5-2020

Understanding the Implications of Anandamide, an Endocannabinoid in an Early Land Plant, Physcomitrella patens

Md Imdadul Haq
East Tennessee State University

Understanding the Implications of Anandamide, an Endocannabinoid in an Early Land Plant, *Physcomitrella patens*

A dissertation

presented to

the faculty of the Department of Biological Sciences

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Doctor of Philosophy in Biomedical Sciences, concentration in Biochemistry

---

by

Md Imdadul Haq

May 2020

Aruna Kilaru, Ph.D., Chair

Cecilia A. McIntosh, Ph.D.

Christopher L. Pritchett, Ph.D.

Douglas Thewke, Ph.D.

Sona Pandey, Ph.D.

Keywords: FAAH, Anandamide, mRNA/lncRNA, Endocannabinoid Signaling

ABSTRACT

Understanding the Implications of Anandamide, an Endocannabinoid in an Early Land Plant, *Physcomitrella patens*

by

Md Imdadul Haq

Endocannabinoid signaling is well studied in mammals and known to be involved in numerous pathological and physiological processes. Fatty acid amide hydrolase (FAAH) terminates endocannabinoid signaling in mammals. In *Physcomitrella patens*, we identified nine orthologs of FAAH (PpFAAH1 to PpFAAH9) with the characteristic catalytic triad and amidase signature sequence. Kinetics of PpFAAH1 showed specificity towards anandamide (AEA) at 37°C and pH 8.0. Further biophysical and bioinformatic analyses revealed that, structurally, PpFAAH1 to PpFAAH4 were closely associated to the plant FAAH whereas PpFAAH6 to PpFAAH9 were more closely associated to the animal FAAH. A substrate entry gate or 'dynamic paddle' in FAAH is fully formed in vertebrates but absent or not fully developed in non-vertebrates and plants. *In planta* analysis revealed that *PpFAAH* responded differently with saturated and unsaturated *N*-acylethanolamines (NAEs). *In vivo* amidohydrolase activity showed specificity associated with developmental stages. Additionally, overexpression of *PpFAAH1* indicated the need for NAEs in developmental transition. To understand and identify key molecules related to endocannabinoid signaling in *P. patens*, we used high-throughput RNA sequencing. We analyzed temporal expression of mRNA and long non-coding RNA (lncRNA) in response not only to exogenous anandamide but also its precursor arachidonic acid and abscisic acid (ABA, a stress hormone). From the 40 RNA-seq libraries generated, we identified 4244 novel lncRNAs. The highest number of differentially expressed genes (DEGs) for both mRNA and lncRNA were detected on short-term exposure (1 h) to AEA. Furthermore, gene ontology enrichment analysis showed that 17 genes related to activation of the G protein-coupled receptor signaling pathway were highly expressed along with a number of genes associated with organelle relocation and localization. We identified key signaling components of AEA that showed significant difference when compared with ABA. This study provides a fundamental understanding of novel

endocannabinoid signaling in early land plants and a future direction to elucidate its functional role.

DEDICATION

To my late grandfather.

ACKNOWLEDGMENTS

First, I would like to thank my parents, especially my mom; without her, I would not be doing what I am doing in life. Even though she does not have the slightest clue of what I am doing in school for so many years, she is always proud of me. A very special acknowledgment goes to my Ph.D. advisor/mentor Dr. Kilaru for her endless support and encouragement from the very beginning to this point; you allowed me to grow as an independent thinker, thank you for that.

Thank you to all my committee members for your guidance, support, and critical suggestions during my graduate school journey. Because you all have helped me to learn and grow into the scientific world, now I consider myself as a better science student and I will be forever grateful.

I would also like to thank my friends back home in Bangladesh and all my friends I made here at ETSU. My friends were always my source of energy and will be forever. I would like to thank my current and previous lab members, who also became my good friends.

I want to thank ETSU for allowing me to pursue my graduate studies. Also, thanks go to ETSU staff, especially current and previous staff in the Biology Department; you always welcomed me with open arms. Maria, Amy, Julie and Wendy are four famous names that I will remember forever.

I am grateful to all the funding (NSF-IOS1456917 and ETSU-RDC award to Dr. Kilaru and Sigma XI GIAR award to me) sources that supported my research work. I would also like to thank the ETSU Office of Equity and Inclusion and Graduate Professional Student Association, the National Institute of Child Health & Human Department, and the National Center for Genomic Resources for their funding in support to attend conferences and certification courses.

I would like to thank Dr. Kumar for letting me use his lab and providing vectors that I used for protein expression. I also thank Drs. Estelle, Chapman and Bezanilla Lab's for providing vectors and clones that I needed in my work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

Mosses have been studied as an experimental organism for over 80 years, and for developmental genetics for almost half a century (Cove et al. 2006). The moss, *Physcomitrella patens* was first introduced as a model organism for developmental genetics in 1968 (Cove 2005). The dominant haploid gametophyte stage in the development of mosses attracted many scientists to use it as a model organism; mutant analyses and genetics studies in haploids are simpler compared with dominant diploid phase in vascular plants (Cove 2005). Moss is also a less complicated candidate for gene targeting and allele modification because of its efficient transformation through homologous recombination, where DNA containing a genomic sequence of interest and corresponding homologous genomic sequence are involved in genetic recombination at a high frequency (Kammerer and Cove 1996). Additionally, mosses, due to their unique position in the evolution of land plants, offer an advantageous system for gaining insight into the functional aspects of complex signaling processes, their evolution, and diversity (Cove et al. 2006). With the availability of a sequenced genome, *P. patens* is an excellent model system for genetic and biochemical investigations. In this study, *P. patens* is a choice model system because of its three unique characters: 1) dominant haploid phase with rapid life cycle and simple morphology, 2) high abiotic stress tolerance (survives up to 92% of water loss) and 3) unique lipid composition (Ruibal et al. 2013; Tan et al. 2017).

Physcomitrella patens and its life cycle

The Gransden strain of *P. patens* has been the focus for genetics study over past two decades and was chosen as one of the first non-flowering plants for genomic sequencing (Cove 2005). Like ferns or seed plants, *P. patens* shows alternate haploid and diploid phases in life cycle, but unlike others its haploid is dominant. The haploid phase starts with the germination of spores that produce protonema (Cove 2005). Protonemal apical cells divide into filamentous chloronemal cells, which are densely packed with large well-developed chloroplasts. Some apical cells of chloronema generate a second type of protonemal filament called caulonema, that contains less developed chloroplasts. Interestingly, some subapical cells of caulonema divide into chloronema cells and few of them develop into either same caulonema cells or gametophores, which initiate the next phase of their life cycle. Gametangia develop from gametophores. Male and female gametes are produced on the same shoot in the antheridia and

archegonia, respectively. Moist conditions help to develop male and female gametes for self-fertilization. A lower temperature, below 18 ºC, is essential to induce gametogenesis and short-day length is essential to increase gametangia. Therefore, late summer produces sporophytes and spores are produced during overwintering. In a laboratory setting, the total life cycle of the Gransden strain is between 3 and 4 months (Cove 2005).

Stress studies in P. patens

Mosses are an attractive plant model organism to study stress because of their evolutionary position and simple cellular structure. Mosses, including *P. patens,* are bryophytes which, separated from their ancestors at least 500 million years ago during the transition from aquatic to terrestrial habitat (Heckman et al. 2001). Because of this transition, early land plants had to confront with the uncertainty of water supply, radiation, and extreme temperatures that led to anatomical, physiological, and biochemical changes and adaptation. Because of their adaptative features, a significant number of abiotic stress studies have been done on *P. patens* in the last decade, including proteomics and RNA-seq analysis on drought, cold, osmotic, and salinity stress tolerance (Wang and He 2009). A proteomic study on cold stress showed a decrease in photosynthetic proteins but increase in catabolic proteins (Wang and He 2009). *P. patens* can tolerate a high concentration of NaCl (350 mM), which up-regulated proteins involved in defense, protein folding, and ionic homeostasis (Wang and He 2009). High salinity also changed the expression of proteins associated with protein synthesis and degradation, metabolism, transportation, cell growth and development, transposons, and signal transduction (Wang et al. 2008). Upon ultra violet-B (UV-B) radiation, approximately 400 genes altered their expression pattern, which revealed distinct and conserved pathways compared with the seed plant *Arabidopsis* (Wolf et al. 2010). *Physcomitrella patens* can also tolerate very high osmotic pressure and dehydration; it survives up to 500 mM sorbitol and 92% water loss by fresh weight (Frank et al. 2005). Based on the molecular characterization of *P. patens* stress responses, it has been proposed that signaling pathways for abiotic stress tolerance may have been changed during their evolution as land plants (Kroemer et al. 2004). Thus, it is pertinent to understand stress tolerance mechanisms in mosses.

Unique lipids in P. patens

Physcomitrella patens possesses a significantly different lipid profile compared with higher plants (Kaewsuwan et al. 2006). Long-chain polyunsaturated fatty acids such as arachidonic acid (AA) and eicosapentaenoic acid are present in mosses or lower organisms unlike in gymnosperms or higher plants where they are absent (Kaewsuwan et al. 2006; Gachet et al. 2017). The presence of unique lipid molecules in these early land plants increases their ability in membrane remodeling and activation of signaling pathways in response to stressors. Arachidonic acid is a precursor of an *N*-acylethanolamine (NAE) family member, anandamide or *N*-arachidonoylethanolamine (AEA, NAE 20:4). Anandamide is a signaling molecule in the endocannabinoid pathway. Gachet and colleagues have recently reported that the presence of AA and AEA is limited to lower plants, but no mechanistic insight was proposed for them (Gachet et al. 2017; Sante 2014). These distinct molecules have profound physiological implications in humans but their function in lower plants is unclear (Kaewsuwan et al. 2006; Horrocks and Yeo 1999; Uauy et al. 2000).

N-acylethanolamines and its biosynthesis

NAEs are a class of lipids with ethanolamine as the head group of fatty acids; the acyl chain can be 12 to 20 carbon long with saturation, monounsaturation, or polyunsaturation. NAEs exist in a wide range of organisms, including early eukaryotes, such as *Chara vulgaris* and *Caenorhabditis elegans* (Gachet et al. 2017; Lucanic et al. 2011), to higher organisms (human, *Homo sapiens*) (Schmid 2000). In 1965, Bachur and his colleagues first discovered NAEs in mammalian tissue during the investigation of ethanolamine metabolism in rat liver microsomes (Bachur et al. 1965) NAEs in mammals are synthesized "on demand" and are maintained by a tight regulation of formative and degradative pathways (Ueda et al. 2010a). In mammalian systems, pharmacological and neurological aspects of NAEs are well studied. The role of NAEs as signaling molecules, however, remained unclear until the finding of anandamide as an endogenous ligand for the cannabinoid binding (CB) receptors 1 and 2 (Devane et al. 1992). However, most of the endogenous NAEs do not bind to the CB receptors. Alternatively, *N*-oleoylethanolamine (OEA) binds peroxisome proliferator-activated receptor α (Fu et al. 2003) and *N*-palmitoylethanolamine (PAE) binds to an unidentified CB2-like receptor (Calignano et al. 1998).

NAEs are synthesized from a minor membrane lipid, *N*-acylphosphatidylethanolamine (NAPE) (Ueda, Tsuboi, and Uyama 2010a). In mammals, NAPE synthesis is catalyzed by a Ca^{2+} dependent or independent *N*-acyltransferase (NAT). In the Ca-dependent pathway, an increase in intracellular Ca^{2+} leads to activation of Ca-NAT, which results in the transfer of a fatty acid chain from a glycerophospholipid, such as phosphatidylcholine (PC) to the amino group of phosphatidylethanolamines (PE) to produce NAPE (Ueda et al. 2010a). In a Ca-independent pathway, HRASLS-5 (Human Ras-like suppressor 5) functions as NAT, where PC and PE work as an acyl donor and acceptor, respectively to synthesize NAPE (Jin et al. 2007). In plants, NAPE synthesis occurs by *N*-acylation of PE using free fatty acid as acyl donor, which is catalyzed by NAPE synthase. However, *in vitro* enzyme assays using *Arabidopsis* NAPE synthase showed preference for acyl-CoA over free fatty acid for NAPE synthesis (Faure et al. 2009). NAPE is then hydrolyzed to produce NAE and phosphatidic acid (PA) by two different pathways- classic and alternative pathways. In the classic pathway, NAPE-PLD, a membrane-associated enzyme hydrolyzes NAPE to NAE and PA (Okamoto et al. 2004). In the alternative pathway, *N*-acyl-lyso-PE is formed from *O*-deacylation of NAPE, which is then hydrolyzed by lysophospholipase D (lyso-PLD) to produce NAEs (Natarajan et al. 1984; Sun et al. 2004). These NAE synthesis pathways are not well understood in plants and a NAPE-specific PLD remains to be discovered.

The biosynthesis of NAEs is species- and tissue-dependent; both content and composition vary with physiological conditions in animals and plants (Ueda et al. 2010b; Chapman 1998; Blancaflor et al. 2014; Venables et al. 2005). In plants, NAEs represent a very small fraction of total lipid (Chapman 1998; Kilaru and Chapman 2012). Among tissues, desiccated seeds contain higher NAE levels in various plant species; however, the concentrations are very low, ranging from 0.5 $\mu g\ g^{-1}$ to 35.0 $\mu g\ g^{-1}$ fresh weight (Kilaru et al. 2007). In seeds, NAE content and type is dependent on the NAPE precursor (Kilaru and Chapman 2012). Additionally, NAE levels decrease during germination and play a fundamental role in seedling development by interacting with the abscisic acid (ABA) signaling pathway (Chapman 1998; Kilaru and Chapman 2012; Teaster et al. 2007; Keereetaweep et al. 2013). However, in vegetative tissues, NAE composition is different and the content is lower than that of seeds (Wang et al. 2006; Kang et al. 2008). The most abundant NAEs of higher plants are NAE 18:2, NAE 18:1 and NAE 16:0, however, NAE

12:0, NAE 14:0, NAE 18:0 and NAE 18:3 are also present at very low levels (Venables et al. 2005; Kilaru and Chapman 2012; Chapman et al. 1999; Zoerner et al. 2011).

Saturated, monounsaturated, and polyunsaturated NAEs (PU-NAEs) are hydrolyzed into corresponding fatty acids by membrane-associated enzymes (Kilaru and Chapman 2012). NAE hydrolytic enzymes are reported to brealdown anandamide to arachidonic acid, and oleamide to oleic acid in neuroblastoma cells and central nervous system (Wang et al. 2006). These enzymes also showed similar properties, including membrane association, tissue distribution and sensitivity to inhibitors of serine and cysteine hydrolases (Maurelli et al. 1995). Among these enzymes, fatty acid amide hydrolase (FAAH) can hydrolyze anandamide to arachidonic acid and ethanolamine.

On the other hand, PU-NAEs can also be subjected to oxygenation of polyunsaturated acyl moieties. The enzymes that are responsible for oxygenation are cyclooxygenase-2 (COX-2), lipoxygenases (LOXs) and cytochrome P-450 (McKinney and Cravatt 2005; Yu et al. 1997; Bornheim et al. 1993). COX-2, LOX and cytochrome P-450 oxygenate anandamide to prostaglandin like ethanolamide, hydroperoxy-eicosatetraenoylethanolamide, and hydroxyl-eicosatetraenoylethanolamide, respectively (Yu et al. 1997; Bornheim et al. 1993; Rahman et al. 2014). While all the three oxygenase enzymes are characterized in animals only LOXs are characterized in plants (Keereetaweep et al. 2013). LOXs are classified based on oxygenation location on the carbon atom of the hydrocarbon backbone of PU-NAEs. For instance, 9-LOX or 13-LOX oxygenate the 9^{th} or 13^{th} carbon atom of hydrocarbon backbone of linoleic or linolenic acid to produce the (9S)- and (13S)-hydroperoxy derivatives of these fatty acids, respectively. In Arabidopsis, AtLOX1 to AtLOX6 have been identified and characterized. Oxidative products of polyunsaturated fatty acids are together referred to as oxylipins. Kilaru et al., also reported that the endogenous level of NAE-oxylipin synthesis increased in FAAH knock outs (KO), relative to the wild-type and overexpressors (OE), which suggests that FAAH and LOXs compete to metabolize PU-NAEs (Kilaru et al. 2011). Moreover, *N*-linolenoylethanolamine (NAE 18:3) oxylipins (9- or 13-hydroxides) bleached the cotyledons and inhibited seedling development (Kilaru et al. 2011). Other studies showed that both NAE 18:2 and NAE 18:3 may be oxidized by LOX during seedling growth; the LOX activity, however, can be inhibited by NAE 12:0 (Keereetaweep et al. 2013).

Fatty acid amide hydrolase (FAAH)

FAAH is an important catalytic enzyme in eukaryotes and a member of a large and diverse enzyme family referred to as amidase signature (AS) family. The AS sequence is approximately 160 amino acid stretch and is highly conserved among FAAH proteins. In eukaryotes, FAAH is involved primarily in the hydrolysis of NAEs (McKinney and Cravatt 2005; Cravatt et al. 1995). FAAH hydrolyze anandamide to ethanolamine and arachidonic acid (Fig. 1.1).

Mammalian FAAH

Human, rat and mouse FAAH were first reported by Giang and Cravatt in 1997 (Giang et al. 1997). FAAH in mammals was reported as an integral membrane enzyme and remains insoluble (Thomas et al. 1997). FAAH is distributed throughout the central nervous system (CNS) of rats, and a progressive accumulation of mRNA of *FAAH* occurs from embryonic stage to postnatal day 30 and decreases in adulthood (Thomas et al. 1997). FAAH hydrolyzes neuromodulatory fatty acid ethanolamides such as anandamide and amides like oleamide, and thus influences sleep, analgesia, and euphoria (Thomas et al. 1997). FAAH also hydrolyzes a variety of other saturated and unsaturated amides such as palmitamide, palmitoleamide, 6Z-octadecanamide, linoleamide, and arachidonamide. Among all NAEs and fatty acid amides, FAAH showed a relatively high rate of hydrolysis on AEA and a low rate on linoelaidamide (Boger et al. 2000).

