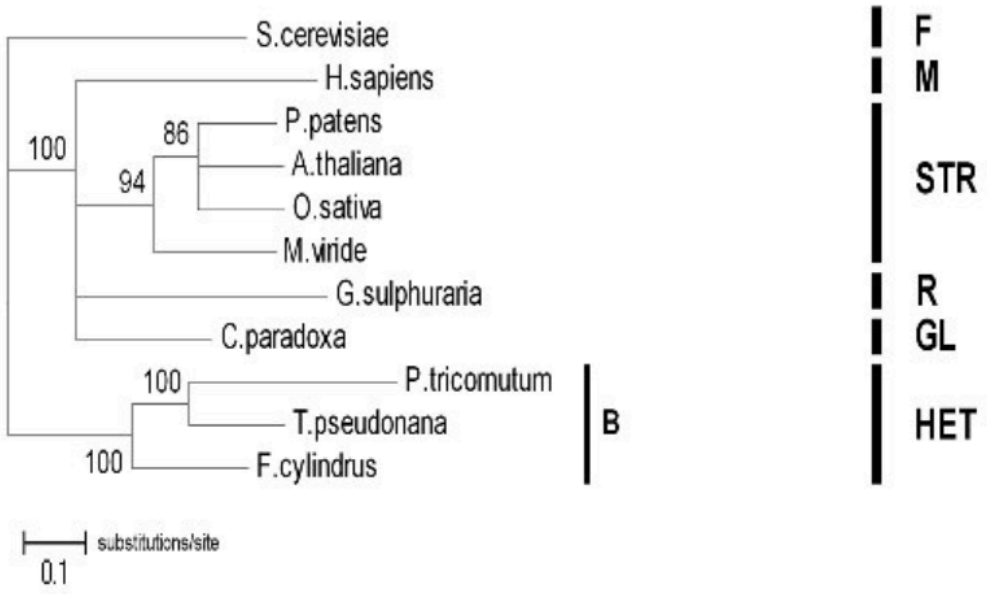
Supplementary Figure 1L

PMVK

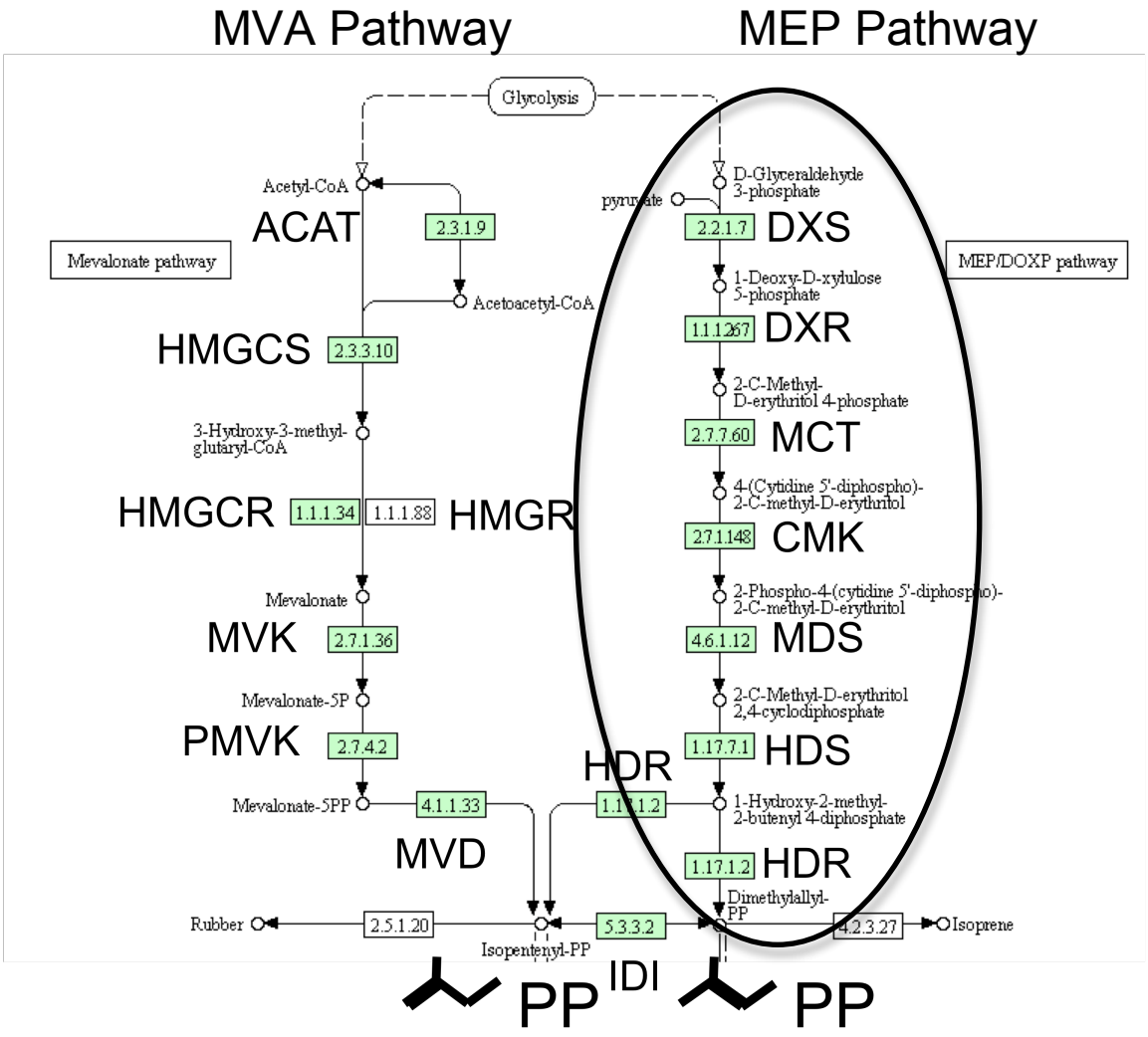


Supplementary Figure 1M

MVD



Supplementary Figure 2
Modified KEGG map showing presence of predicted MEP and MVA pathway proteins in higher plants. Both nucleus-encoded with MVA proteins functioning in the cytosol and MEP chloroplast localized proteins.



Supplementary Table 1

Locations of the genes coding for the MEP pathway. Refer to the main text for abbreviations.

