MAPPING THE DEFLAGELLATION-DEFECTIVE *ADF1* GENE IN CHLAMYDOMONAS REINHARDTII

by

Jaime Ann Kirschner
B.Sc. Archaeology, University of Calgary, 2002
B.Sc. Biology, Okanagan University College, 2000

THESIS
SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

In the
Department of Molecular Biology and Biochemistry

SIMON FRASER UNIVERSITY

Summer 2009

APPROVAL

Name: **Jaime Ann Kirschner**

Degree: **Master of Science**

Title of Thesis: **Mapping the deflagellation-defective *ADF1* gene in *Chlamydomonas reinhardtii***

Examining Committee:

Chair: **Dr. Mark Paetzel**
Associate Professor, Department of Molecular Biology and Biochemistry

_____

Dr. Lynne Quarmby
Senior Supervisor
Professor, Department of Molecular Biology and Biochemistry

_____

Dr. Nancy Hawkins
Supervisor
Assistant Professor, Department of Molecular Biology and Biochemistry

_____

Dr. Esther Verheyen
Supervisor
Associate Professor, Department of Molecular Biology and Biochemistry

_____

Dr. Sherryl Bisgrove
Internal Examiner
Assistant Professor, Department of Biology

Date Defended/Approved: July 27, 2009

Declaration of
Partial Copyright Licence

ABSTRACT

Cilia function in motility, chemo- and mechanosensation and have been linked to the cell cycle and human disease. Stress triggers the calcium-dependent severing of axonemal microtubules at a site near the base of the cilium, resulting in deciliation (also known as deflagellation). Originally isolated as a spontaneous mutation in *Chlamydomonas reinhardtii*, the *adf1-1* mutant is defective in Ca^{2+} influx that triggers deflagellation. Five new alleles of *adf1* were recovered from UV and insertional mutagenesis screens, but none of the alleles were tagged with insertional DNA. I have used PCR based recombination mapping to place *ADF1* in a 394 kb region of linkage group IX. Six BACs spanning the region have been used to transform mutant *adf1* strains, with the goal of rescuing the phenotype. BAC transformations have not yielded a rescue; therefore, several candidate genes have been subcloned and are being used in attempts to identify the elusive gene.

Keywords: *Chlamydomonas reinhardtii*; deflagellation; ADF1; Ca^{2+} signalling; recombination mapping

ACKNOWLEDGEMENTS

I would like to thank everyone who has supported and encouraged me through my education and research. My eternal gratitude goes to Dr. Lynne Quarmby, for taking me on as a graduate student and having so much faith in me. Thank you to Dr. Esther Verheyen and Dr. Nancy Hawkins for supporting me as committee members. I would like to thank Dr. Sherryl Bisgrove for being my internal examiner and Dr. Mark Paetzel for being chairperson at my defense. I would like to thank the members of the Quarmby lab, past and present, for being friends as well as colleagues. Thank you to Dr. Qasim Rasi for introducing me to the Quarmby lab and the excitement of research.

I would also like to thank my friends and family for many years of love and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Increasingly, human diseases are being associated with cilia and flagella, microtubule based organelles, which emanate from the cell body into the extracellular space to function in locomotion, chemo- and mechano-sensation and fluid propulsion. The following introduction discusses: the structure and function of cilia and flagella, (the terms are interchangeable); the expanding category of ciliopathies, diseases of the cilia; the model organism we have used to study the classic stress response deflagellation, triggered by cytosolic acidification; and, the progress made in eliciting the pathway and mechanisms culminating in deflagellation. Experimental inquiries have led to speculations that a proton-sensitive, calcium-dependent pathway could initiate at the plasma membrane, near the base of the flagella. An elusive gene, *ADF1*, most likely encodes a component of this pathway and it is our hypothesis that ADF1 is a calcium channel, possibly directly activated by protons. But, it remains to be seen what *ADF1* encodes, as attempts to identify it have yet to yield either gene locus or clone. Herein please find the story of my struggle to find *ADF1*. Enjoy.

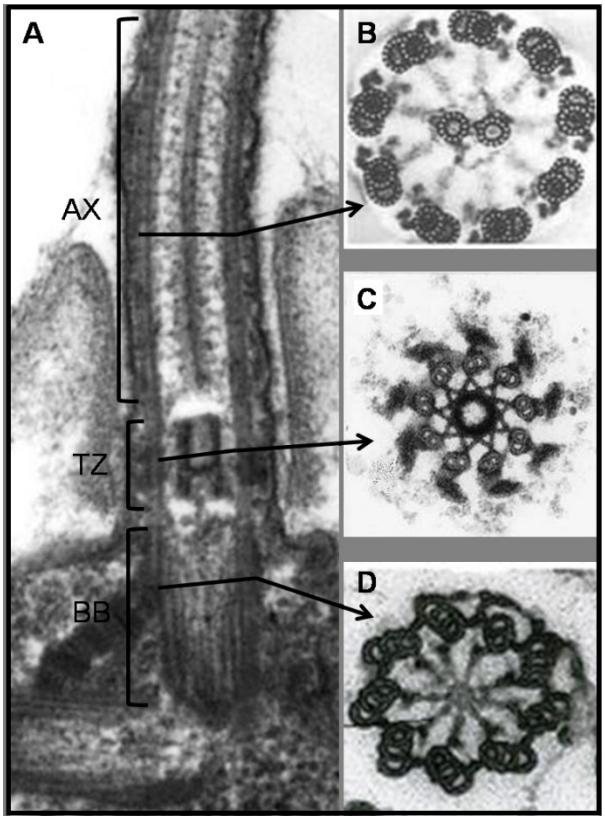
I: Cilia

The cilium, an ancient and highly conserved organelle, evolved over 800 million years ago, before the divergence of the last eukaryotic common ancestor. Cilia act in motile and sensory functions. Presently, cilia are found on almost all eukaryotic organisms with the exceptions of most fungi (including yeast), red

algae and higher plants (Ginger et al., 2008). In mammals, cilia are present on nearly every cell type: either in single copy, such as the primary cilium of a kidney epithelial cell; or in hundreds, such as the motile cilia of the epithelial cells of the lungs (Davenport & Yoder, 2005; see www.bowserlab.org/primarycilia/ciliumpage2.htm for a list of cells displaying cilia). Cilia, found on many different organisms in variable forms, can be described in general terms as including mandatory component parts: the basal body, the transition zone, the axoneme, and the ciliary membrane.

The basal body, a centriole-derived, microtubule organizing centre, anchors the cilium in the cell body, ensuring proper positioning. Basal bodies are comprised of a nine-fold, radial arrangement of triplet microtubules, designated A, B and C. The A tubule is comprised of 13 protofilaments, while the B and C tubules are comprised of 11 protofilaments each. The triplet microtubule architecture of the basal body spans an extent of 0.4 micrometers and then converts to a doublet microtubule arrangement, with the termination of the C tubule. This transfiguration delineates the transition zone boundary (see Figure 1: C and D). Distal appendages anchor the basal body to the plasma membrane, contributing to a network of transition zone fibers known as the ciliary necklace, which separates the cell body compartment from the ciliary compartment. These fibers may act in regulating access to the ciliary compartment (reviewed by Pazour & Bloodgood, 2008).

Figure 1: Microtubule arrangement of basal bodies, transition zone and axoneme.
(A) flagellum of Chlamydomonas reinhardtii, including basal body (BB), transition zone (TZ) and axoneme (AX). (B) Cross section of 9+2 microtubule arrangement of canonical motile flagella, with A and B tubules and central pair. (C) Cross section of 9+0 microtubule arrangement of transition zone. Note stellate arrangement of centrin-containing fibers. (D) Cross section of triplet microtubule arrangement of basal body. (Images compiled from: Geimer et al., 2004; DT Woodrow & RW Linck, in Alberts et al., 2002; E Smith, http://remf.dartmouth.edu/imagesindex.html; and www.cytochemistry.net/Cell-Biology/cilia.htm)



The transition zone is not readily identifiable in many organisms, except by the absence of the C tubule. In *Chlamydomonas* however, an electron-dense pair of central cylinders, which appear as an “H” shape in longitudinal section, are present; and, a stellate arrangement of centrin containing fibers connect the central cylinders to the outer doublets (Figure 1: A and C). The A and B microtubules of the basal body extend through and past the transition zone. Recent indirect immuno-fluorescence experiments are beginning to reveal

proteins that localize to the transition zone in both mammals and humans (Lohret et al., 1999; Mahjoub et al., 2002; Iomini et al., 2006; Fliegauf et al., 2006; Piaskecki & Silflow, 2009; M White, unpublished; and, JDK Parker, unpublished).

Distal to the transition zone, the axoneme projects into the extracellular environment. The basic axonemal structure is composed of nine outer microtubule doublets, tethered together by nexin linkages. The doublet microtubules are continuous with the A and B tubules of the basal body and transition zone. Canonical motile cilia exhibit a central pair of microtubules, enclosed by the nine outer doublets, and have been designated 9+2. Primary cilia have only the nine outer microtubule doublets, are immotile and are designated 9+0. Variations on the canonical descriptions are seen across cell types (reviewed by Fliegauf et al., 2007).

In motile cilia, the ciliary beat is driven by axonemal dyneins. Dynein motor complexes are bound to A microtubules, such that their motor head domains are close to the B tubule of the neighbouring doublet. ATP driven conformational changes allow the motor head domains to transiently bind to the B tubules. Outer dynein arms, controlling beat frequency, and inner dynein arms, controlling beat form, organize the relative sliding of microtubule doublets past each other. Operating as opposing sets, microtubules 1-4 generate the effective stroke, or the principal bend, while microtubules 6-9 generate the recovery stroke, or reverse bend. In “9+2” motile cilia this coordination is aided by the presence of radial spokes, which project from the A tubule inward, so that spokeheads can

interact with projections from the central pair microtubules (Silflow & Lefebvre, 2001).

A great number of accessory proteins are required for building and maintaining the cilia. In most cells, the cycle of ciliogenesis is coordinated with progression through the cell cycle (reviewed by Quarmby & Parker, 2005). Ciliogenesis initiates in G1 of the cell cycle and frequently marks exit from the cell cycle into Gº, a quiescent stage. Following cytokinesis, the mother centriole begins migrating to the apical end of the cell. Golgi derived vesicles dock onto the distal end of the centriole as it migrates, allowing the axoneme to begin elongation within a membrane bound compartment (reviewed by Pedersen et al., 2008). The newly formed ciliary membrane fuses with the plasma membrane, forming the ciliary necklace, and docking the centriole at the apical end of the cell. Once the centriole has docked it is hence forth known as the basal body, and eventually gives rise to the mature cilium as elongation of the axoneme continues via the evolutionarily conserved mechanism of intraflagellar transport (IFT) (reviewed by Rosenbaum & Witman, 2002).

Intraflagellar transport, first described in *Chlamydomonas* and subsequently found in all organisms that have compartmentalized ciliogenesis, is the bidirectional transport of proteins bound to IFT particles along the axoneme. Due to the absence of de novo protein synthesis in the ciliary compartment, assembly of the cilium is accomplished through the delivery of axonemal precursors to the distal ciliary tip, from their site of synthesis in the cell body. Particles are recruited to the base of cilium, near the transition zone, which acts

as a kind of staging area. Here they are bound to particles attached to heterotrimeric Kinesin II motor proteins, which “walk” along the axoneme in an anterograde fashion, away from the cell body proper. Once the kinesin II motors have reached the ciliary tip and released their cargo, they are inactivated. Kinesin return to the cell body is dependent on retrograde trafficking by cytoplasmic dynein motors, which also mediate disassembly at the ciliary tip and resorption of the cilium prior to division (Rosenbaum & Witman, 2002).

IFT particles have been solubilized into two complexes, A and B. The A complex is required for retrograde transport, while the B complex is required for anterograde movement. Complex A is composed of at least 6 subunits and Complex B is composed of at least 10 subunits (reviewed by Pedersen and Rosenbaum, 2008). The components of the complexes exhibit a variety and a variable number of protein-protein interaction sites, such as: N-term WD repeats, C-term TPR (tricopeptide) repeats and coiled-coil domains. These protein-protein binding domains are consistent with the role of cargo transport.

In addition to bringing axonemal precursors to the ciliary tip and back, IFT mediates the localization of flagellar integral membrane proteins to the ciliary compartment. These may include components of signalling pathways, such as: receptors, ion channels, effector proteins, and transcription factors (discussed later). The ciliary membrane is the premier location to initiate signalling due to its exposed nature. Ciliary membrane is continuous with cell membrane, but lacks barriers such as cell wall, ECM components, or tightly adjoined neighbouring cells. As the cilium projects away from the cell body proper, it is cleared of these

encumbrances, which could slow down or prevent environmental cues from reaching the cell.

The sensory organs that allow us to hear, see and smell rely on cilia to transduce mechanical and chemical stimuli. In the auditory canal, sound vibrations are transduced by mechanosensory hair bundles. The hair bundles are composed of sterocilia, or modified microvilli, which derive their orientation from a 9+2 kinocilium that is present during development (Nayak et al., 2007). In the eye, light waves are absorbed by the outer segment of photoreceptor cells, which are modified primary cilia (Insinna & Besharse, 2008). Our sense of smell relies on the multiciliated, dendritic endings of olfactory sensory neurons. Olfactory cilia are immotile, yet have a 9+2 microtubule arrangement (McEwen et al., 2008). Cilia also play important roles in the development and homeostasis of multicellular organisms, interpreting stimuli from the environment, such as morphogenic cues; sensing fluid flow; or producing fluid flow and cell motility.
Cilia are the target compartment for components of several signaling pathways, most notably: hedgehog (Hh), platelet-derived growth factor alpha (PDGFα), and Wnt or planar cell polarity (PCP) pathways.

Hedgehog signaling plays important roles in embryogenesis and post-natal homeostasis (reviewed by Wong & Reiter, 2008). During embryonic development, dorsoventral gradients of the ligand sonic hedgehog (SHh) are established through initial release from the notochord, causing neural tube, floor plate and neuron differentiation in a concentration and time-dependent manner. SHh gradients similarly determine digit identity and number. The hedgehog

pathway is also imperative to left-right asymmetry, and heart development. In adults the Hh pathway maintains stem cell niches and tissue homeostasis. In the absence of Hh ligand, the receptor Patched (PTCH) localizes to the cilium and inhibits the pathway by keeping Smoothened (SMO), a transmembrane mediator of SHh, out of the cilium. The Hh pathway is turned on when ligand binds to PTCH causing it to internalize, therefore, allowing SMO to move into the cilium, where it can activate Gli transcription factors. Finally, all three Gli proteins as well as the repressor, SuFu, localize to cilium prior to their activation.

The PDGFα pathway plays roles in regulation of cell survival, apoptosis, proliferation, migration and angiogenesis during embryonic and postnatal development, as well as growth of connective tissues and wound healing in adults (Christensen et al., 2008). The receptor PDGFα localizes to the ciliary membrane, and upon activation by PDGF-AA, in growth-arrested cells, mediates the activation of Mek1/2-Erk1/2 and Akt pathways, in addition to phosphorylation of RB and CDC2. This marks entry into and progression of the cell cycle.

Wnt signals move through more than one pathway. The canonical pathway induces transcription of genes involved in cell cycle progression, proliferation, cell fate determination and, regulation of embryogenesis (reviewed by Veland et al., 2009). The non-canonical or PCP pathway is responsible for cell polarization and migration, which is necessary for processes such as gastrulation and neurulation. Components of both the canonical and the non-canonical or PCP pathways are localized to the cilium/centrosome axis, including: Vangl2, Inversin, Dvl, Inturned, Fuzzy, APC and β-Catenin, GSK3β, and Fat4 (reviewed

by Christensen et al. 2008). Surprisingly, Wnt receptors have not been found to localize to the ciliary membrane. Interestingly, downstream signals for both pathways are similar to the PDGFα pathway, including Mek1/2-Erk1/2, PI3K-Akt and PKC.

In addition to chemosensory capacities, cilia can also detect mechanical stimuli. Primary cilia of renal epithelial cells are thought to sense the mechanical stimulus of urine flow. Fluid flow strong enough to cause the cilium to bend initiates a calcium influx that is dependent on extracellular calcium levels. A second calcium influx, released from internal stores, is triggered and results in second messenger signalling to neighbouring cells through gap junctions (Praetorius & Spring, 2001). The primary calcium influx is modulated through a TRP channel, TRPP2, which localizes to the primary cilium (reviewed by Qamar et al., 2007). TRPP2 (also known as polycystin-2) forms a complex with polycystin-1, which functions as a G-protein coupled receptor. It has been shown that in the absence of flow, the C-term of polycystin-1 is cleaved and localized to the nucleus, where it may stimulate JAK/STAT, mTor and AP-1 pathways (Wilson, 2001; Weimbs, 2007).

Not to be overshadowed by the importance of chemo- and mechanosensation, motility is a keystone of ciliary function. Motile cilia are required to sweep extracellular fluid across tissue surfaces (reviewed by Fliegauf et al., 2007). The monocilia of the embryonic node generate extra embryonic fluid flow, pushing morphogens in a unilateral distribution breaking left-right symmetry. The 9+2 motile cilia of respiratory epithelial cells are responsible for mucociliary

clearance; similarly, ependymal cilia produce laminar flow of cerebrospinal fluid through brain ventricles. In the female reproductive system, cilia in the oviduct assist in moving both the gamete and embryo to the uterus. Sperm cells, in the male reproductive system, are moved through the efferent duct by cilia, and later navigate through the female reproductive system by flagellar propulsion.

II: Ciliopathies

Defects in cilia lead to an array of diseases, known as ciliopathies. The different syndromes have overlapping phenotypes, which generally include: renal cystic and hepatobiliary defects, laterality defects, retinal degeneration, polydactyly and skeletal malformations. The most common disease traits: renal cysts, retinal degeneration and polydactyly, are prevalent in human populations at rates of 1 in 500 adults, 1 in 3000 and 1 in 500, respectively (Quinlan et al., 2008).

Defects in components of the motile machinery of cilia, such as dynein arms, radial spokes or the central pair, result in primary ciliary dyskinesia (PCD; OMIM 244400), also known as immotile cilia syndrome (reviewed by Chodhari et al., 2004). PCD patients present with symptoms such as: recurrent ear-nose-and-chest infections, which are secondary to defective mucociliary clearance; hydrocephalus, the result of reduced ependymal flow and closure of the cerebral aqueducts; infertility in men, due to immotile sperm or abnormalities of the vas deferens; and finally, patients exhibit laterality defects, such as situs inversus, likely caused by reduction or absence of nodal flow. Symptoms are present from birth, but vary in severity and expressivity. Ciliary ultrastructural defects reported

in PCD patients include: absent central microtubule pair, with transposition of peripheral doublet to centre; peripheral microtubule defects; and, absent radial spokes. The most frequently identified defects are related to the absence or reduction of inner and outer dynein arms, and in fact, the only ultrastructural phenotype for which genetic mutations have been identified, in humans, is of the outer dynein arm. The autosomal recessive condition is most frequently caused by mutations in outer dynein arm components, DNAI1 and DNAH5 (Morillas et al., 2007).

Polycystic kidney disease (PKD), initiating in utero or at birth, is characterized by renal enlargement and biliary dysgenesis. Autosomal dominant polycystic kidney disease (ADPKD) presents with multiple cysts throughout the kidney, resulting in gradual kidney enlargement, and end-stage renal failure by the 6th decade (Chapman et al., 2007). Fluid filled cysts also develop in the liver, pancreas, spleen, thyroid, brain meninges and the seminal vesicles. ADPKD is caused by mutations in two genes: PKD1 (OMIM 601313) and PKD2 (OMIM 173910), encoding for polycystin-1 (PC-1) and polycystin-2 (PC-2), respectively. As discussed earlier, polycystin-1 and polycystin-2 are thought to interact with each other in order to function in Ca^{2+} signalling at the membrane of the primary cilium of renal epithelial cells to transduce fluid flow cues. Mutations in a third gene, PKHD1, which encodes fibrocystin, cause autosomal recessive polycystic kidney disease (ARPKD, OMIM 263200). ARPKD is marked by the rapid progression of collecting duct cysts, which enlarge kidneys and result in renal

failure shortly after birth. Importantly, fibrocystin also localizes to the primary cilium (reviewed Bandano et al., 2006).

Nephronophthisis (NPHP1; OMIM 256100) is another autosomal recessive form of cystic renal disease, and like ARPKD, it is aggressive in its progression. NPHP is distinguished by corticomedullary cysts and tubulointerstitial fibrosis (reviewed by Quinlan et al., 2008). The kidneys do not enlarge, as in PKD, but rather are normal or small in size. Onset of end-stage renal failure subcategorizes NPHP as infantile, juvenile or adolescent. Nine loci, NPHP 1 through 9, have been identified and provide strong links between ciliary function and pathogenesis of disease (Hildebrandt et al., 2009). NPHP gene products localize to adherens junctions, connecting cilia, primary cilia, and basal bodies; they form complexes with focal adhesion proteins, actin cytoskeleton proteins, are suspected to play roles in cell polarity, and are involved in SHh signaling (reviewed by Quinlan et al., 2008).

Some of the genes that cause NPHP are implicated in several other related syndromes. NPHP1, 3, 4 and 5 have been implicated in Senior-Loken syndrome (SLSN; OMIM 266900), an autosomal recessive disease characterized by nephronophthisis and the progressive eye disease, retinitis pigmentosa. NPHP6 is associated with the related ciliopathies SLSN6, Bardet-Biedel syndrome 10 (BBS10), Joubert syndrome 5 (JBTS5), and Meckel-Gruber syndrome 4 (MKS4). These diseases affect the limb and nervous tissue, in addition to kidney and eye (reviewed by Sharma et al., 2008).

Bardet-Biedel syndrome (BBS; OMIM 209900) is a multisystemic disorder characterized by obesity, polydactyly, mental retardation, retinal degeneration and renal and gonadal malformations. Additional phenotypes include: anosmia, asthma, diabetes, situs inversus and congenital heart disease. Twelve genes (BBS1-12) have been associated with Bardet-Biedel syndrome, although the functions of the gene products are not fully understood (reviewed by Quinlan et al., 2008). The BBS proteins localize to the cilium/ basal body/ centrosome complex and participate in IFT, possibly in cargo selection and transport, or in the coordinated movement of the A and B particle complexes. Interestingly, BBS mutant mice develop phenotypes similar to PCP animals, including neural tube defects, open eyelids, and defective hair bundles in cochlea (Bedano et al., 2006).

Joubert syndrome (JBTS; OMIM 213300) features cerebellar vermis hypoplasia, mental retardation, hypotonia, breathing and eye movement abnormalities, in conjunction with retinal degeneration and NPHP (Bedano et al., 2006). Occipital encephalocele, cystic kidneys, polydactyly, and hepatic fibrosis, may also be present. Eight loci have been determined; two are not yet identified and four (JBTS4, 5, 6 and 7) overlap with loci identified in other syndromes, with proteins localizing to cilia/centrosome complex. JBTS8 encodes a Ras-GTPase family member which, when mutated in mice, causes ciliary and SHh signaling defects (reviewed by Quinlan et al., 2008).

Meckel-Gruber syndrome (MKS; OMIM 249000) is a lethal, autosomal recessive condition of cleft palate, renal cysts, hepatic fibrosis, polydactyly, and

occipital encephalocele. Six loci have been found, but only 5 genes have been identified, MKS1 and MKS3 through 5; all are associated with ciliary function, and two overlap with NPHP loci. MKS1 and 3 (meckelin) interact with each other and are required for centriole migration to the apical end of the cell (reviewed by Quinlan et al., 2008).

Several syndromes have no genetic overlap with other ciliopathies. Alstrom syndrome (ALMS; OMIM 203800) is caused by mutations in ALMS1, and is identified by cone and rod dystrophy, neurosensory hearing loss, early onset obesity and insulin resistance, leading to type II diabetes. Oral-facial-digital type 1 syndrome (OMIM 311200), an X-linked disorder characterized by malformations of the face and oral cavity, polydactyly, PKD and central nervous system defects, is caused by mutations in OFD1. Finally, Jeune asphyxiating thoracic dystrophy is characterized by chondrodysplasia, biliary dysgenesis, renal cystogenesis, polydactyly and retinal degeneration (reviewed Quinlan et al., 2008).The identified gene products associated with these diseases all localize to the cilia/ basal body apparatus (reviewed by Bedano et al., 2006).

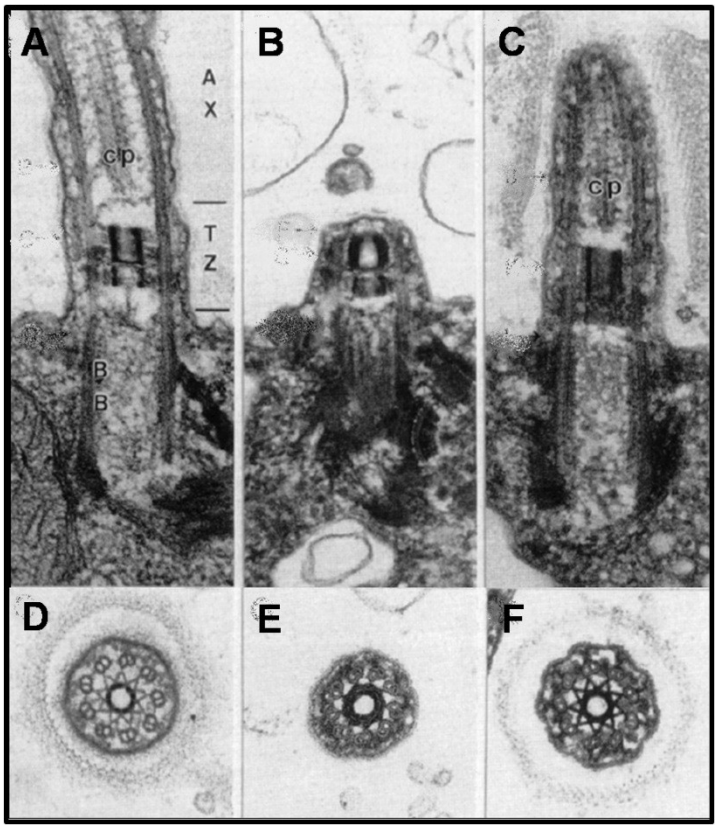
III: Deflagellation

I described above that cells resorb their cilia prior to division and build new cilia after mitosis (Quarmby & Parker, 2005). There is another mechanism, however, that cells employ to remove their flagella. Deflagellation, also known as flagellar excision or flagellar autotomy, refers to the rapid shedding of flagella (reviewed by Quarmby, 2004). Deflagellation occurs at the base of the flagellum, just distal to the transition zone. In *Chlamydomonas*, stressful stimuli (see

below) trigger the contraction of a stellate structure of centrin-containing fibers at the transition zone, creating an inward displacement of the microtubule doublets (see Figure 2: D and E). In conjunction with the alteration of the transition zone structure, the nine outer doublets sever just distal to the transition zone, at the SOFA or site of flagellar autonomy (Mahjoub et al. 2004). After axonemal severing, the cell membrane pinches in, sealing over the transition zone stump, allowing the flagellum proper to separate from the cell body, while maintaining the integrity of the cytoplasm (see Figure 2: B).

Many cell types are reported to undergo deciliation, and it is due to this behaviour that we know as much as we do about cilia. Experimentally, cells can be induced to cast off their cilia through exposure to treatments such as: extremes in pH or heat, chemicals (alcian blue, alcohol, chloral hydrate, dibucaine, and mastoparan) or mechanical shear. The deflagellation behavior has been harnessed for its scientific value for over half a century (Finst et al., 2000; Quarmby, 2009). Diligent investigation has resulted in a host of scientific tools. The rapid and collective shedding of flagella in a population of cells, followed by the synchronous regeneration of this organelle, has allowed researchers to study, in a large cohort, the genesis of an organelle, including: the signaling pathways, gene expression and, posttranslational modification, and transport and assembly mechanisms involved. In addition, the ease of harvesting flagella and sub-fractions of flagella, following deflagellation, has advanced biochemical and structural analysis of flagellar composition, including the development of proteomic collections.

Figure 2: Changes at the transition zone following deflagellation.
(A) The axoneme/basal body complex prior to deflagellation. (B) Transition zone, following deflagellation. Membrane has sealed over the stump. (C) Regeneration of flagella. (D) Stellate arrangement of centrin-containing fibers before deflagellation. (E) Condensed stellate arrangement of centrin-containing fibers, following deflagellation. (F) Expanded stellate arrangement of centrin-containing fibers, after the onset of flagellar regeneration. Axoneme (AX), central pair (cp), TZ (transition zone) and, basal body (BB). Lower panel (D, E, and F) images are cross sections of the transition zone. (Modified from Sanders & Salisbury, 1989).