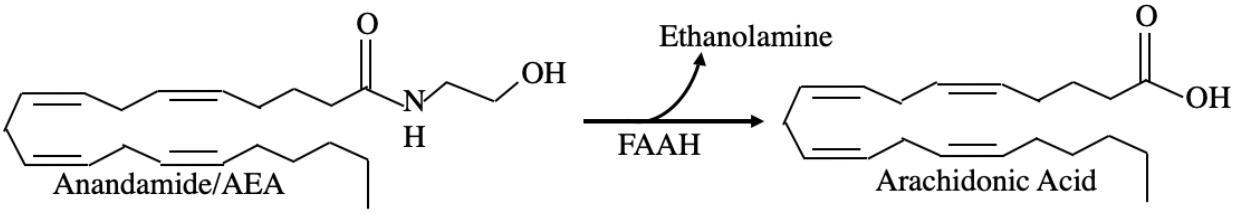


Figure 1.1. Hydrolysis of anandamide into arachidonic acid and ethanolamine

Rat and mouse FAAH showed 91% identity, while human FAAH showed 84% and 81% identity with mouse and rat FAAH, respectively (Giang et al. 1997). Functional analysis of catalytic residues of rat FAAH showed an unusual Ser-Ser-Lys catalytic triad that forms a tetrahedral intermediate through which Ser217 attacks the carbonyl group of the substrate. In the tetrahedral intermediate, Lys142 is a general acid-general base that mediates the deprotonation of

the Ser241 and Ser217 and works as a shuttle for the protonation of the leaving group. The reaction ends when the enzyme-bound acyl intermediate goes through water-mediated deacylation and releases the free fatty acid (Patricelli and Cravatt 1999). FAAH can also act as a bioactive amidase and esterase competitively *in vivo*; a single mutation of a lysine residue to alanine (K142A) in rat FAAH increased esterase hydrolysis over amide hydrolysis by 500-fold (Patricelli and Cravatt 1999). FAAH has a strong preference for hydrophobic substrates (Patricelli and Cravatt 1999) and hydrolyzes all three classes of amidated signaling lipids, NAEs (Devane et al. 1992; Lambert et al. 2002), fatty acid primary amides (oleamide) (Cravatt et al. 1995), and *N*-acyl taurines (Saghatelian et al. 2006).

FAAH has two homologs in humans, FAAH-1, and FAAH-2. FAAH-1 showed preference for polyunsaturated over monounsaturated acyl chains, while FAAH-2 showed the opposite selectivity (Wei et al. 2006). Tissue-specific expression of FAAH-1 (brain, kidney, liver, small intestine, lung, prostate, testis) and FAAH-2 (heart, kidney, liver, lung, prostate, ovary) suggests their importance in fatty acid amide catabolism (Wei et al. 2006). The Ser-Ser-Lys tetrahedral intermediate has been the main target for inhibitor binding sites because of its potential for treatment of anxiety and pain. Inhibition of FAAH increased anandamide production and its oxygenated metabolites (Long et al. 2009). Inhibition of FAAH also lead to overstimulation of CB1, transient receptor potential of vanilloid 1 ion channel and *N*-methyl-D aspartate receptors, which could result in neurotoxicity (Long et al. 2009; Hampson et al. 1998; Stelt et al. 2005). Similarly, FAAH knockout (KO) in mice showed less degradation of fatty acid ethanolamides, specially anandamide and increased the risk of CB1-dependent behavioral responses (B F Cravatt et al. 2001). FAAH KO mice also showed increased endogenous fatty acid amides in nervous system and analgesic, anti-inflammatory, and anxiolytic related peripheral tissues (Lichtman et al. 2002; Kathuria et al. 2003; Cravatt et al. 2004).

FAAH in plants

In plants, FAAH activity was first shown *in vivo* and *in vitro* in *Gossypium hirsutum* (Shrestha et al. 2002). Since then, FAAH studies were mostly limited to *Arabidopsis* where *At*FAAH was identified as a mammalian FAAH ortholog and characterized (Shrestha et al. 2003). *At*FAAH possess 37% residual identity to ratFAAH within the AS region (Shrestha et al. 2003; Haq and Kilaru 2020). The highly conserved amidase region is rich in serine, glycine, and

alanine residues (Patricelli and Cravatt 1999; Neu et al. 2007; Labahn et al. 2002). Homologs of FAAH have also been identified and analyzed in tomato, grape, rice, wheat, and medicago; all identified FAAH have the conserved amidase sequence with catalytic residues (Wang et al. 2006).

In higher plants, FAAH modulates seed germination, seedling growth, and flowering by playing a role in catabolism of NAEs (Shrestha et al. 2003). In Arabidopsis, low concentrations of NAE 12:0 reduced seedling growth in a dose-dependent manner and altered root cell and cytoskeletal organization (Blancaflor et al. 2003; Motes et al. 2005). The *AtFAAH* KO did not show any strong phenotype but exhibited severe effects on root growth and seedling germination in the presence of external NAE 12:0 (Teaster et al. 2007; Wang et al. 2006). On the other hand, OE of *At*FAAH tolerated exogenous NAE and enhanced seedling growth and flowered early compared to wild type (Wang et al. 2006). Early flowering, both during the inductive long day and non-inductive short day in *At*FAAH OEs elevated the expression of *FLOWERING LOCUS T (FT)* gene (Wang et al. 2006). Furthermore, *At*FAAH OEs grown under short day conditions produced about 9% less total NAEs; specifically, NAE 12:0 and NAE 18:2 were reduced by 30% compared to wild type (Teaster et al. 2012). However, *At*FAAH OEs showed increased sensitivity to biotic and abiotic stressors and ABA suggesting an interaction between ABA and NAE signaling pathways (Teaster et al. 2007; Cotter et al. 2011; Kang et al. 2008) Overall, these implications of increased or decreased levels of NAEs in various tissues where *AtFAAH* was either knocked out or overexpressed, respectively, indicate the vital role of FAAH in regulating NAE metabolism (Wang et al. 2006; Teaster et al. 2012).

Interacting proteins in biological processes

Several major biological or cellular processes involve protein-protein interactions (PPI). Processes such as environmental sensing, signal transduction, maintenance of cellular organization, replication, transcription, translation, cell cycle control, regulating metabolic and signaling enzymes include transient PPI. These protein complexes of PPI are distinguished as homo- and hetero-dimeric proteins, antibody-protein complexes and enzyme-inhibitor complexes (Jones and Thornton 1996). Studies on wild type (WT) and predicted N-terminal transmembrane (TM; amino acid 9-29) of rat FAAH suggested that FAAH holds multiple modes of membrane integration (Patricelli et al. 1998). WT-FAAH comprises a heterogeneous mixture of species

with sedimentation coefficients between 15S and 28S, whereas ΔTM-FAAH exists as a single detergent-protein complex; this suggests that TM of FAAH has membrane interacting partners (Bracey et al. 2002). Additionally, when WT-FAAH was expressed in a COS-7 cell line, immunoblot results showed a band corresponding to dimeric and higher-molecular weight forms, possibly trimeric or tetrameric, regardless of whether or not WT-FAAH cell extracts were subjected to boiling and reducing agents. The ΔTM-FAAH cell extract did not the show same immunoblot band pattern which suggests that FAAH interacts with itself or other proteins (Patricelli et al. 1998). In Arabidopsis, AtFAAH was shown to physically interact with HNH endonuclease domain-containing protein (HNH; At3g47490) and RSZ33, Arg/Ser-rich zinc knuckle-containing protein 33 (At2g37340). Interaction of AtFAAH with RSZ33 suggests an involvement of AtFAAH in ABA signaling pathway. A homozygous *RSZ33* knock out plant showed hypersensitivity to growth on ABA (Kim 2010). A mutagenesis study of AtFAAH showed three consecutive arginine residues (Arg-491, Arg-492, and Arg-493) were responsible for AtFAAH protein interaction (Kim et al. 2009).

Anandamide and endocannabinoid system

Anandamide was first isolated from porcine brain and reported as the first endogenous ligand of central CB receptors (Ameri et al. 1999). Endogenous AEA is produced "on demand" and induces similar pharmacological actions as THC upon interaction with CB receptors[76]. For example, AEA controls the neuronal excitability by reducing excitatory neurotransmission at a presynaptic site (Ameri et al. 1999). Anandamide supplementation in the diet increased other bioactive NAEs in brain homogenate, which suggests that long-chain polyunsaturated fatty acids play a role in brain biochemistry (Berger et al. 2001). In rodents, AEA provokes series of behavioral responses such as hypomotility, hypothermia, analgesia, and catalepsy (Crawley et al. 1993).

Anandamide takes part in signal transduction by binding to endocannabinoid receptors (CB1 and CB2), which are G-protein coupled receptor (GPCR) proteins. In mammalian systems, CB1 and CB2 expression is tissue-specific; CB1 receptors are predominantly found in central and peripheral nervous systems; they are however, also expressed in non-neuronal cells and tissues, and play role in immune response, reproduction and pituitary gland (Howlett 2005). CB2 receptors are primarily found in immune cells and tissues. Mechanistically, the binding of AEA

to CB1 or CB2 receptors reduces cyclic AMP, which activates p42/p44 mitogen-activated protein kinase (MAPK), p38 MAPK and Jun N-terminal kinase through the coupled $G_{i/o}$ proteins to regulate nuclear transcription factors (Howlett 2005). Several endogenous agonists in addition to AEA have also been identified thus far; among them, 2-arachidonoylglycerol (2-AG), and 2-arachidonylglyceryl ether (2-AGE) are most notable (Howlett 2005). While AEA is a well-established agonist for CB1, 2-AG is for CB2 (Marzo et al. 1998). Even though CB1 and CB2 have distinct expression patterns, tissue location, and function, they exhibit 48% amino acid sequence identity (Pertwee 2009).

In recent years, plant-derived natural products other than Δ^9-THC that bind to CB receptors have been identified in other plant species (Howlett 2005). Fatty acid derivatives such as *N*-alkylamides, *N*-isobutylamides prefer to bind to CB2 receptor over CB1 (Rahman et al. 2014). A polyacetylenic polyene, falcarinol found in the *Apiaceae* family showed significant interactions with both CB1 and CB2 receptors (Leonti et al. 2010). However, the interaction of NAEs with CB receptors in plants has not been demonstrated (Kilaru et al. 2007; Marzo et al. 1998).

Significance of big data analysis

The detection and quantification of RNA using transcriptome analysis became a dominating field of research in both plant and animal model organisms, especially with the improved and sophisticated methods of the high-throughput cDNA sequencing (RNA-seq) (Zhang et al. 2010). Initially, RNA-seq analysis was limited to most studied model organisms, such as Arabidopsis and/or important crop plants such as maize and rice (Zhang et al. 2010; Lister et al. 2008; Eveland et al. 2010). However, recent advancement on deep sequencing and *de novo* transcriptome assembly opened up the possibility for genomic studies on any organism. In recent years, hundreds of plant organisms were sequenced. Most notably, the project to generate one thousand plant transcriptomes and the phylogenomics of green plants was recently completed and genome information for 1124 plants including green and red algae, glaucophytes, and Archaeplastida was published ("One Thousand Plant Transcriptomes and the Phylogenomics of Green Plants" 2019).

Several transcriptomic studies have been carried out with *P. patens*, as a non-seed model organism to study evolution of mechanisms associated with transition from water to land (Xiao

et al. 2012). Studies to understand the developmental transitions, such as from protoplasts to protonema to gametophores to sporophyte identified key regulatory genes for each transition (Xiao et al. 2012, 2011; Perroud et al. 2018; O'Donoghue et al. 2013). RNA-seq analysis was also performed under conditions of drought, cold, salt, osmotic, and UV-B. The number of differentially expressed genes (DEGs) were identified for each stress condition; there were 3220 and 400 DEGs in response to cold and UV-B stress, respectively (Frank et al. 2005; Beike et al. 2015). With a more targeted approach using microarrays, 130, 10, and 8 genes were identified with dehydration, osmotic and salt stress, respectively (Cuming et al. 2007). In contrast, an RNA-seq study showed the number of DEGs in dehydration and osmotic stresses were 1138 and 789, respectively (Stevenson et al. 2016). The transcriptome in response to ABA treatment in *P. patens* revealed 1073 DEGs (Stevenson et al. 2016). A recent study showed that exogenous ABA enable survival in unfavorable environmental conditions by forming vegetative organs called diaspores or brood cells, in which 1030 DEGs were identified (Arif et al. 2019).

Along with the transcriptome of mRNA, non-coding RNAs (ncRNAs) are receiving attention because of the recent findings of their involvement in responses to various physiological conditions including biotic and abiotic stresses (Arif et al. 2019). Among the ncRNAs, long noncoding RNAs (lncRNAs) that are longer than 200nt constitute a huge percentage of total genome. The research area of lncRNA is new and emerging for both plants and animals. In mammals, lncRNAs were reported to be involved in breast, prostate and pancreatic cancers, diabetes and Alzheimer disease (Zhang et al. 2019; Mitobe et al. 2018; Gao et al. 2019; Li et al. 2019; Tan et al. 2013). In plants, hundreds of novel lncRNAs were identified to be associated with cold, heat, drought and salt stress (Yuan et al. 2018; Zhao et al. 2018). Advanced RNA-seq technology and analytical tools increased the ability to detect and identify lncRNA and their function in plants and animals (Zhao et al. 2018). In Arabidopsis, 6480 lncRNAs were identified from 200 transcriptome data sets, which were tissue specific and/or stress associated (Liu et al. 2012). In rice, over 7000 lncRNA related to cold, heat, drought and salt stresses were identified (Yuan et al. 2018). A recent study in *P. patens* suggested that lncRNA may bind to mRNA or pre-mRNA complementary binding sites and regulate gene expression (Fesenko et al. 2017). Overall understanding of the role of lncRNA is still preliminary; extensive studies are necessary to determine the correlation between mRNA and non-coding RNAs and their role in physiological conditions.

In this study, we aimed to identify the role of anandamide, an endocannabinoid that is exclusive to bryophytes and absent in vascular plants. We focused on identification and characterization of the metabolic enzymes and discovery of the global molecular response to anandamide along with ABA. We used biochemical, biophysical, bioinformatic and molecular approaches to recognize the functional role of FAAH and anandamide and the key regulatory components in the endocannabinoid system of *P. patens*.

Rationale, hypothesis and specific aims

Physcomitrella patens is a well-known model organism for its stress tolerance ability. Recent discovery of endogenous cannabinoids, especially AEA in bryophytes and lower organisms led to hypothesize the involvement of endocannabinoids in stress tolerance. The main objective of this research is to elucidate the functional role of AEA in *P. patens* by identifying its metabolic enzymes, signaling components, and/or involvement in stress-mediated responses.

To identify the functional role of AEA, we hypothesized: 1) A functional catabolic enzyme of AEA, fatty acid amide hydrolase (FAAH) is present in *P. patens*; 2) AEA has a unique global molecular response compared to known signaling molecules, such as ABA or AA; and 3) FAAH plays a role in endocannabinoid signaling by interacting with other proteins. In order to test the hypotheses and determine the role of AEA in development and stress response in *P. patens*, the following specific aims were pursued proposed.

Specific Aim 1: Identify the anandamide catabolic enzyme, FAAH in *P. patens*

In silico analysis was performed to identify FAAH orthologs in *P. patens* using AtFAAH, rtFAAH and hsFAAH. Identified FAAH was then characterized biochemically. The sub-aims were

- Identification of FAAH orthologs in *P. patens*.
- Biochemical characterization of PpFAAH.
- Analysis of the predicted structure of PpFAAH in relation to other FAAH.

Specific Aim 2: Determine the global molecular response to AEA

Protonema of *P. patens* was treated with exogenous AEA in temporal manner. While ABA and AA treatments served as a positive control, DMSO was the solvent/negative control. Specific sub-aims were

- Analyze the RNA-seq data for global response over time.
- Analyze the gene ontology of DEGs.
- Identify the important signaling molecules.
- Predict the regulatory role for lncRNA.

Specific Aim 3: *In vivo* characterization of anandamide catabolic enzyme

The *in vivo* role of FAAH was determined by the amidohydrolase activity and its interaction with other proteins and by utilization of metabolite mutants, Specific sub-aims were

- Determine *in vivo* amidohydrolase activity in different developmental stages.
- Identify interacting proteins of PpFAAH1.
- Generate and characterize the mutants of PpFAAH1.
- Biochemical characterization of ECS components.

CHAPTER 2. AN ENDOCANNABINOID CATABOLIC ENZYME FAAH AND ITS PARALOGS IN AN EARLY LAND PLANT REVEAL EVOLUTIONARY AND FUNCTIONAL RELATIONSHIP WITH EUKARYOTIC ORTHOLOGS

Imdadul Haq and Aruna Kilaru

Department of Biological Sciences and Biomedical Sciences, East Tennessee State University, Johnson City, TN, 37614, USA

kilaru@etsu.edu

An endocannabinoid catabolic enzyme FAAH and its paralogs in an early land plant reveal evolutionary and functional relationship with eukaryotic orthologs

Abstract

Endocannabinoids were known to exist only among Animalia but recent report of their occurrence in early land plants prompted us to study its function and metabolism. In mammals, anandamide, as an endocannabinoid ligand, mediates several neurological and physiological processes, which are terminated by fatty acid amide hydrolase (FAAH). We identified nine orthologs of FAAH in the moss *Physcomitrella patens* (PpFAAH1 to PpFAAH9) with amidase signature and catalytic triad. The optimal amidase activity for PpFAAH1 was at 37ºC and pH 8.0, with higher specificity to anandamide. Further, the phylogeny and predicted structural analyses of the nine paralogs revealed that PpFAAH1 to PpFAAH4 were closely related to plant FAAH while PpFAAH6 to PpFAAH9 were to the rat FAAH, categorized based on the membrane binding cap, membrane access channel and substrate binding pocket. We also identified that a true 'dynamic paddle' that is responsible for tighter regulation of FAAH is recent in vertebrates and absent or not fully emerged in plants and non-vertebrates. These data reveal evolutionary and functional relationship among eukaryotic FAAH orthologs and features that contribute to versatility and tighter regulation of FAAH. Future studies will utilize FAAH mutants of moss to elucidate the role of anandamide in early land plants.