Species	MEP Pathway Gene Locations						
	DXS 2.2.1.7	DXR 1.1.1.267	MCT 2.7.7.60	CMK 2.7.1.148	MDS 4.6.1.12	HDS 1.17.7.1	HDR 1.17.1.2
A.anophagefferens	Auran1scaffold_68:33258-35240	Auran1scaffold_11:984070-985248	Auran1scaffold_5:1078950-1079814	Auran1scaffold_15:338600-340400	Auran1scaffold_4:2046653-2047189	Auran1scaffold_5:261845-272241	Auran1scaffold_9:497989-499191
C.reinhardtii - version4	chromosome_7:5945946-5951103	chromosome_12:7373102-7379873	chromosome_1:3303372-3304946	chromosome_2:9596678-9602933	chromosome_12:2936690-2940785	chromosome_12:6783862-6865526	chromosome_8:2171961-2176321
Chlorella sp. NC64A	ChlNC64A_1scaffold_25:200926-206136	ChlNC64A_1scaffold_3:282752-285133	ChlNC64A_1scaffold_4:225815-227316	ChlNC64A_1scaffold_1:244774-246783	ChlNC64A_1scaffold_8:808337-809855	ChlNC64A_1scaffold_8:173121-23592	ChlNC64A_1scaffold_8:1409019-1412156
Coccomyxa_C-169	scaffold_10:1324614-1329240	scaffold_7:1959550-1964944	scaffold_4:1090552-1092337	scaffold_5:703290-705545	scaffold_14:1533465-1534575	scaffold_3:277842-287711	scaffold_10:1717586-1720830
Emiliania huxleyi CCMP1516	EmihuV1scaffold_106:95671-100053	EmihuV1scaffold_79:138213-141050	EmihuV1scaffold_274:111862-113019	EmihuV1scaffold_1:1699537-1702637	EmihuV1scaffold_4:2008292-2009274 EmihuV1scaffold_1084:791-1590	EmihuV1scaffold_181:99560-100096 EmihuV1scaffold_3735:4615-57533 EmihuV1scaffold_6020:749-1501	EmihuV1scaffold_148:110265-111903 scaffold148:110265-111903 scaffold_172:26120-28176
Micromonas pusilla CCMP1545	MicpuC2scaffold_2:971674-976886	MicpuC2scaffold_6:1177125-1178894	MicpuC2scaffold_1:968949-970360	MicpuC2v5scaffold_14:7333097-734645	MicpuC2v5scaffold_10:194519-195269	MicpuC2scaffold_3:163664-166829	MicpuC2v5scaffold_5:3373713-376881
Micromonas st. RCC299	Chr_01:1619287-1622184	Chr_07:1553648-1554657	Chr_07:7851111-785265 Chr_14:664592-667305	Chr_15:197915-199072	Chr_12:2893316-897150	Chr_14:664592-667305	Chr_06:1051378-1052677
Ostreococcus RCC809	OstRCC809_1scaffold_7:5922298-5925585	OstRCC809_1scaffold_4:5379224-5380630	OstRCC809_2chr_7:645355-646110 scaffold_6:631443-632198	OstRCC809_1scaffold_18:70768-71842	OstRCC809_1scaffold_6:6083137-6086551	OstRCC809_1scaffold_3:132690-134076	OstRCC809_1scaffold_10:177541-178683
Ostreococcus lucimarinus	Ost5901_2chr_2:924826-927185	Ost5901_2chr_4:5711118-5712515	Ost5901_2chr_7:585245-587416	Ost5901_3chr_15:261786-262940	Ost5901_3chr_11:5576-12115	Ost5901_3chr_9:137320-139362	Ost5901_3chr_8:183833-185075
Ostreococcus tauri	Ostta4Chr_02:0001391966-1394191	Ostta4Chr_04:0001529269-1529597	No hits found	Ostta4Chr_16:0001296660-1297705	Ostta4Chr_11:0001:9004-9522	Ostta4Chr_09:0001:136321-137342	Ostta4Chr_08:0001:197366-198956
P.tricornutum	Phatr2_bd0bd_30c35:15762-18609	Phatr2chr_1:2304936-2306581	Phatr2chr_14:469692-481573	Phatr2chr_4:1088916-1100067	Phatr2chr_8:407877-409434	Phatr2chr_5:765110-767756	Phatr2chr_4:980823-982163
Thalassiosira pseudonana	Thaps3chr_7:618633-620856	Thaps3chr_18:502462-504130	Thaps3chr_3:1386903-1387844	Thaps3chr_2:2625219-263557	Thaps3chr_15:273617-274622	Thaps3chr_8:998620-1001699	Thaps3chr_5:1227355-1228798
Volvox carteri	Volca1scaffold_67:394262-401996	Volca1scaffold_37:1134504-1141916	Volca1scaffold_87:344955-345974	no hits found	Volca1scaffold_8:204819-206098	Volca1scaffold_36:1312255-1320564	Volca1scaffold_5:2446445-2448934
Fragilariopsis cylindrus	scaffold_58:143374-145315 scaffold_4:595368-598049	scaffold_1:2688370-2688937	scaffold_1:2062459-2063403	scaffold_2:5587105-5592116 scaffold_69:194974-195483	scaffold_1:5866386-5868360 scaffold_81:77776-78743	scaffold_2:1055332-1057926	scaffold_12:1395204-1398118 scaffold_1:5665461-5667315 scaffold_54:277092-279824
Cyanidioschyzon merolae	cd1f0001:245961>246176	cd1f0002:182200>183904	Location:cd1f0002:180969<182027	c1f90009:710965>712314	c2f01009:229787	c1f20001:730256<731778	c1f0f010:119865>121263

Supplementary Table 2

Locations of the genes coding for the MVA pathway. Refer to the main text for abbreviations.

Species	MVA Pathway Gene Locations						
	AACT_2.3.1.9	HMGCS_2.3.3.10	HMGR_1.1.1.88	HMGR_1.1.1.34	MVK_2.7.1.36	PMVK_2.7.4.2	MVD_4.1.1.33
Aureococcus anophagefferens	scaffold_5:1325769-1326977	scaffold_591:849-1478		scaffold_9:1167186-1168584	scaffold_22:142191-143223	scaffold_6:1983110-1983589	scaffold_227:12658713187
	scaffold_6:1574063-1575535	scaffold_11:1263949-1266578		scaffold_80:25712-37450			
		scaffold_11:1240916-1241465					
Chlamydomonas reinhardtii	chromosome_2:9740346-9744055	chromosome_16:4300361-4301827					
Chlorella sp. NC64A	scaffold_24:409090-411342	scaffold_2:2260770-2264230					
Coccomyxa C-169	scaffold_9:1254942-1255823	scaffold_3:367509-373428					
Emiliania huxleyi CCMP1516	scaffold_121:200160-202037	scaffold_19:280160-282886		scaffold_12:639445-641375	scaffold_185:257367-259442	scaffold_52:193228-196343	scaffold_416:50828-53144
		scaffold_8:1188638-1191282		scaffold_72:79479-82656	scaffold_209:211880-213484		
		scaffold_61:5174285-5186948			scaffold_14:621700-623189		
Micromonas pusilla CCMP1545	scaffold_11:1196119-1197271						
Micromonas strain RCC299	Chr_02:1750907-1752676						
Ostreococcus RCC809	>chr_7:580759|59064						
Ostreococcus lucimarinus	>Chr_6|781022|782203						
Ostreococcus tauri							
Phaeodactylum tricornutum	estExt_gwp_gw1.C_chr_280067	e_gw1.27.40.1		NA	estExt_Phatr1_ua_kg.C_chr_10004	NA	gw1.22.47.1
Thalassiosira pseudonana	estExt_Genewise1.C_chr_60337	thaps1_ua_kg.chr_1000228		NA	fgenesh1_pg.C_chr_2000331	estExt_fgenesh1_pg.C_chr_30609	gw1.1.340.1
Volvox carteri	scaffold_26:475423-485613	scaffold_14:1351838-1356271					
	scaffold_26:458865-463279						
	scaffold_149:456-19285						
Fragilariopsis cylindrus	scaffold_2:1166915-1168581	scaffold_6:2141036-2143064		scaffold_6:1929927-1930533	scaffold_9:1076709-1078067	scaffold_11:650104-651552	scaffold_5:1942173-1943463
	scaffold_3:2521886-2523584	scaffold_46:371150-373133		scaffold_6:605689-606543		scaffold_9:2170582-2172027	scaffold_41:102808-104170
	scaffold_5:1179231-1181026			scaffold_39:10142-10994			
	scaffold_50:140674-150446				scaffold_9:1267142-1268479		

References

Adair, W.S., Hwang, C., and Goodenough, U.W. (1983). Identification and visualization of the sexual agglutinin from the mating-type plus flagellar membrane of *Chlamydomonas*. Cell *33*, 183-193.