Deflagellation can also occur under more natural conditions. Deciliation has been observed in Paramecium during mating. The sperm of many species deflagellate, usually following entry into the oocyte. Respiratory and oviduct epithelium both deciliate in response to infection, and respiratory epithelial cells deciliate in response to smoke (reviewed by Quarmby, 2009). Interestingly, a recent report suggests that deciliation of renal epithelial cells in tissue culture triggers the fortification of tight junctions (Overgaard et al., 2009). *Chlamydomonas* is known to shed its flagella following adhesion to the predatory

heliozoan, Actinophrys (reviewed by Quarmby, 2009), allowing the algal cell body to float away and swim another day. The universality of this behavior begs to question its evolutionary history and physiological roles.

It is obvious that deflagellation is quite useful to scientists, but what does it do for the cell? At first glance it seems that the deflagellation behavior could have developed as a defense mechanism, suggesting positive selection for survival advantages. It has been propounded that, in *Chlamydomonas*, deflagellation is a swift way to reduce exposed, permeable membrane surface area when presented with unfavorable physiochemical conditions, thus increasing survivorship for the free living cells (Lewin et al., 1982). Based on this hypothesis, deflagellation mutants should display poor survivorship when exposed to deflagellation stimuli. However, deflagellation mutants, when exposed to common laboratory noxious stimulants, do not have reduced survivorship (reviewed by Quarmby, 2009).

The most parsimonious explanation is that deflagellation is a conserved behaviour that belonged to the ciliated ancestral cell. Based on the conserved structure of cilia and the conserved mechanism for assembly, and the fact that the deflagellation response itself is conserved, it is LIKELY that the mechanism for deflagellation is conserved and involves orthologous proteins and processes. If a conserved “break point”, a sensitive junction between the transition zone and the cilium proper, was a compulsory constituent of cilia, perhaps playing a role in pre-mitotic re-absorption, then it would provide every cilia-generating cell the foundation necessary to evolve behaviours around that break point (Parker &

Quarmby, 2003). Determining the signaling pathway(s) that lead to deflagellation should aid in understanding the importance of the ubiquitous process. To study this question, we have turned to the premier model system to investigate the deflagellation mechanism, *Chlamydomonas*.

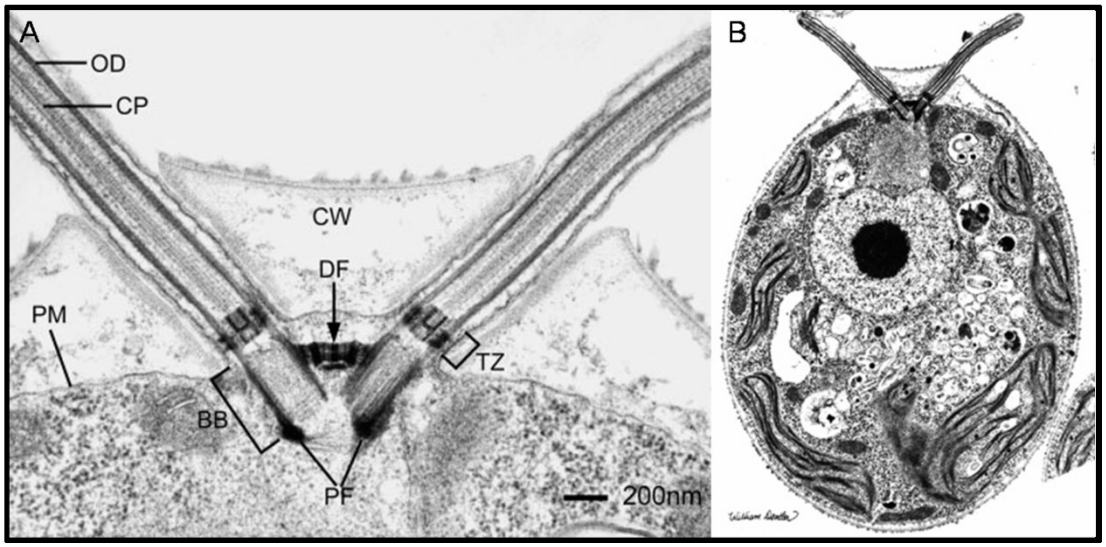
IV: *Chlamydomonas reinhardtii*

Chlamydomonas reinhardtii is a unicellular, green alga. It possesses two canonical, motile flagella, 10-12 um in length, which project through specialized collar regions of the apical cell wall (see Figure 3). Notable features of the cell body include a basal chloroplast, surrounding one or more pyrenoids; contractile vacuoles, important for osmoregulation; an eye spot, required for phototaxis; and basal bodies, which are the rooting place of the flagella (the reader is referred to The *Chlamydomonas* Sourcebook, 2nd Edition (edited by Stern et al., 2009)). The nuclear genome has been sequenced (121 Mb), is approximately 64% GC rich and is arranged on 17 linkage groups and (Merchant et al., 2007).

The *Chlamydomonas* life cycle progresses through haploid and diploid states (see Figure 4). Haploid vegetative cells divide mitotically and generate flagella in liquid media. Vegetative cells, grown in liquid or on agar, have an average doubling time of 6-8 hours allowing vast quantities of *Chlamydomonas* cells to be grown up for biochemical assays. In addition, *Chlamydomonas* is especially amenable to genetic analyses. When placed in reduced nitrogen media vegetative haploid cells will form gametes. Gametes of opposite mating type recognize one another, fuse, and form diploid zygotes. Zygotes form a thick proteinaceous cell wall and can remain inert for months or be induced to proceed

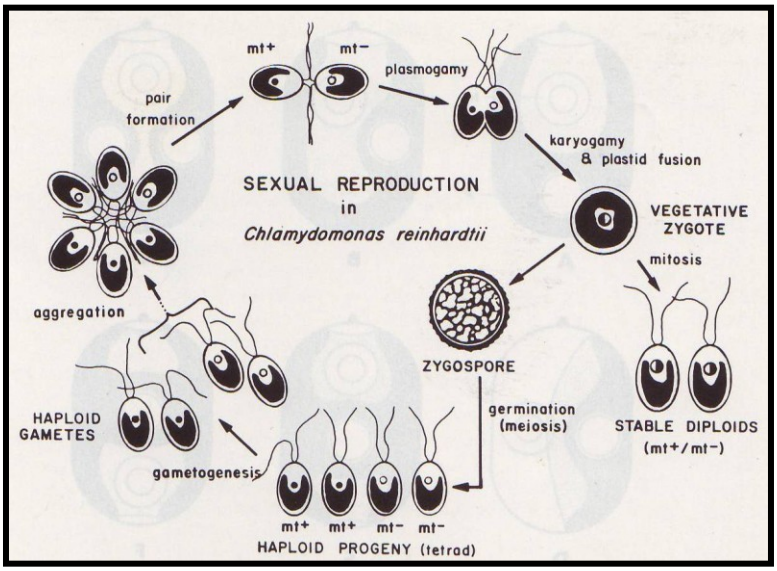
through meiosis several days after formation. Meiosis produces a tetrad, four haploid daughter cells enclosed within one cell wall.

Figure 3: *Chlamydomonas reinhardtii.*
(A) Longitudinal section through flagella and basal bodies of *Chlamydomonas*. Note flagella projecting through specialized collar regions in cell wall. (CW, cell wall; PM, plasma membrane; OD, outer doublet microtubule of the axoneme; CP, central pair microtubules; BB, basal body; DF, distal striated fiber; TZ, transition zone; PF, proximal fiber connecting the two basal bodies). (B) Scanning electron micrograph of Chlamydomonas cell. (Electron micrographs by Dr. William Dentler, adapted from Silflow & Lefebvre, 2001).



Tetrads can be separated and used in genetic analyses, such as genetic mapping. Occasionally during the mating process vegetative zygotes do not go through meiosis, but instead go through mitosis generating stable diploids. Diploid cells can be selected for by mating strains with complementing auxotrophic markers and then plating the mating mixture on media that only supports life for complementing diploids, or meiotic recombinants. Diploids start dividing hours after mating, whereas zygotes take several days to germinate. These cells are useful in complementation assays to determine if mutations are allelic or unrelated.

Figure 4: Life cycle of *Chlamydomonas reinhardtii*.
Alternative fates of mated pairs as meiotic zygotes and as vegetative diploid cells. (Harris, 1989; Figure 1.3 (Courtesy of K Swift)).



As cilia are not essential to viability, vegetative cells have been used in mutation screens to isolate cells with specific flagellar phenotypes, such as non-motility or deflagellation defects. Mutations can be induced with UV, chemical mutagenesis and insertional mutagenesis. Insertional mutagenesis allows the non-homologous integration of exogenous DNA which may disrupt gene function. Interrupted genes can be cloned by using the exogenous sequence as a starting point. In addition to disrupting genes, transformation of exogenous DNA can also be utilized to rescue mutant phenotypes or influence protein expression, through RNAi. Transformation of exogenous DNA has been achieved for nuclear, mitochondrial and chloroplast genomes.

In *Chlamydomonas* we can correlate genetic and biochemical data with cell biological observations, such as IFT, ciliary waveform and beat frequency (reviewed by Pazour & Witman, 2009). Peptide sequences from biochemical

analysis, a multitude of flagellar mutants and a sequenced genome have allowed the identification of many gene loci. Over half of the proteins in the flagellar proteome have human homologues, with BLAST E score of $\leq$ 1e-10, and between 30-90 % conserved identity. These facts make *Chlamydomonas* a clear choice for flagellar studies.

Chlamydomonas deflagellation mutants have been studied for decades. The first *Chlamydomonas* deflagellation–defective mutant, *fa* (flagellar autotomy), was originally generated by UV mutagenesis and does not deflagellate in response to any known stimulus (Lewin & Burrascano, 1983). A second deflagellation mutant, *adf1* (acid deflagellation), was isolated in the stock strains of U. Goodenough as a second unlinked mutation in the *imp4* strain (Finst et al., 1998). This mutant was determined to be defective in deflagellation in response to acid, but unlike *fa1*, it will deflagellate if the membrane is permeabilized in the presence of calcium. Armed with these first two deflagellation mutants Finst et al. (1998) screened 26,000 mutants, generated by insertional mutagenesis and UV mutagenesis, for defective deflagellation. Their specific goal was to identify proteins involved in Ca^{2+} influx regulation and microtubule severing in response to Ca^{2+}. Thirteen deflagellation mutants were isolated in the screen, all of which fall into either the *fa* class or the *adf1* class. Eight of the recovered thirteen mutants displayed the *fa* phenotype; however, complementation tests revealed that two distinct genes were mutated. Four of the mutants were allelic to the original, *FA1*; the other four mutants represented a second gene, *FA2*. Five new

adf1 alleles were isolated, all of which were complementary to *adf1-1* in stable diploids, indicating that only one gene was interrupted.

Following the mutagenesis screen, attempts were made to identify the three genes required for deflagellation. *FA1* was the first to be cloned and encodes a 171 kDa protein, with a large coil-coiled domain and predicted Ca^{2+}/calmodulin binding motifs. Western analysis of whole cell and flagellar fractions suggest that FA1 is a low abundance protein that localizes to the transition zone, at the base of the cilium (Finst et al., 2000). *FA2* encodes a 68 kDa NIMA-related kinase that also localizes to the transition zone. In addition to its role in deflagellation, FA2 also plays a role in cell cycle progression; *fa2* mutants stall at G2/M transition (Mahjoub et al., 2002). *ADF1* has proven much more difficult to identify than either *FA1* or *FA2*.

V: Signaling deflagellation

The pathway that culminates in the severing of the outer microtubule doublets is not well understood. As mentioned earlier, deflagellation can be triggered by many experimental conditions (reviewed by Quarmby, 2004). As far as is known, all involve the generation of a calcium signal. The molecular machinery of deflagellation can be activated in detergent permeabilized cells with the addition of 1μm calcium. In fact, in order to distinguish between *fa* mutants and *adf1* mutants, cells are permeabilized with non-ionic detergent and subjected to 1 mM Ca^{2+}. Under these conditions, *adf1* mutants deflagellate, where the *fa* mutants do not (discussed below).

Deflagellation induced by the addition of weak organic acid requires an extracellular calcium concentration of ~1 mM (Quarmby & Hartzell, 1994). Benzoic acid (50 mM, pH 6.0), or acetic acid (40 mM, pH4.5), are readily membrane permeant in their protonated forms. After moving across the plasma membrane the acids dissociate; the liberated protons acidify the cytosol and trigger a calcium cascade. Influx assays utilizing ^{45}Calcium, developed to study the acid-induced calcium influx, suggest a biphasic response: first, a rapid, initial influx of calcium; followed by a prolonged influx of calcium (Quarmby & Hartzell, 1994). These influx assays revealed that deflagellation–defective *adf1* is also defective in the rapid calcium influx. This result is supported by recent studies by Wheeler et al. (2007). Using biolistically loaded calcium dyes, these researchers found a mean increase of 1.7 +/- 0.12 fold, in internal calcium concentration, in wild type cells exposed to acetate. In response to benzoate, wild type cells exhibited an internal calcium concentration increase of 2.01 +/- 0.19 times higher than baseline. Results for *adf1*, cells that do not deflagellate, showed no increase in internal calcium levels. A diverse set of channel blockers (Cd^{2+}, flufenamic acid, D-600, nifedipine, nicardipine and SKF-96365) have been shown to block the sustained phase, but not the rapid initial phase or deflagellation (Quarmby, 1996). Gadolinium ($Gd^{3+)}$ and Lanthium (La^{3+}) block all acid induced calcium influx and deflagellation. Taken together, these results suggest that the initial rapid calcium influx plays a role in mediating acid-induced deflagellation (reviewed by Quarmby, 2004).

Localization of the acid-induced calcium influx has also fallen under scrutiny. Calcium is a major player in many cellular signaling pathways and therefore needs to be carefully regulated. Calcium channels involved in the regulation of motility in *Chlamydomonas* have been localized to the ciliary membrane. CAV2 is the voltage-gated calcium channel responsible for inducing light or mechanical stimuli into a reorientation of ciliary beat form, causing the cell to move backwards, instead of forwards (reviewed by Quarmby, 2009). Localization of this channel, seen by immunofluorescence, is only to the distal regions of the flagella. This is exciting, as it seems just as likely that other channels could localize solely to the proximal regions or even the transition zone of the flagella.

bld-2 mutants are defective in epsilon tubulin and therefore, lack basal bodies and flagella (Dutcher et al., 2002). However, these cells show nearly normal levels of acid-stimulated Ca^{2+} influx. Based on this finding, two scenarios can be postulated: 1) calcium channels required for this pathway are located on the plasma membrane; or, 2) they are inserted in the plasma membrane only because there are no flagella to localize to. In related experiments, isolated flagella do not accumulate Ca^{2+} in response to pH shock, suggesting that channels are not located in the flagellar membrane. The isolated cell bodies from these experiments are not capable of initiating the rapid calcium influx for approximately 30 minutes post deflagellation; flagella have regrown to ¼ of their length by this time. Cells treated with colchicine do not regenerate flagella after deflagellation, but do recover the ability to have a rapid initial influx of calcium

upon acid treatment, albeit slower than wild type cells. This suggests that flagellar assembly is not required for the reinstatement of the pathway. *fa1-1* cells have levels of calcium influx that exceed that of wild type cells and the rapid initial response is never inactivated. This is evidence for the premise that pathway components localize to the transition zone and that inactivation of the pathway may be the result of a change at the transition zone, following deflagellation, a change that does not occur in the *fa1* mutants. The transition zone and ciliary necklace remain associated with the cell body following deflagellation and undergo structural changes which could affect the operation of ion channels or associated proteins. These data make localization of calcium channels to the transition zone and their involvement in the deflagellation pathway highly plausible.

It is still not known how cytosolic acidification triggers calcium influx or if the initial calcium influx directly activates the machinery of deflagellation. Experimentation on *adf1* cells suggests that the signalling cascade of the acid-deflagellation pathway is initiated at the membrane, with cytosolic acidification triggering the initial, rapid calcium influx. One line of evidence suggests that there are other players in the signalling cascade that results in axonemal severing. Strontium, in place of calcium, can initiate pH-shock induced deflagellation, but not on detergent-permeabilized cells (reviewed by Quarmby, 2006). Presumably, the effector immediately downstream of the first calcium influx cannot tell the difference between calcium and strontium, and therefore induces the second calcium influx, possibly released from internal stores. The severing mechanism,

however, requires more specificity and does not activate in the presence of strontium, in permeabilized cells.

The internal calcium stores of many cells are sensitive to inositol 1, 4, 5-triphosphate (IP_3) and research in *Chlamydomonas* has indicated that phosphoinositide metabolism may be involved in the deflagellation pathway. Deflagellation, either by low pH or mastoparan, a G protein activator, is correlated with increased levels of inositol 1, 4, 5-triphosphate, likely triggered through activation of phospholipase C (PLC). In *Chlamydomonas*, pH-shock induces an IP_3 accumulation 10 times greater than baseline levels. There is an accompanying decrease in phosphotidylinositol 4,5-bisphosphate (PtdIns P2), indicating activation of PLC and, although increased levels of diacylglycerol (DAG) are not seen, an increase in phosphatidic acid suggests DAG kinase is activated (reviewed by Quarmby, 2004). It has been shown in *fa1-1* cells that the activation of PLC is not a consequence of deflagellation, as these cells have normal levels of IP_3 in response to acid, but as we know, do not deflagellate.

It is obvious that there is activation of the PLC pathway in response to acid, however, does the activation of PLC mediate deflagellation or is it an independent response to acidification? There is some data to suggest that PLC can transduce cytosolic acidification into a calcium signal. In experiments using continuous-flow rapid-quench techniques, the early events of deflagellation have been recorded. In the first 400 ms following acidification there is a 1.6 to 2 fold increase in IP_3 (Yueh & Crain, 1993). This increase peaks between 200 and 400 ms, returns to baseline 150 ms post peak and precedes deflagellation. There is a

second IP_3 peak distinct from the initial rapid peak which is smaller and occurs at ~ 50 seconds post acidification. This is interesting considering that calcium also exhibits a biphasic response to pH-shock, with rapid and prolonged phases. An inhibitor of PLC, neomycin, prevents the initial, rapid IP_3 increase and deflagellation in response to acid (Quarmby et al., 1992). Neomycin is an aminoglycoside antibiotic, which binds PtdIns P2, thus preventing its hydrolysis by PLC. Sadly, unpublished data suggests that the block of deflagellation by neomycin is independent of the PLC inhibition. Calcium influx and deflagellation are both blocked by La^{3+}. La^{3+} doesn't inhibit PLC activation following acidification, suggesting that PLC activation must be upstream of the Ca^{2+} influx. Therefore, if PLC was in the pathway, then neomycin would be expected to block calcium influx following pH-shock. It does not; therefore, neomycin must be blocking deflagellation via a PLC-independent mechanism. Despite this data, PLC activation in the deflagellation pathway is still intriguing. Flagellar proteome protein, flagellar associated protein 48 (FAP48) is identified by 33 unique peptides and is similar to an IP_3 Receptor Type 3 (IPTR3), a calcium releasing channel (Pazour et al., 2005). Additionally, unpublished results suggest that *adf1* is deficient in IP_3 accumulation in response to acid-shock (cited in Yueh & Crain, 1993). The activation of PLC and the resulting signal cascade remains an open issue. Determining the protein product of *ADF1* will aid in understanding how PLC activation fits into the deflagellation pathway, as well as unveiling part of the deflagellation pathway itself.

VI: ADF1

As I have already discussed, *adf1* deflagellation mutants are defective in proton-activated calcium influx and as a result do not shed their flagella in an acidic environment. Alleles, *adf1-2* through *adf1-4*, were recovered in a screen for mutants generated through insertional mutagenesis, whereas *adf1-6* was recovered in a screen for mutants generated through UV mutagenesis (Finst et al., 1998). The presumptive, insertional mutants were backcrossed in order to determine linkage of the mutation to the inserted selection marker. Progeny from these crosses did not exhibit linkage. In addition, southern blot analysis performed on backcrossed strains did not show co-segregation of the mutants with insertional DNA-derived probes. There is some chance that the alleles were spontaneous mutants, like *adf1-1*, coincidentally found because a deflagellation screen was performed. It is also possible that the gene would not tolerate the insertional DNA and it was excised, perhaps simultaneously causing a deletion.

Prior to my assignment to the project, Quarmby lab graduate student Jeremy Parker mapped the *ADF1* locus to linkage group IX. Upon my commencement of the project, I had the goals of mapping the locus, cloning and identifying *ADF1*. The reclusive gene has proven difficult to find and I have succeeded in narrowing the search to a 394kb region on linkage group IX, using PCR-based recombinant mapping. Using PCR, amplifying 2 kb fragments and digesting them with restriction enzymes, I attempted to identify polymorphisms. In addition, I have performed transformation of the allelic strains with BACs containing wild type DNA, with the goal of rescuing the mutant phenotype.

Finally, I have done a thorough review of the candidate genes in the 394 kb region, outlining the best candidates for subcloning for ongoing research.

METHODS AND MATERIALS

I: *Chlamydomonas reinhardtii* strains and growth conditions

Chlamydomonas strains required for these experiments, are available from the *Chlamydomonas* Genetics Center (Duke University, Durham, NC), with the exception of B214. Wild type strain B214 was obtained from Dr. Greg Pazour at the University of Massachusetts Medical School. a*df1* mutants, alleles 1-6, were isolated as previously described by Finst et al. (1998), and maintained in the Quarmby laboratory. A standard laboratory wild type strain, 137c (mt+ and mt-) and the B214 strain were used as wild type controls for all experiments. The inter-fertile field-isolate strain, S1D2 (mt-; CC-2290) was used in genetic crosses. Paralyzed flagella mutant, *pf16*, also was used in crosses. Double mutants: 20C, 2, 5 and 57B were generated in this study for PCR-based recombination mapping. Stains R3 and NO were used to produce gametic lytic enzyme (GLE) for use in transformations. *Chlamydomonas* strains were maintained on 1.5% TAP, under light, at 16 ºC. Table 1 lists the *Chlamydomonas* strains required for the mapping of *ADF1,* their phenotypes, and a brief description of what they have been used for.

Table 1: *Chlamydomonas* strains used to map *ADF1*.

Strain	Phenotype	History
137c	wild type	genetic analysis; phenotypic quantification
B214	wild type	genetic analysis; phenotypic quantification
S1D2	wild type	RFLP strain; genetic analysis
pf16	paralyzed flagella	genetic analysis; phenotypic quantification
adf1-1	acid deflagellation mutant	spontaneous mutant; genetic analysis; phenotypic quantification
adf1-2	acid deflagellation mutant	B214 background; genetic analysis; phenotypic quantification
adf1-3	acid deflagellation mutant	B214 background; genetic analysis; phenotypic quantification
adf1-4	acid deflagellation mutant	B214 background; genetic analysis; phenotypic quantification
adf1-5	acid deflagellation mutant	B214 background; genetic analysis; phenotypic quantification
adf1-6	acid deflagellation mutant	g1 background; genetic analysis; phenotypic quantification
20C	*pf16, adf1-2* double mutant	generated in this study (see results);genetic analysis
2	*pf16, adf1-3* double mutant	generated in this study (see results);genetic analysis
5	*pf16, adf1-6* double mutant	generated in this study (see results);genetic analysis
57B	*pf16, adf1-2* double mutant	generated in this study (see results);genetic analysis
R3	wild type	used for GLE preparation
NO	wild type	used for GLE preparation

II: Genetic crosses

Chlamydomonas strains were mated, as described in The *Chlamydomonas* Sourcebook (edited by Stern et al., 2009), with slight modifications. Strains of opposite mating type (+ or -), were plated separately on 1.5% TAP plates and allowed to grow for five days, under bright light, at 25º C. After five days cells were transferred to Low-N (nitrogen reduced) plates, in order to induce gametogenesis (Sears et al., 1980), then kept under light for two more days. On the eighth day the plates were flooded with 5 ml of mating buffer (0.60 mM MgCl2, 1.20 mM HEPES, pH 6.8) and left in the light for 30 minutes. The plates were then scraped with inoculating loops to loosen cells not yet swimming and liquid containing cells was transferred to flasks to shake at 40 rpm, in the light, for two hours. After two hours cells had generated flagella and liquid cultures of opposite mating types were combined into one flask to shake for two

more hours. Combined liquid cultures were assessed for quadraflagellate cells and then 300 µl of the culture was aliquoted onto 4% TAP plates and placed in the dark, at 16 ºC. After two days of incubation in the dark the plates were inverted over chloroform for 30 seconds to speed up the death of vegetative cells, and then placed back in the dark for three more days. After a total of five days in the dark, plates were scraped of dead vegetative cells, while zygotes remained embedded in the 4% agar. Zygotes were handpicked, with a modified glass pipette, into gridded boxes imprinted on 1.5% washed-TAP plates (to prepare washed TAP agar is repeatedly rinsed with deionized H_20 before TAP is autoclaved) and placed in the light overnight in order to induce meiosis. Meiotic tetrads were separated using hand-drawn glass tools under the dissecting microscope. Resulting colonies were grown up on 1.5% TAP. Colonies were grown up to be scored for the ability to swim and the ability to deflagellate in acid, as described below.

III: Phenotypic assays

Assays for flagellar paralysis and acid defective deflagellation were carried out in the following manner: each well of a 96 well plate, containing 150 µl of minimal media, was inoculated with toothpick scrapings of genetic cross progeny strains, and placed under the light. Assays were carried out following a minimum incubation of 2 hours and up to a maximum of 48 hours, after inoculation. Individual assays were carried out by placing two 4 µl aliquots of culture on a microscope slide. One aliquot was observed for swimming behaviour and the other was treated with 4 µl of acid (40 mM Na acetate, pH 4.5, 1 mM $CaCl_2$; Finst

et al. 1998), for 30 seconds, then fixed with 4 µl of 1% glutaraldehyde. Colonies were scored as paralyzed flagella if cells were not swimming, or only twitching slightly. Careful action was taken to observe that the cells had actually grown flagella; they are visible at 10x magnification (objective Achrostigmat 10x/0,25 PH1, Zeiss) under dark field illumination. Acid deflagellation was determined defective if the cells retained flagella after acid treatment, as viewed at 10x magnification. Under 10x magnification flagella floating in the media was a sign of wild type deflagellation, along with flagella-less, "bald", cells observed at 100x (objective Achroplan 100x/1,25 Oil PH 3, Zeiss) magnification. Phenotypes were recorded as: *adf1* only, single mutants; *pf16* only, single mutants; *adf1, pf16* double mutants; or S1D2, wild type.

IV: *Chlamydomonas* genomic DNA isolation

In the pursuit of a reliable, high-throughput, relatively non-toxic protocol for the isolation of genomic DNA, several different methods were employed over the course of this project. Initially, *Chlamydomonas* genomic DNA was isolated using DNAzol-ES (Molecular Research Center Inc.). As this product was designed for use with plant material, a modified protocol was developed in our lab, based on the manufacturer's instructions (DNAzol ES, Manufacturer protocol, 2006). Cells were plated on 1.5 % TAP for five days, or until a healthy, green lawn developed. A loop-full of cells (10 µl loops) was scraped from the plate into 50 µl of dH2O. Cells were pelleted by centrifugation (1 min, 14000 rpm) and the supernatant discarded. Pellets were weighed and adjusted to a range of 50-60 mg. The pellet was then resuspended in 200 µl of DNAzol-ES, by gentle pipetting, and mixed by

gentle rotation for 15 minutes at room temperature. An equal volume (200 µl) of chloroform was added and tubes were inverted for 20 seconds, incubated on the bench top for five minutes, and pelleted by centrifugation (5 minutes, 14000 rpm). The aqueous top layer was removed to a clean 1.5 ml Eppendorf tube and 150 µl of 99.9% ethanol was added. Tubes were inverted to mix and then incubated at -80º C for 15 minutes. The solution was then centrifuged for 5 minutes at 7800 rpm, and the supernatant was discarded. Pellets were resuspended in 150 µl of EDTA (10mM, pH8) and 750 µl DNAzol wash (1 volume DNAzol: 0.75 volume 99.9% EtOH). The mixture was incubated for 5 minutes at room temperature, pelleted by centrifugation at 7800 rpm for five minutes, and supernatant was discarded. The pellet was washed with 150 µl of 70 % EtOH, lightly mixed on the bench top vortex, re-pelleted (30 sec, 7800 rpm) and, the supernatant was discarded. Tubes were inverted to dry upside down for five minutes, then dried right side up for another five minutes. 80 µl of sterile, deionized H_2O was added to the pellet, which was allowed to resuspend over 2 nights at 4 ºC. This method was used to isolate DNA for PCR based recombination mapping as it was less timely, as well as less toxic than the following phenol: chloroform method. The DNAzol method was later replaced because the DNA yield was low (an average of 10 ng/µl) and contaminated with carbohydrate.