Introduction

Endocannabinoids such as anandamide belong to a class of lipid derivatives referred to as *N*-acylethalomines (NAEs). These fatty acid ethanolamides are found in wide range of eukaryotic organisms such as yeast, *Caenorhabditis elegans*, bivalve mollusk, mammals and also plants[1–4]. The fatty acid chain length of these NAEs can range from 12C to 20C and are either saturated or unsaturated[5]. Among these NAEs, only anandamide or *N*-arachidonylethanolamide (NAE 20:4) is known to serve as a ligand to cannabinoid binding receptors to activate endocannabinoid signaling[6]. In mammals, while endocannabinoids are key participants in neural signaling and other physiological events[7–11], other NAEs such as NAE16:0 serve protective function in a receptor-independent manner[12]. Irrespective of their receptor binding capability, the NAE type, content, composition and functions are highly diverse among organisms, in a tissue-

specific manner. In *C. elegans*, endocannabinoids, including anandamide are essential for mobilization of cholesterol from internal reserves[13] and they affect life span[14]. In human, circulating endocannabinoids are positively correlated to visceral adipose tissue mass[15], while in rat, anandamide induces overeating by activating cannabinoid receptor[16]. In higher plants such as *Arabidopsis*, only 12C to 18C NAEs occur and they mediate growth, development and biotic and abiotic stress responses[17–21]. Interestingly, the fatty acid and NAE composition of early land plants such as bryophytes is distinct from that of higher plants[1]. For example, in the moss *Physcomitrella patens* and other bryophytes there is abundance of arachidonic acid and its derivative anandamide, which are absent in vascular plants[1]. Preliminary studies showed that anandamide content is about 20% of the total NAEs in *P. patens* and at higher concentrations (> 10 μM) is a negative inhibitor of growth. However, there is no clear understanding of why anandamide was absent in land plants, which evolved later than bryophytes. Furthermore, the mechanistic role of anandamide, the identity of a potential receptor and the associated signaling network in *P. patens* remains to be discovered[22].

The enzyme fatty acid amide hydrolase (FAAH) is highly conserved among eukaryotes[23–26], which terminates NAE functions, including anandamide resulting in inactivation of the endocannabinoid signaling[23,26–28]. Thus far, FAAH has been extensively studied in *C. elegans*, mammals and model plant *Arabidopsis*. The distinct feature that separates FAAH from other amidase family proteins is a lysine-serine-serine based catalytic motif within the ~130 amino acids long sequence referred to as amidase signature (AS)[29,30]. While only one FAAH protein and its encoding gene are characterized in rodents, two proteins encoded by *FAAH1* and *FAAH2* were characterized in human and in *Arabidopsis*[31–33]. There is 20% identity between human FAAH1 and FAAH2 proteins and they differ in their tissue-specific expression, substrate specificity, and function; FAAH1 has higher specificity for NAE 20:4 while FAAH2 prefers monounsaturated NAE (NAE 18:1)[29]. The human FAAH1 received greater attention of researchers due to its ability to terminate endocannabinoid signaling, which affects a number of physiological conditions including but not limited to Crohn's disease, obesity, gastrointestinal disorder, cardiovascular disease, depression and apathy symptoms[34–38]. In *Arabidopsis*, FAAH influences seedling growth, root length, stomatal closure and abiotic and biotic stress responses[39–42]. In general, when *FAAH* was knocked-out, irrespective of the organism, NAE levels increased but most often an associated phenotype was not obvious. In contrast,

overexpression of *FAAH* enhanced growth and development of arabidopsis seedlings but compromised their ability to respond to stressors and abscisic acid[39,41,43].

For mechanistic understanding of FAAH, crystal structure of mammalian FAAH with different inhibitors[30,44–48], and recently, a plant FAAH, *At*FAAH have been resolved[26]. These crystal structures aided in understanding the primary, secondary, tertiary and quaternary structures of the protein, mechanism of the catalytic reaction and also the chemistry involved in the entry of the substrate and release of the end product, free fatty acid. The catalytic mechanism of FAAH is unique when compared to other AS family proteins[23,49]. Series of mutagenic studies of mammalian FAAH suggested that Lys142, Ser217 and Ser241 are essential catalytic residues; the crystal structure of rat FAAH (Protein Data Bank (PDB) ID: 1MT5) revealed that these catalytic residues compose a novel catalytic triad in the mammalian FAAH[23,30,50].

The crystal structure of AtFAAH (PDB ID 6DII) opened a new avenue for understanding the details of structural mechanism of plant FAAH in general and in relation to other eukaryotic FAAH in terms of substrate accommodation and catabolism[26]. The sequence identity between AtFAAH and RtFAAH is 34% with identical arrangement of the catalytic triad. Also, both Arabidopsis and mammalian FAAH can dimerize[26,30] and share characteristic features such as the occurrence of a membrane binding cap (MBC), membrane access channel (MAC), acyl binding channel (ABC), cytoplasmic access channel and substrate binding pocket (SBP)[26,30], while their secondary structure formation is different. For instance, in RtFAAH, although a transmembrane (TM) domain is predicted in the N-terminus region, its deletion indicated that FAAH still binds to the membrane through MBC in α18 and α19 helices forming helix turn helix motif (amino acids 404-438). This motif basically interrupts the AS fold and forms the hydrophobic plateau domain facilitating the integration of FAAH into one leaflet of the lipid bilayer[23,30]. On the membrane face, an access port defined by Arg486 and Asp403 facilitates the substrate entry to the active site[30]. In contrast, the N-terminus region of AtFAAH is longer compared to RtFAAH and yet lacks a TM domain and its MBC is located in α1 and α2 helices[26]. Having MBC in a different region in relation to AS did not affect AtFAAH function. The MBC is considered to be important for an easy and direct substrate entry from the hydrophobic membrane bilayer[23]. The MAC is amphipathic and accommodates the entry and movement of polar NAE head groups to the active site of FAAH[23,26,30]. The MAC in RtFAAH coordinates

with ABC by a 'dynamic paddle' composed of Phe432 and Trp531[44,51]. The dynamic paddle is not conserved in AtFAAH and its MAC and ABC are predicted to be a one large channel[23,26,30]. The cytoplasmic access channel on the other hand, is proposed to release the products into cytosol after catabolism of the substrate[23]. The SBP in both AtFAAH and RtFAAH is mostly hydrophobic but in AtFAAH, there are a number of hydrophilic residues making it more polar, which perhaps allows for accommodation of diverse substrates with or without the polar functional group[26]. Additionally, in AtFAAH, a distinct 'squeeze and lock' mechanism was proposed for ligand binding and release, which is absent in RtFAAH[26]. Studies thus far indicate some key similarities and differences between plant and animal FAAH, with regards to their structure, mechanism and substrate specificity.

Knowing the structural details of mammalian FAAH has helped a great deal in generating targeted therapeutics[52–54]. Further understanding of a distant plant FAAH, which might have evolved around 500 million years ago is expected to provide evolutionary and functional insights. Bryophytes are the first group of plants that successfully made transition from water to land and are naturally resilient to various stressors[55]. Interestingly, the lipid composition of these early land plants, including *P. patens* is distinct from that of higher plants[56,57]. Specifically, the identification of anandamide, along with other NAEs and its influence on the development prompted us to gain functional insights into the endocannabinoid catabolism and signaling in the moss. In this study, we not only identified a functional FAAH in *P. patens* but also predicted the relationship of its paralogs with other eukaryotic orthologs.

Results and Discussion

Putative moss FAAH with highest identity to its eukaryotic orthologs is an amidase

To identify potential FAAH candidates with the ability to hydrolyze anandamide and other NAEs in the moss, *P. patens* database (v3.3) in Phytozome 12 was searched for homologs of rat, human and *Arabidopsis* FAAH1. Nine moss proteins with high similarity to RtFAAH and AtFAAH were identified, which were considered putative and based on their order of homology and sequence identity were named chronologically, PpFAAH1 to PpFAAH9 (Table S1). The percentage identity of moss FAAH paralogs with AtFAAH, as generated by pairwise alignment ranged from 26% to 47% while with mammalian FAAH it was 28% to 34%. To obtain a more accurate identity among the sequences, percent identity matrix was generated by multiple

sequence alignment using CLUSTAWL (Table S2). These data show that while PpFAAH1 shared similar identity with PpFAAH2 to PpFAAH5 and AtFAAH, which ranged from 46-44%, its identity with PpFAAH6 to PpFAAH9 was less than 26%. With mammalian FAAH, all the nine PpFAAH paralogs shared 18-25% identity. Among FAAH paralogs, PpFAAH2 to PpFAAH5 shared higher identity with each other than with the remaining PpFAAH; and while PpFAAH6 and PpFAAH7 shared highest similarity (85.6%) with each other, both PpFAAH8 and PpFAAH9 remained relatively distant from all other paralogs with <24% identity (Table S2). Interestingly, despite the differences in identity and the position of the AS region among these putative FAAH paralogs, the number of amino acids residues that make up the AS region remained between 122 to 124. All the nine paralogs not only retained the highly conserved AS sequence but also preserved the lysine-serine-serine catalytic triad (Fig. 2.1A). These nine FAAH paralogs of moss also showed diversity in their phylogenetic relationship to other eukaryotic FAAH (Fig. 2.1B). Together, these data suggest possibility for a shared functional relationship among these paralogs with some variations. First, to determine if these nine proteins are indeed paralogs of FAAH, we carried out biochemical characterization of putative PpFAAH1, including cloning, heterologous expression and purification, and radiolabeled enzyme assays (Fig. S1).

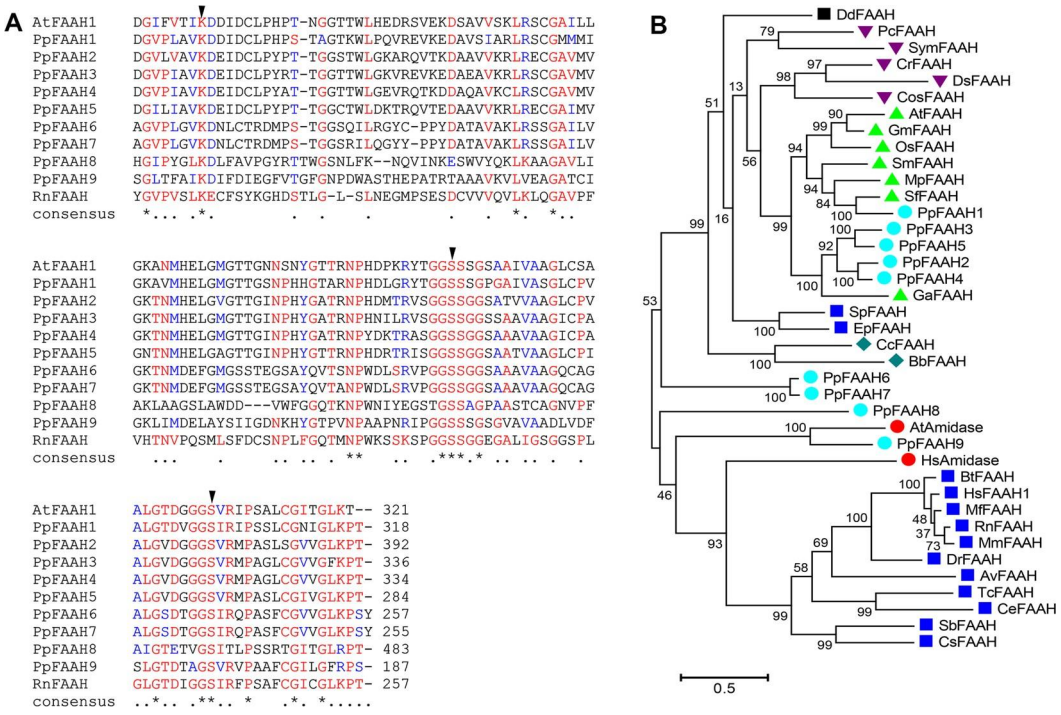


Figure 2.1. Alignment and phylogenetic analysis of FAAH.A). Characteristic amidase signature (AS) of AtFAAH and RtFAAH were compared with nine FAAH candidates (PpFAAH1 to PpFAAH9) of moss. Arrows indicate the conserved catalytic triad of lysine-serine-serine. Numbers at the end of the sequences represents the last amino acid position of the AS. Alignment of full-length sequences is shown in Fig. S2. The symbols: asterisk, dot and gap for the consensus sequence indicate identical, same class and different class of residues at the same position, respectively. Red, green, and black colors also represent the same order of consensus symbols in terms of conserved residues. B). Phylogenetic analysis of PpFAAH in relation to other eukaryotic orthologs (Table S3). Numbers indicate bootstrap values obtained from 500 replicates using the maximum likelihood method. The scale bar represents 0.5 amino acid substitutions per site

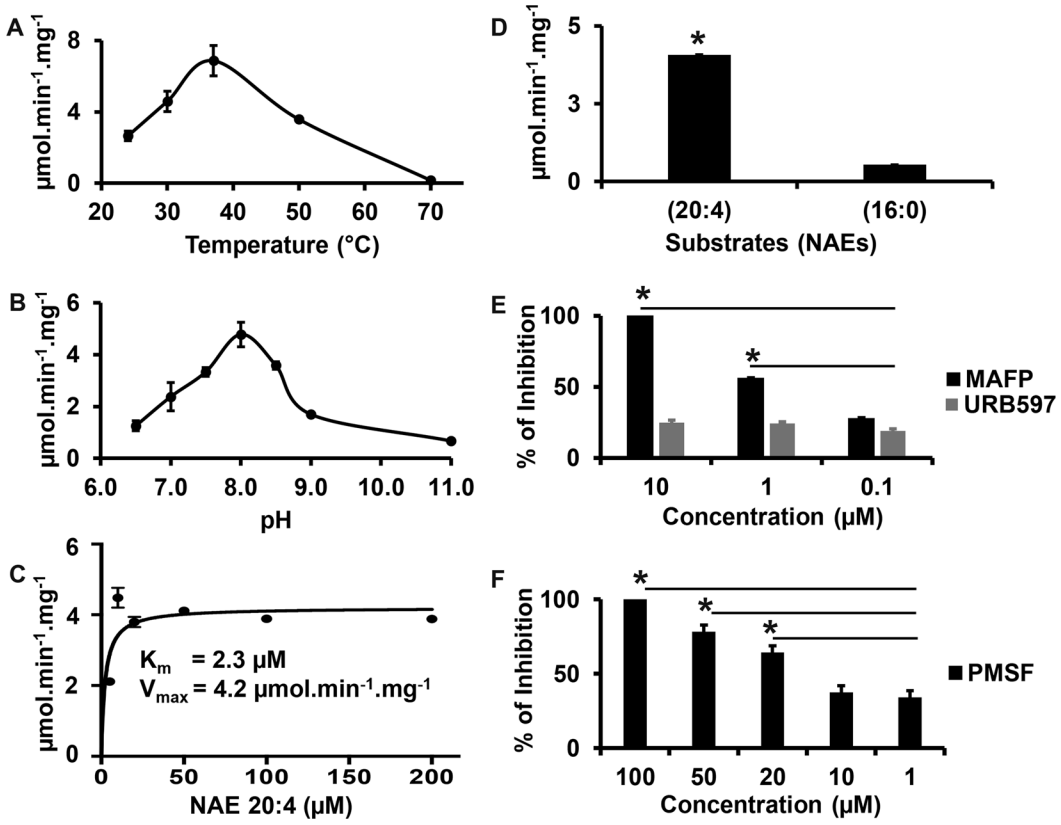


Figure 2.2. Amidohydrolase activity of PpFAAH1. Amidase activity at varying A). temperature and B). pH. C). Saturation kinetics of PpFAAH1 with anandamide (NAE 20:4) as substrate; Km and Vmax values were calculated using Prism GraphPad. D). Activity of PpFAAH1 with substrates anandamide (NAE 20:4; polyunsaturated) and palmitoylethanolamide (NAE 16:0; saturated). E). and F). The percentage of inhibition PpFAAH1 activity against anandamide by inhibitors - methyl arachidonyl fluorophosphonate (MAFP), [3-(3- carbamoylphenyl) phenyl] *N*-cyclohexyl carbamate (URB597) and phenylmethane sulfonyl fluoride (PMSF). Data represents mean values with standard deviation of three biological replicates. Statistical analysis (t-test) was performed using Prism GraphPad 8.0. The asterisk (*) sign and line (-) on top of the bar graphs represent significant difference relative to control without inhibitor (D-F)

Purified PpFAAH1 showed the ability to hydrolyze anandamide (Fig. S1) with an optimal activity at 37°C and pH 8.0 (Fig. 2.2A, B), which was interestingly similar to that of HsFAAH2[58]. In contrast, both mammalian (HsFAAH1, RtFAAH) and Arabidopsis FAAH showed optimal activity at pH 9.0 but the temperature was 37°C[58] and 30°C[24], respectively. Additionally, at least 50% of amidase activity of PpFAAH1 was noted for a wide range of

temperature (~ 25 to 50ºC) and pH (7.0 to 9.0), suggesting its adaptability to varying conditions. Furthermore, PpFAAH1 displayed typical Michaelis-Menten kinetics with a saturation curve (Fig. 2.2C). The kinetic parameters, Km and Vmax for PpFAAH1 with anandamide as a substrate were determined to be 2.3 M and 4.2 μmol.min^{-1} mg^{-1}, respectively (Fig. 2.2C), and with catalytic efficiency (Kcat) of 1.4 S^{-1} and specificity constant (Kcat/Km) of 0.61 μM^{-1}S^{-1}. In case of AtFAAH with NAE 20:4 as substrate, Km and Vmax of AtFAAH1 are 13.6 μM. and 0.29 μmol.min^{-1}.mg^{-1}, respectively with Kcat 0.33 S^{-1} and Kcat/Km of 0.024 μM^{-1}S^{-1} [25]. These data suggest that the catalytic efficiency and specificity of PpFAAH1 towards anandamide was more than four- and 25-fold higher than that of AtFAAH, respectively. Kinetic data for Hs/RtFAAH from various sources differed due the nature of samples and expression systems used and thus was not used for comparison here. Additionally, PpFAAH1 showed higher preference for polyunsaturated substrate (NAE 20:4) relative to a saturated NAE (NAE 16:0); the activity against NAE 16:0 was very low to determine its kinetic parameters (Fig. 2.2D). Interestingly, human FAAH1 also showed preference for polyunsaturated NAEs (NAE 20:4) while FAAH2 preferred monounsaturated NAEs (NAE18:1)[58]. Considering that there are nine potential FAAH candidates in *P. patens*, it would be of significance to identify if all the nine paralogs have amidase activity. Multiple FAAH, functional at different optimal conditions and diverse specificity towards NAEs and/or other substrates would allow for redundancy and flexibility in function in early land plants under varying environmental and physiological conditions.