Aksoy, M. (2008). Cloning a gamete fusion gene in *Chlamydomonas reinhardtii.* PhD Thesis, *Graduate school and University Center of The City University of New York*.

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

Armbrust, E.V., Berges, J.A., Bowler, C., Green, B.R., Martinez, D., Putnam, N.H., Zhou, S., Allen, A.E., Apt, K.E., Bechner, M*., et al.* (2004). The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science, New York, N.Y *306*, 79-86.

Austin, C.R. (1951). Observations on the penetration of the sperm in the mammalian egg. Aust J Sci Res B *4*, 581-596.

Austin, C.R. (1952). The capacitation of the mammalian sperm. Nature *170*, 326.

Austin, C.R., and Bishop, M.W. (1958). Capacitation of mammalian spermatozoa. Nature *181*, 851.

Bach, T.J. (1995). Some new aspects of isoprenoid biosynthesis in plants--a review. Lipids *30*, 191-202.

Barbier, G., Oesterhelt, C., Larson, M.D., Halgren, R.G., Wilkerson, C., Garavito, R.M., Benning, C., and Weber, A.P. (2005). Comparative genomics of two closely related unicellular thermo-acidophilic red algae, *Galdieria sulphuraria* and C*yanidioschyzon merolae*, reveals the molecular basis of the metabolic flexibility of *Galdieria sulphuraria* and significant differences in carbohydrate metabolism of both algae. Plant physiology *137*, 460-474.

Basanez, G. (2002). Membrane fusion: the process and its energy suppliers. Cell Mol Life Sci *59*, 1478-1490.

Bentz, J. (2000). Minimal aggregate size and minimal fusion unit for the first fusion pore of influenza hemagglutinin-mediated membrane fusion. Biophys J *78*, 227-245.

Bentz, J., and Mittal, A. (2000). Deployment of membrane fusion protein domains during fusion. Cell Biol Int *24*, 819-838.

Berditchevski, F. (2001). Complexes of tetraspanins with integrins: more than meets the eye. Journal of cell science *114*, 4143-4151.

Bergman, K., Goodenough, U.W., Goodenough, D.A., Jawitz, J., and Martin, H. (1975). Gametic differentiation in *Chlamydomonas reinhardtii.* II. Flagellar membranes and the agglutination reaction. J Cell Biol *67*, 606-622.

Bigler, D., Takahashi, Y., Chen, M.S., Almeida, E.A., Osbourne, L., and White, J.M. (2000). Sequence-specific interaction between the disintegrin domain of mouse ADAM 2 (fertilin beta) and murine eggs. Role of the alpha(6) integrin subunit. J Biol Chem *275*, 11576-11584.

Blanc, G., Duncan, G., Agarkova, I., Borodovsky, M., Gurnon, J., Kuo, A., Lindquist, E., Lucas, S., Pangilinan, J., Polle, J.*, et al.* (2010). The *Chlorella variabilis* NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. Plant Cell *22*, 2943-2955.

Bloch, K. (1992). Sterol molecule: structure, biosynthesis and function. Steroids *57*, 378-383.
Bochar, D.A., Stauffacher, C.V., and Rodwell, V.W. (1999). Sequence comparisons reveal two classes of 3-hydroxy-3-methylglutaryl coenzyme A reductase. Molecular genetics and metabolism *66*, 122-127.

Bourque, D.P., Boynton, J.E., and Gillham, N.W. (1971). Studies on the structure and cellular location of various ribosome and ribosomal RNA species in the green alga *Chlamydomonas reinhardtii.* Journal of cell science *8*, 153-183.

Bouvier, F., Rahier, A., and Camara, B. (2005). Biogenesis, molecular regulation and function of plant isoprenoids. Progress in lipid research *44*, 357-429.

Buchanan, M.J., Imam, S.H., Eskue, W.A., and Snell, W.J. (1989). Activation of the cell wall degrading protease, lysin, during sexual signalling in *Chlamydomonas*: the enzyme is stored as an inactive, higher relative molecular mass precursor in the periplasm. The Journal of cell biology *108*, 199-207.

Buffone, M.G., Foster, J.A., and Gerton, G.L. (2008). The role of the acrosomal matrix in fertilization. Int J Dev Biol *52*, 511-522.

Carretero-Paulet, L., Ahumada, I., Cunillera, N., Rodriguez-Concepcion, M., Ferrer, A., Boronat, A., and Campos, N. (2002). Expression and molecular analysis of the *Arabidopsis* DXR gene encoding 1-deoxy-D-xylulose 5-phosphate reductoisomerase, the first committed enzyme of the 2-C-methyl-D-erythritol 4-phosphate pathway. Plant physiology *129*, 1581-1591.

Cassab, G.I. (1998). Plant Cell Wall Proteins. Annual review of plant physiology and plant molecular biology *49*, 281-309.

Chang, M.C., and Pincus, G. (1951). Physiology of fertilization in mammals. Physiol Rev *31*, 1-26.

Chatterjee, I., Richmond, A., Putiri, E., Shakes, D.C., and Singson, A. (2005). The *Caenorhabditis elegans* spe-38 gene encodes a novel four-pass integral membrane protein required for sperm function at fertilization. Development (Cambridge, England) *132*, 2795-2808.

Chen, E.H., and Olson, E.N. (2005). Unveiling the mechanisms of cell-cell fusion. Science, New York, N.Y *308*, 369-373.

Cheng, K.-C., Klancer, R., Singson, A., and Seydoux, G. (2009). Regulation of MBK-2/DYRK by CDK-1 and the pseudophosphatases EGG-4 and EGG-5 during the oocyte-to-embryo transition. Cell *139*, 560.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., and Thompson, J.D. (2003). Multiple sequence alignment with the Clustal series of programs. Nucleic acids research *31*, 3497-3500.

Chernomordik, L.V., and Kozlov, M.M. (2005). Membrane hemifusion: crossing a chasm in two leaps. Cell *123*, 375-382.

Chernomordik, L.V., and Kozlov, M.M. (2008). Mechanics of membrane fusion. Nat Struct Mol Biol *15*, 675-683.

Cho, C., Bunch, D.O., Faure, J.E., Goulding, E.H., Eddy, E.M., Primakoff, P., and Myles, D.G. (1998). Fertilization defects in sperm from mice lacking fertilin beta. Science, New York, N.Y *281*, 1857-1859.

Churchward, M.A., Rogasevskaia, T., Hofgen, J., Bau, J., and Coorssen, J.R. (2005). Cholesterol facilitates the native mechanism of Ca2+-triggered membrane fusion. Journal of cell science *118*, 4833-4848.