A second DNA isolation method was utilized in order to yield a higher concentration and better quality DNA for the 2Kb PCR walk, oligonucleotide array Comparative Genomic Hybridization (discussed in Results) and sequencing protocols. Phenol: chloroform: isoamyl alcohol extractions were based on the

method used in Finst et al. (1998). Cells were plated on 1.5 % TAP, and grown for five days, or until a green lawn developed. The full plate of cells was transferred to a 1.5 ml Eppendorf tube and resuspended in 1ml of liquid TAP, and centrifuged for 2 min at 14000 rpm. The supernatant was discarded and the cell pellet was resuspended in 400 µl of TEN, pH8 (10 mM Tris-HCl, 1 mM EDTA, 100 mM NaCl). 40 µl of 20% SDS, 40 µl of 20% SLS and 200 µl of (10 mg/ml) proteinase K (Sigma) were added and then the tube was inverted to mix. The tubes were then incubated at 65º C for 1.75 hours. Samples were phenol-chloroform extracted (Sambrook & Russell, 2001) and treated with RNase A (Sigma). After ethanol precipitation, the DNA was resuspended in 100 µl sterilized, deionized water.

A third method of DNA isolation was used for PCR based recombinant mapping late in the project. This protocol was obtained from the *Chlamydomonas* Genetics Center webpage, courtesy of Steve Pollack of Louisiana State University (http://www.chlamy.org/methods/quick_pcr.html). A 5-10 µl loop-full of *Chlamydomonas* cells was scraped from a fresh plate and resuspended in 50 µl of 10 mM EDTA by vortexing. Tubes were incubated at 100º C for 5 minutes in the thermal cycler, vortexed to mix, and centrifuged at 14000 rpm for 1 minute. The resulting supernatant was used directly in PCR reactions, 1.5 µl in a 25 µl reaction. DNA isolated by this method had to be used immediately.

V: PCR

Primers for PCR were designed from genomic sequence published by the Joint Genome Institute (JGI), *Chlamydomonas reinhardtii* version 3.0

(http://genome.jgi-psf.org/Chlre3/Chlre3.home.html), using PrimerQuest, located on the website of Integrated DNA Technologies (IDT; (http://www.idtdna.com/Scitools/Applications/PrimerQuest/). Oligonucleotide primers were manufactured by either Invitrogen or IDT and shipped in lyophilized form. Lyophilized primers were reconstituted to 250 pmoles/µl concentration and then diluted 1:10 in a master mix, containing both forward and reverse primers, to a final concentration of 25 pmoles/µl for each primer, (50 pmoles/µl total). Hot start PCR was used for all PCR reactions; this allowed the 60% GC rich Chlamydomonas genomic DNA to have a 30 second head start in denaturing before DNA polymerase was added to the reaction mixture. Test PCRs were run, using control DNA from wild type strains, 137c and B214, and polymorphic strain S1D2, to optimize each primer set for annealing temperature and extension time. Following the optimization, PCR was run on test samples. Hot start PCR required that two sets of reaction buffers be assembled. For mapping PCRs, Taq DNA polymerase was used, and buffers specific to that enzyme were generated. The first buffer contained: sterilized, deionized water, 5x Q (5 M Betaine), diluted to 1x concentration; 10x ThermoPol buffer (New England Biolabs; NEB), diluted to 1x concentration; 50mM $MgCl_2$, 1 µl per reaction; primer master mix (50 pmoles/µl), diluted to a final concentration of 4 pmoles/µl; and finally, dNTPs at a concentration of 10mM. I routinely called this Master Mix I (MMI, as seen in Table 1). Master Mix II was made up of sterilized, deionized water, 5x Q (5 M Betaine, Sigma-Aldrich), diluted to 1x concentration; 10x ThermoPol buffer (NEB), diluted to 1x concentration; and Taq DNA polymerase (NEB), as outlined in Table 2.

Table 2: Recipes for Taq DNA polymerase hot start PCR master mixes.
Based on 1 (1x) reaction, 10 (10x) reactions and 15 (15x) reactions.

	1x		10x		15x	
	MMI	MMII	MMI	MMII	MMI	MMII
dH2O	7.25	7.25	72.5	72.5	108.75	108.75
5xQ	2.5	2.5	25	25	37.5	37.5
10x buffer	1.25	1.25	12.5	12.5	18.75	18.75
$MgCl_2$	1		10		15	
primers	0.5		5		7.5	
dNTPs	0.5		5		7.5	
Taq		0.25		2.5		3.75
total	13 ul	11.25 ul	13 ul/ tube	11.25 ul/tube	13 ul/ tube	11.25 ul/tube

A volume of 13 µl of master mix I was added to 1 µl of DNA, aliquoted into the bottom of a thin walled, 200 µl PCR tube. PCR was run in the Mastercycler Gradient (Eppendorf). The PCR program was set up to run with a 1 minute warm up at 95º C, followed by a pause, to allow the samples to be loaded into the machine. After samples were loaded and incubated for 30 seconds at 95º C, their tube tops were flipped open and 11.25 µl of Taq polymerase containing Master Mix II was added. The final reaction volume for each sample was 25.25 µl. The machine was then locked closed to cycle through melting, annealing and extension phases. The first three cycles included: a denaturing phase, of 30 seconds at 95º C; an annealing phase, either on a gradient or at an experimentally determined temperature, for 30 seconds; and then, an extension phase at 74º C, for one to two minutes based on the length of the desired product. The last 30 cycles were completed with a step down of one degree for the annealing temperature. The initial annealing temperature was determined for specificity. Therefore, after the first three rounds, enough desired template had been generated such that, that a drop in annealing temperature would induce

more rapid primer binding in later cycles and greater product yield. Table 3 outlines the generic Mastercycler (Eppendorf) program, as described.

For test PCRs, each primer set was run in triplicate on a gradient that varied the annealing temperature, usually 65 +/- 3º C, while the melting and extension temperatures were invariable, unless a second test PCR was required. In the event that a second test PCR was required annealing time was shortened, extension time was altered, or the gradient was changed to include annealing temperatures as low as 55º C. For sample PCRs, changes to this program were made based on annealing temperatures discovered during optimization PCRs, and on the length of the predicted product. Generally, extension times were determined based on a 1 minute/ 1kb guide for Taq DNA polymerase (http://www.neb.com/nebecomm/products/protocol54.asp).

Table 3: Generic Mastercycler program for PCR.
Temperature step down, as described in the text, starts at step 9.

Step	Task
1	T= 95 ºC t= 1 min
2	Sound 3
3	Pause Press Enter
4	T= 95 ºC t= 30 s
5	T= 95 ºC t= 30 s
6	T= 68 ºC t= 30 s
7	T= 74 ºC t= 2 min
8	Go To 5 Rep 3
9	T= 95 ºC t= 30 s
10	T= 67 ºC t= 30 s
11	T= 74 º C t= 2 min
12	Go To 9 Rep 30
13	T= 74 ºC t= 5 min
14	Hold 4 ºC Enter

PCR reactions for the purpose of sequencing required a higher fidelity enzyme, so the proof reading enzyme Phusion (Finnzymes) was used, in place of Taq polymerase. Hot start PCR was still used and therefore Master Mix solutions had to be amended to reflect that. Table 4 outlines the Master Mix components for Phusion PCR reactions. Extension times were reduced to 30 seconds per 1000 bp amplified for Phusion (http://www.finnzymes.com/pcr/phusion_products.html).

Table 4: Recipes for Phusion hot start PCR master mixes.
Based on 1(1x) reaction, 10 (10x) reactions and 15 (15x) reactions. All quantities are in µl.

	1x		10x		15x	
	MMI	MMII	MMI	MMII	MMI	MMII
dH2O	5.05	5.05	50.5	50.5	75.75	75.75
5xQ	2	2	20	20	30	30
5xHF	2	2	20	20	30	30
MgCl2	1		10		15	
primers	0.4		4		6	
dNTPs	0.5		5		7.5	
Phusion		0.15		1.5		2.25
total	10.95 ul	9.2 ul	10.95/tube	9.2/tube	10.95/tube	9.2/tube

VI: BAC DNA isolation

BAC DNA was isolated using an optimized patchwork protocol based on methods from PhasePrep BAC DNA Kit (Sigma), Large-Construct Kit (Qiagen), and a protocol by Villalobos et al. (2004). BAC-containing bacterial glycerol stocks were streaked onto fresh Luria Bertani (LB) plates containing chloramphenicol (12.5 µg/ml) plates, and incubated at 37º C overnight. The following evening colonies were picked into 5 ml liquid starter cultures of LB with chloramphenicol (12.5 µg/ml), and grown up overnight. The starter culture was

added to 500 ml of LB and chloramphenicol at a 1:500 concentration, placed on a shaker (250 cycles/ minute) and grown overnight to an OD_{600} of 2. The culture was then centrifuged at 5000 *g* for 10 minutes, at 4º C. The supernatant was poured off and the bottles were inverted to allow the last drops to drain away from the bacterial pellet. The pellet was resuspended in 25 ml of ice cold P1 (Qiagen; 50 mM Tris-CL, pH 8.0; 10 mM EDTA; 100 µg/ml RNase A) by gentle pipetting. The bacteria were then lysed with the addition of 25 ml of solution P2 (Qiagen; 200 mM NaOH; 1% SDS) and gentle rotation and inversion, then set on the counter top for a strict four minute incubation. At four minutes the solution was neutralized with 25 ml of P3 (Qiagen; 3.0 M Potassium Acetate, pH 5.5), gently mixed by slow rotation and inversion, then incubated on ice for ten minutes. The insoluble fraction of the lysate was pelleted by centrifugation at 15000 *g*, for 15 minutes, at 4º C and discarded. The supernatant was transferred to a clean bottle and centrifuged again, at the same speed, temperature and time. The cleared supernatant was transferred once more to a clean bottle in preparation for precipitation. A 100 µl aliquot was reserved for ethanol precipitation to confirm the presence of BAC. Isopropanol precipitation was carried out at room temperature with the addition of an equal volume of isopropanol (about 75 ml) and gentle inversion to mix. The solution was then centrifuged for 15 minutes at 15000 *g* at room temperature. The supernatant was poured off and the bottle was inverted to drain the bottle of remaining isopropanol. The pellet was resuspended in 10 ml of 10:50 TE (10 mM Tris-Cl, pH 7.6; 50 mM EDTA), and transferred to a 15 ml disposable Falcon tube.

Protein was precipitated out by the addition of 5 ml of 7.5 M potassium acetate and incubation at -80º C for 30 minutes. The solution was removed from the freezer and left on the counter to thaw, then centrifuged at 5000 *g* for 15 minutes at 4º C. The supernatant was transferred to 50 ml tubes and ethanol precipitated upon the addition of a 2.5x volume of 95 % ethanol, then incubated at -20º C for 30 minutes. The DNA was pelleted by centrifugation at 15000 *g*, for 5 minutes, at 4º C. The DNA pellet was resuspended in 5 ml of 50:50 TE. A second 100 µl aliquot was taken at this point to verify the presence of BAC DNA and likely RNA contamination. BAC DNA isolate was treated with RNase A (10 µg/ml) overnight at room temperature. The DNA was then precipitated for the final time with one volume of isopropanol, incubated at -20º C overnight and then pelleted by centrifugation. The resulting pellet was reconstituted in 500 µl sterile, deionized H_2O and the DNA concentration was measured with the ND-1000 spectrophotometer (NanoDrop). Samples, including the 100 µl aliquots taken throughout the protocol, were run on agarose gel to confirm that BAC DNA was present, and that RNA contamination had been removed.

VII: Bacterial artificial chromosome library screen

It was necessary to screen the *Chlamydomonas* BAC library to find a BAC that spanned the region Sc23: 846721 - 931782. We already had one Easy-to-Screen High-Density Filter, from Incyte Genomics, in the lab. However due to its age, I ordered a second one from Clemson University Genomics institute (http://www.genome.clemson.edu/capabilities/bacCenter.shtml). To screen the library, I made use of the probe hybridization protocol provided by Incyte

Genomics (Easy-To-Screen High-Density Filter: BAC *Chlamydomonas* product Manual, 2000) on the old filter. When this proved successful, I performed the protocol a second time on the new filter, for two sets of results. I prepared the probe template using primer set 171537s, covering Sc23: 887383 -889427, for a product size of 2044 bp, and gel purified the resulting PCR product (QiaQuick, Qiagen). The dry filter was soaked in hybridization solution, rolled, and placed in a bottle containing 60ml of hybridization solution. The bottle was rotated for one hour at 65º C. During the one hour pre hybridization of the filter, the probe was prepared. The probe was radio-labelled according to the protocol provided in the Ladderman Labeling Kit (TakaRa Bio Inc.), and assessed for incorporation of radioactivity by following column filtration of the PCR product. The radio labeled probe was then denatured and added to the hybridization bottle and incubated overnight at 65º C. Following overnight hybridization, the filter was washed with low-stringency wash buffer twice, and then washed with high-stringency wash buffer four times. The bottle, containing 60 ml of high-stringency wash buffer and filter were then placed back in the hybridization oven for one hour at 65º C. The waste buffer was discarded and the filter was exposed to a phosphor screen overnight. The phosphor screen was imaged with the Storm 820 PhosphorImager, and analyzed with ImageQuant TL (GE Healthcare; all courtesy of the Unrau Lab, SFU). Figure 5 shows the blot images with grids overlaid. The circled squares both have two spots of radioactivity. The orientation of those spots to each other indicates that the BAC of interest is in plate 22 and, the grids

indicate row N, well 9 (see Incyte product manual cited above for further explanation).

VIII: Subcloning of amplicons in preparation for sequencing

Vector pGEM-T (Promega) was used to subclone PCR fragments destined for sequencing. Table 5 lists the primer sets used to amplify fragments containing predicted SNP sites and candidate gene 171537. Following PCR, as described above, fragments were separated on agarose gels. Gel slices, containing bands of desired size, were cut out of the gel and purified using QiaQuick Gel Extraction Kit (Qiagen). Following gel extraction an aliquot of purified DNA was run on a gel to confirm that the gel purification was successful. Fragments were then ligated into the pGEM-T Easy vector as described in the pGEM-T and pGEM-T Easy Systems Technical Manual (Promega; http://www.promega.com/tbs/tm042/tm042.pdf). Positive clones were determined based on blue/white selection and grown in overnight cultures of LB containing ampicillin (50 µg/ml). Plasmid DNA was isolated using GeneJet Plasmid Miniprep Kit (Fermentas), digested with EcoRI (NEB) to confirm the insert size and then set for sequencing to Macrogen (Korea; http://www.macrogen.com/eng).

Figure 5: Phosphorimage of BAC library filters.
Images have been overlaid with grids that allow interpretation of hybridization patterns. The image to the left is the older blot, while the image to the right is of the newer blot. Both grids indicate well 9 of row N. The orientation of the small dots inside the circled squares indicate plate 22 – BAC 22N9.

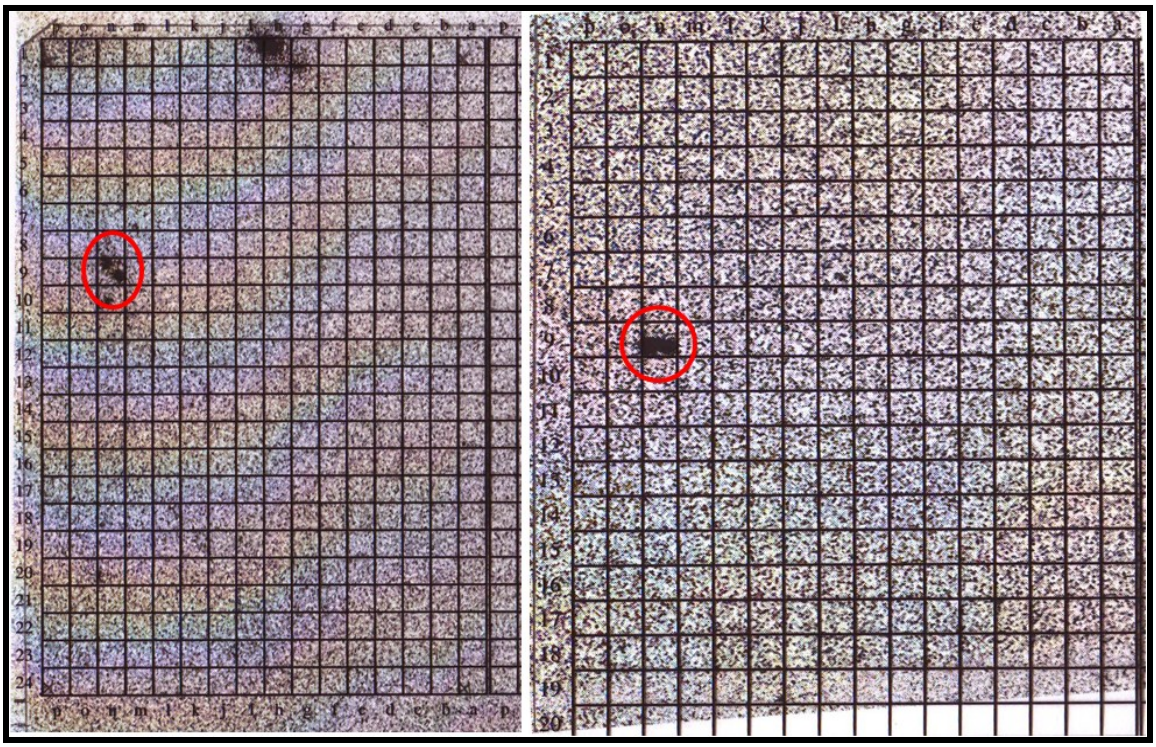


Table 5: Primer sets used to amplify fragments destined for sequencing.

	Primer Set	Predicted Fragment Size
oaCGH Predicted SNPs	171530	1441
	171535t	2283
	171536s	1355
	171537e	2026
	171512h	1644
	171540	1838
	171513SNP	2224
	940041e	1875
Predicted gene 171537	171537s	2045
	171537m	1915
	171537b	1912
	171537t	2100
	171537e	2026
	171509s	2134
	171509e	2013
	171510s	2311

Candidate gene 171510 (predicted adenylyl cyclase,) was subcloned by digesting BAC 22N9 with ApaI and NotI (New England Biolabs). The digested BAC DNA was run on a 0.5% agarose gel. Based on NEBCutter (http://tools.neb.com/NEBcutter2/index.php; Vincze et al., 2003) predictions for fragment sizes following restriction digest, a 7769 bp band was extracted from the gel, purified (as described previously) and ligated into the vector pBluescript. pBluescript DNA was isolated using the GeneJet Plasmid Miniprep Kit and digested with restriction enzyme ApaI. The digested plasmid was subjected to dephosphorylation, using Antarctic Phosphatase (New England Biolabs) to prevent recircularization prior to insert ligation. Ligation reactions were incubated overnight at 4º C, transformed into chemically competent *E.coli* and plated on LB plates containing ampicillin (50 µg/ml). Positive clones were picked based on blue/white selection and grown up in 5 ml overnight cultures (LB and ampicillin, 50 µg/ml). Plasmid DNA was isolated and digested with ApaI and NheI to confirm insert size. The plasmid predicted to contain candidate gene 171510 was then sent for sequencing to Macrogen.

IX: Preparation of gametic lytic enzyme (GLE)

Gametic lytic enzyme is required to digest the cell wall of *Chlamydomonas* cells before transformation. In nature, cells secrete gametic lytic enzyme prior to mating, and so can be induced to synthesize it in the laboratory when placed under mating conditions. The protocol outlined here is modified from the *Chlamydomonas* Sourcebook (Harris, 1989). Wild type cells, R3 (mt+) and NO (mt-), were plated on regular TAP for 5 days. They were then transferred to Low-

N plates for two days to induce gamete formation. Cultures on agar plates were suspended in liquid Low-N and shaken under light for 2 hours. Strains of opposite mating type were mixed together to shake for 30 minutes under the light, in a wide bottom flask. During this time cells release GLE into the media to facilitate fusion and the formation of zygotes. The cells are then harvested by centrifugation (5 minutes, 3000 *g*). The supernatant is aliquoted into 15 ml Falcon tubes, immersed in liquid nitrogen, and stored at -80º C. Freezing the GLE eliminates the possibly of contamination of transformants with vegetative cells. Prior to use GLE must be thawed quickly at 42º C, then incubated on ice.

X: *Chlamydomonas* transformations

Transformation of *Chlamydomonas* cells was carried out by two methods: glass bead bombardment and electroporation. Glass bead transformations were performed as previously described by Kindle et al. (1990), with some modifications. Cells were inoculated from plates into 200 ml liquid TAP and grown overnight. A hematocytometer was used to determine the number of cells in the culture. The culture was then pelleted and resuspended in GLE, which had been previously prepared, and left under the light to shake for 30 minutes. The culture was centrifuged to pellet the cells and the pellet was resuspended in fresh TAP containing 40 mM sucrose, so that the final concentration of cells was 1 x 10^8 cells/ml. One ml of cells was used per transformation. Cells were added to 15 ml Falcon tubes containing 300 µg of glass beads (1mm in diameter; Sigma) and DNA. The cells and DNA were then agitated at maximum speed with the bench top vortex (Fisher Vortex Genie II) for 30 seconds. DNA was added to the glass

beads just prior to the addition of *Chlamydomonas* cells. Co-transformations required the addition of 1 µg of plasmid pSI103 (4982 bp), which confers paromomycin resistance to *Chlamydomonas*, and BAC, or a second plasmid (ie. pBS+171510, subcloned gene) DNA in a 1:1, 1:2 or 1:3 ratio. The amount of BAC DNA used for transformations was determined using the calculation below (modified from pGEM-T and pGEM-T Easy Systems Technical Manual; Promega):

$$\frac{\text{µg of vector x kb size of BAC}}{\text{kb size of vector}} \times \text{BAC: vector molar ratio} = \text{µg of BAC}$$

The cultures were set to shake under light overnight. Cells were then pelleted by centrifugation and spread on selective agar plates using sterile inoculating loops. Paromomycin-resistant colonies were visible by 5 days.

Transformation by electroporation was carried out as described by Brown et al. (1991), with modifications. Cultures were grown, pelleted and resuspended to a concentration of 1 x 10^8 cells/ml as described above. Cells were placed in disposable cuvettes with DNA (same amounts as described above for glass bead transformation) and were pulsed at 1000V (capacitance was 50 µF) once, then immediately again (total of 2 pulses). Cells were then transferred briefly to ice and then to 15 ml Falcon tubes containing 10 ml of fresh TAP media, to shake overnight under light. Cells were plated as described above.

RESULTS

I: Optimization of the deflagellation screen

It was imperative to the mapping strategy that conditions for efficient and accurate assessment of deflagellation and motility be refined in order to screen large numbers of colonies rapidly. I determined that cells grown on solid media could be transferred to liquid media in 96 well plates and grown under the light for up to 48 hours, to provide adequate numbers of flagellated cells for assay. Assays were conducted following a minimum of 2 hours incubation in 96 well plates by placing 5 µl of liquid culture on a slide, adding 5 µl of acid and, fixing at 30 seconds, with 5 µl of 1% glutaraldehyde. The sample was then covered with a cover slip and immediately viewed under the microscope.

II: PCR-based recombination mapping

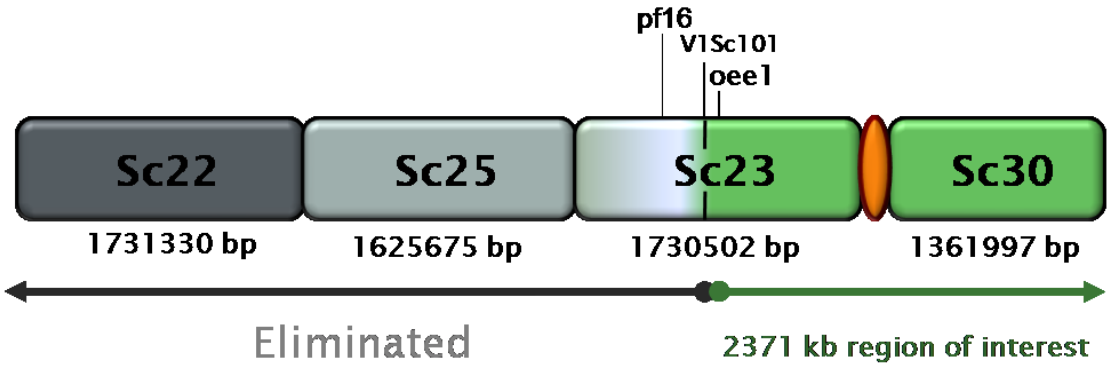
When cloning genes disrupted through insertional mutagenesis, the ideal process for identification of the interrupted gene involves amplifying and sequencing the DNA flanking the insertion. However, with regards to the *adf1* mutants, Southern analysis and backcrosses revealed no linkage to the insertional plasmid, demanding that another method be employed to determine the locus of the mutations. We chose PCR-based recombination mapping. Map-based cloning depends on two determinants: first, the existence of a genetic or physical map; and second, the ability to generate progeny of sexual crosses that segregate for the trait of interest as well as phenotypic or molecular marker

(Rymarquis et al., 2005). With genetic markers on every arm of each of 17 linkage groups, a sequenced genome (JGI), and the ability to cross *C. reinhardtii* to an inter-fertile strain, S1D2 (*C. Grossii*), which has a profusion of sequence tagged sites (STS), cleavable amplified polymorphic sequences (CAPS), single nucleotide polymorphisms (SNP), and RFLP markers (Gross et al., 1988; Vysotskaia et al., 2001; Grossman et al., 2003), *Chlamydomonas* has all the requisites for map-based cloning. My goals, with regards to PCR-based recombination mapping, were: 1) to develop the PCR markers that would allow the visualization of differences between the *adf1* mutants and S1D2; 2) to generate a double mutant strain (*pf16, adf1*) that facilitated the identification of recombinations in the region of interest; and, 3) to use these data to narrow down the region we suspect *ADF1* localizes to.

Large tracts of the *Chlamydomonas* genome have been sequenced (~120 megabases) and organized as scaffolds (Merchant et al., 2007). Each scaffold has been associated with one of *Chlamydomonas*' seventeen linkage groups. Previous work utilized genetic mapping to place the *ADF1* locus on linkage group IX, based on linkage of *adf1* with marker Oee1(see Figure 6; Jeremy Parker, unpublished data). Linkage group IX, based on historical mapping data (Kathir et al., 2003; Rymarquis et al., 2005) and Version 3 of the *Chlamydomonas reinhardtii* genome (JGI), is comprised of four scaffolds: scaffold 22, 1731 kb; scaffold 25, 1626 kb; scaffold 23, 1731 kb; and scaffold 30, 1362 kb. By employing both molecular and genetic markers *ADF1* was mapped to the right of

marker V1Sc101, a region of approximately 2371 kb, and this is the region I started with (see Figure 5).

Figure 6: Linkage group IX.
Includes scaffolds 22, 25, 23 and 30 and markers pf16, V1Sc101 and Oee1. The linkage group is approximately 6450 kb, of which 4079 kb were eliminated, through PCR-based recombinant mapping. Eliminated region is depicted in gray, region of interest is green, and predicted centromere is orange.



PCR-based recombination mapping relies on the generation of a large number of progeny, from crosses between a strain carrying the mutation in question and a polymorphic strain. In *Chlamydomonas*, a three week period is required, from initial plating of parental strains, to generate progeny ready for primary assays. Relevant progeny, those with the desired phenotype, are then grown for genomic DNA isolation, which takes an additional week. It was important, therefore, to be able to determine quickly and visually that recombination had taken place. To ensure rapid diagnosis of phenotype and increased potential for recombinant recovery, I decided to generate a double mutant carrying *adf1* and a second mutation, *pf16*, under the premise that subsequent recombination events between *pf16* and *adf1* would yield single mutants. The *PF16* locus is located on the same chromosome as *ADF1*, on

scaffold 23 (see Figure 6) and cells carrying this mutation can be easily identified by observation in liquid culture. The *pf16* mutants are unable to swim, occasionally twitch, but mostly rest on the bottom of assay wells. Using the *adf1-2* allele, I recovered three double mutant strains: *pf16, adf1-2* (20C, mt+), *pf16, adf1-2* (17A, mt+) and *pf16, adf1-2* (12B, mt-). As two of the new double mutants had opposite mating type (mt+) from the polymorphic strain, S1D2 (mt-), I was able to start crosses for mapping purposes immediately.