The conserved catalytic triad in FAAH orthologs is also responsible for PpFAAH1

Specific endocannabinoid signaling inhibitors such as methyl arachidonyl fluorophosphonate (MAFP) and methyl α-linolenyl fluorophosphonate (MLnFP) interact with the catalytic triad of FAAH to inhibit its amidase activity. Among such analogs, PpFAAH1 activity was greatly affected by MAFP, which was potent at a concentration of 10 nM (Fig. 2.2E). Our subsequent *in silico* analyses showed how the polar head group of MAFP or MLnFP could make covalent bonds with the catalytic residues of PpFAAH1 to inhibit the hydrolysis of NAEs or fatty acid primary amides. Another well-known FAAH inhibitor, [3-(3-carbamoylphenyl) phenyl] *N*-cyclohexyl carbamate (URB597), which has a metabiphenylamide leaving group, a cyclohexyl moiety, and a carbamate reactive group showed an IC_{50} of 4.6 nM for mammalian FAAH[59]; in case of PpFAAH1, only 20% inhibition was achieved with 10 nM

concentration (Fig. 2.2E). Structure analysis of URB597 bound Hs/RtFAAH revealed how a water molecule is likely involved with a different active site to inactivate enzyme by substrate diacylation[44]. Low inhibition of PpFAAH1 by URB597 indicates the possibility for different structural properties relative to Hs/RtFAAH. A serine protease, phenylmethanesulfonyl fluoride (PMSF) was also used to test the inhibition of PpFAAH1 as serine contributes to two residues in the catalytic triad. A 100% inhibition was accomplished with 100 μM concentration of PMSF (Fig. 2.2F). Although, similar concentration range of MAFP, PMSF and URB597 completely inhibited the activity of human[45,48] or Arabidopsis FAAH[25], only MAFP and PMSF but not URB597 inactivated PpFAAH1 at a similar concentration. Inhibition by PMSF confirms that PpFAAH1 is also a serine hydrolase; inhibition by an anandamide analog, MAFP indicates the specificity of PpFAAH1 towards NAE 20:4; URB597 perhaps lacks appropriate structural interaction with its three benzene rings to successfully inhibit the activity of PpFAAH1. Nevertheless, these data unequivocally reveal that PpFAAH1 is indeed a hydrolase with ability to catabolize anandamide.

Potential paralogs of moss FAAH1 reveal association with plant and animal FAAH

Considering the early evolution of mosses, relative to mammals or higher plants we attempted to understand the evolutionary relationship of the nine FAAH paralogs with other eukaryotic orthologs. To this extent, putative or known FAAH from 28 different species representing wide range of phyla from the five eukaryotic Kingdoms Protozoa, Chromista, Fungi, Animalia and Plantae were analyzed (Table S3)[60] using Maximum likelihood method. The unrooted phylogenetic tree suggests an independent evolution of various FAAH orthologs, possibly from a common ancestral amidase protein that evolved much earlier to FAAH (Fig. 2.1B). The most ancestral FAAH to be characterized was that of *Dityostelium*, a protozoan and was separated from the FAAH of Chromista, Fungi, Plantae, and two Cnidarians from Animalia, by a common ancestor (Fig. 2.1B). Most of the Animalia FAAH likely diversified from the orthologs of an amidase to form an independent clade. Separation of the cnidarian FAAH with the non-Animalia clade suggests that they evolved earlier and separately from the FAAH in higher phyla of Animalia. Interestingly, PpFAAH6 and PpFAAH7, which reflect duplication, AtAmidase and PpFAAH9, and PpFAAH8 show early and independent divergence from an ancestral protein. Also, not surprisingly, paralogs of PpFAAH1 to 5 were closest to orthologs

from other bryophytes *Marchantia* (Mp) and *Sphagnum* (Sf)), followed by *Selaginella* (Sm), and then late vascular plants such as arabidopsis, rice (Os) and soybean (Gm). It appears that an early duplication of plant clade separated PpFAAH1 with other plant FAAH, which underwent a further duplication that separated the bryophyte and lycophyte FAAH from angiosperms. The second plant clade included PpFAAH2 to PpFAAH5 and FAAH from cotton. While PpFAAH1, which separated from the rest of its paralogs clearly showed an amidase activity, we speculate that the other paralogs perhaps have a related or redundant function. We carried out comprehensive predictive structural analyses to further assess the possible function of PpFAAH1 paralogs, which evolved at different time periods, in termination of endocannabinoid signaling.

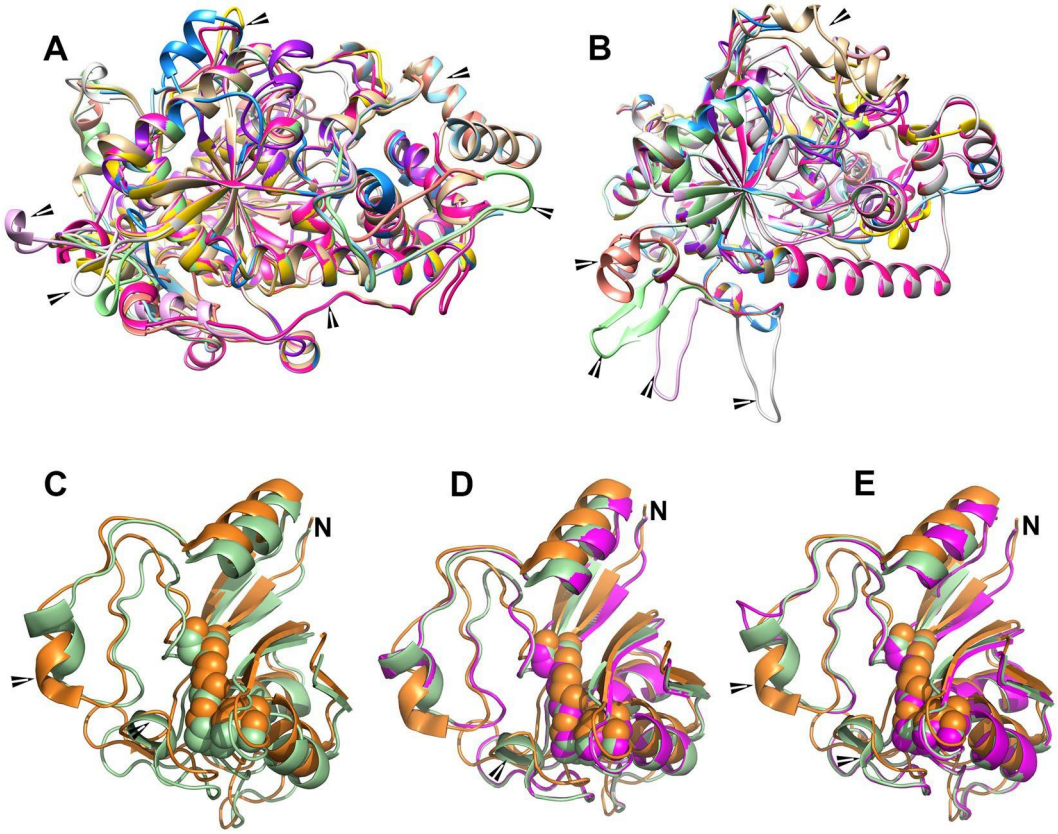


Figure 2.3. Comparison of predicted secondary structures of nine PpFAAH paralogs. An overlay of predicted secondary structures of nine PpFAAH with A). AtFAAH (PDB ID: 6DII) and B). RtFAAH (PDB ID: 1MT5) as templates. Structures were generated using Chimera 5 software. The comparisons predict that the secondary structures with both templates were similar with alpha-helices and loops surrounding the beta sheets core. Arrows point the differences on the aligned structure; Comparison of the amidase signature (AS) regions of C). AtFAAH (green) and

RtFAAH (orange) with arrow pointing to an additional α-helix in AtFAAH, and the two templates compared with D). PpFAAH1 (purple) and E). PpFAAH9. Arrows show the differences in the AS core. PpFAAH9 was predicted to make one less α- helix in the AS region compared to PpFAAH1. Catalytic triad is shown as sphere and N indicates the N-terminus of the AS region. Structural prediction and comparison for PpFAAH2 to PpFAAH8 is presented in Fig. S3

Paralogs of moss FAAH reveal unique and diverse features in relation to arabidopsis and mammalian FAAH

In comparison to AtFAAH and RtFAAH, paralogs of moss FAAH varied in sequence length, identity, predicted protein size (58 to 81 KDa), isoelectric point and number of TM domains (Table S1 and Table S2). Nevertheless, higher identity was noted in the AS region of PpFAAH paralogs, AtFAAH and RtFAAH (Table S1 and Table S2), than in their N- or C-termini (Fig. 2.1A, Fig. S2). Furthermore, in reference to the AS region, the N-terminus is extended in PpFAAH1 to PpFAAH4, like in AtFAAH and shortened in PpFAAH6 and PpFAAH7, as in RtFAAH. Unlike the other FAAH, PpFAAH8 has the longest N-terminus (Table S1, Fig. S2). Thus, the overall identity of N-terminus of PpFAAH1 to PpFAAH4 was higher with AtFAAH and PpFAAH6 and PpFAAH7 with RtFAAH. Both PpFAAH8 and PpFAAH9 showed less similarities because of their extended or shortened N-termini, respectively; PpFAAH5 on the other hand retained the conserved AS region but the remaining sequence lacked significant identity with either AtFAAH or RtFAAH (Fig. S1), and therefore, was not considered for comprehensive structural analysis. Although, the phylogenetic tree indicates PpFAAH5 and PpFAAH3 are likely a result of duplication, it is curious that they depart from each other significantly in predicted structural analyses. Additionally, PpFAAH8 includes fasciclin domain that attaches to the membrane with a lipid link[61], and PpFAAH9 has a tetratricopeptide repeat domain, which serves as mediator of multiprotein complex. These interaction modules regulate diverse physiological processes in various organisms[62], and thus suggests a unique role for PpFAAH8 and PpFAAH9 in moss. Using TMpred[63], PpFAAH3, PpFAAH8 and PpFAAH9 were predicted to contain one TM domain from residues 39-61, 13-32, and 7-25, respectively in the N-terminus region. The probability for TM domain, however, was low at 0.2 for PpFAAH1 and none for the remaining paralogs. In contrast to AtFAAH1[26], although RtFAAH possess a N-terminus TM[30], its deletion did not interfere with the enzyme activity. In case of PpFAAH paralogs, the overall prediction analysis suggests that some share

features of RtFAAH, while the others are more similar to that of AtFAAH. Such separation was also evident from the phylogenetic tree where only a few of the PpFAAH paralogs were closely related to AtFAAH (Fig. 2.1B). The occurrence of a number of potential amide hydrolases in the moss with features similar to both plant and animal FAAH suggests possible variability in terms of their function, substrate specificity and catabolism. Proteins with AS signature are typically involved in hydrolyzing a variety of substrates, in addition to NAEs such as, 2-arachidonoylglycerol, fatty acid primary amides and alkamides[6,23,31,64–66]. Such role for PpFAAH paralogs is expected to affect diverse physiological processes including growth, development, and responses to stress in early land plants[17–21].

Predicted structural variability of PpFAAH suggests evolutionary plasticity

Since protein structure is functionally more conserved than sequence similarity, the nature of secondary and tertiary structures of PpFAAH were also evaluated. The secondary structures were generated using PhyRe2.0 Protein Folding Recognition Server[67] and with both AtFAAH (PDB ID: 6DII) and RtFAAH (PDB ID: 1MT5) as templates. The confidence of generating predicted secondary structures for all the nine PpFAAH paralogs with either template was 100%. The percentage coverage of the residues, however, varied with each comparison depending on the identity between template and the query (PpFAAH). For instance, when 6DII was used as a template, the sequence coverage of PpFAAH1 was 99.2% while it was only 72% with 1MT5 template (Table S4). Similarly, with 6DII template, the amino acid coverage for PpFAAH2 to PpFAAH6 was above 90%, while PpFAAH8 had the least with only 57% (Table S4). The coverage reduced from 85% to 57%, when 1MT5 template was used; interestingly, PpFAAH7 to PpFAAH9 showed similar coverage with both the templates. The secondary structure of AtFAAH1 contained 23 α-helices and 17 β-sheets where as RtFAAH has 22 α-helices and 11 β-sheets[26,30]. The number of α-helices in PpFAAH paralogs ranged from 25 to 16 and 8-10 β-sheets, when modeled with AtFAAH; this number varied in relation to RtFAAH (Table S4). Furthermore, the tertiary structure of PpFAAH was similar to that of AtFAAH and RtFAAH where the core is composed of a number of β-sheets, which are surrounded by α-helices and loops[26,30]. Since the length of the N- and C-termini of PpFAAH paralogs varied, the number of α-helices and loops surrounding the β-sheets also varied accordingly (Fig. 2.3A, 2.3B). In general, predicted secondary and tertiary structures of PpFAAH paralogs were less identical with each other relative to their comparison with AtFAAH and RtFAAH.

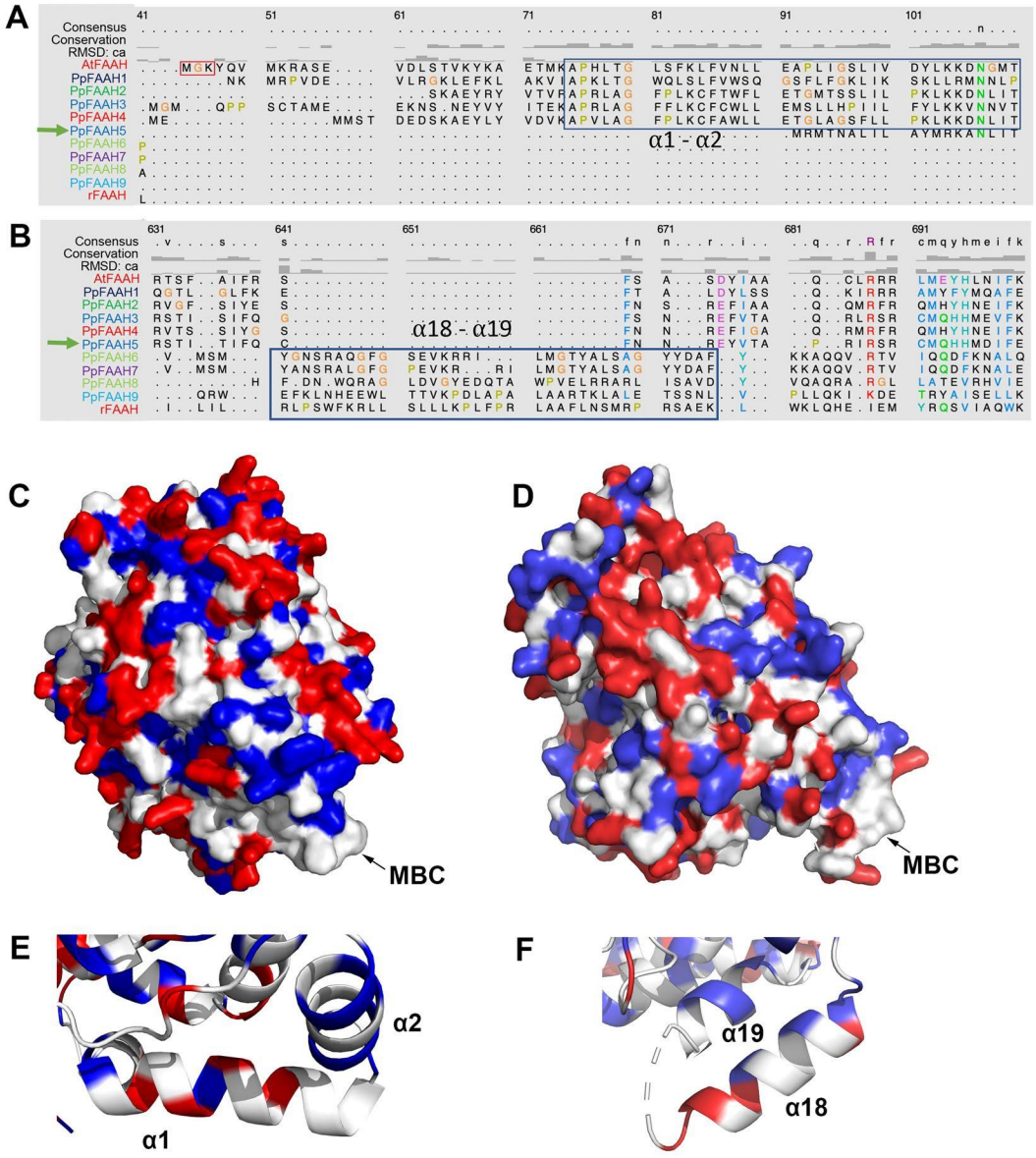


Figure 2.4. The membrane binding cap (MBC) properties of PpFAAH. Sequence alignment of PpFAAH with AtFAAH and RtFAAH using Chimera 5 to identify potential MBC in A). N-terminus and B). C-terminus. Boxes indicate conserved residues of PpFAAH paralogs either with AtFAAH to form the α1 and α2 or with RtFAAH to form α18 and α19, for MBC; Spatial surface structures showing MBC of C). PpFAAH1, D). PpFAAH6, and the close up their respective alpha helices in E). and F). White, red and blue colors represent the hydrophobic, charged and polar residues, respectively

In addition to the highly conserved AS region, FAAH proteins also have characteristic features such as the MBC, ABC and MAC, which differed between AtFAAH and RtFAAH. The AS region of AtFAAH and RtFAAH makes four β-sheets but differ in the number of α-helices, which are five and four, respectively (Fig. 2.3C). The additional ⟨-helix in AtFAAH is made by His252, Glu153, Leu154, Gly155 and Met256, whereas in RtFAAH, those residues are not conserved and instead make a loop. Similarly, the arrangement of α-helices, β-sheets and turns and loops of AS region of PpFAAH paralogs were almost identical to At/RtFAAH, even though the sequence identity between them ranged between 26% to 47% (Fig. 2.3 C-E, and S3). Additionally, like RtFAAH, the AS region of PpFAAH1 to PpFAAH5 also lacked the additional ⟨-helix observed in AtFAAH (Fig. 2.3D, S3A-D). In contrast, PpFAAH6 to PpFAAH9 make one less ⟨-helix in their AS region compared to other PpFAAH (Fig. 2.3E, S3E-G) and two less ⟨-helices, relative to AtFAAH. Furthermore, although the position of the characteristic catalytic triad Lys-Ser-Ser in the primary sequence varied among all the paralogs of PpFAAH (Table S3), depending on the size of the protein they were spaced out with same distance within the sequence and closely match the fold of either AtFAAH or RtFAAH (Fig. 2.3 and S3). For instance, the distance between the nitrogen atom of lysine and oxygen of serine is ~2.6 Å and nitrogen of serine to oxygen of serine is ~3.0 Å. The catalytic triad position of PpFAAH1 (K202, S278, S302) is very similar to AtFAAH (K205, S281, S305)[26] but PpFAAH6 (K139, S215, S239) and PpFAAH7 (K142, S217, S241) have positioning similar to that of RtFAAH (K142, S217, S241)[30].