Claes, H. (1971). Autolysis of the cell wall of gametes of *Chlamydomonas reinhardtii*. Archiv fur Mikrobiologie *78*, 180-188.

Clark, G.F., and Dell, A. (2006). Molecular models for murine sperm-egg binding. J Biol Chem *281*, 13853-13856.

Cooper, J.B., Adair, W.S., Mecham, R.P., and Heuser, J.E. (1983). *Chlamydomonas* agglutinin is a hydroxyproline-rich glycoprotein. Proc Natl Acad Sci USA *80*, 5898-5901.

Corcoran, J.A., Salsman, J., de Antueno, R., Touhami, A., Jericho, M.H., Clancy, E.K., and Duncan, R. (2006). The p14 fusion-associated small transmembrane (FAST) protein effects membrane fusion from a subset of membrane microdomains. J Biol Chem *281*, 31778-31789.

Cunningham, F.X., Jr., and Gantt, E. (2000). Identification of multi-gene families encoding isopentenyl diphosphate isomerase in plants by heterologous complementation in *Escherichia coli*. Plant & cell physiology *41*, 119-123.

Curtis, B.A., Tanifuji, G., Burki, F., Gruber, A., Irimia, M., Maruyama, S., Arias, M.C., Ball, S.G., Gile, G.H., Hirakawa, Y.*, et al.* (2012). Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. Nature *492*, 59-65.

Derelle, E., Ferraz, C., Rombauts, S., Rouze, P., Worden, A.Z., Robbens, S., Partensky, F., Degroeve, S., Echeynie, S., Cooke, R.*, et al.* (2006). Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. Proc Natl Acad Sci USA *103*, 11647-11652.

Detmers, P.A., Goodenough, U.W., and Condeelis, J. (1983). Elongation of the fertilization tubule in *Chlamydomonas*: new observations on the core microfilaments and the effect of transient intracellular signals on their structural integrity. J Cell Biol *97*, 522-532.

Eddy, E., and O'Brien, D. (1994). The spermatozoon. In The Physiology of reproduction (ed. E.Knobil and J.D Neil) *Raven Press: New York*, 29-68.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research *32*, 1792-1797.

Eto, K., Huet, C., Tarui, T., Kupriyanov, S., Liu, H.Z., Puzon-McLaughlin, W., Zhang, X.P., Sheppard, D., Engvall, E., and Takada, Y. (2002). Functional classification of ADAMs based on a conserved motif for binding to integrin alpha 9beta 1: implications for sperm-egg binding and other cell interactions. J Biol Chem *277*, 17804-17810.

Evans, J.P. (2012). Sperm-Egg Interaction. Annu Rev Physiol.

Felsenstein, J. (1989). Mathematics vs. Evolution: Mathematical Evolutionary Theory. Science New York, N.Y *246*, 941-942.

Ferris, P.J., Armbrust, E.V., and Goodenough, U.W. (2002). Genetic structure of the mating-type locus of *Chlamydomonas reinhardtii.* Genetics *160*, 181-200.

Ferris, P.J., Waffenschmidt, S., Umen, J.G., Huawen, L., Jae-Hyeok, L., Ishida, K., Kubo, T., Lau, J., and Goodenough, U.W. (2005). Plus and Minus Sexual Agglutinins from *Chlamydomonas reinhardtii.* Plant Cell *17*, 597-615.

Ferris, P.J., Woessner, J.P., and Goodenough, U.W. (1996). A sex recognition glycoprotein is encoded by the plus mating-type gene fus1 of *Chlamydomonas reinhardtii.* Molecular biology of the cell *7*, 1235-1248.

Florman, H., and Ducibella, T. (2006). Fertilization in Mammals. Knobil and Neill's Physiology of Reproduction *1*, 55-112.

Forest, C.L. (1983). Specific contact between mating structure membranes observed in conditional fusion-defective *Chlamydomonas* mutants. Experimental cell research *148*, 143-154.

Forest, C.L. (1987). Genetic control of plasma membrane adhesion and fusion in *Chlamydomonas* gametes. Journal of cell science *88 (Pt 5)*, 613-621.

Forest, C.L., Goodenough, D.A., and Goodenough, U.W. (1978). Flagellar membrane agglutination and sexual signaling in the conditional GAM-1 mutant of *Chlamydomonas*. J Cell Biol *79*, 74-84.

Forest, C.L., and Togasaki, R.K. (1975). Selection for conditional gametogenesis in *Chlamydomonas reinhardtii.* Proc Natl Acad Sci U S A *72*, 3652-3655.

Friedmann, I., Colwin, A.L., and Colwin, L.H. (1968). Fine-structural aspects of fertilization in *Chlamydomonas reinhardtii.* Journal of cell science *3*, 115-128.

Frommolt, R., Werner, S., Paulsen, H., Goss, R., Wilhelm, C., Zauner, S., Maier, U.G., Grossman, A.R., Bhattacharya, D., and Lohr, M. (2008). Ancient recruitment by chromists of green algal genes encoding enzymes for carotenoid biosynthesis. Mol Biol Evol *25*, 2653-2667.

Fujihara, Y., Murakami, M., Inoue, N., Satouh, Y., Kaseda, K., Ikawa, M., and Okabe, M. (2010). Sperm equatorial segment protein 1, SPESP1, is required for fully fertile sperm in mouse. Journal of cell science *123*, 1531-1536.

Gao, A.C., Lou, W., Dong, J.T., and Isaacs, J.T. (1997). CD44 is a metastasis suppressor gene for prostatic cancer located on human chromosome 11p13. Cancer Res *57*, 846-849.

Gao, B., Klein, L.E., Britten, R.J., and Davidson, E.H. (1986). Sequence of mRNA coding for bindin, a species-specific sea urchin sperm protein required for fertilization. Proc Natl Acad Sci U S A *83*, 8634-8638.

Gertz, E.M., Yu, Y.K., Agarwala, R., Schaffer, A.A., and Altschul, S.F. (2006). Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. BMC biology *4*, 41.

Gilbert, S. (2003). Developmental Biology. Seventh Edition. Sinauer Associates, Inc., Sunderland, MA.

Glabe, C.G. (1985a). Interaction of the sperm adhesive protein, bindin, with phospholipid vesicles. I. Specific association of bindin with gel-phase phospholipid vesicles. J Cell Biol *100*, 794-799.

Glabe, C.G. (1985b). Interaction of the sperm adhesive protein, bindin, with phospholipid vesicles. II. Bindin induces the fusion of mixed-phase vesicles that contain phosphatidylcholine and phosphatidylserine in vitro. J Cell Biol *100*, 800-806.

Glabe, C.G., Grabel, L.B., Vacquier, V.D., and Rosen, S.D. (1982). Carbohydrate specificity of sea urchin sperm bindin: a cell surface lectin mediating sperm-egg adhesion. J Cell Biol *94*, 123-128.

Glabe, C.G., and Lennarz, W.J. (1979). Species-specific sperm adhesion in sea urchins. A quantitative investigation of bindin-mediated egg agglutination. J Cell Biol *83*, 595-604.