Double mutant strains, 20C (mt+) and 17B (mt+), were mated with mapping strain S1D2 (mt-); meiotic tetrads were dissected, grown and assayed. Colonies, with single mutant phenotypes, indicated that a recombination event had taken place between the *PF16* and *ADF1* loci, and were therefore reserved for genomic DNA isolation and PCR. Experimental groups 1 and 2 were produced from these initial efforts. However, strains within this group were generated solely from the 20C (mt+) double mutant, as the double mutant 17B(mt+) did not mate efficiently. Subsequent crosses between *pf16, adf1* (20C, mt+) and S1D2 (mt-) resulted in groups: 3, 7 and 7II. Table 6 is a summary of: the crosses of the first double mutant, 20C (mt+), with S1D2 (mt-); the number of meiotic cells; the progeny assayed for paralyzed flagella and the ability to deflagellate in acid; the experimental group (strains that were subjected to PCR, following assays); and the number of recombinant strains recovered in those experimental groups. Note that tetrads frequently went through a round of mitosis before separation, resulting in octets and, that recovery of complete tetrads was rare. I assayed from one to eight progeny colonies from each meiosis.

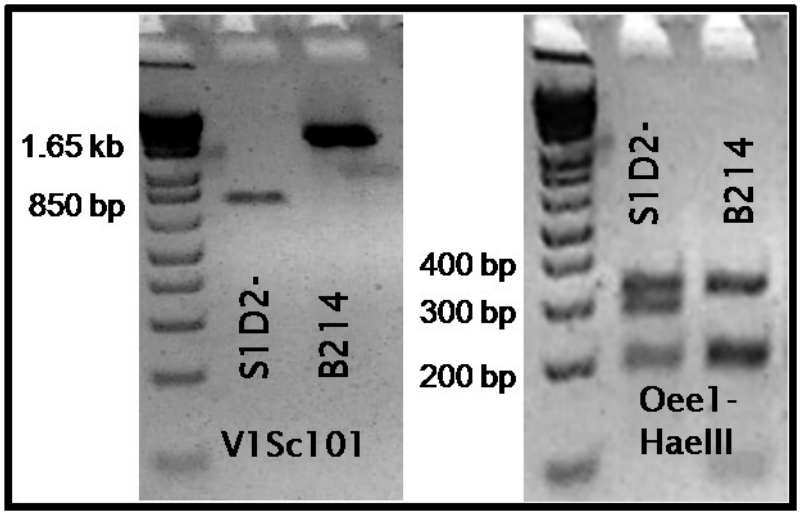
Table 6: Crosses of *pf16, adf1* (20C, mt+) with S1D2.
Experimental groups represent progeny from five separate crosses, performed on five distinct days, and therefore, may have had variable conditions.

Cross	Meiotic Cells	Progeny Assayed	PCR Group	Recombinant Strains
pf16,adf1-2 (20C+) x S1D2-	277	352	Group 1	6
pf16,adf1-2 (20C+) x S1D2-	158	225	Group 2	10
pf16,adf1-2 (20C+) x S1D2-	93	205	Group 3	13
pf16,adf1-2 (20C+) x S1D2-	58	144	Group 7	4
pf16,adf1-2 (20C+) x S1D2-	101	292	Group 7II	27

Once the recombinant strains had been identified (by virtue of their single mutation), the cells grown and the DNA isolated, it was time to determine where recombinations occurred, using PCR. Initially, established primer sets developed for linking the molecular map to the genetic map of the *Chlamydomonas* genome (Kathir et al., 2003) were utilized in order to map recombination events in *adf1* crosses. However, these primer sets provided limited coverage and eventually it was necessary to design new primers specific to the narrowing region. Primer sets were designed using the online software PrimerQuest (IDT), from DNA sequence published by the Joint Genome Institute (JGI). Primers sets generally amplified intronic sequence, as the S1D2 strain varies from the usual laboratory strains in these unexpressed regions (Kathir et al., 2003). Test PCRs were run to optimize cycle temperatures and extension times, as well as to establish whether the fragments generated represented polymorphisms, either by the appearance of an aberrant fragment size for the S1D2 strain, or a difference in band patterns

after restriction digest. Figure 7 shows examples of aberrant fragment size, as seen with bands generated by primers for marker V1sc101 and, the difference in band patterns, as results with a HaeIII digest of PCR product from primers for marker Oee1. Details of the primers sets developed and used in these experiments, can be found in APPENDIX I.

Figure 7: Examples of PCR products from two of the primer sets used to map *ADF1.*
Primer set for marker V1Sc101 shows a size discrepancy between controls S1D2 and B214, while the primer set for marker Oee1 reveals a difference in banding pattern following restriction digest of PCR products with HaeIII.



Following primer optimization, PCR reactions were performed using genomic DNA of progeny recovered from crosses of mutant and polymorphic strains. Each set of PCRs included genomic controls from several strains: S1D2, the polymorphic strain; 137c or B214, wild type strains; and, the parental mutant strain, either a double or single mutant. PCR fragments were run on agarose gels, and banding patterns were imaged. Results for progeny strains were recorded in Excel spreadsheets as “RFLP”, indicating that the colony PCR

duplicated the S1D2 (RFLP strain) control PCR, or “MTNT”, indicating that the colony PCR duplicated the parental control (double or single mutant) PCR, which also is derived from and yields the same PCR bands as B214 and 137c. Additionally, categories were colour coded, “RFLP” in blue and “MTNT” in yellow, to ease visualization of the recombinations in the spreadsheet.

As previously mentioned, progeny had been subjected to deflagellation and swimming assays. Within the experimental groups, these original phenotypes were used as subgroups: *adf1*, single mutant; *pf16*, single mutant; *pf16, adf1*, double mutant; and, S1D2 (i.e. cells with neither mutation). There was a general expectation that strains phenocopying S1D2 in assays , would display as “RFLP” following PCR and, strains phenocopying the parental double mutant, *pf16, adf1*, would display as “MTNT”. This is the general trend, reported from progeny that were included in PCR, despite not having a single mutation phenotype (for first example see Figure 9). The most efficient explanation of the results is that yellow sections contain *ADF1,* for any colony that is in an *adf1,* single mutant or, *adf1, pf16,* double mutant subgroup. Vice versa, *ADF1* may be found in blue sections, for colonies sub-grouped as, *pf16,* single mutant, or S1D2.

It is important to note that the generation of experimental groups and primer sets occurred over a period of several years. Recording PCR data was increasingly difficult as the number of progeny and primers increased. In the text and figures that follow, PCR data is not presented for some recombinant strains in the early tables. This reflects technical difficulties that were subsequently

overcome. In order to provide the best flow in describing this work, these data appear in a later spreadsheet (Figure 20) that summarizes the strains with the most important recombinations.

Mapping data for groups 1 and 2 (subsequently combined) and 3 are summarized in Figure 8 and Figure 9, respectively. Figure 8 reveals 5 recombinant progeny strains: 23, 94, 130, 131 and 160. Progeny 23, 94, 130 and 160 recombined near marker V1Sc101, while progeny 131 had a much more telling recombination near 171503b. As all of these strains are *adf1*, single mutants, these results suggest that *ADF1* is to the right of 171503b. There are six recombinations shown for the progeny of group 3, as seen in Figure 9. Strains 33B and 96A have recombinations to the presumptive left of *ADF1*, near marker V1Sc101, while four of the recombinations took place to the presumptive right of *ADF1* and will be discussed below.

In the middle of Figure 9, in the subgroups *pf16*, single mutant and S1D2, are some lonely “patches” of MTNT yellow. Generally, these sorts of patches (they also occur as sad patches of blue) occur in early records of PCR results. Although theoretically they could have indicated double recombination events, all were subsequently determined to be a consequence of unreliable primer sets that were eventually discarded. They have been left in just so that the results, as they were originally recorded, are not misrepresented. In most cases, the neighbouring results represented a common continuous theme and therefore I felt safe in disregarding these “patches”. Additionally, blank cells are indicative of PCR reactions that were never set up, while blank cells marked with “N/A”, or “?”

are representative of PCR reactions that never came to fruition. Finally, some cells contain the word “both” which indicates that both S1D2 and the parental double mutant were represented in PCR results for those strains. This could indicate contamination of genomic DNA or incomplete separation of tetrads – which eventually grow up to be genomic preps. This overview applies to all of the figures which report PCR results.

Following the completion of PCRs on groups 1, 2 and 3, a pattern began to emerge that allowed me to reduce the number of PCRs I ran on each sample. For instance, for *adf1* and *pf16* single mutants, I ran PCR using only the markers that had previously revealed recombinations to the right of *ADF1*, usually: V1Sc101, Oee1, and 171503b. Similarly, I ran PCRs on progeny which displayed parental phenotypes, double mutant or S1D2, only for markers which are associated with the walk boundary on the left side of adf1: 169190, 167281 and Sc1530030.

Mapping data for groups 7 and 7II are in Figures 10 and 11, respectively. Figure 10 shows us that group 7 was unyielding, with only one recombinant strain, 6C, which only indicates that *ADF1* is to the right of V1Sc101, which is not novel. Fortunately, group 7II was more productive.

Figure 8: PCR results for group 1/2.
Blue indicates that the colony PCR matched control S1D2, while yellow indicates that the colony PCR matched the double mutant, parental control, as well as the wild type control, B214. White boxes are explained in the text.

	PRIMER SET																	
COLONY	V1Sc101	Oee1 - HaeIII	171503b-StuI	C940010	C940041	Sc94_ssr_272kb	C6830002-Msp I	C239008 - MspI	Sc2440003a/b	Sc2440006a/b	Sc1530030 - MspI	Sc23_1410	Sc23_1510	Centromere	Sc52_536_ssr	Sc52_368_caps_Pst	Sc52_243950	MBO2
adf1 (single mutants)																		
23	RFLP	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT
44B	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT
48	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT
62B	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT
73A	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT
94	RFLP	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT
119A	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT
130	RFLP	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT
131	RFLP	RFLP	RFLP	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT
150	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT
160	RFLP	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT
219A	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT
pf16 (single mutants)																		
5B	RFLP	RFLP		RFLP	RFLP	RFLP	RFLP	RFLP	RFLP	RFLP	RFLP	RFLP	RFLP		RFLP	RFLP	RFLP	RFLP
39B	RFLP	RFLP		RFLP	RFLP	RFLP	RFLP	RFLP	RFLP	RFLP	RFLP	RFLP	RFLP		RFLP	RFLP	RFLP	RFLP
135	RFLP	RFLP		RFLP	RFLP	RFLP	RFLP	RFLP	RFLP	RFLP	RFLP	RFLP	RFLP		RFLP	RFLP	RFLP	RFLP

Figure 9: PCR results for group 3.
Blue indicates that the colony PCR matched control S1D2, while yellow indicates that the colony PCR matched the double mutant, parental control, as well as the wild type control, B214. For this group extra colonies with parental phenotypes were used as controls. White boxes are explained in the text.

	PRIMER SET																		
COLONY	V1 Sc101	oee1 - HaeIII	171503b-StuI	C940010	C940041	sc94_ssr_272kb	C6830002-Msp I	C2390006 - MspI	C239008 - MspI	sc2440003a/b	sc2440006a/b	Sc1530030 - MspI	Sc23_1410	Sc23_1510	Centromere	Sc52_536_ssr	Sc52_368_caps_Pst	Sc52_243950	MBO2
adf1 (single mutants)																			
4	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT
33B	RFLP	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT
45B	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT
46C	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT
67B	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT
75A	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT
86A	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT
96A	RFLP	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT
100B	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT
of16 (single mutants)																			
29B	RFLP	RFLP		RFLP	RFLP	RFLP	MTNT	RFLP	RFLP	RFLP	RFLP	RFLP	RFLP	RFLP		RFLP	RFLP	RFLP	RFLP
59A	RFLP	RFLP		RFLP	RFLP	RFLP	RFLP		RFLP	RFLP	RFLP	RFLP	RFLP	RFLP		RFLP	RFLP	RFLP	RFLP
70	RFLP	RFLP		RFLP	RFLP	RFLP	RFLP		RFLP	RFLP	RFLP	RFLP	MTNT	RFLP		RFLP	RFLP	RFLP	RFLP
S1D2-																			
2A	RFLP	RFLP		RFLP	RFLP	RFLP	MTNT	RFLP	RFLP	RFLP	RFLP	RFLP	RFLP	RFLP		RFLP	RFLP	RFLP	RFLP
12	RFLP	RFLP		RFLP	RFLP	RFLP	MTNT	RFLP	RFLP	RFLP	RFLP	RFLP	RFLP	RFLP		RFLP	RFLP	RFLP	RFLP
31A	RFLP	RFLP		RFLP	RFLP	RFLP	MTNT	RFLP	RFLP	RFLP	RFLP	RFLP	RFLP	RFLP		RFLP	RFLP	RFLP	RFLP
43B	RFLP	RFLP		RFLP	RFLP	RFLP	MTNT	RFLP	RFLP	RFLP	RFLP	RFLP	RFLP	RFLP		RFLP	RFLP	RFLP	RFLP
82C	RFLP	RFLP		RFLP	RFLP	RFLP	RFLP		RFLP	RFLP	RFLP	RFLP	RFLP	RFLP		RFLP	RFLP	RFLP	RFLP
of16, adf1 (double mutants)																			
5C	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT
13	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT
15C	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT
17C	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT
21A	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT
25B	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	RFLP
57B	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT		MTNT	RFLP	RFLP	RFLP	RFLP	RFLP		RFLP	RFLP	RFLP	RFLP
66B	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	RFLP	RFLP
78A	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT
82D	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	N/A		MTNT	MTNT	RFLP	RFLP
86B	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT
95	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT

Figure 10: PCR results for group 7.
Blue indicates that the colony PCR matched control S1D2, while yellow indicates that the colony PCR matched the double mutant, parental control, as well as the wild type control, B214.

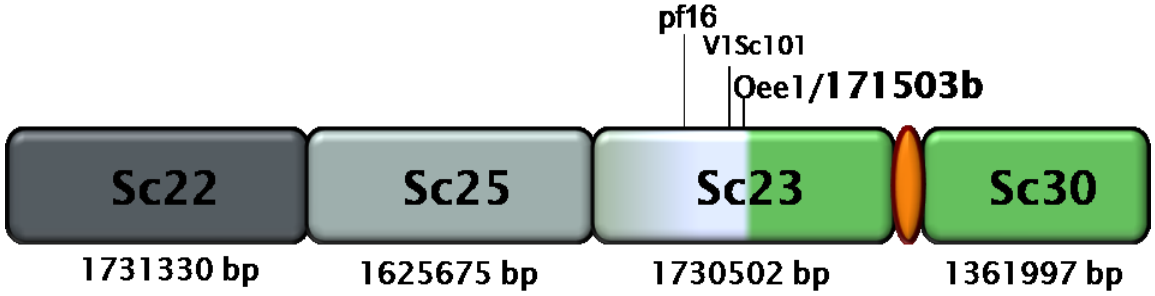
	PRIMER SET				
COLONY	V1Sc101	oee1-HAeIII	167281	1530030-MspI	Centromere
adf1 (single mutants)					
24	MTNT	MTNT		MTNT	
33	MTNT	MTNT		MTNT	
58A	MTNT	MTNT		MTNT	
pf16 (single mutants)					
6C	MTNT	RFLP	RFLP		
adf1,pf16 (double mutants)					
2A			MTNT		
3C			MTNT		
5C				MTNT	
6A				MTNT	
8A				MTNT	
9A			MTNT		
10A				MTNT	
12C				MTNT	
14D				MTNT	
16C			MTNT		
18A				MTNT	
19B			MTNT		
20A				MTNT	
25				MTNT	
35A				MTNT	
42			MTNT		
43			MTNT		
46				MTNT	
47A				MTNT	
50B				MTNT	
51				MTNT	
57A				MTNT	

Figure 11: PCR results for group 7II.
Blue indicates that the colony PCR matched control S1D2, while yellow indicates that the colony PCR matched the double mutant, parental control, as well as the wild type control, B214.

	PRIMER SET						PRIMER SET			PRIMER SET						PRIMER SET					
COLONY	V1Sc101	oee1-HaeII	171503b-StuI	171505-MspI	C1530030-MspI	Centromere	COLONY	C1530030-MspI	Centromere	COLONY	V1Sc101	167281	C1530030-MspI	Centromere	Sc52_243950	COLONY	V1Sc101	167281	C1530030-MspI	Centromere	Sc52_243950
adf1 (single mutants)							*pf16, adf1*			*pf16, adf1* (double mutants)						S1D2-					
2B	MTNT	MTNT					6	MTNT		66A			MTNT			14D			RFLP		
3A	MTNT	MTNT					7C	MTNT		67A			MTNT			15B			RFLP		
8C	MTNT	MTNT					11B	MTNT		69C			MTNT			17B			RFLP		
9D	RFLP	RFLP	both	MTNT			16B	MTNT		70C			MTNT			22D			RFLP		
12C	MTNT	MTNT					19B	MTNT		73A			RFLP			24A			RFLP		
18C	RFLP	RFLP	RFLP	both			21B	MTNT		74C			MTNT			26B			RFLP		
20A	MTNT	MTNT					23B	MTNT		76B			MTNT			27B			RFLP		
31A	MTNT	MTNT					25	MTNT		77A			MTNT			28C			RFLP		
42	MTNT	MTNT					29C	MTNT		78A			MTNT			34C			RFLP		
43B	MTNT	MTNT					32	MTNT		80A			MTNT			38			RFLP		
46A	RFLP	MTNT					35B	MTNT		81A			MTNT			45			RFLP		
58B	MTNT	MTNT					36	MTNT		82A			MTNT			53C			RFLP		
75A	RFLP	MTNT					37	MTNT		85D			MTNT			59			RFLP		
83B	RFLP	RFLP	both	MTNT			39C	RFLP		86C			MTNT			60			RFLP		
pf16 (single mutants)							40C	MTNT		87B			MTNT			63A			RFLP		
10A	RFLP	RFLP					44C	MTNT		88C			MTNT			68B			RFLP		
30B	RFLP	RFLP					47B	MTNT		90A			MTNT			89D	RFLP	RFLP	RFLP		RFLP
33	RFLP	RFLP					49B	MTNT		91C			MTNT			93B	RFLP	RFLP	RFLP		RFLP
41B	RFLP	RFLP					50B	MTNT		92A			MTNT			94B	RFLP	RFLP	RFLP		RFLP
48C	RFLP	RFLP					51C	MTNT		95A			MTNT			101	RFLP	RFLP	RFLP		RFLP
54B	RFLP	RFLP					52C	MTNT		96A			MTNT								
71B	RFLP	RFLP					55C	MTNT		97A			MTNT								
72C	RFLP	RFLP					56	MTNT		99C			MTNT								
79A	RFLP	RFLP					57E	MTNT		100D			MTNT								
84	RFLP	RFLP					61A	MTNT		S1D2-											
98B	RFLP	RFLP					62A	MTNT		1C	RFLP	RFLP	RFLP								
pf16, adf1 (double mutants)							64C	MTNT		4C			RFLP		RFLP						
5B					MTNT		65C	MTNT		13C			RFLP								

Group 7II produced seven recombinant strains (Figure 11). Five of these were of *adf1* single mutant phenotype and collectively, indicated that *ADF1* was to the right of marker 171503b. The strain 18C has been categorized as "both" for marker 171505, which is notable only because its brother strains, 9D and 83B, have similar categorizations for marker 171503b. What's interesting is that on either side of the "both" there is a different parental phenotype represented. As I mentioned earlier, this could simply be genomic contamination. Future PCRs, amplifying away from the discrepancy, could be performed to determine what this means. In the meantime, Figure 12 represents the collective findings for most of the data represented up to this point. After starting with the region to the right of marker V1Sc101 (a region of 2371 kb), the border of the region has moved to marker 171503b, decreasing the suspect region to about 2275 kb.

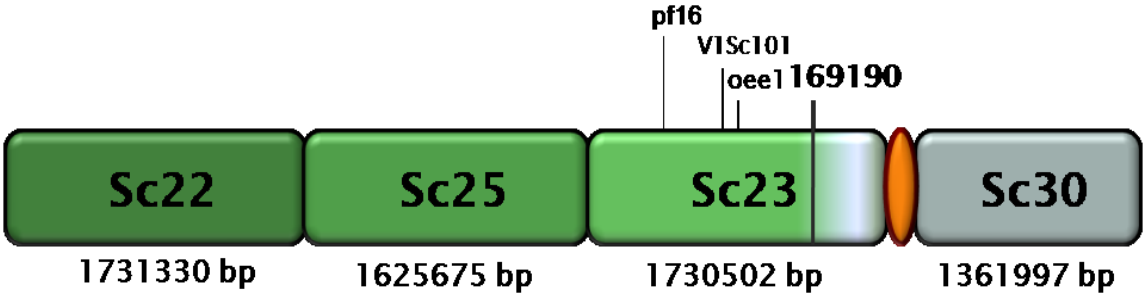
Figure 12: Representation of linkage group IX with recombination at marker 171503b. Grey indicates sequence that has been eliminated. Green represents region of interest, specifically, the region to the right of marker 171503b-StuI.



As depicted in Figure 9, a subset of colonies from group 3, with parental phenotypes, was subjected to DNA isolation, as extra controls. This action yielded a surprising result. Strain 57B (*pf16, adf1*, mt+) had a very valuable cross over event, which effectively eliminated one arm of the chromosome. This

recombination is diagrammed in Figure 13, at marker 169190. From this point onward, at least one colony from each dissected tetrad was grown up for DNA isolation and PCR, in order to map recombinations on either side of *ADF1*.

Figure 13: Representation of linkage group IX with recombinations at marker 169190. Grey indicates sequence eliminated by the recombination event that produced 57B. Green represents region of interest, for strain 57B (pf16, adf, mt+), of group 3.



After the recovery of strain 57B (*pf16, adf1*, mt+) I decided to cross the *adf1* alleles to S1D2 without first generating the *pf16* double mutant. The rationale behind this action was that the double mutants I had already generated were not yielding large progeny numbers when crossed to the polymorphic strain, and I thought the single mutants might mate better, as they are motile. Additionally, as it was now apparent that the region could be narrowed from either side of *ADF1*, the *pf16* mutation wasn't needed to eliminate the region to the right of *ADF1*. Table 7 summarizes the crosses; the number of meiotic cells; the tetrad progeny assayed for the ability to deflagellate in acid; the experimental group (strains that were subjected to PCR, following assays); and the number of recombinant strains recovered in those experimental groups.

Table 7: Cross of adf1-2 with S1D2, which generated experimental group 4, and cross of adf1-5 with S1D2, which generated experimental group 5.

Cross	Meiotic Cells	Progeny Assayed	Experimental Group	Recombinant Strains
adf1-2+ x S1D2-	75	118	Group 4	1
adf1-5+ x S1D2-	7	14	Group 5	0

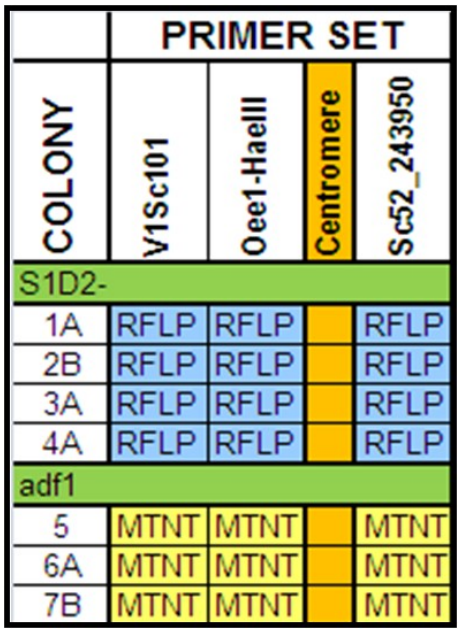
These crosses were not highly successful, as only one recombinant strain that would later be included in calculations for recombination frequency was generated. This could have been due to the use of older mutant strains in the crosses; they may have needed to be backcrossed prior to mating with the inter-fertile strain. In addition, crosses with the inter-fertile strain have been recognized as yielding low numbers of progeny. Five recombinants were recovered from group 4 (see Figure 14), but as I had already established the right boundary at marker 169190, four of them were not considered, although they could have been used to calculate recombination frequencies to the right of *ADF1*. This is certainly an activity that could be considered for the future, but would involve many more PCR reactions for markers on the right side of the centromere, for all groups, not just group 4. Additionally, further mapping on the right side of the centromere shouldn't take precedence over more important efforts, such as subcloning of candidate genes (see below). The one strain that was used in calculations, 39, did not contribute to narrowing the region. A cross of *adf1-5*, an allele that has not been thus far utilized for mapping, to S1D2, did not yield many

zygotes; only seven became meiotic and there were no recombinants, as illustrated by Figure 15.

Figure 14: PCR results for group 4.
Blue indicates that the colony PCR matched control S1D2, while yellow indicates that the colony PCR matched the single mutant, parental control, *adf1-2*, as well as the wild type control, B214.

	PRIMER SET													
COLONY	V1Sc101	oee1-HaeIII	C940010	Sc94_ssr_272kb	C6830002-MspI	Sc_2390006_Msp	Sc_239008-Msp	Sc2440003a/b	C1530030-MspI	Sc23_1410 - Msp I	Sc23_1510	Centromere	Sc52_536	Sc52_243950
adf1														
1	MTNT	?	MTNT	MTNT?	MTNT	MTNT	MTNT	MTNT	?	?	?		MTNT	?
2	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT
3	MTNT	MTNT	MTNT	MTNT?	MTNT	MTNT	?	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT
4B	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT
5A	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT
6	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT
12	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	RFLP
18A	MTNT	?	MTNT	MTNT?	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT
28	MTNT	MTNT	MTNT	?	MTNT	MTNT	MTNT	MTNT	MTNT	?	MTNT		?	MTNT
29	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	RFLP	?	MTNT	MTNT		MTNT	MTNT
30	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	?	MTNT	?	MTNT	MTNT		MTNT	MTNT
33	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	?	MTNT	MTNT		MTNT	MTNT
36	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	?	MTNT	MTNT		MTNT	MTNT
44A	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT
47B	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	?	?		MTNT	MTNT
48	?	?	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	?		RFLP	RFLP
49	?	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT
55A	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	RFLP
57	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	?	MTNT	MTNT		MTNT	MTNT
59	MTNT	?	MTNT	MTNT	MTNT	MTNT	MTNT	RFLP	MTNT	MTNT	MTNT		MTNT	MTNT
62A	MTNT	?	both	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	?	?		MTNT	MTNT
63A	MTNT	?	MTNT	MTNT	MTNT	MTNT	MTNT	RFLP	?	?	?		MTNT	MTNT
67B	?	MTNT	both	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT
68B	MTNT	?	MTNT	MTNT	MTNT	MTNT	?	RFLP	?	?	?		MTNT	MTNT
71A	?	?	both	MTNT	MTNT	MTNT	?	MTNT	MTNT	?	?		RFLP	RFLP
75	?	?	both	MTNT	MTNT	MTNT	?	RFLP	MTNT	?	?		MTNT	MTNT
76	MTNT	?	MTNT	MTNT	MTNT	MTNT	?	MTNT	MTNT	?	?		MTNT	MTNT
77B	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT
S1D2-														
23	RFLP	?	RFLP	RFLP	RFLP	RFLP	RFLP	MTNT		?	?			RFLP
39	MTNT	RFLP	RFLP	RFLP	RFLP	RFLP	RFLP	RFLP	RFLP	RFLP	RFLP			RFLP

Figure 15: PCR results for group 5.
Blue indicates that the colony PCR matched control S1D2, while yellow indicates that the colony PCR matched the single mutant, parental control, *adf1-5*, as well as the wild type control, B214.



COLONY	PRIMER SET: V1Sc101	Oee1-HaeIII	Centromere	Sc52_243950
S1D2-				
1A	RFLP	RFLP		RFLP
2B	RFLP	RFLP		RFLP
3A	RFLP	RFLP		RFLP
4A	RFLP	RFLP		RFLP
adf1				
5	MTNT	MTNT		MTNT
6A	MTNT	MTNT		MTNT
7B	MTNT	MTNT		MTNT

Strain 57B (*pf16, adf1*; mt+) was backcrossed with the polymorphic strain, S1D2 (mt-), with the goal of recovering a single mutant. Table 8 shows the details of the cross, including: the number of meiotic cells, progeny assayed and the resulting number of recombinant strains.