Furthermore, in comparing the features a dimer, both AtFAAH and RtFAAH show a symmetric pattern by which the orientation of the protein subunit align in a way that MBC and MAC of each subunit are placed on the same face of the dimer[23,26,42]. Several amino acid residues that interact by hydrogen bond and Van der Waals interactions for dimerization were also identified; in AtFAAH, four residues are present in N-terminus and two in α17 and α20[26]. Structural comparison of PpFAAH1 to PpFAAH4 with AtFAAH revealed that either identical or same class of residues are present in the same regions (Table S5), suggesting the potential of PpFAAH to function as a homodimer. However, with a number of FAAH candidates in *P. patens*, forming heteromers to attain functional diversity is also a possibility. Overall, PpFAAH1 to PpFAAH5 have more structural similarities with AtFAAH, while PpFAAH6 to PpFAAH9

showed better match with RtFAAH, which agrees with the way they clustered in phylogenetic tree and thus suggesting a possible functional similarity. The subtle structural variabilities among PpFAAH paralogs also points to likely flexibility in substrate utilization that is perhaps associated with plasticity in evolutionary adaptation of mosses to environmental variation.

The membrane binding cap in PpFAAH shares similarities with mammalian and plant FAAH

In AtFAAH, MBC is formed in the α1 and α2 helices of the N-terminus, of which 21/34 residues are hydrophobic and are arranged like teeth on a comb[26]. In RtFAAH, since the N-terminus is shorter and has a TM domain, MBC is formed in α18 and α19 helices of C-terminus with 23/34 hydrophobic amino acid residues[30]. Like AtFAAH, PpFAAH1 to PpFAAH4 have MBC in their relatively longer N-terminus (Fig. 2.4A, C, E and 3C), while PpFAAH6 and PpFAAH7, like RtFAAH with shorter N-terminus have their MBC in the C-terminus (Fig. 2.4B, D, F and 3D). Both PpFAAH8 and PpFAAH9, with their longest and shortest N-terminus, respectively also have their MBC in the C-terminus. Similar to AtFAAH, the MBC in both PpFAAH1 and PpFAAH2 is predicted in α1 and α2 helices with 24/41 and 19/37 hydrophobic residues arranged from L21-P61 and A94-L127, respectively (Fig. 2.4A, Table S6), and are arranged like teeth on a comb. For PpFAAH3 and PpFAAH4, a TM integrated MBC is predicted in the α1 and α2 helices in N-terminus with higher ratio of hydrophobic residues19/23 and 24/31, respectively (Table S6). The N-termini of PpFAAH6 and PpFAAH7 do not align with that of AtFAAH and are predicted to have MBC in the α18 and α19 helices of the C-terminus, but with a high ratio of hydrophobic residues that align with RtFAAH (Fig. 2.4B, Table S6). Although the coverage of predicted secondary structure for PpFAAH8 and PpFAAH9 was relatively low, by comparing them to AtFAAH and RtFAAH, even with the limited coverage, it is predicted that C-terminus could make the MBC. Residues A602-V616 in α12 and V258-L271 in α9 have the potential to make MBC for PpFAAH8 and PpFAAH9, respectively. Predicted secondary structure of PpFAAH5, on the other hand, did not reveal any hydrophobic plateau in either termini, and thus the MBC region was not predicted (Table S6). Wherever the MBC was predicted for PpFAAH, irrespective of the termini, they could be membrane integrated or arranged as teeth on a comb.

The moss FAAH paralogs split to share similarities with AtFAAH and RtFAAH for substrate interaction

In both AtFAAH and RtFAAH, to access substrates such as NAE from the membrane, the MAC starts at the edge of the MBC in α1 and α2, and α18 and α19 helices, respectively[26,30]. The entrance of ligand binding pocket for AtFAAH is constituted with a number of hydrophobic (A27, P28, L30, P38, I51, and L55) and two charged (K26 and D58) residues. Similar to AtFAAH, the entrance of the ligand binding pocket for PpFAAH1-PpFAAH4 are also found in the α1 and α2 helices and have all the hydrophobic residues conserved, except PpFAAH4 has leucine instead of isoleucine (Table S7, Fig. S4). Among the two charged residues, both are conserved in PpFAAH4, but valine is replaced aspartic acid in PpFAAH3 and lysine with arginine in PpFAAH2; there are no charged residues in PpFAAH1 and instead two hydrophobic residues (isoleucine and methionine) are present in the same structural position (Table S7, Fig. S4). Nevertheless, PpFAAH1 to PpFAAH4 made the same predicted secondary structure as AtFAAH (Fig. 2.5A).

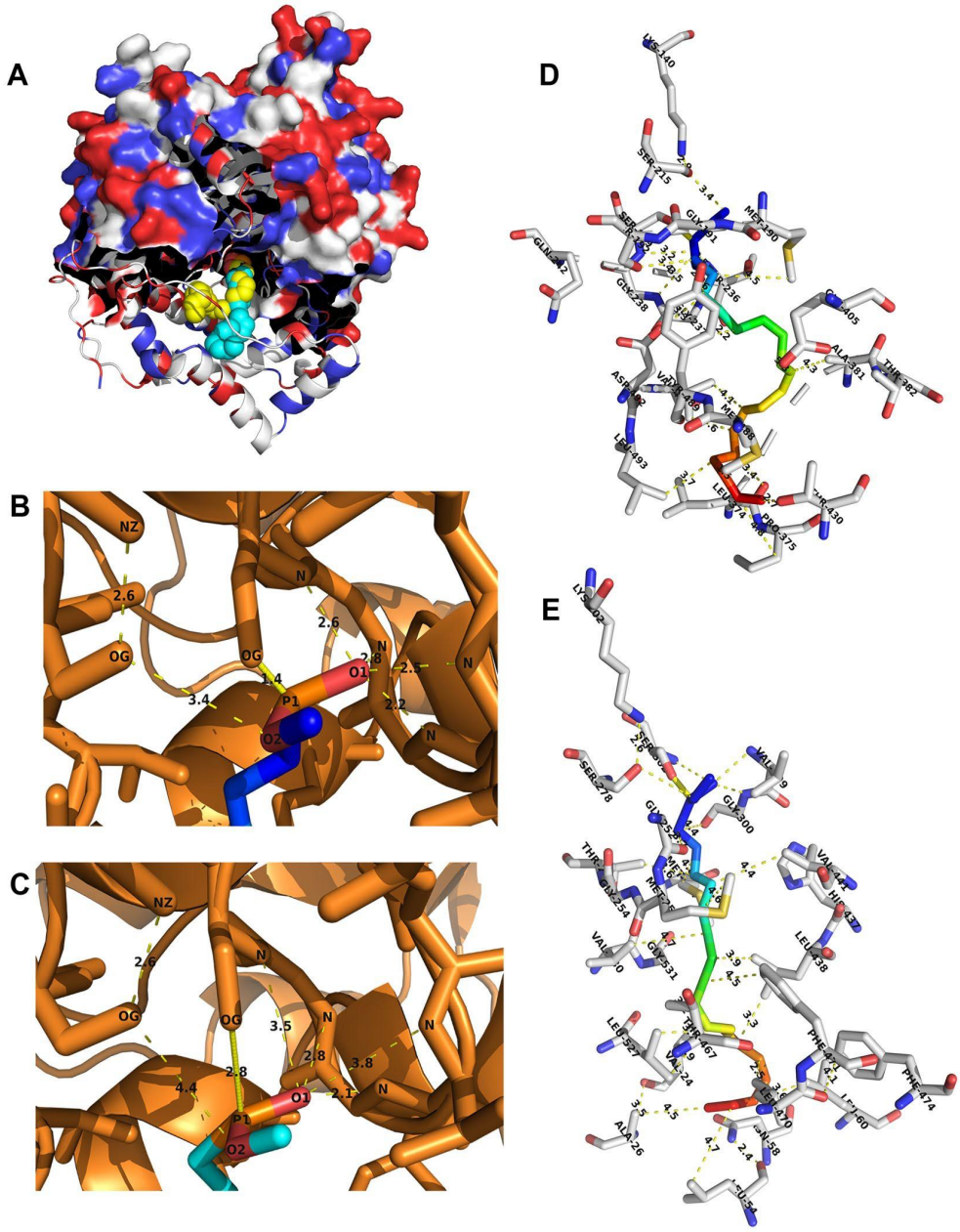


Figure 2.5. Substrate docking of PpFAAH1 and PpFAAH6. A). The structure of PpFAAH1 along with docked substrate analogs, MAFP (yellow) and MLnFP (green) are presented. The PpFAAH1 structure is shown in a partial space-filling model with secondary structures as ribbon, interacting with substrate shown as sphere. B). Polar interaction of MAFP head group with catalytic residues of PpFAAH1. C). The polar interaction between MLnFP head group and catalytic residues of PpFAAH1. The distance (in angstrom) between the atoms were shown with yellow dotted lines. The nucleophilic attack on the phosphorus of substrate by Ser302 shown as solid yellow line. Atoms are labeled with atomic symbols (N, nitrogen; O, oxygen; P,

phosphorus). D). Substrate MAFP is shown in the substrate binding pocket of PpFAAH1 and E). PpFAAH6. Van der Waals interaction between residues and acyl chain of MAFP (in rainbow color) are shown with dashed yellow lines. Amino acid residues are notated with their three-letter code and position while their carbon, oxygen and nitrogen are represented as gray, red and blue respectively

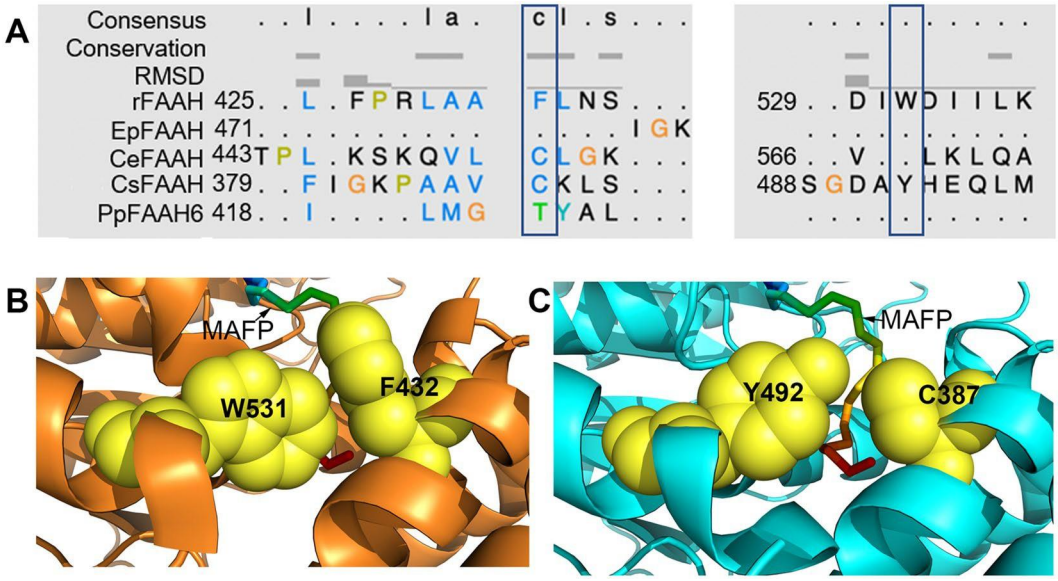


Figure 2.6 The dynamic paddle of FAAH. A). Structural alignment of EpFAAH, CeFAAH, CsFAAH and PpFAAH6 with F432 and W531 regions of RtFAAH; boxes represent the position associated with dynamic paddle residues of RtFAAH. B). Formation of dynamic paddle by W531 and F432 of RtFAAH with docked MAFP. C). Formation of dynamic paddle like structure in CsFAAH with substituted residues of Y492 and C387 in place of W531 and F432, respectively. Substrate, MAFP is shown as a stick with rainbow color. Residues of dynamic paddle are shown as sphere and yellow in color. Formation of dynamic paddle in these five FAAH orthologs shown in the alignment are presented in Fig. S7

The SBP in both AtFAAH and RtFAAH are also composed of mostly hydrophobic and few hydrophilic residues[26,30]. The acyl chain of the substrate interacts with hydrophobic residues by Van der Waals interaction, whereas hydrophilic residues in the SBP helps to move the polar head group of the substrate deeper into the pocket towards the catalytic triad[23]. Analyzing the SBP of PpFAAH1 to PpFAAH4 revealed that the hydrophobic residues that form the SBP in AtFAAH (M25, A27, L55, M61, G255, M256, G257, V442, I445, I475, F476, F479, I532 and

M539) are either conserved or replaced with another hydrophobic residue (Table S8). The hydrophilic residues (N59, T258, H441, S472, T535, and T536) of AtFAAH make its SBP relatively more polar than that of RtFAAH. Out of the six hydrophilic residues in AtFAAH, three of them (H441, S472 and T535) are also conserved among other higher plants[26]; however, that is not the case for PpFAAH1 to PpFAAH4. While N59, T258 and H441 are conserved T535 and T536 are replaced by hydrophobic residues (Gly, Val or Ala) in all four PpFAAH; S472 was replaced with the same class threonine for PpFAAH1, PpFAAH3 and PpFAAH4 (Table S8). These reduced number of hydrophilic residues suggest that the SBP of PpFAAH1 to PpFAAH4 is less polar than AtFAAH but remains more polar than RtFAAH.

In case of PpFAAH6 to PpFAAH9, since their MBC is predicted in the C-terminus, the entrance of the ligand binding pocket, MAC and SBP depart from AtFAAH, and instead they are similar with RtFAAH. Structural alignment of PpFAAH6 to PpFAAH9 with RtFAAH revealed that the positioning of the residues that contribute to MAC and SBP are mostly conserved with a few exceptions where they are either replaced with same or different class of residues, and thus retain the structural identity (Table S5, Fig. 5SB). Specifically, of the 29 residues of RtFAAH SBP, nine are substituted with different class of residues; six of nine were hydrophobic and three hydrophilic. In SBP of PpFAAH6 and PpFAAH7, six were hydrophilic and three were hydrophobic making them more polar than RtFAAH and thus flexible to accept substrates of that nature. As mentioned earlier, the coverage of the predicted secondary structure of PpFAAH8 is relatively low, which is reflected in the predicted SBP and their secondary structure arrangement (Table S9). In conclusion, the MAC and SBP for PpFAAH1 to PpFAAH4 and PpFAAH6 to PpFAAH9 are predicted to be similar to AtFAAH and RtFAAH, respectively.

Previously, crystal structures of AtFAAH and RtFAAH were generated using analogs MLnFP (analog of 18C NAE) and MAFP (analog of anandamide), respectively. To identify potential substrate preference of PpFAAH, both the analogs were docked using Auto Dock Vina in PyRx software. Substrate docking analysis for PpFAAH paralogs revealed that substrates interact with residues in the predicted SBP *via* Van der Waals forces or polar-covalent bonds. Van der Waals interactions occur between hydrophilic acyl chain of the substrate and residues in the SBP. The polar-covalent interaction occurs between the polar head of the substrate and catalytic residues. Among the nine paralogs PpFAAH1 and PpFAAH6, were selected as the

representatives that were predicted to be closer to AtFAAH and RtFAAH, respectively and analyzed (Fig. 2.5). The head group of both substrates are identical, but the tails or acyl chains are different in number of carbon and double bonds. Because of their structural differences in tails, entry and accommodation of the substrates in to the SBP of the enzyme are different as the residues of SBP that interact by Van der Waals force with each of the substrates vary. The polar interaction between the head group and catalytic residues are the same positional wise, however, the distance between the catalytic residues and the polar head groups of the substrates are different which determines how efficiently substrate will undergo a nucleophilic attack by the catalytic residue of S302 of PpFAAH1 (Fig. 2.5A). In contrast, PpFAAH6 docking revealed that the positioning of the substrate inside its secondary structure was different from that of AtFAAH and PpFAAH1 to PpFAAH4 but rather similar to RtFAAH (Fig. 2.5C, Table S9). Docking analysis of MLnFP or MAFP with PpFAAH1 showed that catalytic nucleophile Ser302 makes polar-covalent bond with phosphorus of the substrate, and oxygen atoms make polar covalent bonds with nitrogen atom of Ser278, Val299 and Gly300 (Fig. 2.5A, 5B). On the other hand, for PpFAAH6 substrate docking showed that catalytic nucleophile Ser239 makes polar bond with phosphorus of MAFP (Fig. 2.5C). In RtFAAH, Ser241 and in AtFAAH Ser305 form covalent bond with phosphorous of MAFP and MLnFP, respectively[26,30] This analysis suggested that even though the path of MAC or SBP is different among rat, arabidopsis or moss FAAH, the catalytic mechanism remained identical.