Glabe, C.G., and Vacquier, V.D. (1977). Isolation and characterization of the vitelline layer of sea urchin eggs. J Cell Biol *75*, 410-421.

Glaser, R.W., Grune, M., Wandelt, C., and Ulrich, A.S. (1999). Structure analysis of a fusogenic peptide sequence from the sea urchin fertilization protein bindin. Biochemistry *38*, 2560-2569.

Goodenough, U., Lin, H., and Lee, J.H. (2007). Sex determination in *Chlamydomonas*. Seminars in cell & developmental biology *18*, 350-361.

Goodenough, U.W. (1989). Cyclic AMP enhances the sexual agglutinability of *Chlamydomonas* flagella. J Cell Biol *109*, 247-252.

Goodenough, U.W., Armbrust, E.V., Campbell, A.M., and Ferris, P.J. (1995). Molecular genetics of sexuality in *Chlamydomonas*. Annu. Rev. Plant Physiol. Plant Mol. Biol., 21-44.

Goodenough, U.W., Detmers, P.A., and Hwang, C. (1982). Activation for cell fusion in *Chlamydomonas*: analysis of wild-type gametes and nonfusing mutants. J Cell Biol *92*, 378-386.

Goodenough, U.W., Hwang, C., and Warren, A.J. (1978). Sex-limited expression of gene Loci controlling flagellar membrane agglutination in the *Chlamydomonas* mating reaction. Genetics *89*, 235-243.

Goodenough, U.W., and Weiss, R.L. (1975). Gametic differentiation in *Chlamydomonas reinhardtii.* III. Cell wall lysis and microfilament-associated mating structure activation in wild-type and mutant strains. J Cell Biol *67*, 623-637.

Grauvogel, C., and Petersen, J. (2007). Isoprenoid biosynthesis authenticates the classification of the green alga *Mesostigma viride* as an ancient streptophyte. Gene *396*, 125-133.

Grossman, A.R., Lohr, M., and Im, C.S. (2004). *Chlamydomonas reinhardtii* in the landscape of pigments. Annu Rev Genet *38*, 119-173.

Han, C., Choi, E., Park, I., Lee, B., Jin, S., Kim do, H., Nishimura, H., and Cho, C. (2009). Comprehensive analysis of reproductive ADAMs: relationship of ADAM4 and ADAM6 with an ADAM complex required for fertilization in mice. Biology of reproduction *80*, 1001-1008.

Harada, Y., Kawazoe, M., Eto, Y., Ueno, S., and Iwao, Y. (2011). The Ca2+ increase by the sperm factor in physiologically polyspermic newt fertilization: its signaling mechanism in egg cytoplasm and the species-specificity. Developmental biology *351*, 266-276.

Harris, E.H., Burkhart, B.D., Gillham, N.W., and Boynton, J.E. (1989). Antibiotic resistance mutations in the chloroplast 16S and 23S rRNA genes of *Chlamydomonas reinhardtii:* correlation of genetic and physical maps of the chloroplast genome. Genetics *123*, 281-292.

Harrison, S.C. (2008). Viral membrane fusion. Nat Struct Mol Biol *15*, 690-698.

Horsley, V., and Pavlath, G.K. (2004). Forming a multinucleated cell: molecules that regulate myoblast fusion. Cells Tissues Organs *176*, 67-78.

Hsieh, M.H., and Goodman, H.M. (2005). The *Arabidopsis* IspH homolog is involved in the plastid nonmevalonate pathway of isoprenoid biosynthesis. Plant physiology *138*, 641-653.

Huang, T.T., Jr., and Yanagimachi, R. (1985). Inner acrosomal membrane of mammalian spermatozoa: its properties and possible functions in fertilization. Am J Anat *174*, 249-268.

Hudock, G.A., and Rosen, H. (1976). Formal Genetics of *Chlamydomonas reinhardtii.* inThe Genetics of Algae, R.A. Lewin, ed., *Blackwell Scientific Publications*, 29-48.

Huelsenbeck, J.P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics (Oxford, England) *17*, 754-755.

Inoue, N., Kasahara, T., Ikawa, M., and Okabe, M. (2010). Identification and disruption of sperm-specific angiotensin converting enzyme-3 (ACE3) in mouse. PloS one *5*, e10301.

Inoue, N., Yamaguchi, R., Ikawa, M., and Okabe, M. (2007). Sperm-egg interaction and gene manipulated animals. Soc Reprod Fertil Suppl *65*, 363-371.

Jaffe, L.A., and Cross, N.L. (1986). Electrical regulation of sperm-egg fusion. Annu Rev Physiol *48*, 191-200.

Jahn, R., Lang, T., and Sudhof, T.C. (2003). Membrane fusion. Cell *112*, 519-533.

Jahn, R., and Scheller, R.H. (2006). SNAREs--engines for membrane fusion. Nature reviews *7*, 631-643.

Johnson, M.A., von Besser, K., Zhou, Q., Smith, E., Aux, G., Patton, D., Levin, J.Z., and Preuss, D. (2004). *Arabidopsis* hapless mutations define essential gametophytic functions. Genetics *168*, 971-982.

Jovine, L., Park, J., and Wassarman, P.M. (2002). Sequence similarity between stereocilin and otoancorin points to a unified mechanism for mechanotransduction in the mammalian inner ear. BMC Cell Biol *3*, 28.

Kaji, K., Oda, S., Shikano, T., Ohnuki, T., Uematsu, Y., Sakagami, J., Tada, N., Miyazaki, S., and Kudo, A. (2000). The gamete fusion process is defective in eggs of Cd9-deficient mice. Nat Genet *24*, 279-282.

Kanandale, P., Stewart-Michaelis, A., Gordon, S., Rubin, J., and Klancer, R. (2005). The egg surface LDL receptor repeat-containing proteins EGG-1 and EGG-2 are required for fertilization in *Caenorhabditis elegans*. Curr. Biol. *15*, 2222-2229.

Kates, J.R., Chiang, K.S., and Jones, R.F. (1968). Studies on DNA replication during synchronized vegetative growth and gametic differentiation in *Chlamydomonas reinhardtii.* Experimental cell research *49*, 121-135.

Kinoshita, T., Fukuzawa, H., Shimada, T., Saito, T., and Matsuda, Y. (1992). Primary structure and expression of a gamete lytic enzyme in *Chlamydomonas reinhardtii:* similarity of functional domains to matrix metalloproteases. Proc Natl Acad Sci U S A *89*, 4693-4697.

Kontani, K., Moskowitz, I.P., and Rothman, J.H. (2005). Repression of cell-cell fusion by components of the *C. elegans* vacuolar ATPase complex. Developmental cell *8*, 787-794.

Kresge, N., Vacquier, V.D., and Stout, C.D. (2001). The crystal structure of a fusagenic sperm protein reveals extreme surface properties. Biochemistry *40*, 5407-5413.

Kroft, T.L., Gleason, E.J., and L'Hernault, S.W. (2005). The spe-42 gene is required for sperm-egg interactions during *C. elegans* fertilization and encodes a sperm-specific transmembrane protein. Developmental biology *286*, 169-181.