Table 8: Crosses of *pf16, adf1* (57B, mt+) with S1D2, which generated experimental group 6.

Cross	Meiotic Cells	Progeny Assayed	Experimental Group	Recombinant Strains
pf16, adf1-2 (57B+) x S1D2-	133	442	Group 6	2

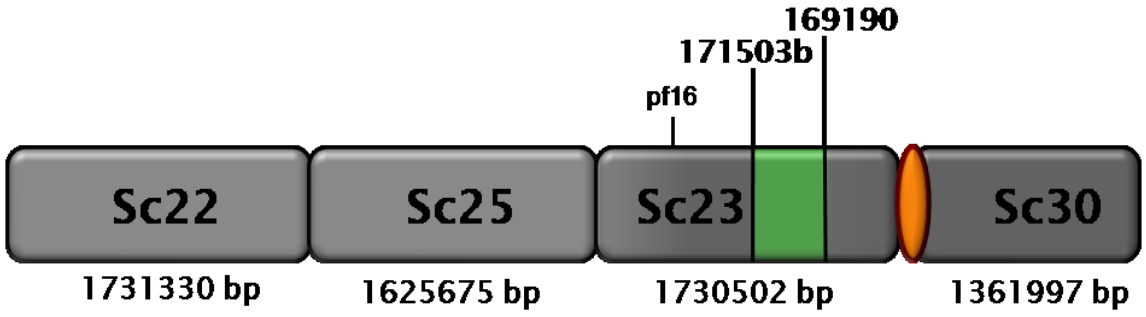
Figure 16: PCR results for group 6.
Blue indicates that the colony PCR matched control S1D2, while yellow indicates that the colony PCR matched the single mutant, parental control, *pf16, adf1* (57B) as well as the wild type control, B214.

PRIMER SET								PRIMER SET						PRIMER SET					PRIMER SET				
COLONY	v1Sc101	oee1-HaeIII	C_6830002 -Msp1	C_239008 - MspI	Sc23_1410 - Msp	Centromere	sc52_243950	COLONY	v1Sc101	oee1-HaeIII	C_239008 - MspI	Sc23_1410 - Msp	Centromere	COLONY	v1Sc101	oee1-HaeIII	C_239008 - MspI	Centromere	COLONY	v1Sc101	oee1-HaeIII	C_239008 - MspI	Centromere
adf1								*adf1*						*adf1*					*adf1*				
1c	MTNT	MTNT	MTNT	MTNT			RFLP	36a	MTNT	MTNT	MTNT	RFLP		77b	MTNT	MTNT	MTNT		117b	MTNT	MTNT	MTNT	
3d	MTNT	MTNT	MTNT	MTNT			RFLP	37b	MTNT	MTNT	MTNT	?		78b	MTNT	MTNT	MTNT		118b	MTNT	MTNT	MTNT	
4b	MTNT	MTNT	MTNT	MTNT			RFLP	38c	MTNT	MTNT	MTNT	?		79a	MTNT	MTNT	MTNT		119a	MTNT	MTNT	MTNT	
5a	MTNT	MTNT	MTNT	MTNT			RFLP	39a	MTNT	MTNT	MTNT	RFLP		81c	MTNT	MTNT	MTNT		121b	MTNT	MTNT	MTNT	
6e	MTNT	MTNT	MTNT	MTNT			RFLP	40c	MTNT	MTNT	MTNT	?		84a	MTNT	MTNT	MTNT		122b	MTNT	MTNT	MTNT	
7a	MTNT	MTNT	MTNT	MTNT			RFLP	43a	MTNT	MTNT	MTNT	RFLP		86a	MTNT	MTNT	MTNT		123a	MTNT	MTNT	MTNT	
9b	MTNT	MTNT	MTNT	MTNT			RFLP	44a	MTNT	MTNT	MTNT	RFLP		89	MTNT	MTNT	MTNT		124b	MTNT	MTNT	MTNT	
12b	MTNT	MTNT	MTNT	MTNT			RFLP	45b	MTNT	MTNT	MTNT	RFLP		92	MTNT	?	MTNT		126a	MTNT	MTNT	MTNT	
14a	MTNT	MTNT	MTNT	MTNT			RFLP	46d	MTNT	MTNT	MTNT			93a	MTNT	MTNT	MTNT		126b	RFLP	RFLP	MTNT	
15b	MTNT	MTNT	MTNT	MTNT			RFLP	47b	MTNT	MTNT	MTNT	RFLP		94a	MTNT	MTNT	MTNT		127c	MTNT	MTNT	MTNT	
17a	MTNT	MTNT	MTNT	MTNT			RFLP	51a	MTNT	MTNT	MTNT	RFLP		95d	MTNT	MTNT	MTNT		130b	MTNT	MTNT	MTNT	
18a	RFLP	RFLP	MTNT	MTNT			RFLP	52	MTNT	MTNT	MTNT			96a	MTNT	MTNT	MTNT		131b	MTNT	MTNT	MTNT	
19b	MTNT	MTNT	MTNT	MTNT			RFLP	55a	MTNT	MTNT	MTNT			97b	MTNT	MTNT	MTNT						
20a	MTNT	MTNT		MTNT				56b	MTNT	MTNT	MTNT			98b	MTNT	MTNT	MTNT						
21b	MTNT	MTNT	MTNT	MTNT			RFLP	57c	MTNT	MTNT	MTNT			100c	MTNT	MTNT	MTNT						
22a	MTNT	?		MTNT				58b	MTNT	MTNT	MTNT			101b	MTNT	MTNT	MTNT						
23a	MTNT	MTNT		MTNT	RFLP			59d	MTNT	MTNT	MTNT			102c	MTNT	MTNT	MTNT						
24a	MTNT	MTNT		MTNT	RFLP			61a	MTNT	MTNT	MTNT			103b	MTNT	MTNT	MTNT						
25a	MTNT	MTNT		MTNT	RFLP			63c	MTNT	MTNT	MTNT			104a	MTNT	MTNT	MTNT						
26a	MTNT	MTNT		MTNT	RFLP			67	MTNT	MTNT	MTNT			105b	MTNT	MTNT	MTNT						
27a	MTNT	MTNT		MTNT	RFLP			68c	MTNT	MTNT	MTNT			106b	MTNT	MTNT	MTNT						
29b	MTNT	MTNT		MTNT	RFLP			69d	MTNT	MTNT	MTNT			108a	MTNT	MTNT	MTNT						
30a	MTNT	MTNT		MTNT	RFLP			70c	MTNT	MTNT	MTNT			110c	MTNT	MTNT	MTNT						
31a	MTNT	MTNT		MTNT	RFLP			71e	MTNT	MTNT	MTNT			111a	MTNT	MTNT	MTNT						
32a	MTNT	MTNT		MTNT	RFLP			74b	MTNT	MTNT	MTNT			113	MTNT	MTNT	MTNT						
34b	MTNT	MTNT		MTNT	RFLP			75a	MTNT	MTNT	MTNT			114	MTNT		MTNT						
35a	MTNT	MTNT		MTNT	RFLP			76b	MTNT	MTNT	MTNT			116b	MTNT	MTNT	MTNT						

Keeping in mind that the parental, 57B, strain was comprised, to the left of 169190 of “MTNT” DNA and, to the right of 169190, of “RFLP” DNA, I was really only looking for a recombination that would eliminate the left arm of the chromosome and sandwich *ADF1* between the polymorphic “RFLP” DNA, and which would hopefully recombine past marker 171503b. Experimental group 6 yielded two single mutants. Both recombined near marker Oee1 (as seen in Figure 16), but subsequent PCR (see Figure 20) has shown that the recombination event was near 171503b, therefore (unfortunately) not eliminating any more of the region. The low recombination at this side of the interval might indicate tighter linkage with *ADF1*.

Figure 17 depicts the 394 kb region defined as containing the *ADF1* locus, between markers 171503b and 169190. This figure is the culmination of all the PCR data reported above.

Figure 17: Representation of linkage group IX with recombination at markers 171503b and 169190.
Grey indicates sequence that has been eliminated. Green represents region of interest, as defined by group 6 strains 18A and 126B.



Double mutants were also generated using *ADF1* alleles, *adf1-3* and *adf1-6,* crossed to *pf16.* Double mutants were subsequently crossed to S1D2 (mt-), for mapping purposes as described above. These crosses and progeny are summarized in Table 9.

Table 9: Crosses of double mutants, *pf16, adf1-6* (5, mt+) or *pf16, adf1-3* (13A, mt+) with S1D2, which generated experimental groups 8 and 9.

Cross	Meiotic Cells	Progeny Assayed	Experimental Group	Recombinant Strains
pf16, adf1-6 (5+) x S1D2-	55	131	Group 8	19
pf16, adf1-3 (13A+) x S1D2-	48	75	Group 9	5

Figures, 18 and 19, report the PCR results for Groups 8 and 9, respectively. Group 8 produced six recombinant strains, but all of the recombinations occurred at or near previously determined boundaries, Oee1 and 169190. Results recovered for group 9 also did not contribute to narrowing the region. Three progeny strains had recombination events near marker V1Sc101 and the fourth had a recombination near 169190.

Figure 18: PCR results for group 8.
Blue indicates that the colony PCR matched control S1D2, while yellow indicates that the colony PCR matched the double mutant, parental control, *pf16, adf1-6* (5, mt+) as well as the wild type control, B214.

	PRIMER SET				PRIMER SET				PRIMER SET				PRIMER SET		
COLONY	oee1-HaeIII	169190 - SacI	Centromere	COLONY	oee1-HaeIII	169190 - SacI	Centromere	COLONY	oee1-HaeIII	169190 - SacI	Centromere	COLONY	oee1-HaeIII	169190 - SacI	Centromere
adf1 (single mutants)				*adf1,pf16* (double mutants)				*S1D2*				*S1D2*			
6	MTNT	MTNT		29B	MTNT	MTNT		22	RFLP	RFLP		51	RFLP	RFLP	
15	RFLP	MTNT		30A	MTNT	MTNT		23	RFLP	RFLP		53A	RFLP	RFLP	
16	MTNT	MTNT		31B	MTNT	MTNT		24C	RFLP	RFLP					
45D	MTNT	MTNT		32D	MTNT	MTNT		25B	RFLP	RFLP					
52A	?	MTNT		34A	MTNT	MTNT		26A	RFLP	RFLP					
55	MTNT	MTNT		35A	MTNT	MTNT		28A	RFLP	RFLP					
pf16 (single mutants)				37A	MTNT	MTNT		29A	RFLP	RFLP					
33A	RFLP	RFLP		38A	MTNT	MTNT		30C	RFLP	MTNT					
41D	RFLP	RFLP		39A	MTNT	MTNT		31A	RFLP	RFLP					
45B	RFLP	RFLP		40A	MTNT	RFLP		32A	RFLP	RFLP					
adf1,pf16 (double mutants)				42B	MTNT	MTNT		32C	RFLP	MTNT					
2A	MTNT	MTNT		44A	MTNT	MTNT		33B	RFLP	RFLP					
4	MTNT	MTNT		45C	MTNT	MTNT		34B	RFLP	RFLP					
5	MTNT	MTNT		46A	MTNT	MTNT		35D	RFLP	RFLP					
7	MTNT	RFLP		47A	MTNT	MTNT		36	RFLP	RFLP					
9	MTNT	MTNT		49A	MTNT	MTNT		37B	RFLP	RFLP					
10B	MTNT	MTNT		54	MTNT	RFLP		38C	RFLP	RFLP					
11	MTNT	MTNT		*S1D2*				39B	RFLP	RFLP					
13	MTNT	MTNT		1	RFLP	RFLP		40B	RFLP	RFLP					
14	MTNT	MTNT		2B	RFLP	RFLP		41B	RFLP	RFLP					
17	MTNT	MTNT		3	RFLP	RFLP		43A	RFLP	RFLP					
18B	MTNT	MTNT		8A	RFLP	RFLP		44B	RFLP	RFLP					
19A	MTNT	MTNT		10A	RFLP	RFLP		45A	RFLP	RFLP					
24A	MTNT	MTNT		12A	RFLP	RFLP		46C	RFLP	RFLP					
25A	MTNT	MTNT		18A	RFLP	RFLP		48	RFLP	RFLP					
26B	MTNT	MTNT		20A	RFLP	RFLP		49B	RFLP	RFLP					
27A	MTNT	MTNT		21A	RFLP	RFLP		50	RFLP	RFLP					

Figure 19: PCR results for group 9.
Blue indicates that the colony PCR matched control S1D2, while yellow indicates that the colony PCR matched the double mutant, parental control, *pf16, adf1-3* (13A, mt+) as well as the wild type control, B214.

	PRIMER SET							
COLONY	V1Sc101	oee1-HaeIII	171503b	171505-MspI	C239008-MspI	167281	169190-SacI	Centromere
adf1 (single mutants)								
16A	RFLP	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	
30A	RFLP	MTNT	MTNT	MTNT		MTNT	MTNT	
36A	RFLP	MTNT		MTNT		MTNT	MTNT	
pf16 (single mutants)								
18A	RFLP	RFLP	RFLP	RFLP		RFLP	RFLP	
adf1,pf16 (double mutants)								
2B			MTNT	MTNT	MTNT	MTNT	MTNT	
5D						MTNT	MTNT	
9A		MTNT	MTNT		MTNT		MTNT	
9B		MTNT	MTNT		MTNT		MTNT	
10C						MTNT	MTNT	
11A						MTNT	MTNT	
11B						MTNT	MTNT	
14A						MTNT	MTNT	
16B						MTNT	MTNT	
17A						MTNT	MTNT	
19A						MTNT	MTNT	
21A						MTNT	MTNT	
22B						MTNT	MTNT	
22C						MTNT	MTNT	
23A						MTNT	MTNT	
28A						MTNT	MTNT	
28C						MTNT	MTNT	
35A						MTNT	MTNT	
38A						MTNT	MTNT	
39A						MTNT	MTNT	
47B						MTNT	MTNT	

	PRIMER SET				
COLONY	V1Sc101	C239008-MspI	167281	169190-SacI	Centromere
S1D2-					
1A			RFLP	RFLP	
1B			RFLP	RFLP	
2A		RFLP		RFLP	
3A		RFLP		RFLP	
4A		RFLP		RFLP	
5A		RFLP	RFLP	RFLP	
5B		RFLP		RFLP	
5C		RFLP	RFLP	RFLP	
5E			RFLP	RFLP	
6A		RFLP	RFLP	RFLP	
7A			RFLP	RFLP	
8A		RFLP	RFLP		
10A			RFLP	RFLP	
10B			RFLP	RFLP	
12A			RFLP	RFLP	
15A			RFLP	RFLP	
20A			RFLP	RFLP	
21B			RFLP	RFLP	
22A			RFLP	RFLP	
24A	RFLP	RFLP		RFLP	
25A			RFLP	RFLP	
26A		RFLP		RFLP	
27A			RFLP	RFLP	
28B			RFLP	RFLP	
28D				RFLP	
29A			RFLP	RFLP	
29B			RFLP	RFLP	

	PRIMER SET			
COLONY	C239008-MspI	167281	169190-SacI	Centromere
S1D2-				
29C		RFLP	RFLP	
31A		RFLP	RFLP	
33B		RFLP	RFLP	
34A		RFLP	RFLP	
37A	RFLP	RFLP	MTNT	
38B		RFLP	RFLP	
40A		RFLP	RFLP	
42A		RFLP	RFLP	
42B		RFLP	RFLP	
43A			RFLP	
44A			RFLP	
45A		RFLP	RFLP	
46A		RFLP	RFLP	
46B		RFLP	RFLP	
47A		RFLP	RFLP	
48A		RFLP	RFLP	

Figure 20: Summary: PCR results for all strains with boundary defining recombinations.
Blue indicates that the colony PCR matched control S1D2, while yellow indicates that the colony PCR matched the mutant, parental control, as well as the wild type control, B214.

	PRIMER SET																		
COLONY	V1Sc101	oee1-HaeIII	171503b-Stu I	171505-MspI	C940010	C940041	Sc94_ssr_272kb	C6830002-Msp	Sc_239008-Msp	167281	169190	Sc2440003a/b	Sc2440006a/b	C1530030-MspI	Sc23_1410 - Msp I	Sc23_1510	Centromere	Sc52_536	Sc52_243950
adf1 (single mutants)																			
g1/2:23	RFLP	MTNT			MTNT	MTNT	MTNT	MTNT	MTNT				MTNT	MTNT	MTNT	MTNT		MTNT	MTNT
g1/2:94	RFLP	MTNT			MTNT	MTNT	MTNT		MTNT				MTNT	MTNT	MTNT	MTNT		MTNT	MTNT
g1/2:130	RFLP	MTNT			MTNT	MTNT	MTNT	MTNT	MTNT				MTNT	MTNT	MTNT	MTNT		MTNT	MTNT
g1/2:131	RFLP	RFLP	RFLP	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT			MTNT	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT
g1/2:160	RFLP	MTNT			MTNT	MTNT	MTNT	MTNT	MTNT				MTNT	MTNT	MTNT	MTNT		MTNT	MTNT
g3:33B	RFLP	MTNT			MTNT	MTNT	MTNT		MTNT				MTNT	MTNT	MTNT	MTNT		MTNT	MTNT
g3:96A	RFLP	MTNT			MTNT	MTNT	MTNT		MTNT				MTNT	MTNT	MTNT	MTNT		MTNT	MTNT
g6:18a	RFLP	RFLP	RFLP	MTNT	MTNT			MTNT	MTNT										RFLP
g6:126b	RFLP	RFLP	RFLP	MTNT		MTNT			MTNT										
g7II:9D	RFLP	RFLP	both	MTNT															
g7II:18C	RFLP	RFLP	RFLP	both															
g7II:46A	RFLP	MTNT																	
g7II:75A	RFLP	MTNT																	
g7II:83B	RFLP	RFLP	both	MTNT															
g9:16A	RFLP	MTNT	MTNT	MTNT					MTNT	MTNT	MTNT								
g9:30A	RFLP	MTNT	MTNT	MTNT						MTNT	MTNT								
g9:36A	RFLP	MTNT		MTNT						MTNT	MTNT								
pf16 (single mutants)																			
g4:39	MTNT	RFLP			RFLP				RFLP					RFLP					
g6:41c		MTNT	RFLP	RFLP	RFLP	RFLP		RFLP	RFLP						RFLP				
g7:6C	MTNT	RFLP		RFLP	RFLP	RFLP		RFLP		RFLP									
g8:52B	MTNT	RFLP									RFLP								
adf1,pf16 (double mutants)																			
g3:57B	MTNT	MTNT			MTNT	MTNT	MTNT	MTNT	MTNT		RFLP	RFLP	RFLP	RFLP	RFLP	RFLP		RFLP	RFLP
g7II:39C	MTNT				MTNT	MTNT			MTNT	MTNT	RFLP			RFLP					RFLP
g7II:73A	MTNT				MTNT	MTNT			MTNT	MTNT	RFLP			RFLP					RFLP
g8:7		MTNT		MTNT						MTNT	RFLP								
g8:40A		MTNT		MTNT						MTNT	RFLP								
g8:54		MTNT		MTNT						MTNT	RFLP								
S1D2-																			
g8:30C		RFLP		RFLP						RFLP	MTNT								
g8:32C		RFLP		RFLP						RFLP	MTNT								

Figure 20 summarizes the most interesting recombinant strains and PCR results for those strains. Most of the progeny that determined the 394 kb region and the recombination frequencies (discussed below), are represented here.

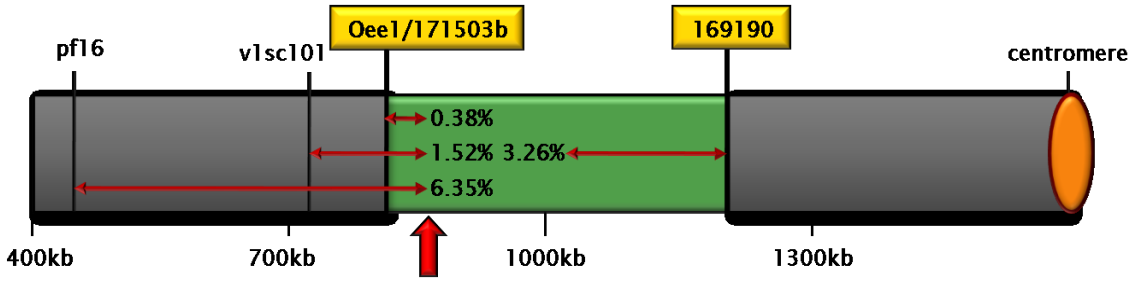
The final combination of data from all of the crosses, assays, and PCRs allowed the relative frequency of recombination to be calculated between markers: pf16, V1Sc101, Oee1, 171503b and *ADF1*; as well as *ADF1* and 169190. The infrequent recovery of full tetrads prevented an accurate calculation of recombination frequency however, an inaccurate, relative recombination frequency was determined, only to be used as a guide. The total number of meiotic cells was tabulated for each experimental group. For some markers only subsets of progeny were assayed with PCR. In these instances, the number of progeny assayed was used in place of the number of meiotic progeny. For instance, progeny showing parental phenotypes were used as controls in experimental group 3. However, PCRs for marker 169190 indicated that a recombination event had occurred in strain 57B. PCR for marker 169190 was performed on only ten colonies from this group, and one yielded a recombination, this was therefore calculated as 1 out of 10. Table 10 shows the ratios of recombinants over meiotic progeny, or progeny assayed. These ratios multiplied by 100% give the relative recombination frequencies. Based on these frequencies and, assuming linear recombination, a center was established for the mapping region, which is diagrammed in Figure 21. The mapping center is at Scaffold_23: 856760 bp, which corresponds to intergenic region between gene predictions 171530 and 171531(see APPENDIX 3). The recombination

frequencies at marker Oee1 and 171503b have been combined as they are in very near proximity to each other, with a difference of only about 1.3 kb.

Table 10: Ratios of recombinant progeny strains to meiotic cells for boundary markers, used to determine the recombination frequencies (in map units) between markers and *ADF1*.

	Recombinations per Marker				
Group	**Past V1sc101**	**V1Sc101**	**Oee1-HaeIII**	**171503b-StuI**	**169190-SacI**
1	2/275	4/275	0/275	0/275	N/A
2	8/158	0/158	0/158	1/158	N/A
3	10/93	2/93	0/93	0/93	1/10
4	N/A	1/30	0/30	0/30	0/30
5	N/A	0/7	0/7	0/7	N/A
6	N/A	0/95	1/95	2/95	N/A
7	3/58	1/58	0/58	0/58	0/58
7II	20/101	2/101	2/101	1/101	2/75
8	6/55	1/55	0/55	0/55	5/55
9	1/48	3/48	0/48	0/48	1/48
Total	50/788	14/920	3/920	4/920	9/276
Map Units	6.35 mu	1.52 mu	0.33 mu	0.43 mu	3.26 mu

Figure 21: Centre of mapping as determined from relative recombination frequencies, on linkage group IX.
Grey indicates sequence that has been eliminated. Green represents region of interest, and includes relative recombination frequencies between several markers and *ADF1*. The center of mapping, marked by the large red arrow, is at 856760 bp.



In parallel to PCRs of potential recombinant progeny, PCR was also performed on all of the *adf1* alleles, on the off-chance that there was a difference in banding pattern possibly indicative of the site of the causal mutation. Figure 22

depicts the PCR results for the alleles with a variety of primer sets. No aberrations were detected in the products of these PCRs, but this approach would be continued and expanded upon for the 2Kb PCR walk (discussed below).

Figure 22: PCR results for adf1 alleles, 1 through 6, using primer sets designed for the purpose of mapping the region boundaries.
MTNT indicates that the allelic PCR product matched the wild type control.

	PRIMER SET																
	V1 Sc101	oee1 - HaeIII	171505-MspI	171530	171530b	C940010	C940038-MspI	C940041	C940017 - MspI	sc94_ssr_272kb	Star 2	C_7680001	C_17420001	C940052	C_6830002-MspI	C_940023-MspI	C_2390006 - MspI
137c+	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT
adf1-1	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT
adf1-2	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT
adf1-3	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT
adf1-4	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT
adf1-5	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT
adf1-6	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT

	PRIMER SET																
	C2390007-MsspI	C239008 - MspI	167281	167281b	Ooooh	161655	160774	sc2440003a/b	sc2440006a/b	Sc1530030 - MspI	Sc23_1410	Sc23_1510	Sc52_536_ssr	Sc52_368_caps_Pst	Sc52_243950	MBO2
137c+	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT
adf1-1	MTNT	MTNT	MTNT	MTNT		MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT
adf1-2	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT
adf1-3	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT
adf1-4	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT
adf1-5	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT
adf1-6	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT	MTNT

III: Review of candidate genes in 394 kb region

As the region slowly narrowed, it became apparent that PCR-based mapping would yield a limited amount of progress, and I started to look to other methods to identify the gene locus. Specifically, I looked at the predicted genes in the 394 kb region, in order to determine the best candidate genes. Published on JGI, are 62 predicted genes for this region (Version 2 predictions have been used). I used the DNA sequence in the region of each of these predicted genes for new predictions using GreenGenie2 (Kwan et al., 2009; http://bifrost.wustl.edu/GreenGenie2/), new software that has improved gene-finding capabilities, specific to *Chlamydomonas*. Following gene prediction, cDNA sequences were translated into proteins using ExPasy (http://ca.expasy.org/) and finally, the sequences were compared using BLAST software from NCBIs (http://blast.ncbi.nlm.nih.gov/Blast.cgi) public data base, to determine similarity with known proteins. Hits, including conserved domains, were recorded (APPENDICES 3 and 4) and can now be used as guides to choose candidate genes for subcloning for future transformation into *adf1* cells and, hopefully, rescue of the *adf1* phenotype. Proteomic databases were also searched for the inclusion of the predicted genes. APPENDIX 5 reports the search results of: the flagellar proteome (Pazour et al., 2005), the centriole proteome (Keller & Marshall, 2008), the flagellar and basal body (FBB) proteome (Li et al., 2004) and the eyespot proteome (Wagner et al., 2008). Most of the 62 predicted genes in the 394kb region were not featured in any of these proteomic collections, and those that were represent genes that can be eliminated, based

on identities such as “ribosomal subunit”, which probably indicates contamination, as de novo protein synthesis does not occur in the flagella. In addition, I have compiled annotations from JGI of conserved domains and flagellar proteome peptides, not encountered in previous BLAST results or proteome inspections (APPENDIX 4). The most interesting predicted genes are listed in Table 11, and will be addressed in the Discussion.

Table 11: Best candidate genes based on BLAST E-values and protein predictions, for *ADF1*.

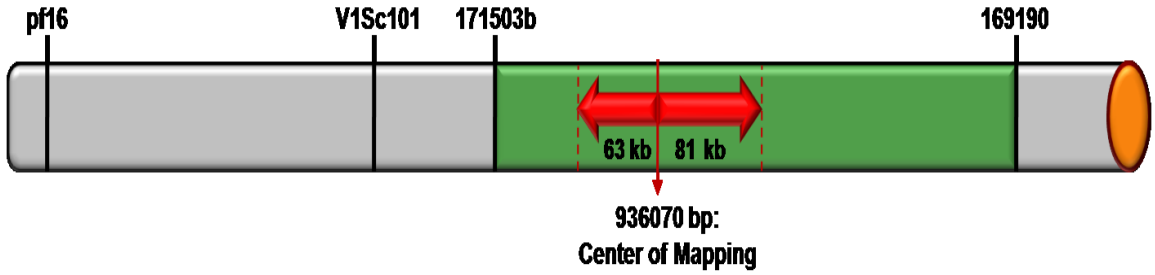
Gene Identity Version 2	Scaffold 23 Location	BLAST E-Value	Description
171534	874240-875141	8.00E-05	ANK Superfamily; ion channel nompc
171510	902809:907622	3.00E-13	receptor type adenylate cyclase
171541	938343:940636	1.00E-24	TBC superfamily; RabGAP/ TBC domain-containing protein; growth hormone regulated TBC protein (1.00E-18)
158447	987960:991490	3.00E-77	Chlamydomonas predicted protein; calmodulin-binding protein trp1 (2.00E-04
171549	1030743:1036247	4.00E-32	ATP binding site; catalytic loop; PI3Kc_like Superfmaily; DNA dependent protein kinase

IV: 2Kb PCR Walk

After the boundaries of the walk had been established as 171503b and 169190, I began a PCR walk with the rationale that a deletion or insertion footprint, left over from the insertional mutagenesis, should be apparent in PCR products. I hypothesized that PCR products, generated from genomic DNA of those *adf1* alleles generated by insertional mutagenesis, would display as aberrantly sized fragments or with discordant banding patterns, following

restriction digest, in comparison to controls, in the event that primers amplified through the mutation site. New primers were designed in PrimerQuest, so that successive amplicons of an average 2 kb were generated, with an average of 50 bp overlaps. I began the 2 kb PCR walk prior to finalizing the region boundaries at 171503b and 169190 and, prior to finishing PCR-based mapping. The mapping centre, following the addition of data from each new experimental group (ie. group 7II, group 8, group 9, etc.) fluctuated over a range of ~80 kb, as the recombination frequencies fluctuated. At the time that I started the 2kb walk, the mapping centre existed at 936070 bp on Scaffold 23 and, primers were designed to amplify approximately 63 kb to the left of this center, and approximately 81 kb to the right, for a total of ~145 kb (see Figure 23; APPENDIX 2 for primer details).