Fully conserved 'dynamic paddle' feature is likely limited to Phylum Chordata

Key difference of RtFAAH from that of PpFAAH and AtFAAH is the presence of 'dynamic paddle' residues, F432 and W531 in the C terminus of 18 to 21α helices that separate SBP/ABC and MAC. The 'dynamic paddle' of FAAH enzyme plays an essential role in terms of substrate selectivity and its accommodation from MAC to SBP/ABC pocket[51]. Double mutation of Phe432 and Trp531 to alanine in RtFAAH, along with microsecond-long molecular dynamic simulations study revealed the significance of these two residues in substrate selection and catabolic rate[51]. Change in specificity of substrate selection and the rate of catabolism by human FAAH are associated with a number of disorders, including but not limited to weight gain, energy balance, food intake, and anxiety control, etc[68–70]. Therefore, 'dynamic paddle' residues are considered crucial in tight regulation of substrate selection, entrance and its catabolism and thus, prevention of various diseases. To further understand the role and evolution of dynamic

paddle, predicted FAAH structure of 28 organisms, representing select phyla of Kingdom Protozoa, Chromista, Fungi, Plantae and Animalia were analyzed. Unlike in RtFAAH and HsFAAH, none of the plant FAAH analyzed were predicted to make the α helices with W531 to constitute the 'dynamic paddle' (Fig. S6B). Among those examined, only PpFAAH6 to PpFAAH9 were predicted to make the α helix in the region where F432 of RtFAAH occurs, but phenylalanine was not conserved (Fig. S6A, S7). Among the FAAH orthologs that were analyzed, tryptophan and phenylalanine were only conserved in the Phylum Chordata; organisms analyzed were- human, rat, mouse, cow, rock chuck and zebrafish; both the residues were not conserved in any other phyla of the five Kingdoms examined (Fig. 2.6, S6A).

In Kingdom Animalia, among the 13 species that were tested for similarities in dynamic paddle structure, lower organisms like *Stylophora pistillata* and *Exaiptasia pallida* from Cnidaria Phylum did not make the predicted α helices in that region in which W531 of RtFAAH is present (Fig. 2.6, S6A). Nematodes such as *Toxocara canis* and *C. elegans*, however, made the α helices but tryptophan was not conserved. Among the Platyhelminthes, *Clonorchis sinensis* and *Schistosoma bovis* were also predicted to make α helices, but tryptophan was substituted with same class of aromatic amino acid, tyrosine (Fig. 2.6A, 2.6C, S6A, S7E). The occurrence of another residue in dynamic paddle, phenylalanine also showed the same trend in Kingdom Animalia. Organisms from Nematoda and higher Phyla, including Platyhelminthes and Arthropoda were predicted to make the α helices but the position aligned with F432 of RtFAAH was substituted by a different class of amino acid (Fig. 2.6A, S6A). These analyses indicate that dynamic paddle-like structure slowly evolved from lower to higher organisms. For example, among the organisms analyzed, Phylum Cnidaria was not predicted to make any of the α helices in which W531 and F432 are present. Phylum Nematoda could make the α helices corresponding to the region of F432 and W531 but without the two residues. Phylum Platyhelminthes was predicted to make the α helices associated but with substituted residues to Phe and Trp (Fig. 2.6, S6, S7). Moreover, when substrate docking (MAFP) on FAAH from respective organisms was observed, we detected that FAAH from *Clonorchis sinensis*, a Platyhelminth forms a structure similar to that of dynamic paddle but the space that accommodates substrate entry through MBC to the SBP is wider compared with RtFAAH (Fig. 2.6B, S7D,E); recent study suggested that this space is important for a tight regulation of substrate entry[51]. Interestingly, although the occurrence of endocannabinoids is widely reported among eukaryotes, the identification of

endocannabinoid signaling system, including the cannabinoid-binding receptors are thus far limited to chordates. This observation led us to hypothesize that organisms in Animalia kingdom evolved dynamic paddle-like structure in FAAH, as complexity increased and tighter regulation of substrates became crucial, leading to the true dynamic-paddle. As such, it is likely that the complexity in regulation of the enzyme evolved as a necessity associated with functional implications of the signaling network. Among the nine PpFAAH, only PpFAAH6 to PpFAAH9 were predicted to make the α helices in the region corresponding to RtFAAH F432. This observation further justifies the versatility of FAAH in *P. patens*, which is an important transitional species in evolution.

Conclusions

Structural variations of a highly conserved enzyme across diverse phyla are critical for understanding functional evolution. Crystal structures of mammalian and plant FAAH served as templates to predict and understand structural details of uncharacterized amidase family of proteins. In the moss, *P. patens*, it is evident that there is a functional FAAH that can not only metabolize anandamide but does so with higher specificity and catalytic efficiency of AtFAAH or RtFAAH. Additionally, ability of PpFAAH1 to retain more than 50% activity at a wide range of temperature and pH also suggests the adaptability of the enzyme to varying conditions. Such adaptations by enzymes are likely to rely on structural alterations that affect ligand binding and catalytic rate. The enzyme FAAH being a highly conserved protein across the eukaryotic lineage that evolved about 1.5 billion years ago, along with multiple paralogs of FAAH in moss provide a unique resource to further explore their successful functional and structural adaptability to shifting environmental conditions during evolution.

Our comprehensive and systematic *in silico* analyses of identification and evaluation of structural details of PpFAAHs also alluded to divergence of FAAH between the plant and animal lineages. While the AS region and catalytic triad appear to be universally conserved among the FAAH orthologs, variations in other key functional features such as the MBC, MAC, SBP, ABC and extended or shortened N or C termini evidently contributed to their structural and perhaps functional divergence. Of interest here is the clear splitting of moss FAAH paralogs into either *Arabidopsi*s or rat FAAH related, based on these additional key features. While the moss FAAH1 paralogs might in general possess the potential for amidase activity, as indicated by the

interaction of their catalytic site with substrate analogs such as MAFP and MLnFP, their differentiation into two lineages provides us additional tools to evaluate the subtle variations that are predicted to affect specificity, catabolic rate, enzyme-ligand interaction and regulation and varying physiological conditions for optimal activity.

Recent studies have shown the significance of two gating residues Phe432 and Trp531 in forming the catalytic site of FAAH, which selectively accommodates anandamide and orients for efficient hydrolysis. Comparative analysis of dynamic paddle region among FAAH orthologs from various phyla, curiously revealed that the true dynamic paddle region with Phe and Trp residues was limited to orthologs from Chordates, which might be associated with the need for evolution of tighter regulation of the endocannabinoid signaling system. Like in mammals, because of the occurrence of anandamide in mosses, we expected PpFAAH paralogs to show a conserved dynamic paddle region. However, the absence of such region among PpFAAH paralogs suggests that they might be more versatile in their lipid selection for substrates. In mosses, the class of NAEs, in addition to anandamide include other saturated and unsaturated NAEs ranging from 16C to 20C. Thus, it is plausible that these paralogs are more attuned to accept these diverse range of substrates.

Our studies conclude that NAE-mediated functions in moss can be complex and diverse and various FAAH paralogs are likely to play a role in their hydrolysis. The structural similarities and dissimilarities identified among the various orthologs provides us basis to understand the diversity among them. Finally, the role of anandamide in mammals and its regulation by FAAH implicates that such a system in mosses is also likely to affect various physiological aspects and address how early land plants might deal with extreme biotic and abiotic stressors.

Materials and Methods

Identification and in silico analyses of FAAH homologs

Protein sequence of FAAH from human (NP_001432.2), rat (NP_077046) and Arabidopsis (AT5G64440.1) were used as query sequences to search for homologs in *Physcomitrella patens* using *Phytozome 12 Physcomitrella patens v3.3* database. Best hits for putative PpFAAH (1-9) thus obtained were further used for multiple sequence alignment and to obtain percent matrix identity against rat and Arabidopsis FAAH using Clustal Omega Multiple

Sequence Alignment [71]. The MEGA7 software was used to generate phylogenetic tree using FAAH protein sequences from 29 different organisms (Table S3) representing five eukaryotic kingdoms. For this MEGA and Maximum Likelihood method based on Subtree-Pruning-Regrafting (SPR) algorithm with search label 1 was used to generate the phylogenetic tree. For the test of phylogeny, bootstrap method with 500 replicates was used; substitute type was amino acid; Gaps/Missing Data Treatment was used for partial deletion with 95% coverage cutoff [72]. A human amidase (NP_777572) and arabidopsis amidase (AT1G08980.1) sequences that were not FAAH served as outgroups.

Cloning, expression, and purification of heterologous PpFAAH1

The moss *Physcomitrella patens* (ecotype Gransden 2004) was cultured on BCD media at 25^{o}C with light intensity of 17.45 W/m^2 under 16-h light and 8-h dark cycle. Total RNA was extracted using plant mini RNA kit (Qiagen) and cDNA was synthesized using reverse transcriptase (Promega) following manufacturer's protocol. Full length *PpFAAH1* was amplified using forward (5'-AAAAAGCAGGCTCGATGGCGCAAAATAAGATGCGAC-3') and reverse primers (5'-AGAAAGCTGGGTCTTACTTCAAGATGTTATAGAATG-3') including STOP codon. Amplified *PpFAAH1* was then cloned into pDEST15 vector following Gateway cloning protocol. For expression, *PpFAAH1* containing vector was then transformed into *E. coli* host BL21(DE3) cell line. The 1 mM IPTG induced cultures were incubated for 4 h at 37^{o}C and harvested cells were French pressed with lysis buffer (50 mM Tris-Cl pH 8.0, 100 mM NaCl, 1% Triton-100), and then purified by GST-tagged Spin Purification Column Kit (Thermo-Fisher). Purified PpFAAH1 was further concentrated by Amicon Ultra-15 (Ultracel -30K), quantified by using Nanodrop (ND-1000) and further confirmed by Western blot (Text S1).

Fatty acid amide hydrolase (FAAH) assays

For FAAH amidohydrolase assay, 3 μg of purified PpFAAH1 was used for all experiments. To determine optimal conditions, pH range from 6.5 to 11, and a temperature range of 24^{o}C to 70^{o}C were used. For kinetic assays, pH 8.0 and 37^{o}C were used as optimal conditions. To determine substrate specificity for PpFAAH1, [1-^{14}C] NAE20:4 and [1-^{14}C] NAE16:0 were used. While AtFAAH was used as positive control, GST protein and no enzyme were used as negative controls. Since optimal condition for AtFAAH activity is different[26], kinetic comparisons were not made between AtFAAH and PpFAAH1 in this current study.

Amidohydrolase reaction assay and lipid extraction for product analysis were carried out as previously described[73]. Thin layer chromatography (TLC) plate (TLC Silica Gel 60, Millipore) was used to separate lipid soluble reaction products and a radiometric scanner (Eckert & Ziegler, AR2000) was used to quantify the FAAH activity.

For FAAH inhibition assays, the above protocol was followed except that the reaction mixture with enzyme was incubated with an inhibitor for 10 min prior to the addition of 100 μM substrate [1-^{14}C]NAE20:4. Three inhibitors, PMSF, MAFP and URB597 were used in the assays, with varying concentrations (1 to 100 μM for PMSF; 0.1 to 10 nM for MAFP and URB597).

Secondary structure prediction and molecular docking

Secondary structures of putative PpFAAH were generated using Phyre 2.0 Protein Folding Recognition Server[67] and using both RtFAAH (PDB ID 1MT5) and AtFAAH (PDB ID 6DII) as templates. For FAAH from 27 different organisms (Table S3), secondary structures were generated using either RtFAAH or AtFAAH, depending on their respective Kingdom. The PDB format of secondary structure for PpFAAH were further analyzed using Chimera 5.0[74] and/or PyMOL software. For quality control, the predicted secondary structures of RtFAAH and AtFAAH were also generated using the same resource, Phyre 2.0. For quality check, we have also used PDBeFold (PDB in Europe, Structure Similarity)[75] by which several criteria of 3D structure between template and query were analyzed. We reported Q and root-mean standard deviation (RMSD) scores ranging from 0-1, where higher Q score and lower RMSD indicates better quality.

Crystal structures of AtFAAH (PDB ID 6DII) and RtFAAH (PDB ID 1MT5) were previously generated with inhibitors MLnFP[26] and MAFP[30], respectively. These structures, obtained from PDB database were used for substrate docking analyses of PpFAAH, and further understanding of plant and animal FAAH. The PDB format PpFAAH structures with substrates were docked using Auto Dock Vina of PyRx software[76]. Docked structures were then visualized using PyMOL.

REFERENCES

1. Gachet, M. S., Schubert, A., Calarco, S., Boccard, J. & Gertsch, J. Targeted metabolomics shows plasticity in the evolution of signaling lipids and uncovers old and new endocannabinoids in the plant kingdom. *Sci. Rep.* (2017). doi:10.1038/srep41177

2. Schmid, H. H. O., Schmid, P. C. & Natarajan, V. N-Acylated glycerophospholipids and their derivatives. *Progress in Lipid Research* (1990). doi:10.1016/0163-7827(90)90004-5

3. Berrendero, F., Sepe, N., Ramos, J. A., Di Marzo, V. & Fernández-Ruiz, J. J. Analysis of cannabinoid receptor binding and mRNA expression and endogenous cannabinoid contents in the developing rat brain during late gestation and early postnatal period. *Synapse* **33**, 181–91 (1999).

4. Bíró, T., Tóth, B. I., Haskó, G., Paus, R. & Pacher, P. The endocannabinoid system of the skin in health and disease: novel perspectives and therapeutic opportunities. *Trends in Pharmacological Sciences* (2009). doi:10.1016/j.tips.2009.05.004

5. Chapman, K. D. Phospholipase activity during plant growth and development and in response to environmental stress. *Trends Plant Sci.* **3**, 419–426 (1998).

6. Devane, W. A. *et al.* Constituent Receptor. *Lancet, The* (1992). doi:10.1126/science.1470919

7. Alger, B. E. Endocannabinoids at the synapse a decade after the dies mirabilis (29 March 2001): What we still do not know. *Journal of Physiology* (2012). doi:10.1113/jphysiol.2011.220855

8. Chevaleyre, V., Takahashi, K. A. & Castillo, P. E. Endocannabinoid-mediated synaptic plasticity in the CNS. *Annu. Rev. Neurosci.* (2006). doi:10.1146/annurev.neuro.29.051605.112834

9. Freund, T. F., Katona, I. & Piomelli, D. Role of endogenous cannabinoids in synaptic signaling. *Physiol. Rev.* **83**, 1017–66 (2003).

10. Kano, M., Ohno-Shosaku, T., Hashimotodani, Y., Uchigashima, M. & Watanabe, M. Endocannabinoid-Mediated Control of Synaptic Transmission. *Physiol. Rev.* (2009). doi:10.1152/physrev.00019.2008

11. Katona, I. & Freund, T. F. Multiple Functions of Endocannabinoid Signaling in the Brain. *Annu. Rev. Neurosci.* (2012). doi:10.1146/annurev-neuro-062111-150420

12. Garg, P., Duncan, R. S., Kaja, S. & Koulen, P. Intracellular mechanisms of N-acylethanolamine-mediated neuroprotection in a rat model of stroke. *Neuroscience* (2010). doi:10.1016/j.neuroscience.2009.11.069

13. Galles, C. *et al.* Endocannabinoids in Caenorhabditis elegans are essential for the mobilization of cholesterol from internal reserves. *Sci. Rep.* (2018). doi:10.1038/s41598-018-24925-8

14. Lucanic, M. *et al.* N-acylethanolamine signalling mediates the effect of diet on lifespan in Caenorhabditis elegans. *Nature* **473**, 226–9 (2011).

15. Blüher, M. *et al.* Dysregulation of the peripheral and adipose tissue endocannabinoid system in human abdominal obesity. *Diabetes* **55**, 3053–60 (2006).

16. Gaetani, S., Kaye, W. H., Cuomo, V. & Piomelli, D. Role of endocannabinoids and their analogues in obesity and eating disorder. **13**, 42–48 (2008).

17. Blancaflor, E. B. *et al.* N-Acylethanolamines: Lipid metabolites with functions in plant growth and development. *Plant Journal* (2014). doi:10.1111/tpj.12427

18. Keereetaweep, J., Blancaflor, E. B., Hornung, E., Feussner, I. & Chapman, K. D. Ethanolamide oxylipins of linolenic acid can negatively regulate Arabidopsis seedling development. *Plant Cell* **25**, 3824–40 (2013).

19. Keereetaweep, J., Blancaflor, E. B., Hornung, E., Feussner, I. & Chapman, K. D. Lipoxygenase-derived 9-hydro(pero)xides of linoleoylethanolamide interact with ABA signaling to arrest root development during Arabidopsis seedling establishment. *Plant J.* (2015). doi:10.1111/tpj.12821

20. Iannotti, F. A., Di Marzo, V. & Petrosino, S. Endocannabinoids and endocannabinoid-related mediators: Targets, metabolism and role in neurological disorders. *Prog. Lipid Res.* **62**, 107–28 (2016).

21. Schwartz, G. J. *et al.* The lipid messenger OEA links dietary fat intake to satiety. *Cell Metab.* **8**, 281–288 (2008).

22. Chilufya, J. Y., Devaiah, S. P., Sante, R. R. & Kilaru, A. Endocannabinoid-Like Lipids in Plants. in *eLS* 1–9 (John Wiley & Sons, Ltd, 2015). doi:10.1002/9780470015902.a0024632

23. McKinney, M. K. & Cravatt, B. F. Structure and function of fatty acid amidy hydrolase. *Annu. Rev. Biochem.* **74**, 411–432 (2005).

24. Shrestha, R., Noordermeer, M. A., van der Stelt, M., Veldink, G. A. & Chapman, K. D. N-acylethanolamines are metabolized by lipoxygenase and amidohydrolase in competing pathways during cottonseed imbibition. *Plant Physiol.* **130**, 391–401 (2002).