Kuhn, R.J., Zhang, W., Rossmann, M.G., Pletnev, S.V., Corver, J., Lenches, E., Jones, C.T., Mukhopadhyay, S., Chipman, P.R., Strauss, E.G.*, et al.* (2002). Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. Cell *108*, 717-725.

La, T. (1987). Master Thesis. Brooklyn College *The City University of New York.*

Lai, T. (2011). Master Thesis. Brooklyn College *The City University of New York.*

Lam, J. (1991). Master Thesis. Brooklyn College *The City University of New York.*

Lange, B.M., and Ghassemian, M. (2003). Genome organization in *Arabidopsis thaliana*: a survey for genes involved in isoprenoid and chlorophyll metabolism. Plant molecular biology *51*, 925-948.

Lange, B.M., Rujan, T., Martin, W., and Croteau, R. (2000). Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. Proc Natl Acad Sci U S A *97*, 13172-13177.

Laule, O., Furholz, A., Chang, H.S., Zhu, T., Wang, X., Heifetz, P.B., Gruissem, W., and Lange, M. (2003). Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis in *Arabidopsis thaliana*. Proc Natl Acad Sci U S A *100*, 6866-6871.

Le Naour, F., Rubinstein, E., Jasmin, C., Prenant, M., and Boucheix, C. (2000). Severely reduced female fertility in CD9-deficient mice. Science (New York, N.Y *287*, 319-321.

Lentz, B.R., Malinin, V., Haque, M.E., and Evans, K. (2000). Protein machines and lipid assemblies: current views of cell membrane fusion. Curr Opin Struct Biol *10*, 607-615.

Lichtenthaler, H.K. (1999). The 1-Deoxy-D-Xylulose-5-Phosphate Pathway of Isoprenoid Biosynthesis in Plants. Annual review of plant physiology and plant molecular biology *50*, 47-65.

Lichtenthaler, H.K., Schwender, J., Disch, A., and Rohmer, M. (1997). Biosynthesis of isoprenoids in higher plant chloroplasts proceeds via a mevalonate-independent pathway. FEBS letters *400*, 271-274.

Lin, H., and Goodenough, U.W. (2007). Gametogenesis in the *Chlamydomonas reinhardtii* minus mating type is controlled by two genes, MID and MTD1. Genetics *176*, 913-925.

Liu, Y., Tewari, R., Ning, J., Blagborough, A.M., Garbom, S., Pei, J., Grishin, N.V., Steele, R.E., Sinden, R.E., Snell, W.J., and Billker, O. (2008a). The conserved plant sterility gene HAP2 functions after attachment of fusogenic membranes in *Chlamydomonas* and *Plasmodium* gametes. Genes & development *22*, 1051-1068.

Liu, Y., Tewari, R., Ning, J., Blagborough, A.M., Garbom, S., Pei, J., Grishin, N.V., Steele, R.E., Sinden, R.E., Snell, W.J., and Billker, O. (2008b). The conserved plant sterility gene HAP2 functions after attachment of fusogenic membranes in *Chlamydomonas* and *Plasmodium* gametes. Genes & development.

Lopo, A.C., Glabe, C.G., Lennarz, W.J., and Vacquier, V.D. (1982). Sperm-egg binding events during sea urchin fertilization. Ann N Y Acad Sci *383*, 405-425.

Martin, N.C., and Goodenough, U.W. (1975). Gametic differentiation in *Chlamydomonas reinhardtii.* I. Production of gametes and their fine structure. J Cell Biol *67*, 587-605.

Maruyama, R., Velarde, N.V., Klancer, R., Gordon, S., Kadandale, P., Parry, J.M., Hang, J.S., Rubin, J., Stewart-Michaelis, A., Schweinsberg, P.*, et al.* (2007). EGG-3 regulates cell-surface and cortex rearrangements during egg activation in *Caenorhabditis elegans*. Curr Biol *17*, 1555-1560.

Matsuda, Y., Saito, T., Yamaguchi, T., and Kawase, H. (1985). Cell wall lytic enzyme released by mating gametes of *Chlamydomonas reinhardtii* is a metalloprotease and digests the sodium perchlorate-insoluble component of cell wall. The Journal of biological chemistry *260*, 6373-6377.

Matsuzaki, M., Misumi, O., Shin, I.T., Maruyama, S., Takahara, M., Miyagishima, S.Y., Mori, T., Nishida, K., Yagisawa, F., Nishida, K.*, et al.* (2004). Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. Nature *428*, 653-657.

Matthews, P.D., and Wurtzel, E.T. (2000). Metabolic engineering of carotenoid accumulation in *Escherichia coli* by modulation of the isoprenoid precursor pool with expression of deoxyxylulose phosphate synthase. Applied microbiology and biotechnology *53*, 396-400.

McClure, M.O., Marsh, M., and Weiss, R.A. (1988). Human immunodeficiency virus infection of CD4-bearing cells occurs by a pH-independent mechanism. Embo J *7*, 513-518.

Mesland, D.A., Hoffman, J.L., Caligor, E., and Goodenough, U.W. (1980). Flagellar tip activation stimulated by membrane adhesions in *Chlamydomonas* gametes. J Cell Biol *84*, 599-617.

Misamore, M.J., Gupta, S., and Snell, W.J. (2003). The *Chlamydomonas* Fus1 protein is present on the mating type plus fusion organelle and required for a critical membrane adhesion event during fusion with minus gametes. Molecular biology of the cell *14*, 2530-2542.

Miyado, K., Yamada, G., Yamada, S., Hasuwa, H., Nakamura, Y., Ryu, F., Suzuki, K., Kosai, K., Inoue, K., Ogura, A.*, et al.* (2000). Requirement of CD9 on the egg plasma membrane for fertilization. Science (New York, N.Y *287*, 321-324.

Moellering, E.R., and Benning, C. (2009). RNA interference silencing of a major lipid droplet protein affects lipid droplet size in *Chlamydomonas reinhardtii.* Eukaryotic cell *9*, 97-106.

Mori, T., Kuroiwa, H., Higashiyama, T., and Kuroiwa, T. (2006). GENERATIVE CELL SPECIFIC 1 is essential for angiosperm fertilization. Nature cell biology *8*, 64-71.

Mori, T., Kuroiwa, H., Higashiyama, T., and Kuroiwa, T. (2006). GENERATIVE CELL SPECIFIC 1 is essential for angiosperm fertilization. Nature cell biology *8*, 64-71.

Nishimura, H., Cho, C., Branciforte, D.R., Myles, D.G., and Primakoff, P. (2001). Analysis of loss of adhesive function in sperm lacking cyritestin or fertilin beta. Developmental biology *233*, 204-213.

Page, R., and Holmes, E. (2005). Molecular Evolution: A Phylogenetic Approach. Wiley-Blackwell.

Palenik, B., Grimwood, J., Aerts, A., Rouze, P., Salamov, A., Putnam, N., Dupont, C., Jorgensen, R., Derelle, E., Rombauts, S.*, et al.* (2007). The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. Proc Natl Acad Sci USA *104*, 7705-7710.

Parry, J., Velarde, N., Lefkovith, A., Zegarek, M., and Hang, J. (2009). EGG-4 and EGG-5 link events of the oocyte-to-embryo transition with meiotic progression in *C. elegans*. Cuur. Biol. *19*, 1752.