Figure 23: Coverage of 2kb walk primer sets.
Primers were designed, working outward from the mapping centre, then at 936070, to cover 63 kb to the left and 81 kb to the right.



DNA from the six alleles of *adf1* and wild type control B214, were used in PCR reactions. The resulting fragments were restriction enzyme digested, with a variety of enzymes, usually chosen for the ability to cut at the nucleotides guanine and cytosine and, run on agarose gels and imaged. The alleles were compared to wild type strains, and the results were recorded in Excel

spreadsheets. Figure 24 shows the PCR results; green boxes indicate that PCR products from *adf1* genomic DNA phenocopied the wild type controls, and white boxes indicate that PCR results were never produced, either due to the inability of the primer set to amplify or the lack of PCR product in individual reactions.

Figure 24: PCR results for the 2kb walk.
Green boxes indicate that allelic PCR products phenocopied wild type control PCR products. White boxes indicate PCR that did not yield product.

			Allele						
Primer	**Location Scaffold 23**	**Restriction Enzyme Digest**	**137c+/ Fud44**	**adf1-1-**	**adf1-2+**	**adf1-3-**	**adf1-4+**	**adf1-5+**	**adf1-6-**
171534	872909-875573	Nae I							
		Pvu II							
		Dde I							
171535s	875550-877558	Dde I							
		Sac II							
		Mbo II							
		Taq 1							
171535e	877522-879626	Mbo I							
		Nae I							
		Taq 1							
171535t	879567-881849	Mbo I							
		Eag I							
		Dde I							
171536s	881820-883174								
Sc940036	883166-884650	Pvu II							
		Mbo I							
		Sal I							
171536	884630-885605	Dde I							
		Sac II							
		Pvu II							
171537g	885582-887590	Hae III							
		Sac I							
		Msp I							
		Dde I							
		Taq I							
171537s	887383-889427	Mbo I							
		Taq I							

Primer	Location Scaffold 23	Restriction Enzyme Digest	Allele						
			137c+/ Fud44	adf1-1-	adf1-2+	adf1-3-	adf1-4+	adf1-5+	adf1-6-
		Pst I							
		Dde I							
171537m	889404-891318	Pvu II							
		Sac II							
		Dde I							
171537b	891292-893203	Ear I							
		Pst I							
		Eag I							
171537t	893180-895279	Pvu II							
		Sac II							
		Pst I							
171537e	895256-897281	Pvu II							
		Nae I							
		Pst I							
171509s	897264-899397	Fsp I							
		Nae I							
		Pst I							
171509e	889348-901360	Nhe I							
		Pvu II							
		Taq I							
171510s	901263-903573	Sac I							
		Taq I							
		Dde I							
171510b	903570-905642	Dde I							
		Sac II							
		Pvu II							
171510e	905619-907735	Fsp I							
		Pvu II							
		Sac I							
171511s	907706-910181	Taq I							
		Pvu II							
		Pst I							
171511e	910146-912311	Blp I							
		Dde I							
		Not I							
171512s	912060-914082	Dde I							
		Mbo I							
		Pst I							
171512e	914057-916177	Mbo I							
		Pvu II							

Primer	Location Scaffold 23	Restriction Enzyme Digest	Allele						
			137c+/ Fud44	adf1-1-	adf1-2+	adf1-3-	adf1-4+	adf1-5+	adf1-6-
		Nae I							
171512g	916154-918283	Pst I							
		Pvu II							
		Mbo I							
171512h	918138-919781	Fsp I							
		Pst I							
		Mbo I							
171538	919758-921966	Sac II							
		Fsp I							
		Dde I							
		Pst I							
940038	921379-921620	Msp I							
		Pst I							
		Eag I							
171538N	921546-927638								
171539	927615-929553	Xho I							
		Mbo I							
		Pvu II							
171539g	929543-931495	Nae I							
		Nhe I							
		Dde I							
171540	931286-933123	Taq I							
		Blp I							
		Xma I							
		Pst I							
171513	933100-934299	Sac II							
		Blp I							
		Nae I							
171513m	934276-936274								
171513e	936251-938515								
940041	938496-939743	Msp I							
		Hha I							
		Stu I							
940041e	939267-941141	Taq I							
		Mbo I							
		Pst I							
940041g	941104-943527	Dde I							
		Mbo I							
		Pst I							
171514g	943504-945571	Stu I							

			Allele						
Primer	**Location Scaffold 23**	**Restriction Enzyme Digest**	**137c+/ Fud44**	**adf1-1-**	**adf1-2+**	**adf1-3-**	**adf1-4+**	**adf1-5+**	**adf1-6-**
		ApaL I							
		Pst I							
171514	945548-947481	Eag I							
		Nae I							
		Pst I							
171514b	947401-949526	Sac II							
		Pvu II							
		Pst I							
171514t	949503-951537	Pst I							
		Mbo I							
		Dde I							
		Sac II							
		Pvu II							
171542s	951514-953515	Taq I							
		Mbo I							
		Dde I							
171542b	953495-955418								
GAP	955418-955711								
171542III	955711-957316	Not I							
		Sac I							
		Dde I							
		Eag I							
171542t	957293-958962	Nae I							
		Mbo I							
		Sac II							
171515	958934-961202	Dde I							
		Taq I							
		Msp I							
171517	961179-963227	Mbo II							
		Nae I							
		Taq I							
171516N	963206-969073								
171516e	969051-971214								
171543g	971191-973204								
171543s	973182-975529								
171544s	975506-977781								
171544m	977756-979869								
171544e	979851-982048								
171546s	981992-984009								

Primer	Location Scaffold 23	Restriction Enzyme Digest	Allele						
			137c+/ Fud44	adf1-1-	adf1-2+	adf1-3-	adf1-4+	adf1-5+	adf1-6-
171546m	983991-986328								
171546n	986304-988377								
158447s	988354-990637								
158447e	990614-992163								
171546N	992113-1003191								
171546g	1002991-1003996								
171546n2	1003968-1006345								
171546e	1006201-1008468								
171518g	1008426-1010566								
171518	1010560-1012979								
171520g	1012950-1015440								
170630	1015391-1017623	Mbo I							
		Nae I							

The 63 kb to the left of the mapping centre, at 936070 bp of Scaffold 23, was successfully amplified and subjected to restriction digest. Despite this, the numerous restriction digests did not reveal a deletion or obvious polymorphism. The 81 kb to the right of the centre was not as productive, in generating negative data. The first 27 kb to the right of the mapping centre was covered by primer sets that produced product less effectively and no aberrations were detected following restriction digest. Primer sets covering the remaining 54 kb to the right of the mapping centre were difficult to optimize and resulted in gaps in the walk. Effort exerted toward this end did not yield positive clues to the location of *ADF1*. After separating many digested PCR-generated fragments on agarose gels, it became apparent that a deletion or insert of less than 50 bp could be easily overlooked. This realization, coupled with the consideration that the *adf1* alleles could have been spontaneous mutants, unrelated to the mutagenesis (see

discussion), made an opportunity to collaborate with the Moerman lab at UBC attractive.

V: Sequencing

In collaboration with the Moerman Lab, a custom targeted array method was used to interrogate the 394 kb region of linkage group IX. Oligonucleotide array Comparative Genomic Hybridization (oaCGH; Nimblegen; http://www.nimblegen.com) was used to identify single nucleotide polymorphisms (SNPs) in the GC rich genome of *Chlamydomonas reinhardtii*, as has been done in *C.elegans* (Jones et al. 2007). The microarray chip contained overlapping, oligonucleotide probes of approximately 50 bp, designed from the 394kb region, between markers 171503 and 169190. DNA from control B214 and mutant *adf1-2* was labelled with different colours and hybridized to the chip. Regional differences in the fluorescence ratios were detected and used to identify abnormal regions in the 394 kb interval. Ten potential SNPs were identified and eight of these sites were used as starting points for sequencing in the 394 kb region. Table 12 lists the SNP sites, as located in our region.

Table 12: Predicted oaCGH SNP sites.
Deletions are determined by negative log2 ratios; insertions are determined by positive log2 ratios.

Position on Scaffold 23	Details of oaCGH Analysis
~855936	weak evidence for SNP
~877905	obvious signal but positive log2 ratio
~882275	obvious signal but positive log2 ratio
~895280	weak evidence for SNP
~919594	obvious signal but positive log2 ratio, and only on minus strand
~932039	good evidence for SNP on both strands
~937301	okay evidence for SNP
~939647	obvious signal but positive log2 ratio
~1063600	weak evidencd for SNP
~1141600	obvious signal but positive log2 ratio

Through communication with our collaborators, it was understood that the SNPs should be within ten bases of the identified site. As I already had numerous primer sets, overlapping the regions indicated by the SNP data, I was able to PCR amplify an average of 2kb around each potential SNP. In addition, following a review of mapping data, the predicted genes in the 394 kb region and the oaGCH data, we decided to also sequence 16kb of the candidate gene 171537. This candidate was large, with 18 exons and, near the center of the mapping region as indicated by recombination, and revealed a number of oaGCH hits. In addition, it was an attractive candidate for a signaling protein because of similarity to a human G-protein coupled receptor (G. Pazour, personal communication). Primers from the 2kb walk were reused.

The length of sequencing reads from Macrogen was variable, ranging from 600 bp to 900 bp, and just as variable, were the quality of the reads. Frequently, a read would come back with 850 bp, only 600 of which were free of

ambiguities. Due to this, it was necessary to reprocess trace files with an online software system, LongTrace (http://www.nucleics.com/longtrace-sequencing/index.php), which improves sequencing read length by up to 30%. Following reprocessing, sequences were assembled into contigs with Geneious (http://www.geneious.com; Biomatters Ltd.). The resulting contigs were evaluated, in comparison to the wild type sequence published in JGI, using the following criteria: 1) were discrepancies associated with unreliable reads, and resulting gaps (for example see Figure 25); 2) were discrepancies located at the end of reads, where incomplete PCR or ligation could have been influential (see Figure 26); and, 3) were discrepancies repeated in contigs for numerous alleles, including the control B214. The contigs for candidate gene 171537 were evaluated on a fourth premise: were unique discrepanices present in coding or non-coding regions? Finally, oaCGH contigs generated from *adf1-2* were evaluated for discrepancies at the predicted oaCGH SNP sites.

Contigs that were associated with poor reads were generally disregarded. It was impossible to determine if discrepancies in the read were genuine, even after LongTrace analysis. Despite this, I have still carefully recorded the nature of the discrepancies, for future reference. Similarily, discrepancies at the beginning of reads have also been disregarded. With regards to the sequencing of candidate gene 171537, I was able to align contigs, as the primer sets were designed to overlap, and therefore produce overlapping fragments. Mutations that appeared near the beginning of one contig would not exist in the neighbouring contig (see Figure 26). I therefore concluded that it was likely that

mutations at the beginning of the sequence reads were not genuine, although I had no confirmation for oaCGH examples, as they were not amplified as sequential fragments. I have diligently recorded discrepancies associated with the beginning of reads, also for future reference (Tables 13 and 14).

Figure 25: Example of a poor sequencing read.
Peaks are not sharp and frequently overlap, resulting in N's in the text.

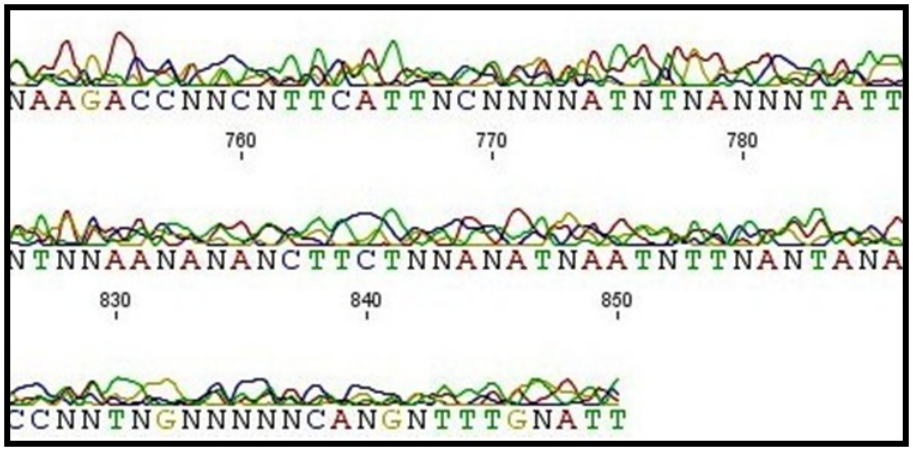


Figure 26: Example of contig ends for overlapping primer sets 171537e and 171509s.
There are discrepancies in the end of the sequence read for *adf1-4* (54A-SP), for the fragment generated with primers 171509s, but not for the fragment generated from adf1-4 (27D-T7) for 171537e.

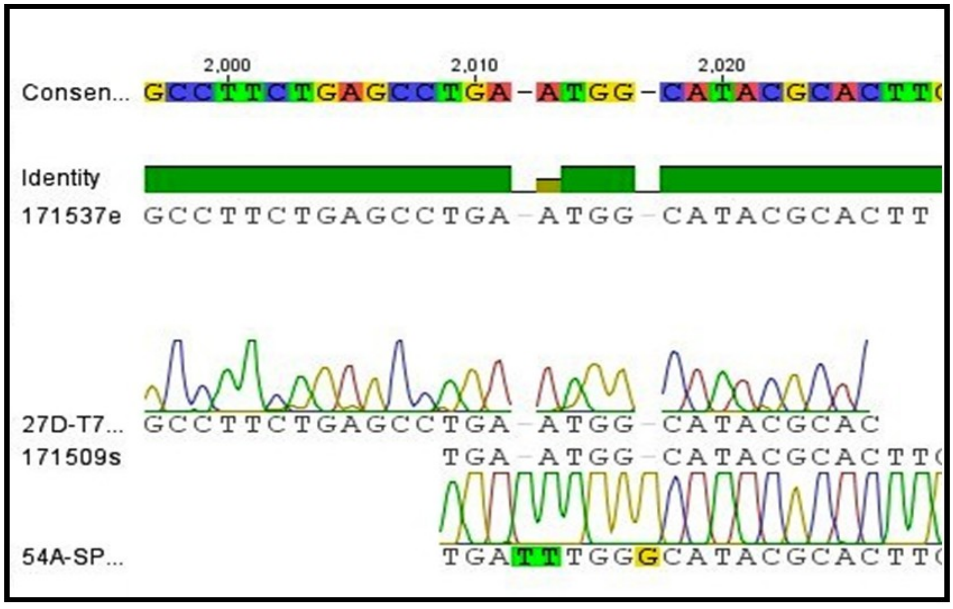


Table 13: Discrepancies in genomic DNA sequence of candidate gene 171537.
Under the heading Location, substitutions have been listed as wild type – location (bp) – mutant substitution (ie. T1915A – a T has been substituted for an A at bp 1915).

Predicted gene 171537 Sequencing Results					
Poor reads associated with gaps and/or multiple discrepancies	**Total = 11**	**Allele**	**Location (bp)**	**Characterization of Discrepancy**	**Exon vs. Intron**
171537t	Contig 271	B214	662-993	gap and poor read	N/A
	Contig 281	B214	795-1036	gap and poor read	N/A
	Contig 288	adf1-4	663-665	gap	N/A
	Contig 289	adf1-5	655-957	gap and poor read	N/A
	Contig 291	adf1-6	667-866	gap and poor read	N/A
171537m	Contig 256	adf1-4	2-424	poor read	N/A
171537e	Contig 183	adf1-2	385-1041	poor read	N/A
	Contig 184	adf1-3	717-870	poor read	N/A
171509e	Contig 312	adf1-3	940-946	poor read	N/A
Discrepancies associated with ends of reads	**Total = 17**	**Allele**	**Location (bp)**	**Characterization of Discrepancy**	**Exon vs. Intron**
171537m	Contig 243	B214	T1915A	substitution	N/A
	Contig 244	B214	G2T	substitution	N/A
171537e	Contig 177	B214	T2026A	substitution	N/A
	Contig 186	adf1-4	T2025A	substitution	N/A
		adf1-4	T2026A	substitution	N/A
171537b	Contig 242	adf1-2	G5T	substitution	N/A
		adf1-2	C6del	deletion	N/A
	Contig 234	adf1-4	in1912C	insertion	N/A
171509s	Contig 205	B214	G2T	substitution	N/A
	Contig 208	adf1-1	G2T	substitution	N/A
	Contig 213	adf1-4	in4T	insertion	N/A
		adf1-4	A5T	substitution	N/A
		adf1-4	in9G	insertion	N/A
171537t	Contig 281	B214	A2T	substitution	N/A
171510s	Contig 219	adf1-2	T2308A	substitution	N/A
		adf1-2	T2309del	deletion	N/A
	Contig 225	adf1-5	C1G	substitution	N/A
Repeat Discrepancies	**Total = 8**	**Allele**	**Location (bp)**	**Characterization of Discrepancy**	**Exon vs. Intron**
171537t	Contig 283	B214	in656G	insertion	N/A
	Contig 274	adf1-1	in656G	insertion	N/A
	Contig 281	B214	in656G	insertion	N/A
171510s	Contig 218	adf1-1	G2255del	deletion	Intron
	Contig 219	adf1-2	G2255del	deletion	Intron
	Contig 221	adf1-3	G2255del	deletion	Intron
	Contig 223	adf1-4	G2255del	deletion	Intron
	Contig 225	adf1-5	G2255del	deletion	Intron

Predicted gene 171537 Sequencing Results					
Unique Discrepancies	**Total = 15**	**Allele**	**Location (bp)**	**Characterization of Discrepancy**	**Exon vs. Intron**
171537e	Contig 176	B214	C1989T	substitution	N/A
171537m	Contig 245	B214	C605T	substitution	N/A
	Contig 247	adf1-2	C359T	substitution	Intron
	Contig 248	adf1-3	C169T	substitution	Intron
	Contig 250	adf1-5	C1589A	substitution	Intron
171537t	Contig 274	adf1-1	in657G	insertion	Exon
		adf1-1	in658G	insertion	Exon
		adf1-1	G660T	substitution	Exon
		adf1-1	T661G	substitution	Exon
		adf1-1	T740C	substitution	Exon
	Contig 288	adf1-4	in602G	insertion	Exon
171510s	Contig 221	adf1-3	G2256del	deletion	Intron
	Contig 225	adf1-5	C1790T	substitution	Intron
	Contig 226	adf1-6	C423T	substitution	Intergenic
		adf1-6	G522A	substitution	Intergenic

The sequencing of candidate gene 171537 has exposed two nonsense mutations that lead to truncated predicted proteins. The fragments generated by primer set 171537t, for alleles *adf1-1* and *adf1-4*, both harbour mutations in exon 14. The wild type protein prediction is 2438 residues long, while the *adf1-1* mutant protein truncates at 1251 residues and, the *adf1-4* mutant protein is 1253 residues. These results suggest that a deficient protein is produced; however, there are several caveats. All of the mutations listed above for contig 274 and contig 288 (Table 13) are associated with a stretch of repetitive guanines (G), and insertions of extra Gs may simply be sequencing error. The likelihood of sequencing error is increased by that fact that both of these reads were initially poor and had to be processed with LongTrace. As helpful as this program has been, it is not 100 % accurate; it seems to occasionally get confused and

mistranslate peaks on the chromatogram. In addition, there are six alleles and only two of them have legitimate mutations in coding region. Do the other four alleles have mutations upstream, perhaps in the promoter region? Based on this data predicted gene 171537 has not been eliminated as a candidate for *ADF1*.

To summarize Table 14, the SNPs predicted by oaCGH did not correspond to SNPs in sequencing reads of *adf1-2*. It is possible that discrepanices in oaCGH influenced reads of the other alleles could lead to the identification of *ADF1*, however, that investigation will have to fall to future researchers.

Table 14: Discrepancies in genomic DNA sequence of oaCGH predicted SNPs.
Under the heading Location, substitutions have been listed as wild type – location (bp) – mutant substitution (ie. T1915A – a T has been substituted for an A at bp 1915).

oaCGH Predicted SNPs Sequencing Results					
Poor reads associated with gaps and/or multiple discrepancies	**Total = 14**	**Allele**	**Location (bp)**	**Characterization of Discrepancy**	**oaCGH SNP Prediction**
171536s	Contig 124	adf1-2	G975C	poor read	NO
		adf1-2	A996C	poor read	NO
		adf1-2	T1026C	poor read	NO
	Contig 125	adf1-3	G975C	poor read	N/A
		adf1-3	T1026C	poor read	N/A
171512h	Contig 74	adf1-4	C908T	poor read	N/A
		adf1-4	G949C	poor read	N/A
		adf1-4	G954C	poor read	N/A
	Contig 86	adf1-6	880-1070	gap and poor read	N/A
171513 SNP	Contig 156	adf1-1	1143-1325	gap and poor read	N/A
	Contig 311	adf1-2	786-2225	gap and poor read	NO
	Contig 162	adf1-3	1402-1478	gap and poor read	N/A
	Contig 172	adf1-5	1405-1417	gap and no read	N/A
Discrepancies associated with ends of reads	**Total = 9**	**Allele**	**Location (bp)**	**Characterization of Discrepancy**	**oaCGH SNP Prediction**
171536s	Contig 120	adf1-1	G1T	substitution	N/A
	Contig 129	adf1-4	T1355A	substitution	N/A
171540	Contig 91	adf1-2	T1837del	deletion	NO
		adf1-2	T1838del	deletion	NO
171513 SNP	Contig 156	adf1-1	C1A	substitution	N/A
		adf1-1	A2T	substitution	N/A
		adf1-1	A3T	substitution	N/A
	Contig 171	adf1-4	in2225C	insertion	N/A
940041e	Contig 145	B214	A1T	substitution	N/A
Repeat Discrepancies	**Total = 8**	**Allele**	**Location (bp)**	**Characterization of Discrepancy**	**oaCGH SNP Prediction**
171536s	Contig 124	adf1-2	G975C	substitution	NO
		adf1-2	T1026C	substitution	NO
	Contig 125	adf1-3	G975C	substitution	N/A
		adf1-3	T1026C	substitution	N/A
171512h	Contig 58	B214	G1116A	substitution	N/A
	Contig 65	adf1-2	G1116A	substitution	NO
	Contig 69	adf1-3	G1118A	substitution	N/A
	Contig 81	adf1-5	G1118A	substitution	N/A

oaCGH Predicted SNPs Sequencing Results					
Unique Discrepancies	**Total = 21**	**Allele**	**Location (bp)**	**Characterization of Discrepancy**	**oaCGH SNP Prediction**
171535t	Contig 139	adf1-1	G777T	substitution	N/A
171536s	Contig 124	adf1-2	in1349C	insertion	NO
	Contig 129	adf1-4	T494A	substitution	N/A
	Contig 134	adf1-6	C568T	substitution	N/A
		adf1-6	C731T	substitution	N/A
171512h	Contig 52	B214	G1114A	substitution	N/A
		B214	C1333T	substitution	N/A
	Contig 65	adf1-2	G954del	deletion	NO
		adf1-2	T955del	deletion	NO
	Contig 86	adf1-6	G547A	substitution	N/A
171540	Contig 98	adf1-5	C762T	substitution	N/A
		adf1-5	G922A	substitution	N/A
171513 SNP	Contig 154	B214	C2016T	substitution	N/A
	Contig 162	adf1-3	A8del	deletion	N/A
	Contig 171	adf1-4	C216T	substitution	N/A
		adf1-4	G1644A	substitution	N/A
	Contig 172	adf1-5	T2150C	substitution	N/A
	Contig 174	adf1-6	T1098A	substitution	N/A
940041e	Contig 143	B214	C1083T	substitution	N/A
	Contig 149	adf1-3	T565del	deletion	N/A
	Contig 150	adf1-5	C404T	substitution	N/A

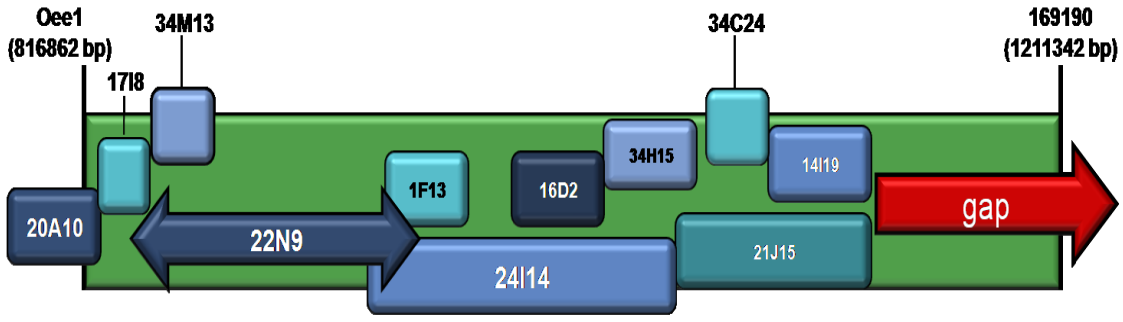
VI: Attempts to rescue the *adf1* phenotype by transforming cells with BAC clones

Rescue of the mutant phenotype is the definitive evidence needed to confirm that a gene has been identified. I decided, therefore, to start transformations of the *adf1* mutants with the bacterial artificial chromosomes (BACs) that span the 394 kb region. The region is incompletely spanned by

eleven BACs (see Figure 27), as annotated on JGI (http://genome.jgi-psf.org). While not revealing the exact gene locus, a rescue could in the very least indicate the BAC that the *ADF1* locus is on, narrowing the region.

BAC DNA was isolated following the hybrid protocol outlined in the Methods section, for BACs: 20A10, 17I8, 34M13, 22D15, 24I14, and 21J15. To confirm that the DNA isolation had been successful aliquots of BAC were subjected to restriction digest and PCR, and run on agarose gels. We wanted to start transformations with the BAC that contained the center of mapping, 22D15, which also contained several good candidate genes and potential (at that time) SNP sites predicted by oaCGH. Unfortunately, BAC 22D15 would not yield PCR fragments, additionally, restriction enzyme band sizes did not match predictions and it was concluded that the BAC was mis-plated or mis-labeled, in the library. It was therefore necessary to probe the BAC library with radioactive PCR fragments in order to determine which BAC actually covered the region from 846721 bp to 931782 bp of Scaffold 23. Two separate blots independently confirmed that the missing BAC was 22N9.

Figure 27: Bacterial artificial chromosomes spanning the 394 kb region between Oee1 and 169190.
Blue indicates a BAC, red indicates a gap (no annotated BAC) and arrow heads indicate unknown size.



After replacing BAC 22D15 with 22N9, and following BAC DNA isolation, BACs were used in co-transformation experiments, with the goal of rescuing the *adf1* mutant phenotype. Table 15 shows the number of colonies assayed for deflagellation following selection on TAP plates containing paromomycin. None of the colonies assayed had a rescue of the *adf1* phenotype. Other dominant phenotypes may have been present, such as paralyzed flagella, but these were not specifically scored for.

Table 15: Colonies assayed for deflagellation following BAC DNA transformation.

BAC	Number of Transformed colonies Assayed for Deflagellation
20A10	330
17I8	35
22N9	391
34M13	1078
24I14	736
21J15	1220

Previously published results reported that there should be a 1-5 % rescue rate in colonies grown on selection (Nguyen et al., 2005). Based on this rate we decided that we needed to complete one thousand assays for each BAC, with an expected ten to fifty rescues, in order to be confident, especially in the case of a negative result. As I learned during these initial experiments, working with BAC DNA is not trivial. I encountered problems with poor yield of BAC DNA and shearing of the BAC DNA. Regrettably, my first rounds of BAC rescue experiments were conducted with considerably less intact BAC than I had

thought I was working with. To circumvent the issues of trying to transform cells with BAC DNA, I modified my approach.