25. Shrestha, R., Dixon, R. A. & Chapman, K. D. Molecular identification of a functional homologue of the mammalian fatty acid amide hydrolase in Arabidopsis thaliana. *J. Biol. Chem.* (2003). doi:10.1074/jbc.M305613200

26. Aziz, M., Wang, X., Tripathi, A., Bankaitis, V. A. & Chapman, K. D. Structural analysis of a plant fatty acid amide hydrolase provides insights into the evolutionary diversity of bioactive acylethanolamides. *J. Biol. Chem.* (2019). doi:10.1074/jbc.RA118.006672

27. Hayes, A. C., Stupak, J., Li, J. & Cox, A. D. Identification of N-acylethanolamines in Dictyostelium discoideum and confirmation of their hydrolysis by fatty acid amide

hydrolase. *J. Lipid Res.* **54**, 457–66 (2013).

28. Ueda, N., Puffenbarger, R. A., Yamamoto, S. & Deutsch, D. G. The fatty acid amide hydrolase (FAAH). *Chem. Phys. Lipids* **108**, 107–121 (2000).

29. Wei, B. Q., Mikkelsen, T. S., McKinney, M. K., Lander, E. S. & Cravatt, B. F. A second fatty acid amide hydrolase with variable distribution among placental mammals. *J. Biol. Chem.* **281**, 36569–78 (2006).

30. Bracey, M. H., Hanson, M. A., Masuda, K. R., Stevens, R. C. & Cravatt, B. F. Structural adaptations in a membrane enzyme that terminates endocannabinoid signaling. *Science (80-.).* (2002). doi:10.1126/science.1076535

31. Cravatt, B. F. *et al.* Molecular characterization of an enzyme that degrades neuromodulatory fatty-acid amides. *Nature* **384**, 83–7 (1996).

32. M-H Wei, O Toure, G M Glenn, M Pithukpakorn, L Neckers, C Stolle, P Choyke, *et al.* Novel mutations in FH and expansion of the spectrum of phenotypes expressed in families with hereditary leiomyomatosis and renal cell cancer. *J. Med. Genet.* (2006). doi:10.1136/jmg.2005.033506

33. Giang, D. K. & Cravatt, B. F. Molecular characterization of human and mouse fatty acid amide hydrolases. *Proc. Natl. Acad. Sci.* (2002). doi:10.1073/pnas.94.6.2238

34. Boileau, P. *et al.* Arthroscopic repair of full-thickness tears of the supraspinatus: does the tendon really heal? *J. Bone Joint Surg. Am.* **87**, 1229–40 (2005).

35. Camilleri, M., Kerstens, R., Rykx, A. & Vandeplassche, L. A placebo-controlled trial of prucalopride for severe chronic constipation. *N. Engl. J. Med.* **358**, 2344–54 (2008).

36. Storr, M., Devlin, S., Kaplan, G. G., Panaccione, R. & Andrews, C. N. Cannabis use provides symptom relief in patients with inflammatory bowel disease but is associated with worse disease prognosis in patients with Crohn's disease. *Inflamm. Bowel Dis.* **20**, 472–80 (2014).

37. Papatheodorou, K., Papanas, N., Banach, M., Papazoglou, D. & Edmonds, M. Complications of Diabetes 2016. *Journal of Diabetes Research* (2016). doi:10.1155/2016/6989453

38. Sarzani, R., Salvi, F., Dessì-Fulgheri, P. & Rappelli, A. Renin-angiotensin system, natriuretic peptides, obesity, metabolic syndrome, and hypertension: an integrated view in humans. *J. Hypertens.* **26**, 831–43 (2008).

39. Wang, Y.-S. *et al.* Manipulation of Arabidopsis fatty acid amide hydrolase expression modifies plant growth and sensitivity to N-acylethanolamines. *Proc. Natl. Acad. Sci.* (2006). doi:10.1073/pnas.0603571103

40. Blancaflor, E. B., Hou, G. & Chapman, K. D. Elevated levels of N-lauroylethanolamine, an endogenous constituent of desiccated seeds, disrupt normal root development in

Arabidopsis thaliana seedlings. *Planta* **217**, 206–17 (2003).

41. Kang, R. *et al.* Neural palmitoyl-proteomics reveals dynamic synaptic palmitoylation. *Nature* **456**, 904–9 (2008).

42. Kang, L. *et al.* Overexpression of a fatty acid amide hydrolase compromises innate immunity in Arabidopsis. *Plant J.* (2008). doi:10.1111/j.1365-313X.2008.03603.x

43. Teaster, N. D. *et al.* Overexpression of Fatty Acid Amide Hydrolase Induces Early Flowering in Arabidopsis thaliana. *Front. Plant Sci.* (2012). doi:10.3389/fpls.2012.00032

44. Mileni, M. *et al.* Structure-guided inhibitor design for human FAAH by interspecies active site conversion. *Proc. Natl. Acad. Sci.* (2008). doi:10.1073/pnas.0806121105

45. Min, X. *et al.* Discovery and molecular basis of potent noncovalent inhibitors of fatty acid amide hydrolase (FAAH). *Proc. Natl. Acad. Sci. U. S. A.* **108**, 7379–84 (2011).

46. Mileni, M. *et al.* Fluoride-mediated capture of a noncovalent bound state of a reversible covalent enzyme inhibitor: X-ray crystallographic analysis of an exceptionally potent α-ketoheterocycle inhibitor of fatty acid amide hydrolase. *J. Am. Chem. Soc.* **133**, 4092–100 (2011).

47. Otrubova, K. *et al.* Rational design of fatty acid amide hydrolase inhibitors that act by covalently bonding to two active site residues. *J. Am. Chem. Soc.* (2013). doi:10.1021/ja4014997

48. Bertolacci, L. *et al.* A binding site for non-steroidal anti-inflammatory drugs in FAAH. *J. Am. Chem. Soc.* (2013). doi:10.1021/ja308733u

49. Tsuboi, K., Takezaki, N. & Ueda, N. The N-acylethanolamine-hydrolyzing acid amidase (NAAA). *Chemistry and Biodiversity* **4**, 1914–1925 (2007).

50. Patricelli, M. P. & Cravatt, B. F. Fatty acid amide hydrolase competitively degrades bioactive amides and esters through a nonconventional catalytic mechanism. *Biochemistry* **38**, 14125–30 (1999).

51. Palermo, G. *et al.* Keys to Lipid Selection in Fatty Acid Amide Hydrolase Catalysis: Structural Flexibility, Gating Residues and Multiple Binding Pockets. *PLoS Comput. Biol.* (2015). doi:10.1371/journal.pcbi.1004231

52. Ahn, K., Johnson, D. S. & Cravatt, B. F. Fatty acid amide hydrolase as a potential therapeutic target for the treatment of pain and CNS disorders. *Expert Opin. Drug Discov.* **4**, 763–784 (2009).

53. Wortley, M. A. *et al.* Targeting fatty acid amide hydrolase as a therapeutic strategy for antitussive therapy. *Eur. Respir. J.* **50**, 1700782 (2017).

54. Bedse, G. *et al.* Therapeutic endocannabinoid augmentation for mood and anxiety disorders: comparative profiling of FAAH, MAGL and dual inhibitors. *Transl. Psychiatry* **8**, 92 (2018).

55. Shaw, A. J., Szövényi, P. & Shaw, B. Bryophyte diversity and evolution: windows into the early evolution of land plants. *Am. J. Bot.* **98**, 352–69 (2011).

56. Schaefer, D. G. & Zrÿd, J. P. Efficient gene targeting in the moss Physcomitrella patens. *Plant J.* **11**, 1195–206 (1997).

57. Kroemer, K., Reski, R. & Frank, W. Abiotic stress response in the moss Physcomitrella patens: Evidence for an evolutionary alteration in signaling pathways in land plants. *Plant Cell Rep.* (2004). doi:10.1007/s00299-004-0785-z

58. Wei, B. Q., Mikkelsen, T. S., McKinney, M. K., Lander, E. S. & Cravatt, B. F. A second fatty acid amide hydrolase with variable distribution among placental mammals. *J. Biol. Chem.* **281**, 36569–78 (2006).

59. Mor, M. *et al.* Cyclohexylcarbamic acid 3'- or 4'-substituted biphenyl-3-yl esters as fatty acid amide hydrolase inhibitors: synthesis, quantitative structure-activity relationships, and molecular modeling studies. *J. Med. Chem.* **47**, 4998–5008 (2004).

60. Ruggiero, M. A. *et al.* A higher level classification of all living organisms. *PLoS One* **10**, (2015).

61. Zinn, K., McAllister, L. & Goodman, C. S. Sequence analysis and neuronal expression of fasciclin I in grasshopper and drosophila. *Cell* (1988). doi:10.1016/0092-8674(88)90574-0

62. Zeytuni, N. & Zarivach, R. Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module. *Structure* (2012). doi:10.1016/j.str.2012.01.006

63. Hofmann., K. & Stoffel., W. TMBase - A database of membrane spanning protein segments. **374**, 166–166 (1993).

64. Lambert, D. M., Vandevoorde, S., Jonsson, K.-O. & Fowler, C. J. The palmitoylethanolamide family: a new class of anti-inflammatory agents? *Curr. Med. Chem.* **9**, 663–74 (2002).

65. Rodríguez de Fonseca, F. *et al.* An anorexic lipid mediator regulated by feeding. *Nature* **414**, 209–12 (2001).

66. Cravatt, B. F. *et al.* Chemical characterization of a family of brain lipids that induce sleep. *Science* **268**, 1506–9 (1995).

67. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* (2015). doi:10.1038/nprot.2015.053

68. Balsevich, G. *et al.* Role for fatty acid amide hydrolase (FAAH) in the leptin-mediated effects on feeding and energy balance. *Proc. Natl. Acad. Sci.* (2018). doi:10.1073/pnas.1802251115

69. Quarta, C. *et al.* CB(1) signaling in forebrain and sympathetic neurons is a key determinant of endocannabinoid actions on energy balance. *Cell Metab.* **11**, 273–85

(2010).

70. Cardinal, P. *et al.* Cannabinoid type 1 (CB1) receptors on Sim1-expressing neurons regulate energy expenditure in male mice. *Endocrinology* **156**, 411–8 (2015).

71. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, (2011).

72. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).

73. Kim, S.-C., Faure, L. & Chapman, K. D. Analysis of Fatty Acid Amide Hydrolase Activity in Plants. in *Methods in molecular biology (Clifton, N.J.)* **1009**, 115–127 (2013).

74. Pettersen, E. F. *et al.* UCSF Chimera--A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

75. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2256–2268 (2004).

76. Dallakyan, S. & Olson, A. J. Small-molecule library screening by docking with PyRx. *Methods Mol. Biol.* **1263**, 243–50 (2015).

SUPPLEMENTARY INFORMATION

Methods

Western blot

To confirm GST-tagged PpFAAH1 expression in *E. coli*, protein extract was first separated by SDS-PAGE (10%) and then transferred to polyvinylidene difluoride membrane, blocked with 1% milk and 3% BSA. The membrane blot was incubated overnight at 4°C with monoclonal anti-GST antibody (1:1000 dilution). The blot was then washed three times with phosphate-buffer saline (PBS) followed by 1X PBS with 3% Tween-20 (PBST), and PBS. Anti-mouse secondary antibody (1:3000 dilution) was added to the blot and incubated for one hour at room temperature. Blot was then washed sequentially with PBS, PBST and PBS, and was then subjected to enhanced chemiluminescent (ECL) HRP and AP substrates, and finally exposed on x-ray film (Fig. S1).

In silico analysis

To estimate the molecular weight and calculate the isoelectric point (pI), UniPort Knowledgebase (Swiss-Port or TrEMBL)[1] was used. The Clustal Omega Multiple Sequence Alignment online tool was used for multiple alignment and BoxShade (https://embnet.vital-it.ch/software/BOX_form.html)[2] was used to obtain print quality alignment file. For transmembrane domain analysis TMHMM2.0[3] and TMpred[4] were used.

REFERENCES

1. Gasteiger, E. *et al.* Protein Identification and Analysis Tools on the ExPASy Server. in *The Proteomics Protocols Handbook* 571–607 (Humana Press, 2005). doi:10.1385/1-59259-890-0:571
2. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, (2011).
3. Sonnhammer, E. L., von Heijne, G. & Krogh, A. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proceedings. Int. Conf. Intell. Syst. Mol. Biol.* **6**, 175–82 (1998).
4. Ikeda, M., Arai, M., Okuno, T. & Shimizu, T. TMPDB: a database of experimentally-characterized transmembrane topologies. *Nucleic Acids Res.* **31**, 406–9 (2003).

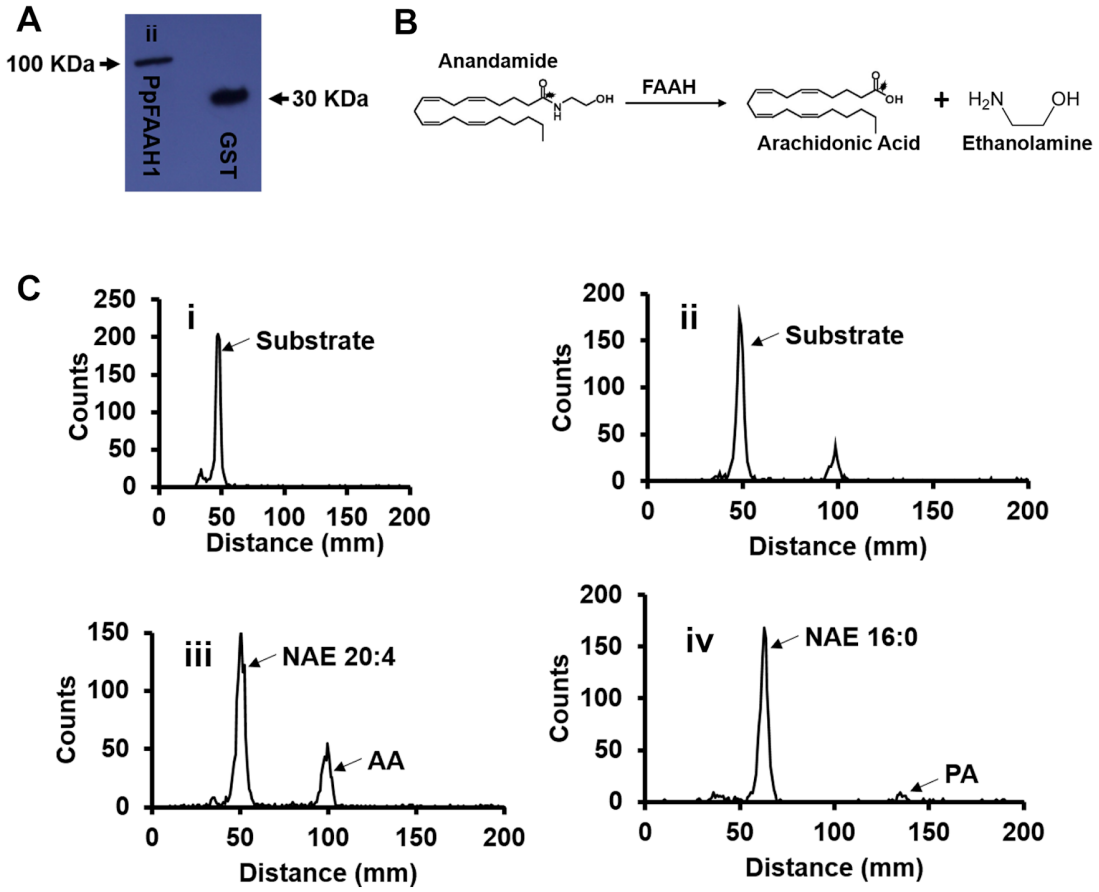
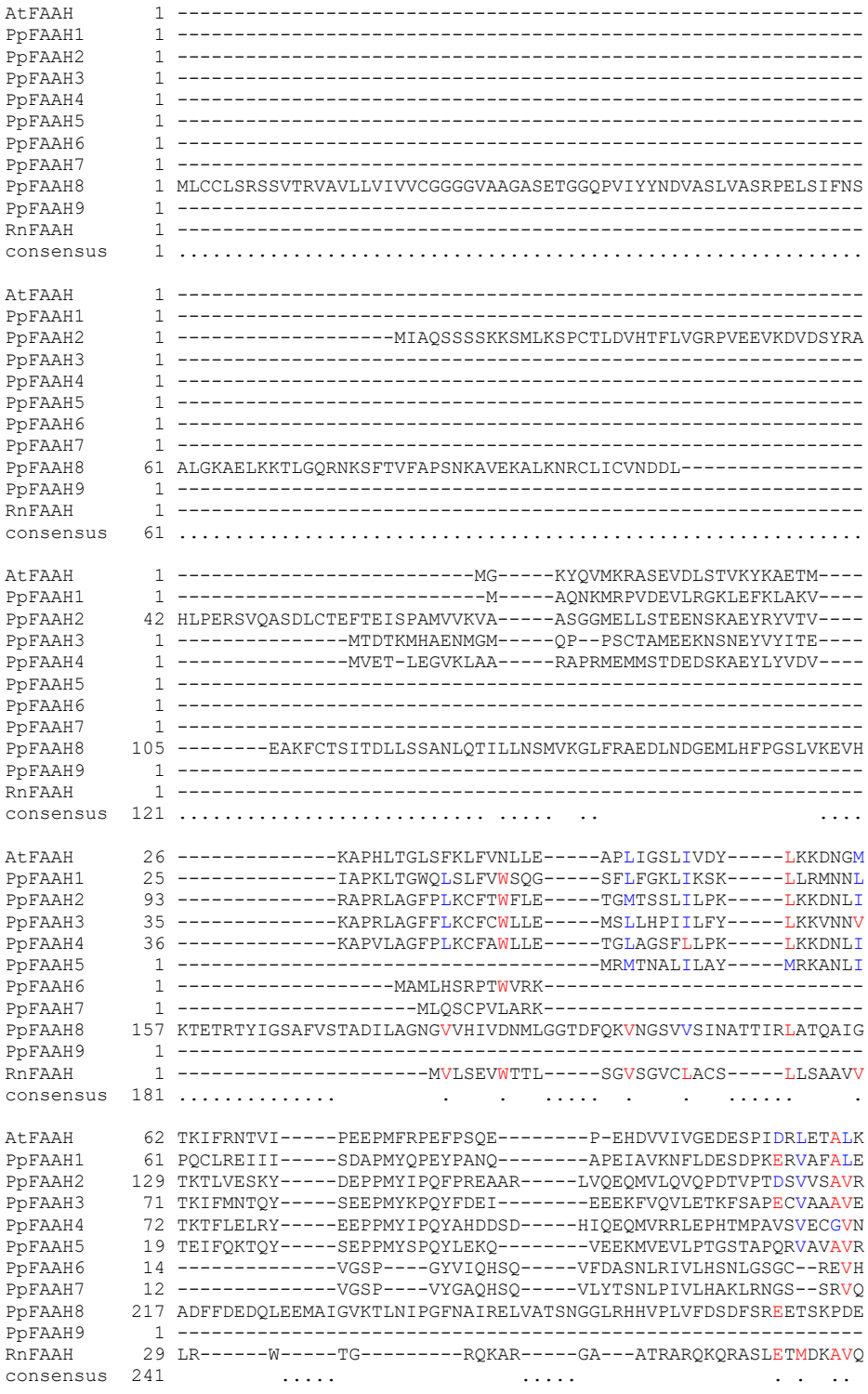