Pasquale, S.M., and Goodenough, U.W. (1987). Cyclic AMP functions as a primary sexual signal in gametes of *Chlamydomonas reinhardtii.* J Cell Biol *105*, 2279-2292.

Podbilewicz, B., Leikina, E., Sapir, A., Valansi, C., Suissa, M., Shemer, G., and Chernomordik, L.V. (2006). The *C. elegans* developmental fusogen EFF-1 mediates homotypic fusion in heterologous cells and in vivo. Developmental cell *11*, 471-481.

Ramos, A., Coesel, S., Marques, A., Rodrigues, M., Baumgartner, A., Noronha, J., Rauter, A., Brenig, B., and Varela, J. (2008). Isolation and characterization of a stress-inducible *Dunaliella salina* Lcy-beta gene encoding a functional lycopene beta-cyclase. Applied microbiology and biotechnology *79*, 819-828.

Rodriguez-Concepcion, M., and Boronat, A. (2002). Elucidation of the methylerythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids. A metabolic milestone achieved through genomics. Plant physiology *130*, 1079-1089.

Rohmer, M. (1999). The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. Natural product reports *16*, 565-574.

Rohmer, M., Knani, M., Simonin, P., Sutter, B., and Sahm, H. (1993). Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate. The Biochemical journal *295 (Pt 2)*, 517-524.

Rubinstein, E., Ziyyat, A., Prenant, M., Wrobel, E., Wolf, J.P., Levy, S., Le Naour, F., and Boucheix, C. (2006). Reduced fertility of female mice lacking CD81. Developmental biology *290*, 351-358.

Sager, R., and Granick, S. (1954). Nutritional control of sexuality in *Chlamydomonas reinhardtii.* The Journal of general physiology *37*, 729-742.

Salsman, J., Top, D., Barry, C., and Duncan, R. (2008). A virus-encoded cell-cell fusion machine dependent on surrogate adhesins. PLoS Pathog *4*, e1000016.

Sapir, A., Choi, J., Leikina, E., Avinoam, O., Valansi, C., Chernomordik, L.V., Newman, A.P., and Podbilewicz, B. (2007). AFF-1, a FOS-1-regulated fusogen, mediates fusion of the anchor cell in *C. elegans*. Dev Cell *12*, 683-698.

Schmidt, M., Gessner, G., Luff, M., Heiland, I., Wagner, V., Kaminski, M., Geimer, S., Eitzinger, N., Reissenweber, T., Voytsekh, O*., et al.* (2006). Proteomic analysis of the eyespot of *Chlamydomonas reinhardtii* provides novel insights into its components and tactic movements. Plant Cell *18*, 1908-1930.

Schotz, F., Bathelt, H., Arnold, C.G., and Schimmer, O. (1972). [The architecture and organization of the *Chlamydomonas* cell. Results of serial-section electron microscopy and a three-dimensional reconstruction]. Protoplasma *75*, 229-254.

Schwender, J., Gemunden, C., and Lichtenthaler, H.K. (2001). *Chlorophyta* exclusively use the 1-deoxyxylulose 5-phosphate/2-C-methylerythritol 4-phosphate pathway for the biosynthesis of isoprenoids. Planta *212*, 416-423.

Schwender, J., Seemann, M., Lichtenthaler, H.K., and Rohmer, M. (1996). Biosynthesis of isoprenoids (carotenoids, sterols, prenyl side-chains of chlorophylls and plastoquinone) via a novel pyruvate/glyceraldehyde 3-phosphate non-mevalonate pathway in the green alga *Scenedesmus obliquus*. The Biochemical journal *316 (Pt 1)*, 73-80.

Shemer, G., Suissa, M., Kolotuev, I., Nguyen, K.C., Hall, D.H., and Podbilewicz, B. (2004). EFF-1 is sufficient to initiate and execute tissue-specific cell fusion in *C. elegans*. Curr Biol *14*, 1587-1591.

Shmulevitz, M., and Duncan, R. (2000). A new class of fusion-associated small transmembrane (FAST) proteins encoded by the non-enveloped fusogenic reoviruses. Embo J *19*, 902-912.

Shrestha, R. (2011). Master Thesis. Brooklyn College. *The City University of New York.*

Sinangil, F., Loyter, A., and Volsky, D.J. (1988). Quantitative measurement of fusion between human immunodeficiency virus and cultured cells using membrane fluorescence dequenching. FEBS letters *239*, 88-92.

Singson, A., Hang, J.S., and Parry, J.M. (2008). Genes required for the common miracle of fertilization in *Caenorhabditis elegans*. Int J Dev Biol *52*, 647-656.

Singson, A., Mercer, K.B., and L'Hernault, S.W. (1998). The *C. elegans* spe-9 gene encodes a sperm transmembrane protein that contains EGF-like repeats and is required for fertilization. Cell *93*, 71-79.

Skehel, J.J., and Wiley, D.C. (2000). Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. Annual review of biochemistry *69*, 531-569.

Snell, W.J., Buchanan, M., and Clausell, A. (1982). Lidocaine reversibly inhibits fertilization in *Chlamydomonas*: a possible role for calcium in sexual signalling. J Cell Biol *94*, 607-612.

Snell, W.J., Eskue, W.A., and Buchanan, M.J. (1989). Regulated secretion of a serine protease that activates an extracellular matrix-degrading metalloprotease during fertilization in *Chlamydomonas*. The Journal of cell biology *109*, 1689-1694.

Snell, W.J., Pan, J., and Wang, Q. (2004). Cilia and flagella revealed: from flagellar assembly in *Chlamydomonas* to human obesity disorders. Cell *117*, 693-697.

Sollner, T., Whiteheart, S.W., Brunner, M., Erdjument-Bromage, H., Geromanos, S., Tempst, P., and Rothman, J.E. (1993). SNAP receptors implicated in vesicle targeting and fusion. Nature *362*, 318-324.

Solter, K.M., and Gibor, A. (1977). Evidence for role of flagella as sensory transducers in mating of *Chlamydomonas reinhardtii.* Nature *265*, 444-445.

Spurgeon, S.L., and Porter, J.W. (1981). Biosynthesis of Isoprenoid Compounds Biosynthesis of carotenoids. In: Porter JW, Spurgeon SL, eds. Biosynthesis of isoprenoid compounds *2*, 1-122.

Stein, B.S., Gowda, S.D., Lifson, J.D., Penhallow, R.C., Bensch, K.G., and Engleman, E.G. (1987). pH-independent HIV entry into CD4-positive T cells via virus envelope fusion to the plasma membrane. Cell *49*, 659-668.

Stitzel, M.L., Cheng, K.C., and Seydoux, G. (2007). Regulation of MBK-2/Dyrk kinase by dynamic cortical anchoring during the oocyte-to-zygote transition. Curr Biol *17*, 1545-1554.

Tachibana, I., and Hemler, M.E. (1999). Role of transmembrane 4 superfamily (TM4SF) proteins CD9 and CD81 in muscle cell fusion and myotube maintenance. J Cell Biol *146*, 893-904.

Takahashi, Y., Bigler, D., Ito, Y., and White, J.M. (2001). Sequence-specific interaction between the disintegrin domain of mouse ADAM 3 and murine eggs: role of beta1 integrin-associated proteins CD9, CD81, and CD98. Mol Biol Cell *12*, 809-820.