One alternate approach was to do the transformations using digested BACs, with the goals of: reducing the size of the DNA fragments being incorporated, reducing the number of fragments that are incorporated into each cell, and eliminating the random cleavage due to shearing. None of these experiments yielded rescue, but nor were they definitive. Because these experiments do not rule out any of the candidate BACs, the data is not included here.

VII: Subcloning

Having not found a rescued phenotype from whole or digested BAC transformations, and still concerned with technical issues surrounding either of these approaches, I decided that the best way to continue/advance would be with single gene subcloning and transformation. Ultimately, each of the candidate genes within the region will be subcloned until one provides rescue and the mutations are revealed for the six alleles. From sequencing reads designed to examine the predicted gene 171537 (see above), potential existed for deletions in a stretch of G's in an intron of 171510 – a predicted adenylyl cylcase (for mutations see Table 13, 171510s). This was possible because the predicted genes 171537 and 171510 code in opposite directions and DNA, upstream of 171537, that had been amplified for its potential as promoter sequence, overlapped with the predicted coding region of 171510. Based on this it was decided that the adenylyl cyclase would be the first predicted gene to go under

the knife. BAC 22N9 was digested with ApaI and NotI (NEB) and the digest was run on agarose gel. A fragment of predicted size 7769 bp, and flanked by ApaI sites, was cut out of the gel and purified. The purified fragment was then ligated into pBluescript, linearized with ApaI and transformed into E.coli. Colonies were selected based on blue white and grown up for minipreps. Isolated plasmid DNA was digested with NheI and ApaI (NEB) to confirm that the insert was the appropriate size and the plasmid was sent for sequencing. Following positive sequencing results, the plasmid pBluescript+171510 was co-transformed into *adf1* mutants with pSI103, a vector that confers paromomycin resistance to *Chlamydomonas*. Colonies were grown on selective media and then picked and grown on regular TAP. With the expectation of at least 10% co-transformation (Kindle, 1990), assays were conducted for 493 colonies and no rescue was found. From these data it is tempting to conclude that a second candidate gene has been eliminated. However, it remains possible that the promoter sequence was not included in the 7769 bp fragment following restriction digest of BAC 22N9.

DISCUSSION

The vast majority of eukaryotic cells deflagellate. It is a fundamental process about which we know very little. The mechanism for the severing of the microtubule based organelle has to some extent been unveiled, with the identification and localization of proteins such as FA1, FA2 and katanin to the transition zone (Lohret et al., 1999; Mahjoub et al., 2004; JDK Parker, unpublished results). The ultimate trigger for severing of the axoneme is a calcium signal, but the signalling pathway that leads to this signal has not been elucidated. The only member of the signalling pathway that has been discovered is ADF1 and we do not yet know its identity. Cilia, their regulation and dysfunction, have been implicated in many diseases (Badano et al., 2006). Understanding the pathway that triggers deflagellation and the signalling components involved may lead to better understanding of some disease conditions. It is important, therefore, to identify the *ADF1* gene because of the implication that it could be the ortholog of a causative factor in disease or a good target for therapeutic measures.

Although, we do not know much about the pathway signaling deflagellation, we do understand some of its attributes. We know that calcium is the ultimate signal: it is required in the extracellular media, and influx is essential to deflagellation. The calcium influx is bipartite, with an initial rapid influx, followed by a prolonged influx (Quarmby & Hartzell, 1994). One report has shown that the calcium influx needs to occur in the apical end of the cell in order

for deflagellation to occur (Wheeler et al., 2007). Second, there is an intracellular proton-activated step. It is not enough for the cell to be in an acidic environment, acidification of the cytosol is required to trigger the calcium influx (Hartzell et al., 1993). There is also evidence that IP_3 may play a role in this pathway. Following acidification and, prior to deflagellation, there is an increase in intracellular IP_3 levels (Quarmby et al., 1992, Yueh & Crain, 1993). The role of IP_3 is, however, unclear. It may induce deflagellation or it may play a role in the signalling pathway that is required to rebuild the flagella after it has been shed. Until we identify at least one of the components of this pathway it will be difficult to determine much more than we already know.

This thesis reports my attempts to use positional cloning to identify *ADF1.* My approach has also included: sequential PCR and restriction enzyme digest of approximately 90 kb, with the goal of revealing an insertion or deletion; sequencing several regions, comparing genomic DNA from the alleles and a control, with the aim of finding a polymorphism that would debilitate the gene product; and, co-transformations of a selective marker and either, one of the five BACs spanning the region, or the subcloned, predicted adenylyl cyclase, with the goal of rescuing the mutant phenotype. Ultimately, I did not succeed in identifying the gene. However, I have contributed to the future identification of *ADF1* by mapping the locus to a 394 kb region on linkage group IX.

adf1 mutations were induced with UV and insertional DNA mutagenesis (Finst et al., 1998). Four *adf1* alleles were recovered in the screen for insertional mutants and ideally, should have contained some exogenous DNA.

Unfortunately, there has been no association of the insertional DNA with the mutation, suggesting that the gene does not tolerate insertions or deletions and that the insertional DNA has been lost. It is strange that no footprint of the insertional DNA could be found and this has led to the hypothesis that these mutants were spontaneous, picked up in the screen because researchers were looking for mutants defective in deflagellation.

The isolation of spontaneous mutants could partially explain why PCR and restriction enzyme digest were not able to detect an insert or deletion. A single base pair change or small insert or deletion would not be obvious, and quite likely undetectable. My experience from designing primers and determining the best enzymes to use for restriction digest has led me to understand that I would only resolve differences of about 20 bp, when amplified bands were under 500 bp and up to, maybe 50 bp differences when amplified bands were over 500 bp. This method, therefore, would not work if the mutations were small deletions, insertions or conversions. Additionally, the entire region was never investigated using this method. As I moved to the right of the mapping center, the primer sets I designed increasingly failed to amplify. On JGI, this region is marked by more and more unsequenced spans of DNA and I can't help but speculate that the looming nearness of the predicted centromere may be causing some difficulties in PCR, just as it may have caused in sequencing reactions (centromeres are notoriously rich in repetitive DNA) (reviewed by Lamb et al., 2004). Perhaps the *ADF1* locus falls within the region I could not amplify.

Similarly, sequencing efforts were restricted to less than 25 kb of the 394 kb region. While several mutations were found and, could still be meaningful, there is a substantial expanse of sequence left to be investigated. The gene candidate, 171537 (the putative G protein-coupled receptor), does exhibit mutations in the coding sequence of two of the alleles, *adf1-1* and *adf1-4.* Both of these mutations result in truncated predicted proteins. Although none of the other alleles exhibited mutations in coding sequence, it remains possible that they carry non-coding mutations that affect expression of this gene. Therefore, this gene remains a good candidate for subcloning. In addition, despite that we were looking for specific mutations with the oaCGH sequences, which we did not find, there were still many mutations. It will be important for future investigators to determine if these mutations are in coding sequence or no. Many of these mutations seemed legitimate and could still lead to the identification of *ADF1*.

Future work will need to include a rescue of the *adf1* phenotype to confirm the gene locus. BAC transformations were attempted with the goal of rescuing the mutant phenotype, however due to many difficulties, this resulted in a non-experiment. We have had trouble accurately measuring the concentration of our BAC DNA following isolation. This has translated into uncontrolled amounts of BAC DNA in transformations and the discovery that we were most likely using insufficient amounts to produce a transformation or a rescue. For co-transformation we wanted to use a ratio, of selection plasmid to BAC, of at least 1:1. The amount of BAC to add to the selection plasmid was determined based on predicted size (kb) of the BAC and concentration. As I have already

discussed, with concentrations being difficult to determine accurately, these calculations would also have been inaccurate.

The reported co-transformation success rate of 1-5% is based on a BAC of 70 kb (Nguyen et al., 2005); however, the BACs we have used are generally larger than this (80-150kb). The larger the BAC the more difficult it is to isolate in a whole state, as well as handle it without shearing it. BAC DNA, transformed in a state where it is fragmented, may easily mutagenize the *Chlamydomonas* genome, which incorporates exogenous DNA non-homologously. Also, what starts out as the functional copy of the gene may be incorporated incompletely due to random shear, and therefore, be incapable of producing a functional protein.

It is my recommendation that transformations for the purpose of rescue are undertaken with individually subcloned candidate genes. Co-transformation, of two vectors with similar size, has a success rate of 10-50% (Kindle, 1990). Also, it is far easier to determine the concentration of smaller plasmids, and therefore calculate the amount required for 1:1, 1:2, or 1:3 ratios. The incorporation of linearized plasmid is also advantageous and, with a smaller plasmid it is easier to find a restriction enzyme that will cut the vector, but not the insert DNA. This allows the subcloned gene to be incorporated with plasmid DNA flanking on either side. Linearized plasmids are incorporated into the genome with the target gene intact at a higher rate, and can therefore, increase the frequency of rescue. The 394 kb region that contains *ADF1* is predicted to encode sixty-two proteins.

Based on what we already know about ADF1, we have predicted a membrane associated protein that may be involved in calcium influx and possibly IP_3 production. A thorough review of the predicted genes in this region has presented several attractive candidates, including an adenylyl cyclase, a G-protein activator, a TRP (transient receptor potential) related gene, and a couple of kinase-like genes. These types of candidates all fit well with our predictions; however, we really don't have supporting evidence to bias us toward any of the gene predictions. *ADF1* could be one of the attractive candidates, or it could be one of the many other predicted genes that have no identifying features. Our best course of action is to continue with the individual subcloning, working toward transformations with all sixty-two predicted genes.

Cloning of the *ADF1* gene has proven to be much more difficult than anticipated. The biggest factor is that, unlike other *Chlamydomonas* genes cloned in the Quarmby lab, the *adf1* alleles generated by insertional mutagenesis are not tagged with the exogenous DNA. Without this handle, getting a grasp on the gene is substantially more difficult. Nevertheless, cloning *ADF1* is an important goal and the work outlined in my thesis has brought us closer to that goal.

APPENDICES

APPENDIX 1: PCR-based recombination mapping primers

Primer Set Name	Location Version 3	Protein ID	Forward Sequence	Reverse Sequence	Expected Band Size
pf16	Sc_23:453374 - 456514		N/A	N/A	
V1sc101-45336	Sc_23:719530 - 721040	170953	ccaccagtttgagggtctcg	tatacgcgcgcacatgtttc	1511
calkish (C_940057)	Sc_23:786087 - 787445	171557	ttgaatctgctggcctggat	aggagccaccctgatgatga	1360
PSBO	Sc_23:811622 - 814128	171502	gtcgaccgctgcgaggagaga	cgagcgccgtatcatccggctta/ cgcatgcacgacgagaagcgag	510
oee1-HaeIII	Sc_23:813076 - 813994	171502	ggagctgctgaaggagaacg	agcgagtcggggaaataggg	200, 350
171503b	Sc_23:815331-816862	171503	ACGTGCAAAGGCATTACTCGAACC	GTGCTGTTGTGGGAAACCATGACA	1532
171503c	Sc_23:8123524-825092	171503	atgtacacggactgctcacacagt	acatgcacactcttgacgcaacac	1568
171503	Sc_23:818263-819602	171503	CAAGAACACGCTGAAGAACGCCAA	GAGTGTTCGTTGCATTGGAAGCGT	1338
171504	Sc_23:831520-832046	171504	attcaaggtttccaggagcggaga	gcgcaagactgcctgtaacacatt	526
171505	Sc_23:840196-842565	171505	AAGCCGCCAATTCCTTATTCCAGC	GTTGCACGTGGGACAAAGTTGACA	924
171530	Sc_23:848754-859292	171530	TGTGCAGGTAGGGTTTCCCTGAAT	AGTAACAGTAGTGGCAGCAGCAGT	1441
171530b	Sc_23:848754-859292	171530	AGGAAACGAAAGCCCGTTTGTGTG	TCCAGAAGGTGGACAAGCACTTCA	1201
171532	Sc_23:862756-866140	171532	AACCGGACAACAGACCTTGTGGTA	AAATGCAAAGGAGCGAGGAGGAGA	1480
171536	Sc_23:884630-885605	171536	GAAGAGTTTCATGCCCACGCACAA	AAGCTGGATGCAAAGCTGAAGGTG	936
Sc940036	Sc_23:883166 - 884649	171536	aacaagcatctgcgacgaca	gcgtgggcatgaaactcttc	1484
Sc940010	Sc_23:902809 - 907622	171510	aagtaagcggtcgtggctga	ggaggcacgaaggttgacac	1137
C_940038	Sc_23: 921379 - 921621	171538	ATGCTGTGTCCACTTGGTAGCGTA	AATAGGCCACTCAATCGTGACCGT	242
c940041	Sc_23:938343 - 940636	171541	cacaaacacgcacccaaaga	gggatactggaaggggcttg	1248
C_940017	Sc_23: 960456 - 971161	171517	AAACTACGAGATCGAGCAGGCCAA	AGCGAGCTTAGCGTGTGACTGTAA	315
sc94_ssr_272	Sc_23:979689 - 979712	171544	agtgtaccctgcaacactgtgctc	agagttgatttcggctgtgaggag	272
Star 1	Sc_23: 987960 - 991490	158447	ACATACATACACGCGGCTCTCGTT	CAGTGCATCCTGACGCAACCAAAT	663
Star 2	Sc_23: 987960 - 991490	158447	CCCTCGAGTGTTCCATGTTT	ACACGCAACACACACACCTT	200

Primer Set Name	Location Version 3	Protein ID	Forward Sequence	Reverse Sequence	Expected Band Size
C_940021	Sc_23: 1022122 - 1025978	171521	GCCCGTGGCAAAGTTTCTAAACCA	ACACACTCACACACTCACACCT	933
171549	Sc_23:1030743 - 1036247	171549	ACAACATGGCGGATAGACAAGGGA	TTGTGGCTGCTGTGTTGTGTCTTG	1307
171549b	Sc_23:1030743 - 1036247	171549	GACTTCAACGCGCACAACTGAACA	TATGGCGGGTCACAAATACCGTGA	1237
C_24100001	Sc_23:1040262 - 1043827	160721	GACGGTAGCCTGGTGATCCT	AAGGTTCCGTTGAGTGATGC	248
C_7680001	Sc_23: 1055258 - 1061332	169570	TGGGTAGGCGTGTGTATGGC	TGTTGTTGTTGTTGGTGGTGGCAG	204
C_17420001	Sc_23: 1061052 - 1064667	157554	ATGGATTTAGCAGAACAGGCCGGA	AGATGGGCTTGCAAAGATGGTTGG	366
1063600	Sc_23:1062623-1064191	157554	GAGCATCCACACAGGCACACGCACACG	TTTGGTACCGGCTTCGTCAGCGGTC	1568
C_940052	Sc_23:1074259 - 1077442	171552	TGCTTTACTTGTGCGCATTC	ACTCTCAGGCTGCTCCTCTG	226
C_6830002	Sc_23: 1083332 - 1084255	168578	AACCACGTATTTGAGGTGGACGTG	GGCGCCACAAAATCAACATATCGAC	204
C_940023	Sc_23: 11027300-1028031	171523	AAACGTACCGAGGTGTGGGTGTAT	TCCTAGGCCGTTCCATGCAATACT	610
C_2390006	Sc_23: 11270007-1130592	160531	CCTCCACCTGCAAGGTATGT	AGGGGCTTTGAATGTGAGTG	241
C_2390007	Sc23:1132688 - 1135220	160532	GCACTAGCCTTACCGTCTGC	GACAAGGCAAGTCAAGCACA	196
160532	Sc23:1132688 - 1135220	160532	CTCTAACCAACTTGTTGCTCGCGT	CATTGAGTGGTGCGGTAATTGCGA	592
C239008	Sc_23:1136313 - 1139725	160533	cccaggtataacgcttggaa	gcaaggtcagagcgattttc	750
1141600	Sc_23:1140512-1142185	intergenic	CGGGCAGGTCCGGGTCGTAATACGGTAG	CGTACGAACCGGGTCGCCTAGGAACG	1673
160529	Sc_23:1149484 - 1151454	160529	AGGAGGAAGAAGAGCAGGCAAAGT	TACAATGGCAAACACGCACTCAGC	350
C_580021	Sc23: 1160123 - 1162092	167281	CGATGACTGGATGGAGGAGT	GGGTGGAGGACGTATGAAGA	225
167281	Sc23: 1160123 - 1162092	167281	GCATCTCCTTTCAAAGCGGCAGTA	CATGACTGCTGGCGTGATGTTGAA	559
167281b	Sc23: 1160123 - 1162092	167281	AGCCAGATGATGACACTGAAGACG	ATTGCTGCTACTGCCATGACTGCT	299
169364	Sc_23:1207933 - 1211333	169364	TCTCTCCAAGTGCTACGCCTTCAT	TGCCGAGTAACCGTAACCATAACC	391
169190	Sc_23:1211342 - 1211847	169190	TGAGGATGAGGACGACGATGGAAA	AAACCGGGTGAAGCCAACATGAAC	506
169190b	Sc_23:1211334 - 1215233	169190	AGTCAGGTTGTATCCGTGCCTAAC	TGTTGAGCCACTCCCTCCAAAGAA	563
Ooooh	Sc_23: 1230845 - 1230997	160775	taaggaaacagtcgcgctgcaa	agttacatgcccgagtccatcgaa	260
161655	Sc_23:1239799 - 1274128	161655	AAGAAGGTATCCAATGCGGCCACT	GTGCAACAGCGAACGTGACTCATA	276
161655b	Sc_23:1239799 - 1274128	161655	GGCGCTGACAATAAGGCAATCGTT	ATCTCCACGCGTTTCCGTGTTAGA	987

Primer Set Name	Location Version 3	Protein ID	Forward Sequence	Reverse Sequence	Expected Band Size
C_2440008	Sc23: 1241262 - 1250229	160778	ACATGTGCTGAAGACGGTTG	CAGGAGGTCGTGTGTTTCAG	229
160778	Sc23: 1241262 - 1250229	160778	CAAGCCTACTGCAACGACTTGTCA	GACTGGCGGCTGTAGCTCTT	1430
160774	Sc_23:1251032 - 1251841	160774	AAGGCACGGTCTGGAGTCTGAAAT	ACGGTGTATCCCATGAAGTTGTGC	461
C2440003a	Sc_ 23:1255890 - 1264474	160773	cgaacttcccagctcgatac	tggcgcgtatgtaagcatag	191
C2440003b	Sc_23:1255890 - 1264474	160773	ggggaatctccattcagtg	agcaaagcaagaggaagcat	279
C2440006a	Sc_23:1289454 - 1291166	160776	taagactaggcgcaacatgc	gggataaggagggatggtgt	308
C2440006b	Sc_23:1289454 - 1291166	160776	gaatgggtgggaggtatgtg	tctacctcgcgtgtgttgac	421
D1bLIC	Sc_23:1384918 - 1389360	156307	aagaagaagctgggctcgac	ctatctgcggctgcatcttc	1050
C1530030	Sc_23:1401728 - 1404848	156326	gggcctgtatgtagccttga	tgaggtaaggcttggacacc	745
C1530029	Sc_23:1408345 - 1409183	156325	catggcggtgttaacttg	ttgaggatgagccagtctcc	766
sc23_1410	Sc_23:1410452 - 1411307	156324	agcaactgtgcgtttgccgt	aagcgattgcgcagcgtgtt	856
C1530007	Sc_23:1422450 - 1428026	156303	ctggtgctgccgtacttctc	ttgaggatgagccagtctcc	945
C1530005	Sc_23:1458819 - 1462704	156301	cacctgcagaccgaagtcatt	tggctgaactccagtgtggat	925
sc23_1510	Sc_23:1510323 - 1511331	156312	tgtagttgagggtgttggcgca	atcggctccatgctgggcaata	993
Centromere					
Sc52_536_ssr	Sc_30:7252 - 7272	166630	actcatcagacccaccccgtact	aacgtgttgtgaaaccaggtgaaa	
Sc52_368_Pst	Sc_30:193887 - 193910	166581	gacggaccaaatgcctaaaacatt	gaaaaagcgtttccacggttcc	
sc52_243950	Sc_30:336328 - 336351	161012	actccccaaatacgacacaccatt	ttgttgcagaagaccgtgtcagag/ agttgctaacagggcaggagacag	
MBO2 (c_16660006)	Sc_30:1216407 - 1225078	157052	cgttaacagccctgaactcggccg	atgcgccaaacccggagctacc/ tcacgccacacctgtacgtgcaa	
GP1	sc_25:38990-40189		AGG TCT CCG TTC AAT CTC GCA TGA	AGG AAG ACT TTG CTG TTT GCT GCG	1200
PPX1	Sc_22:1325147-1332391		CAA CGC GCG ACC AAT TTG CAT TTC	AGC CCT CGA TTT CCC TCG TTT ACA	1244
YPTC4	Sc_24:1596472-1596832		CGC CGT GAT ACG CAG CAA CAA GC	TCC ACA TGA TGG CTA GTG CGG AGC	360
CPX1	Sc_7:1604072-1604401		TTG CGT GCT AGC AGG CGT GGT G	GCT CCA AAC CTG CTG CGG TCA GTC	329

Primer Set Name	Location Version 3	Protein ID	Forward Sequence	Reverse Sequence	Expected Band Size
LC1-2	Sc_7:2525408-2527053		acaatggccaaggcaactac	ttcctgggaatccacctatg	1645
RB60	Sc_7:1931747-1931975		gccaaagagggacgctgtccacag	gcgttcggaaccacgcacatcc	228
CALKish l	Sc_5:615830-617195		TCG TGT ACC GTG TGA AAT CCA CCT	ACA TCA ACA TCA TGG CCA CAA GCC	1366
CALKish s	Sc_5:615830-616327		TCG TGT ACC GTG TGA AAT CCA CCT	AGT TCA CAC ACG AAC ACC ACA AGC	498

APPENDIX 2: 2 kb PCR walk primers

Primer Name	Location on Scaffold 23	Expected Band Size	Forward	Reverse
171534	872909-875573	2665	AGCTAGAATGATGGCCGCTAGTGT	TACGTTGCAGCTTCTCCTATCGCT
171535 s	875550-877558	2009	AGCGATAGGAGAAGCTGCAACGTA	TGCACGGCATTGATTCTGTGGTTC
171535 e	877522-879626	2105	ACTGTCATGACCGGAACCACAGAA	TTACTCTATGTCGGTGTGGCGTGT
171535 t	879567-881849	2283	A CGTCTTGATCACCACCAGCATCA	TACCGGTATTCTGCACGGCTTGAA
171536 s	881820-883174	1355	GCAACGTTCAAGCCGTGCAGAATA	C GCCTTCGAGATGCTA ACAAGCAT
Sc940036	883166-884650	1485	aacaagcatctgcgacgaca	gcgtgggcatgaaactcttc
171536	884630-885605	976	GAAGAGTTTCATGCCCACGCACAA	AAGCTGGATGCAAAGCTGAAGGTG
171537 g	885582-887590	2050	CACCTTCAGCTTTGCATCCAGCTT	TGTCTGGTGTGAGTGTGATGGGTT
171537 s	887383-889427	2045	ACGACAAATGCGAGCTTACACACC	GAAATGCTGCAAATGCCGGAGACA
171537 m	889404-891318	1915	TGTCTCCGGCATTTGCAGCATTTC	ACCATCAAACGGCGATCCACT
171537 b	891292-893203	1912	TGATGCAGTGGATCGCCGTTTGAT	AGGCTTCACACTGTGTCCCGTTTA
171537 t	893180-895279	2100	TAAACGGGACACAGTGTGAAGCCT	TGACACCCAGACCATGCAAGACAT
171537 e	895256-897281	2026	ATGTCTTGCATGGTCTGGGTGTCA	AAGTGCGTATGCCATTCAGGCT
171509 s	897264-899397	2134	TGAATGGCATACGCACTTGCACTG	ATTCCTGCTCTGGTTATGTGCGGT
171509 e	899348-901360	2013	CAGCTGCCAGCCATTCAATATGGT	TGGCAGGTCCATATGTCTCACAGT
171510 s	901263-903573	2311	CGTTGATTCAGCAAGCCTTCAGCA	ATAATGATGAGGAGGCACCAGCCA
171510 b	903570-905642	2073	TTATATGCGAGCACTGAGGCCCAT	ACATGCGATTCAGCACAGAGTCCT
171510 e	905619-907735	2117	AGGACTCTGTGCTGAATCGCATGT	TCCATGGTGAACGAGTGCTCATC
171511 s	907706-910181	2476	TGATGCGGATGAGCACTCGTTCA	GCACAGCGCTTATGGTTGTTGCTA
171511 e	910146-912311	2166	GCGCTGTCATGTTAGCAACAACCA	ATCCTGCGATGAACAGCAGGTAGG
171512 s	912060-914082	2023	ATTGTGGTGGCGAAGGGACAAAGA	AAACTTTCCTGCGAGTTCACCAGC
171512 e	914057-916177	2121	AGGCTGGTGAACTCGCAGGAAAGT	GAAAGGACAAGAAAGCGCAAGCGA
171512 g	916154-918283	2130	TCGCTTGCGCTTTCTTGTCCTTTC	CGTGCATGGTCAGTAGACTGCAAA

Primer Name	Location on Scaffold 23	Expected Band Size	Forward	Reverse
171512 h	918138-919781	1644	ACAAGCTAGTTGCGGGCATATCGT	GTGCGTGAAATGGCGAGTGTATGA
171538	919758-921966	2209	TCATACACTCGCCATTTCACGCAC	AGCGCATGACTACCTCACACTACA
940038	921379-921620	242	ATGCTGTGTCCACTTGGTAGCGTA	AATAGGCCACTCAATCGTGACCGT
171538 Ns	921546-927638	6093	ATTAAATGCCTGCCGCGAGGTTTC	TCATCCAGCTGTGTCGCATGTACT
171539	927615-929553	1939	AGTACATGCGACACAGCTGGATGA	GTCATTCTGTTGCGGTGCGTACAA
171539 gap	929543-931495	1953	AACAGAATGACGCTCTTGCCAACG	TGTGGCGAAGTGTAGCTTCCGATT
171540	931286-933123	1838	GGAAGCCCTTTCGTATAGCTGCAA	AATCCGTCTGCTCTCAACTTGCCT
171513	933100-934299	1200	AGGCAAGTTGAGAGCAGACGGATT	ACATGAGTTGTAGAGCCGAAGCCA
171513 m	934276-936274	1999	TGGCTTCGGCTCTACAACTCATGT	AGAGGTTGGTGTGCGTTGTTGTTG
171513 e	936251-938515	2265	CAACAACAACGCACACCAACCTCT	C GAAC ACAAACACGCACCCAAAGA
171513 SNP	936251-938475	2225	CAACAACAACGCACACCAACCTCT	TGCGCGTGTAGGATGTAAGGTGAA
940041	938496-939743	1248	cacaaacacgcacccaaaga	gggatactggaaggggcttg
940041 e	939267-941141	1875	ATCAGGCCGAGCGCATACAATACA	TTTCAGCGTGGCAACACAGGCATA
940041 g	941104-943527	2424	ATGTGGTGTGCGTGTATGCCTGT	TTCCTCGGGCATCATCAACGGAAT
171514 gap	943504-945571	2068	ATTCCGTTGATGATGCCCGAGGAA	TCTCTGCTCTTCTTTGCGAGGGTT
171514	945548-947481	1934	AACCCTCGCAAAGAAGAGCAGAGA	CGCTAAGGCAGCAGCGATGAATTT
171514 b	947401-949526	2126	CCCAAATTCATCGCTGCTGCCTTA	AAGTGGATTGCACAGAACCTCCGT
171514 t	949503-951537	2035	ACGGAGGTTCTGTGCAATCCACTT	AGGTTGTCAGCAAGCACAGAATGC
171542 start	951514-953515	2002	GCATTCTGTGCTTGCTGACAACCT	TCAGCTTCAGACAGCGTTGGA
171542 body II F	953421-955732	2312	TGTCGCCAAATGTGACCTCGAT	A TGATTCGGACAGCCCGTCAAT
171542 body	953495-955418	1924	TCCAACGCTGTCTGAAGCTGA	AGTGCGCTACACCGCCTTTGA
171542 leg	955179-957316	2138	ACACACCCTCCCGCAGCCAATATC	TGTGTGCCGATGCTACGGTAATCT
171542 leg II F	955711-957316	1606	ATGATTCGGACAGCCCGTCAAT	A GATTACCGTAGCATCGGCACACA
171542 tail	957293-958962	1670	AGATTACCGTAGCATCGGCACACA	AGCCAAAGAAGACGAACATGGCCT