Tan, G., Gao, Y., Shi, M., Zhang, X., He, S., Chen, Z., and An, C. (2005). SiteFinding-PCR: a simple and efficient PCR method for chromosome walking. Nucleic acids research *33*, e122.

Toshimori, K., Saxena, D.K., Tanii, I., and Yoshinaga, K. (1998). An MN9 antigenic molecule, equatorin, is required for successful sperm-oocyte fusion in mice. Biology of reproduction *59*, 22-29.

Trimmer, J.S., Schackmann, R.W., and Vacquier, V.D. (1986). Monoclonal antibodies increase intracellular Ca2+ in sea urchin spermatozoa. Proc Natl Acad Sci U S A *83*, 9055-9059.

Trimmer, J.S., and Vacquier, V.D. (1986). Activation of sea urchin gametes. Annu Rev Cell Biol *2*, 1-26.

Ulrich, A.S., Otter, M., Glabe, C.G., and Hoekstra, D. (1998). Membrane fusion is induced by a distinct peptide sequence of the sea urchin fertilization protein bindin. J Biol Chem *273*, 16748-16755.

Vacquier, V.D., and Moy, G.W. (1977). Isolation of bindin: the protein responsible for adhesion of sperm to sea urchin eggs. Proc Natl Acad Sci USA *74*, 2456-2460.

von Besser, K., Frank, A.C., Johnson, M.A., and Preuss, D. (2006). *Arabidopsis* HAP2 (GCS1) is a sperm-specific gene required for pollen tube guidance and fertilization. Development (Cambridge, England) *133*, 4761-4769.

Waffenschmidt, S., and Jaenicke, L. (1987). Assay of reducing sugars in the nanomole range with 2,2'-bicinchoninate. Analytical biochemistry *165*, 337-340.

Wang, Q., Pan, J., and Snell, W.J. (2006). Intraflagellar transport particles participate directly in cilium-generated signaling in *Chlamydomonas*. Cell *125*, 549-562.

Wang, Q., and Snell, W.J. (2003). Flagellar adhesion between mating type plus and mating type minus gametes activates a flagellar protein-tyrosine kinase during fertilization in *Chlamydomonas*. J Biol Chem *278*, 32936-32942.

Wang, Z.T., Ullrich, N., Joo, S., Waffenschmidt, S., and Goodenough, U. (2009). Algal lipid bodies: stress induction, purification, and biochemical characterization in wild-type and starchless *Chlamydomonas reinhardtii.* Eukaryotic cell *8*, 1856-1868.

Wassarman, P.M. (1987a). The biology and chemistry of fertilization. Science (New York, N.Y *235*, 553-560.

Wassarman, P.M. (1987b). The zona pellucida: a coat of many colors. Bioessays *6*, 161-166.

Wassarman, P.M. (2002). Sperm receptors and fertilization in mammals. Mt Sinai J Med *69*, 148-155.

Wassarman, P.M., Jovine, L., and Litscher, E.S. (2004). Mouse zona pellucida genes and glycoproteins. Cytogenet Genome Res *105*, 228-234.

Wassarman, P.M., Jovine, L., Qi, H., Williams, Z., Darie, C., and Litscher, E.S. (2005). Recent aspects of mammalian fertilization research. Mol Cell Endocrinol *234*, 95-103.

Wassarman, P.M., and Litscher, E.S. (2009). The multifunctional zona pellucida and mammalian fertilization. J Reprod Immunol *83*, 45-49.

Wassarman, P.M., and Mortillo, S. (1991). Structure of the mouse egg extracellular coat, the zona pellucida. Int Rev Cytol *130*, 85-110.

Weiss, R.L., Goodenough, D.A., and Goodenough, U.W. (1977). Membrane differentiations at sites specialized for cell fusion. J Cell Biol *72*, 144-160.

White, J., and Helenius, A. (1980). pH-dependent fusion between the Semliki Forest virus membrane and liposomes. Proc Natl Acad Sci USA *77*, 3273-3277.

White, J.M. (1990). Viral and cellular membrane fusion proteins. Annu Rev Physiol *52*, 675-697.

White, J.M., Danieli, T., Henis, Y.I., Melikyan, G., and Cohen, F.S. (1996). Membrane fusion by the influenza hemagglutinin: the fusion pore. Society of General Physiologists series *51*, 223-229.

White, J.M., Delos, S.E., Brecher, M., and Schornberg, K. (2008). Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme. Crit Rev Biochem Mol Biol *43*, 189-219.

Wickner, W., and Schekman, R. (2008). Membrane fusion. Nat Struct Mol Biol *15*, 658-664.

Wilson, K.L., Fitch, K.R., Bafus, B.T., and Wakimoto, B.T. (2006). Sperm plasma membrane breakdown during *Drosophila* fertilization requires sneaky, an acrosomal membrane protein. Development (Cambridge, England) *133*, 4871-4879.

Wilson, N.F., Foglesong, M.J., and Snell, W.J. (1997). The *Chlamydomonas* mating type plus fertilization tubule, a prototypic cell fusion organelle: isolation, characterization, and in vitro adhesion to mating type minus gametes. J Cell Biol *137*, 1537-1553.

Wolkowicz, M.J., Shetty, J., Westbrook, A., Klotz, K., Jayes, F., Mandal, A., Flickinger, C.J., and Herr, J.C. (2003). Equatorial segment protein defines a discrete acrosomal subcompartment persisting throughout acrosomal biogenesis. Biology of reproduction *69*, 735-745.

Wong, J.L., and Johnson, M.A. (2010). Is HAP2-GCS1 an ancestral gamete fusogen? Trends Cell Biol *20*, 134-141.

Xu, X.Z., and Sternberg, P.W. (2003). A *C. elegans* sperm TRP protein required for sperm-egg interactions during fertilization. Cell *114*, 285-297.

Yamatoya, K., Yoshida, K., Ito, C., Maekawa, M., Yanagida, M., Takamori, K., Ogawa, H., Araki, Y., Miyado, K., Toyama, Y., and Toshimori, K. (2009). Equatorin: identification and characterization of the epitope of the MN9 antibody in the mouse. Biology of reproduction *81*, 889-897.

Yanagimachi, R. (1970). The movement of golden hamster spermatozoa before and after capacitation. J Reprod Fertil *23*, 193-196.

Yanagimachi, R. (1994). Stability of the mammalian sperm nucleus. Zygote *2*, 383-384.

Yanagimachi, R. (1998). Intracytoplasmic sperm injection experiments using the mouse as a model. Hum Reprod *13 Suppl 1*, 87-98.

Zhu, X., Bansal, N.P., and Evans, J.P. (2000). Identification of key functional amino acids of the mouse fertilin beta (ADAM2) disintegrin loop for cell-cell adhesion during fertilization. J Biol Chem *275*, 7677-7683.

Zimmerman, S.W., Manandhar, G., Yi, Y.J., Gupta, S.K., Sutovsky, M., Odhiambo, J.F., Powell, M.D., Miller, D.J., and Sutovsky, P. (2011). Sperm proteasomes degrade sperm receptor on the egg zona pellucida during mammalian fertilization. PLoS One *6*, e17256.