Primer Name	Location on Scaffold 23	Expected Band Size	Forward	Reverse
171515	958934-961202	2269	ACGACAGGCCATGTTCGTCTTCTT	TGTGCTTCAAACTGAAACAGGGCG
171517	961179-963227	2049	CGCCCTGTTTCAGTTTGAAGCACA	ATCATTGCCAGCTTGTGGCCTA
171516 N	963206-969073	5868	TAGGCCACAAGCTGGCAATGAT	TTAGCGGCGGTCTATGGAATGGA
171516 e	969051-971214	2164	TCCATTCCATAGACCGCCGCTAA	TAAACGTCTCTCGGCCACCTCATT
171543 g	971191-973204	2014	AATGAGGTGGCCGAGAGACGTTTA	TATTGCAGGGTGTGGCGTCGTTT
171543 s	973182-975529	2348	AAACGACGCCACACCCTGCAATA	AACTAGCCTCAGGTTGCTGAACCA
171544 s	975506-977781	2275	TGGTTCAGCAACCTGAGGCTAGTT	ATGCGATGTGTGTGTGTGTGTGTG
171544 m	977756-979869	2113	ACACACACACACACACACACATCG	TGTTCGAGACCATGCGAATGGAGA
171544 e	979851-982048	2197	ATTCGCATGGTCTCGAACAGGT	GAAAGCGGCCGGTTGCTTGTTTAT
171546 s	981992-984009	2018	TTTGGCGTACAGATTGCCATCAGGTG	AGCAGCAGCATCATGTACAGCAC
171546 m	983991-986328	2338	TGTACATGATGCTGCTGCTGCT	ATCTATGTGCGTAACAAGGACGCC
171546 n	986304-988377	2073	AGGCGTCCTTGTTACGCACATAGA	AGAGAGGGCACAAGGAAGGAAACA
158447a	987957-989242	1285	GCTCAGACCAAAGGTGAGCACTGAGAAC	CTAAGAAGAAGGTGCGGGTGCTTG
158447 s	988354-990637	2284	TGTTTCCTTCCTTGTGCCCTCTCT	TGGGTGCATGTGTTTGTGTGTCAG
158447 b	989227-990537	1310	CGCACCTTCTTCTTAGCGGCTGCCTTG	CGGCGTCGCTATGCTATTGGCCTTGTG
158447 e	990614-992163	1550	CTGACACACAAACACATGCACCCA	AGCAAAGCATTGCGCTATACTGCC
158447 c	990744-992241	1497	CGGGATATGGGTGCTACTCAGCAGGGTC	GTTGAGTGTAGTGCCGGTGGGCAATCG
171546 N	992113-1003191	11078	TCCTAAGGTGGGTGGAACTTGTGT	AAACCCTATTGGTGCAAGCAACCC
171546 g	1002991-1003996	1006	TCAATCTGCTGCTGCACCTCCT	GGAGGGTTGTTCTCACCACTCCTA
171546 n2	1003968-1006345	2378	GCCGTTAGGAGTGGTGAGAACAA	GTGCATGTGTTCACTTGTGCGACT
171546 e	1006201-1008468	2267	GTGTGTGTGTGTGTGTATTGAGGG	TGTGCACAAGGACGCTTGCATTAC
171518 g	1008426-1010566	2141	CAGTTTCATCGCCCACGCTGTAAT	ACATGTAAGCTGGTGCGCATGTTG
171518	1010560-1012979	2420	TACATGTGGCGACTGTGATGCGTT	TGTCGGCACTACGGCTGATATAC
171520 g	1012950-1015440	2491	AGCTGCTGTATATCAGCCGTAGTG	ACGGTGCTTGCAGCGTGTATGTTA
170630	1015391-1017623	2233	GCACACACACGTCCTGCTTACTAT	AACACACAGGCGATAGTGGTGACA

APPENDIX 3: Candidate Genes

Gene Identity Version 2	Scaffold_23 Location	BLAST E-Value	Description
171502	813076-813994	N/A	
171503	815314-825349	1.00E-25	glutathione reductase; Pyr_redox_dim superfamily
171528	825656-830908	0	Chlamydomonas predicted protein; no other hits
171504	829944-839704	N/A	proline binding motif; proline interaction residues; GYF Superfamily
171529	833231-839484		
171505	840196-842565	4.00E-21	predicted DnaJ domain containing; Hsp70 interaction site
171506	843432-845619	2.00E-84	TP_methylase Superfamily; uroporphyrin-III C-methyltransferase
171507	847265-849579	N/A	no hits
171530	848754-859292	0	Chlamydomonas predicted protein; no hits
171531	860979-861537	1.00E-10	Ring domain ligase 2; ubiquitin-protein ligase
171532	862756-866140	0	Chlamydomonas predicted protein; no hits
171533	868838-872482	0	Chlamydomonas predicted protein; no hits
171534	874240-875141	8.00E-05	ANK Superfamily; ion channel nompc
171535	876303-880772	1.00E-13	Chlamydomonas predicted protein; no hits
171536	881665-886689	1.00E-15	ANK Superfamily; cross brace motif; RING superfamily; Chlamydomonas predicted protein; no hits
171537	887935:900124	0	Chlamydomonas predicted protein; no hits
171509	891002:900351		
171510	902809:907622	3.00E-13	receptor type adenylate cyclase
171511	908919:910402	0.057	Chlamydomonas predicted protein; adenylate cyclase
171512	912146:915716	0	Chlamydomonas predicted protein; no hits
171538	920375:921687	N/A	multidomain PrmA

Gene Identity Version 2	Scaffold_23 Location	BLAST E-Value	Description
171539	927577:930346	9.00E-105	peptidase_MII superfamily; matrix metalloproteinase; gametolysin
171540	932398:934532	N/A	no coding region found; no hits
171513	934354:934850	N/A	no hits; poor gene prediction
171541	938343:940636	1.00E-24	TBC superfamily; RabGAP/ TBC domain-containing protein; growth hormone regulated TBC protein (1.00E-18)
171514	947119:952289	8.00E-136	Chlamydomonas predicted protein; no hits
171542	952684:958131	0	Chlamydomonas predicted protein; no hits
171515	958944:961028	7.00E-22	Chlamydomonas Cytochrome P450; P450 superfamily
171517	960456:971161	4.00E-21	thromboxane A synthase I; P450 Superfamily
171516	961837:971128	2.00E-34	thromboxane A synthase I; P450 Superfamily
171543	972996:974781	0	Chlamydomonas predicted protein; no hits
171544	976167:981677	4.00E-34	P450 Superfamily; cytochrome P450
171546	983169:1007956		
158447	987960:991490	3.00E-77	Chlamydomonas predicted protein; calmodulin-binding protein trp1 (2.00E-04)
171518	1010703:1011729	N/A	no hits
171519	1012889:1013374	N/A	no hits
171520	1014183:1020803	4.00E-35	cytochrome P450, family 3; P450 superfamily
170630	1014550:1018827	0	CypX superfamily; multidomain P450; cytochrome p450, CYP711
171521	1022122:1025978	2.00E-33	FAD linked oxidase-like; FAD_binding_4 Superfamily
171547	1022748:1024352	4.00E-23	FAD_biding_4 Superfamily; FAD linked oxidase-like
171549	1030743:1036247	4.00E-32	ATP binding site; catalytic loop; PI3Kc_like Superfmaily; DNA dependent protein kinase
171551	1037776:1041543	5.00E-24	Chlamydomonas DNA dependent protein kinase catalytic subunit
160721	1040262:1043827	5.00E-16	Chlamydomonas DNA dependent protein kinase catalytic subunit

Gene Identity Version 2	Scaffold_23 Location	BLAST E-Value	Description
169570	1055258:1061332	2.00E-62	Chlamydomonas predicted protein; no hits
154109	1056108:1060173		
157554	1061052:1064667	8.00E-81	Chlamydomonas predicted protein; no hits
171522	1074010:1077005	0	Chlamydomonas predicted protein; no hits
171552	1074259:1077442		
168579	1077057:1077867	8.00E-163	Chlamydomonas predicted protein; no hits
168577	1078471:1080886	7.00E-101	Chlamydomonas predicted protein; DNA dependent protein kinase catalytic subunit
168578	1083332:1084255	6.00E-150	Chlamydomonas predicted protein; no hits
160534	1085352:1085652	N/A	no hits
160528	1089144:1092707	4.00E-12	fibroin
160526	1096102:1098406	5.00E-29	Chlamydomonas predicted protein; armadillo/ beta-catenin repeat family protein (7E-06)
171523	1102730:1108031	N/A	no hits
171524	1111295:1114996	4.00E-35	Chlamydomonas predicted protein; no hits
160527	1120316:1122087		
160530	1123795:1124737	2.00E-106	active site; ATP binding site; PKc-like superfamily; Chlamydomonas predicted protein
160531	1127007:1130592	2.00E-10	Chlamydomonas fatty acid desaturase; many hits
160532	1132688:1135220	7.00E-12	Chlamydomonas glyoxal or galactose oxidase
160533	1136313:1139725	3.00E-143	Ribosomal L4 superfamily; Chlamydomonas 60s ribosomal protein L4
160529	1149484:1151454	N/A	no hits
167281	1160123:1162092	N/A	no hits
169364	1207933:1211333		
169190	1211334:1215233		

APPENDIX 4: Conserved domains and peptides annotated on JGI.

Gene Identity Version 2	Sc23 Location	Conserved Domains Annotated on JGI Version 3	Peptides Annotated on JGI Version 3
171502	813076-813994		
171503	815314-825349	PF03128.4: CXCXC CXCXC repeat	
171528	825656-830908	PF03128.4: CXCXC CXCXC repeat	
171504	829944-839704	PF02083.5: Urotensin_II Urotensin II	CrFP_peptides_11533
171529	833231-839484		
171505	840196-842565		
171506	843432-845619		
171507	847265-849579		CrFP_peptides_5754
171530	848754-859292	PF03128.4: CXCXC CXCXC repeat	CrFP_peptides_6430
171531	860979-861537	PF03854.3: zf-P11 P-11 zinc finger	
171532	862756-866140	PF02809.9: UIM Ubiquitin interaction motif	CrFP_peptides_5753
171533	868838-872482	PF00097.10: zf-C3HC4 Zinc finger, C3HC4 type (RING finger)	
171534	874240-875141	PF00023.15: Ank Ankyrin repeat	
171535	876303-880772	PF00446.7: GnRH Gonadotropin-releasing hormone; PF02083.5: Urotensin_II Urotensin II	
171536	881665-886689	PF00023.15: Ank Ankyrin repeat; PF06994.1: Involucrin2 Involucrin	
171537	887935:900124	PF00249.16: Myb_DNA-binding Myb-like DNA-binding domain; PF00446.7: GnRH Gonadotropin-releasing hormone	
171509	891002:900351	PF00446.7: GnRH Gonadotropin-releasing hormone	
171510	902809:907622		phosphoproteome_169

Gene Identity Version 2	Sc23 Location	Conserved Domains Annotated on JGI Version 3	Peptides Annotated on JGI Version 3
171511	908919:910402		Intergenic: CrFP_peptides_16441
171512	912146:915716	PF02012.8: BNR BNR/Asp-box repeat	
171538	920375:921687		
171539	927577:930346	PF02044.6: Bombesin Bombesin-like peptide; PF07846.1: Metallothio_7 Metallothionein family 7; PF02756.4: GYR GYR motif	CrFP_peptides_478; CrFP_peptides_5105; CrFP_peptides_13914; CrFP_peptides_5121; CrFP_peptides_13913
171540	932398:934532		CrFP_peptides_2380; CrFP_peptides_2444; CrFP_peptides_15400
171513	934354:934850		
171541	938343:940636		
171514	947119:952289	PF00187.8: Chitin_bind_1 Chitin recognition protein	
171542	952684:958131		
171515	958944:961028		
171517	960456:971161		
171516	961837:971128		
171543	972996:974781		
171544	976167:981677		
171546	983169:1007956		CrFP_peptides_17825
158447	987960:991490		
171518	1010703:1011729		
171519	1012889:1013374	PF00646.18: F-box F-box domain	
171520	1014183:1020803		
170630	1014550:1018827		

Gene Identity Version 2	Sc23 Location	Conserved Domains Annotated on JGI Version 3	Peptides Annotated on JGI Version 3
171521	1022122:1025978		
171547	1022748:1024352		
171549	1030743:1036247	PF02260.9: FATC FATC domain	
171551	1037776:1041543		
160721	1040262:1043827	PF00746.10: Gram_pos_anchor Gram positive anchor; PF02985.7: HEAT HEAT repeat	Intergenic: CrFP_peptides_12393
169570	1055258:1061332	PF02985.7: HEAT HEAT repeat; PF00514.10: Arm Armadillo/beta-catenin-like repeat; PF02985.7: HEAT HEAT repeat	
154109	1056108:1060173		CrFP_peptides_2374; CrFP_peptides_15409
157554	1061052:1064667		CrFP_peptides_6951; CrFP_peptides_2453; CrFP_peptides_9029; CrFP_peptides_13273; CrFP_peptides_15426
171522	1074010:1077005		
171552	1074259:1077442		
168579	1077057:1077867		
168577	1078471:1080886	PF00646.18: F-box F-box domain	
168578	1083332:1084255		
160534	1085352:1085652		
160528	1089144:1092707		
160526	1096102:1098406	PF00514.10: Arm Armadillo/beta-catenin-like repeat; PF00627.16: UBA UBA/TS-N domain	

Gene Identity Version 2	Sc23 Location	Conserved Domains Annotated on JGI Version 3	Peptides Annotated on JGI Version 3
171523	1102730:1108031	PF07004.1: DUF1309 Protein of unknown function (DUF1309)	CrFP_peptides_13367
171524	1111295:1114996		Intergenic: CrFP_peptides_9836; CrFP_peptides_6040; CrFP_peptides_12184
160527	1120316:1122087		
160530	1123795:1124737		
160531	1127007:1130592		
160532	1132688:1135220	PF00573.9: Ribosomal_L4 Ribosomal protein L4/L1 family	CrFP_peptides_2971; CrFP_peptides_13419
160533	1136313:1139725	PF00573.9: Ribosomal_L4 Ribosomal protein L4/L1 family	CrFP_peptides_13715
160529	1149484:1151454		CrFP_peptides_8989; CrFP_peptides_15374 Intergenic: CrFP_peptides_3659; CrFp_peptides_5719; CrFP_peptides_5951; CrFP_peptides_493; CrFP_peptides_5029; CrFP_peptides_16807; CrFP_peptides_6951
167281	1160123:1162092	PF00560.18: LRR_1 Leucine Rich Repeat	Intergenic:

Gene Identity Version 2	Sc23 Location	Conserved Domains Annotated on JGI Version 3	Peptides Annotated on JGI Version 3
169364	1207933:1211333	PF07004.1: DUF1309 Protein of unknown function (DUF1309); PF04886.2: PT PT repeat	CrFP_peptides_493; CrFP_peptides_5029; CrFP_peptides_16807
169190	1211334:1215233		

APPENDIX 5: Proteome Search Results

Gene Identity Version 2	Flagellar Proteome	FBB Proteome	EyeSpot Proteome	Centriole Proteome
171502	N	N	7 peptides	N
171503	N	N	N	N
171528	N	N	N	N
171504	N	N	N	N
171529	N	N	N	N
171505	N	N	N	N
171506	N	N	N	N
171507	N	N	N	N
171530	N	N	N	N
171531	N	N	N	N
171532	N	N	N	N
171533	N	N	N	N
171534	N	N	N	N
171535	N	N	N	N
171536	N	N	N	N
171537	N	N	N	N
171509	N	N	N	N
171510	N	N	N	N
171511	N	N	N	N
171512	N	N	N	N
171538	N	N	N	N
171539	2 peptides	N	N	N
171540	N	N	N	N
171513	N	N	N	N
171541	N	N	N	N
171514	N	N	N	N
171542	N	N	N	N
171515	N	N	N	N
171517	N	N	N	N
171516	N	N	N	N
171543	N	N	N	N
171544	N	N	N	N
171546	N	N	N	N
158447	N	N	N	N
171518	N	N	N	N
171519	N	N	N	N
171520	N	N	N	N

Gene Identity Version 2	Flagellar Proteome	FBB Proteome	EyeSpot Proteome	Centriole Proteome
170630	N	N	N	N
171521	N	N	N	N
171547	N	N	N	N
171549	N	Y	N	N
171551	N	N	N	N
160721	N	N	N	N
169570	N	N	N	N
154109	N	N	N	N
157554	N	N	N	N
171522	N	N	N	N
171552	N	N	N	N
168579	N	N	N	N
168577	N	N	N	N
168578	N	N	N	N
160534	N	N	N	N
160528	N	N	N	N
160526	N	N	N	N
171523	N	N	N	N
171524	N	N	N	N
160527	N	N	N	N
160530	N	N	N	N
160531	N	N	N	N
160532	1 peptide	N	N	N
160533	1 peptide	N	5 peptides	N
160529	N	N	N	N
167281	N	N	N	N
169364	N	N	N	N
169190	N	N	N	N

REFERENCES LIST

Badano, J. L., N. Mitsuma, et al. (2006). "The ciliopathies: an emerging class of human genetic disorders." Annu Rev Genomics Hum Genet **7**: 125-148.

Brown, L. E., S. L. Sprecher, et al. (1991). "Introduction of exogenous DNA into Chlamydomonas reinhardtii by electroporation." Mol Cell Biol **11**(4): 2328-2332.

Chapman, A. B. (2007). "Autosomal dominant polycystic kidney disease: time for a change?" J Am Soc Nephrol **18**(5): 1399-1407.

Chodhari, R., H. M. Mitchison, et al. (2004). "Cilia, primary ciliary dyskinesia and molecular genetics." Paediatr Respir Rev **5**(1): 69-76.

Christensen, S. T., S. F. Pedersen, et al. (2008). "The primary cilium coordinates signaling pathways in cell cycle control and migration during development and tissue repair." Curr Top Dev Biol **85**: 261-301.

Davenport, J. R. and B. K. Yoder (2005). "An incredible decade for the primary cilium: a look at a once-forgotten organelle." Am J Physiol Renal Physiol **289**(6): F1159-1169.

Dutcher, S. K., N. S. Morrissette, et al. (2002). "Epsilon-tubulin is an essential component of the centriole." Mol Biol Cell **13**(11): 3859-3869.

Finst, R. J., P. J. Kim, et al. (2000). "Fa1p is a 171 kDa protein essential for axonemal microtubule severing in Chlamydomonas." J Cell Sci **113 (Pt 11)**: 1963-1971.

Finst, R. J., P. J. Kim, et al. (1998). "Genetics of the deflagellation pathway in Chlamydomonas." Genetics **149**(2): 927-936.

Fliegauf, M., T. Benzing, et al. (2007). "When cilia go bad: cilia defects and ciliopathies." Nat Rev Mol Cell Biol **8**(11): 880-893.

Geimer, S. and M. Melkonian (2004). "The ultrastructure of the Chlamydomonas reinhardtii basal apparatus: identification of an early marker of radial asymmetry inherent in the basal body." J Cell Sci **117**(Pt 13): 2663-2674.

Ginger, M. L., N. Portman, et al. (2008). "Swimming with protists: perception, motility and flagellum assembly." Nat Rev Microbiol **6**(11): 838-850.

Gross, C. H., L. P. Ranum, et al. (1988). "Extensive restriction fragment length polymorphisms in a new isolate of Chlamydomonas reinhardtii." Curr Genet **13**(6): 503-508.

Grossman, A. R., E. E. Harris, et al. (2003). "Chlamydomonas reinhardtii at the crossroads of genomics." Eukaryot Cell **2**(6): 1137-1150.

Harris, E. H. (1989). "The Chlamydomonas Sourcebook: A Compehensive Guide to Biology and Laboratory Use." 1st ed. San Diego: Academic Press Inc.

Hildebrandt, F., M. Attanasio, et al. (2009). "Nephronophthisis: disease mechanisms of a ciliopathy." J Am Soc Nephrol **20**(1): 23-35.

Insinna, C. and J. C. Besharse (2008). "Intraflagellar transport and the sensory outer segment of vertebrate photoreceptors." Dev Dyn **237**(8): 1982-1992.

Kathir, P., M. LaVoie, et al. (2003). "Molecular map of the Chlamydomonas reinhardtii nuclear genome." Eukaryot Cell **2**(2): 362-379.
Keller, L. C., E. P. Romijn, et al. (2005). "Proteomic Analysis of Isolated *Chlaymdomonas* Centrioles Reveals Orthologs of Ciliary-Disease Genes." Current Biology **15**: 1090-1098.
Kindle, K. L. (1990). "High-frequency nuclear transformation of Chlamydomonas reinhardtii." Proc Natl Acad Sci U S A **87**(3): 1228-1232.
Kwan, A. L., L. Li, et al. (2009). "Improving gene-finding in Chlamydomonas reinhardtii:GreenGenie2." BMC Genomics **10**: 210.
Lamb, J. C., J. Theuri, et al. (2004). "What's in a centromere?" Genome Biol **5**(9): 239.
Lewin, R. A., T. H. Lee, et al. (1982). "Effects of various agents on flagellar activity, flagellar autotomy and cell viability in four species of Chlamydomonas (chlorophyta: volvocales)." Symp Soc Exp Biol **35**: 421-437.
Li, J. B., J. M. Gerdes, et al. (2004). "Comparative Genomics Identifies a Flagellar and Basal Body Proteome that Includes the BBS5 Human Disease Gene." Cell **117**: 541-552.
Lohret, T. A., L. Zhao, et al. (1999). "Cloning of Chlamydomonas p60 katanin and localization to the site of outer doublet severing during deflagellation." Cell Motil Cytoskeleton **43**(3): 221-231.
Mahjoub, M. R., B. Montpetit, et al. (2002). "The FA2 gene of Chlamydomonas encodes a NIMA family kinase with roles in cell cycle progression and microtubule severing during deflagellation." J Cell Sci **115**(Pt 8): 1759-1768.
Mahjoub, M. R., M. Qasim Rasi, et al. (2004). "A NIMA-related kinase, Fa2p, localizes to a novel site in the proximal cilia of Chlamydomonas and mouse kidney cells." Mol Biol Cell **15**(11): 5172-5186.
McEwen, D. P., P. M. Jenkins, et al. (2008). "Olfactory cilia: our direct neuronal connection to the external world." Curr Top Dev Biol **85**: 333-370.
Merchant, S. S., S. E. Prochnik, et al. (2007). "The Chlamydomonas genome reveals the evolution of key animal and plant functions." Science **318**(5848): 245-250.
Morillas, H. N., M. Zariwala, et al. (2007). "Genetic causes of bronchiectasis: primary ciliary dyskinesia." Respiration **74**(3): 252-263.
Nayak, G. D., H. S. Ratnayaka, et al. (2007). "Development of the hair bundle and mechanotransduction." Int J Dev Biol **51**(6-7): 597-608.
Nguyen, R. L., L. W. Tam, et al. (2005). "The LF1 gene of Chlamydomonas reinhardtii encodes a novel protein required for flagellar length control." Genetics **169**(3): 1415-1424.
Overgaard, C. E., K. M. Sanzone, et al. (2009). "Deciliation is associated with dramatic remodeling of epithelial cell junctions and surface domains." Mol Biol Cell **20**(1): 102-113.
Parker, J. D. and L. M. Quarmby (2003). "Chlamydomonas fla mutants reveal a link between deflagellation and intraflagellar transport." BMC Cell Biol **4**: 11.

Pazour, G. J., N. Agrin, et al. (2005). "Proteomic analysis of a eukaryotic cilium." J Cell Biol **170**(1): 103-113.
Pazour, G. J. and R. A. Bloodgood (2008). "Targeting proteins to the ciliary membrane." Curr Top Dev Biol **85**: 115-149.
Pazour, G. J. and G. B. Witman (2009). "The Chlamydomonas Flagellum as a Model for Human Ciliary Disease. In: Witman G. B., The *Chlamydomonas* Sourcebook. 2nd ed. San Diego: Academic Press Inc. p 445 - 478.
Pedersen, L. B. and J. L. Rosenbaum (2008). "Intraflagellar transport (IFT) role in ciliary assembly, resorption and signalling." Curr Top Dev Biol **85**: 23-61.
Pedersen, L. B., I. R. Veland, et al. (2008). "Assembly of primary cilia." Dev Dyn **237**(8): 1993-2006.
Qamar, S., M. Vadivelu, et al. (2007). "TRP channels and kidney disease: lessons from polycystic kidney disease." Biochem Soc Trans **35**(Pt 1): 124-128.
Quarmby, L. (2009). "Ciliary ion channels: location, location, location." Curr Biol **19**(4): R158-160.
Quarmby, L. (2009). "Deflagellation." In: Witman G. B., The *Chlamydomonas* Sourcebook. 2nd ed. San Diego: Academic Press Inc. p 43 - 69.
Quarmby, L. M. (2004). "Cellular deflagellation." Int Rev Cytol **233**: 47-91.
Quarmby, L. M. and H. C. Hartzell (1994). "Two distinct, calcium-mediated, signal transduction pathways can trigger deflagellation in *Chlamydomonas* reinhardtii." J Cell Biol **124**(5): 807-815.
Quarmby, L. M. and J. D. Parker (2005). "Cilia and the cell cycle?" J Cell Biol **169**(5): 707-710.
Quarmby, L. M., Y. G. Yueh, et al. (1992). "Inositol phospholipid metabolism may trigger flagellar excision in *Chlamydomonas reinhardtii*." J Cell Biol **116**(3): 737-744.
Quinlan, R. J., J. L. Tobin, et al. (2008). "Modeling ciliopathies: Primary cilia in development and disease." Curr Top Dev Biol **84**: 249-310.
Rosenbaum, J. L. and G. B. Witman (2002). "Intraflagellar transport." Nat Rev Mol Cell Biol **3**(11): 813-825.
Rymarquis, L. A., J. M. Handley, et al. (2005). "Beyond complementation. Map-based cloning in Chlamydomonas reinhardtii." Plant Physiol **137**(2): 557-566.
Sanders, M. A. and J. L. Salisbury (1989). "Centrin-mediated microtubule severing during flagellar excision in Chlamydomonas reinhardtii." J Cell Biol **108**(5): 1751-1760.
Sears, B. B., J. E. Boynton, et al. (1980). "The Effect of Gametogenesis Regimes on the Chloroplast Genetic System of CHLAMYDOMONAS REINHARDTII." Genetics **96**(1): 95-114.
Sharma, N., N. F. Berbari, et al. (2008). "Ciliary dysfunction in developmental abnormalities and diseases." Curr Top Dev Biol **85**: 371-427.
Silflow, C. D. and P. A. Lefebvre (2001). "Assembly and motility of eukaryotic cilia and flagella. Lessons from Chlamydomonas reinhardtii." Plant Physiol **127**(4): 1500-1507.

Veland, I. R., A. Awan, et al. (2009). "Primary cilia and signaling pathways in mammalian development, health and disease." Nephron Physiol **111**(3): p39-53.
Vincze, T., J. Posfai, et al. (2003). "NEBcutter: A program to cleave DNA with restriction enzymes." Nucleic Acids Res **31**(13): 3688-3691.
Vysotskaia, V. S., D. E. Curtis, et al. (2001). "Development and characterization of genome-wide single nucleotide polymorphism markers in the green alga Chlamydomonas reinhardtii." Plant Physiol **127**(2): 386-389.
Weimbs, T. (2007). "Polycystic kidney disease and renal injury repair: common pathways, fluid flow, and the function of polycystin-1." Am J Physiol Renal Physiol **293**(5): F1423-1432.
Wheeler, G. L., I. Joint, et al. (2008). "Rapid spatiotemporal patterning of cytosolic Ca2+ underlies flagellar excision in Chlamydomonas reinhardtii." Plant J **53**(3): 401-413.
Wilson, P. D. (2001). "Polycystin: new aspects of structure, function, and regulation." J Am Soc Nephrol **12**(4): 834-845.
Wong, S. Y. and J. F. Reiter (2008). "The primary cilium at the crossroads of mammalian hedgehog signaling." Curr Top Dev Biol **85**: 225-260.
Woodrow, D.T. and R. W. Linck (2002). "Internal Organization of the cell." In: Alberts, B., editor. Molecular Biology of the Cell. 4th ed.New York: Garland Science. Figure 16-80.
Yueh, Y. G. and R. C. Crain (1993). "Deflagellation of Chlamydomonas reinhardtii follows a rapid transitory accumulation of inositol 1,4,5-trisphosphate and requires Ca2+ entry." J Cell Biol **123**(4): 869-875.