Fig. S1. Amide hydrolase activity and representation of radio-chromatograms. A). Western blot of purified GST-tagged PpFAAH1 (~100 KDa) and GST protein (~30 KDa) alone (as a control). B). Hydrolysis of anandamide by FAAH generates a free fatty acid, arachidonic acid and ethanolamine as products. The substrate is radiolabeled with ^{14}[C] and upon hydrolysis, the radiolabel is retained by the free fatty acid product, allowing for quantification of FAAH activity. C). Representative chromatograms generated by the TLC bio-scanner. Substrate peaks are retained around 50 mm and product at ~100 to 150 mm from the point of origin on TLC plate. i) negative control with no enzyme and/or GST protein, ii) positive control using AtFAAH as the enzyme source, and iii) and iv) are with substrate NAE 20:4 and NAE 16:0 using PpFAAH1 as the enzyme source. Peaks AA (iii) and PA (iv) represent free arachidonic acid palmitic acid as product of anandamide activity of PpFAAH1


AtFAAH 1 ----------
PpFAAH1 1 ----------
PpFAAH2 1 ----------
PpFAAH3 1 ----------
PpFAAH4 1 ----------
PpFAAH5 1 ----------
PpFAAH6 1 ----------
PpFAAH7 1 ----------
PpFAAH8 1 MLCCLSRSSVTRVAVLLVIVVCGGGGVAAGASETGGQPVIYYNDVASLVASRPELSIFNS
PpFAAH9 1 ----------
RnFAAH 1 ----------
consensus 1 ....
AtFAAH 1 ----------
PpFAAH1 1 ----------
PpFAAH2 1 ----------MIAQSSSSKKSMLKSPCTLDVHTFLVGRPVEEVKDVDSYRA
PpFAAH3 1 ----------
PpFAAH4 1 ----------
PpFAAH5 1 ----------
PpFAAH6 1 ----------
PpFAAH7 1 ----------
PpFAAH8 61 ALGKAELKKTLGQRNKSFTVFAPSNKAVEKALKNRCLICVNDDL----------
PpFAAH9 1 ----------
RnFAAH 1 ----------
consensus 61 ....
AtFAAH 1 ----------MG-----KYQVMKRASEVDLSTVKYKAETM----
PpFAAH1 1 ----------M-----AQNKMRPVDEVLRGKLEFKLAKV----
PpFAAH2 42 HLPERSVQASDLCTEFTEISPAMVVKVA-----ASGGMELLSTEENSKAEYRYVTV----
PpFAAH3 1 ----------MTDTKMHAENMGM-----QP--PSCTAMEEKNSNEYVYITE----
PpFAAH4 1 ----------MVET-LEGVKLAA-----RAPRMEMMSTDEDSKAEYLYVDV----
PpFAAH5 1 ----------
PpFAAH6 1 ----------
PpFAAH7 1 ----------
PpFAAH8 105 --------EAKFCTSITDLLSSANLQTILLNSMVKGLFRAEDLNDGEMLHFPGSLVKEVH
PpFAAH9 1 ----------
RnFAAH 1 ----------
consensus 121 ....
AtFAAH 26 ----------KAPHLTGLSFKLFVNLLE-----APLIGSLIVDY-----LKKDNGM
PpFAAH1 25 ----------IAPKLTGWQLSLFVWSQG-----SFLFGKLIKSK-----LLRMNNL
PpFAAH2 93 ----------RAPRLAGFPLKCFTWFLE-----TGMTSSLILPK-----LKKDNLI
PpFAAH3 35 ----------KAPRLAGFFLKCFCWLLE-----MSLLHPIILFY-----LKKVNNV
PpFAAH4 36 ----------KAPVLAGFPLKCFAWLLE-----TGLAGSFLLPK-----LKKDNLI
PpFAAH5 1 ----------MRMTNALILAY-----MRKANLI
PpFAAH6 1 ----------MAMLHSRPTWVRK----------
PpFAAH7 1 ----------MLQSCPVLARK----------
PpFAAH8 157 KTETRTYIGSAFVSTADILAGNGVVHIVDNMLGGTDFQKVNGSVVSINATTIRLATQAIG
PpFAAH9 1 ----------
RnFAAH 1 ----------MVLSEVWTTL-----SGVSGVCLACS-----LLSAAVV
consensus 181 .....
AtFAAH 62 TKIFRNTVI-----PEEPMFRPEFPSQE--------P-EHDVVIVGEDESPIDRLETALK
PpFAAH1 61 PQCLREIII-----SDAPMYQPEYPANQ--------APEIAVKNFLDESDPKERVAFALE
PpFAAH2 129 TKTLVESKY-----DEPPMYIPQFPREAAR-----LVQEQMVLQVQPDTVPTDSVVSAVR
PpFAAH3 71 TKIFMNTQY-----SEEPMYKPQYFDEI--------EEEKFVQVLETKFSAPECVAAAVE
PpFAAH4 72 TKTFLELRY-----EEPPMYIPQYAHDDSD-----HIQEQMVRRLEPHTMPAVSVECGVN
PpFAAH5 19 TEIFQKTQY-----SEPPMYSPQYLEKQ--------VEEKMVEVLPTGSTAPQRVAVAVR
PpFAAH6 14 ----------VGSP----GYVIQHSQ-----VFDASNLRIVLHSNLGSGC--REVH
PpFAAH7 12 ----------VGSP----VYGAQHSQ-----VLYTSNLPIVLHAKLRNGS--SRVQ
PpFAAH8 217 ADFFDEDQLEEMAIGVKTLNIPGFNAIRELVATSNGGLRHHVPLVFDSDFSREETSKPDE
PpFAAH9 1 ----------
RnFAAH 29 LR------W-----TG---------RQKAR-----GA---ATRARQKQRASLETMDKAVQ
consensus 241 ....

```
AtFAAH      108 CLPQYDPSRSL-----HADPVSSFRYWKIRDYAYAYRSKLTTPLQVAKRIISIIEEFGY-
PpFAAH1     108 HLPAR---QSH-----FQDDSPPFLYWKIRDYVEAYTTGRVTPTQVAERIIATVEEAKK-
PpFAAH2     179 CLPPQALDKVYNE---VEDKQYSFRHWTIRDYAQAYISGRLTPIQVAERFISAVEDSQN-
PpFAAH3     118 CLPPYLSNTLTYCGKEDQIDVRKFQYATIRDYADAYTSGRVTPTEVAKRFLTSIEDSRK-
PpFAAH4     122 CLPPYVARNINKP----VGGVKGFGHRTIRDYAHAYSSGRITPTQVAEQFISAIEDSRK-
PpFAAH5      66 CLPPYVPKPSINGGNHDHIDAQKFQYATVRDYAHAYSSGRFTPTQVAERFLAAVEETQK-
PpFAAH6      49 RSATP--SKSVSQ---CSEGAQQRPVSVIRKIRESLVRREKSAVEVTEAYLVKLRE----
PpFAAH7      47 CSTAA--STTMSE---SSLASRQSPVSVIRKIRESLVTGEKSAVEITETYLEKLKE----
PpFAAH8     277 KFKYPALEDVQ-----RPKSDEDLAFMSVLQLGSLIKSKKVTSVELVKLYIARLKKVDY-
PpFAAH9       1 ---------------MPND---KIL---ARSL------GVLFGLGLAGIFILRRNLFRLT
RnFAAH       61 RFRLQ------N----PDLDSEALLTLPLLQLVQKLQSGELSPEAVFFTYLGKAWEVNK-
consensus   301              ...           .       .. ...  .   ..    .   .

AtFAAH      162 DKPPTPFLIRFDANEVIKQAEASTRRFEQGNPISVLDGIFVTIKDDIDCLPHPTNGGT-T
PpFAAH1     159 RSPSLVYFISFLSEDVRKQATESTERYKKGTALSVLDGVPLAVKDDIDCLPHPSTAGT-K
PpFAAH2     235 --QGMNLFSAMDSRDVLAQAAESAARYKKGQPLSVLDGVPIAVKDEIDCLPYATTGGT-T
PpFAAH3     177 MSRGLNLFISLDFHDVISQAAAATERYRQGKPLSVLDGVLVAVKDEIDCLPYPTTGGS-T
PpFAAH4     177 --AGMNLFIAMDSGDVLSQAAAATERYKQGKPLSVLDGVPIAVKDEIDCLPYRTTGGT-T
PpFAAH5     125 MSPALNLFIPLDAQNVLSQAAASTERYREGGQLSVLDGILIAVKDEIDCLPYPTTGGC-T
PpFAAH6     100 QEPHVRSFLH-VSEKALQDAQELDRRVAEGREIGPLAGVPLGVKDNLCTRDMPSTGGS-R
PpFAAH7      98 REPHVRSFLH-VSEKALQDAQELDRRIAEGGEVGPLAGVPLGVKDNLCTRDMPSTGGS-Q
PpFAAH8     331 V---LKAVVTFTEELALEQAVAADRLLTKGVYLGPLHGIPYGLKDLFAVPGYRTTWGS-N
PpFAAH9      34 GRKDNGAFIEYFE--LL-----PPPPPPPPSAPHPLSGLTFAIKDIFDIEGFVTGFGNPD
RnFAAH      110 GTNCVTSYLT-D----------CETQLSQAPRQGLLYGVPVSLKECFSYKGHDSTLGL-S
consensus   361     .  ..  .                .    .* *... .*.        .. * ..

AtFAAH      221 WLHEDRSVEKDSAVVSKLRSCGAILLGKANMHELGMGTTGNNSNYGTTRNPHDPKRYTGG
PpFAAH1     218 WLPQVREVKEDAVSIARLRSCGMMMIGKAVMHELGMGTTGSNPHHGTARNPHDLGRYTGG
PpFAAH2     292 WLGEVRQTKDDAQAVKCLRSCGAVMVGKTNMHELGMGTTGINPHYGATRNPYDKTRASGG
PpFAAH3     236 WLGKARQVTKDAAVVKRLRECGAVMVGKTNMHELGVGTTGINPHYGATRNPHDMTRVSGG
PpFAAH4     234 WLGKVREVKEDAEAVKRLRSCGAVMVGKTNMHELGMGTTGINPHYGATRNPHNILRVSGG
PpFAAH5     184 WLDKTRQVTEDAAVVKRLRECGAIMVGNTNMHELGAGTTGINPHYGTTRNPHDRTRISGG
PpFAAH6     158 ILQGYY-PPYDATAVAKLRSSGAILVGKTNMDEFGMGSSTEGSAYQVTANPWDLSRVPGG
PpFAAH7     156 ILRGYC-PPYDATAVAKLRSSGAILVGKTNMDEFGMGSSTEGSAYQVTSNPWDLSRVPGG
PpFAAH8     387 LFKNQV-INKESWVYQKLKAAGAVLIAKLAAGSLAWDD---VWFGGQTKNPWNIYEGSTG
PpFAAH9      87 WASTHEPATRTAAAVKVLVEAGATCIGKLIMDELAYSIIGDNKHYGTPVNPAAPNRIPGG
RnFAAH      158 LNEGMP-SESDCVVVQVLKLQGAVPFVHTNVPQSMLSFDCSNPLFGQTMNPWKSSKSPGG
consensus   421           .   .  *.  *..    ...    .     .. .. . **   ..  .*

AtFAAH      281 SSSGSAAIVAAGLCSAALGTDGGGSVRIPSALCGITGLKTTYGRTDMTGSLCEGGTV---
PpFAAH1     278 SSSGPGAIVASGLCPVALGTDVGGSIRIPSSLCGNIGLKPTFGRITNEGLFCVGWSM---
PpFAAH2     352 SSGGSAAAVAAGICPAALGVDGGGSVRMPAGLCGVVGLKPTFGRTSNVGVLPLNWTV---
PpFAAH3     296 SSGGSATVVAAGLCPVALGVDGGGSVRMPASLSGVVGLKPTFGRIAKSGLLPLNWTI---
PpFAAH4     294 SSGGSSAAVAAGICPAALGVDGGGSVRMPAGLCGVVGFKPTFGRTSNAGVLPLNWTV---
PpFAAH5     244 SSGGSAATVAAGLCPIALGVDGGGSVRMPASLCGIVGLKPTFGRITKSGLLPLNWTI---
PpFAAH6     217 SSGGSAAAVAAGQCAGALGSDTGGSIRQPASFCGVVGLKPSYGRVSRFGLMAYASSL---
PpFAAH7     215 SSGGSAAAVAAGQCAGALGSDTGGSIRQPASFCGVVGLKPSYGRVSRFGLMAYASSL---
PpFAAH8     443 SSAGPAASTCAGNVPFAIGTETVGSITLPSSRTGITGLRPTFGMVGRSWAMSLSESL---
PpFAAH9     147 SSSGSGVAVAADLVDFSLGTDTAGSVRVPAAFCGILGFRPSHGAVSTVGVTPMAQSL---
RnFAAH      217 SSGGEGALIGSGGSPLGLGTDIGGSIRFPSAFCGICGLKPTGNRLSKSGLKGCVYGQTAV
consensus   481 **.* .. .. .  . ..*.. .**.. *   .*. *....  .... ..    .

AtFAAH      338 -EIIGPLASSLEDAFLV------YAAILGSSSADRYNLKP-----SPPCFPKLLSH----
PpFAAH1     335 -ESSGPIASSTEDALLV------YAAMVGSHPTDKLHSWP-----LPPCVPDLNVG----
PpFAAH2     409 -GMLGTLTGTVEDALIM------YAAIQGALPHDHIVSFP-----PPANFPLLMDSHEE-
PpFAAH3     353 -GMVGTLTGTVEDAYIV------YAAVQGHLPSDKLVSIP-----PPATLPLLNDVQME-
PpFAAH4     351 -GMLGTLTGSVEDALIM------YAAMHGPRPDDDIVSFP-----PPANLPLLKDPDEA-
PpFAAH5     301 -GMLGTLTGSVEDAYII------YAAKQGHLPSDKLLSFP-----PPATLPLLNDSQTE-
PpFAAH6     274 -DVVGCFGNSVMDAAIL------LEAIAGADKNDSTCS--------PQDVPDYTSNLLP-
PpFAAH7     272 -DVVGCFGNSVMDAAIL------LEAIAGADKNDSTCS--------PQGVPDYTSKLLP-
PpFAAH8     500 -DKVGPLCRYAADCALV------LDAIRGKDPRDWSSKNIHLEDPFSVDITKLTVGYLPD
PpFAAH9     204 -DTVGCFARDPAILRQVGHILLQLPYMDVRQPRRFFIA--------DDCFKI--------
RnFAAH      277 QLSLGPMARDVESLALC------LKALLCEHL------FT-----LDPTVPPLPFREE--
consensus   541    .* ..  ..   . ......  ..             .....    .. .      .
```

```
AtFAAH      382 ---NGSNAIGSLRLGKYTKWFNDVSSSDISDKCEDILKL---LSNNHGCKVVEIVVPELE
PpFAAH1     379 ----FQAIMGSLKLGKYSDWFNSTFEKEVAEVCNKSLNL---IFETFGTETKEIILPELD
PpFAAH2     456 -TMRRGKLMGNLKFAKFSEWFNDSDE-PIRKACCRAVRL---VQQLYDTQVVEVTIPELE
PpFAAH3     400 -PNIMAKLIEEIKLAKFSKWYNDSDD-SVWRVCDSALRL---IQDTYGCKVVDASIPDLD
PpFAAH4     398 -STNTAKVMGDLKFAKFSKWFDDCDE-PVRNACHRALQL---VQTTFNTKVVEVTIPEIE
PpFAAH5     348 -STNMAKLIGDVKFAKFPKWFNDTDD-PVGRICDKALQL---VQGTYGCKVVDVSIPELQ
PpFAAH6     318 IDDLGSKPLAGIRFGIISETIADGVVEDVLSAVKQAVT----HLESLGASVREVAMPNFS
PpFAAH7     316 VDNLDSRPLAGIRFGIITETVGDGVDEDVVSAIRQAAA----HLESLGASVREVAMPNFS
PpFAAH8     553 ADMGVVKALAGLKVKMVPFKLNYTVPSA-------------------------------E
PpFAAH9     247 -----SLIPTELSLGTVVKSIQK--------LLGRQVLQ---H-INLGD-YVARTVPSLK
RnFAAH      318 ----VYRSSRPLRVGYYETD-NYTMPSP---AMRRALIETKQRLEAAGHT----LIPFLP
consensus   601 .     .    .... .    . .          ...          .       .. .

AtFAAH      436 EMRAA-H-VISIGSP-----TLSSLTPYCEAG-KNSK-----------------LSYD--
PpFAAH1     432 EIKIA-H-LCTVGSE-----YVSGLAGYDTK--LLKE-----------------LSLE--
PpFAAH2     511 EMRLA-H-FVTIGSE-----CFTSLGMDYQQS-GLEA-----------------SGGD--
PpFAAH3     455 KMRLA-H-YITMGGE-----LTASLGVEYENL-GRKT-----------------TGGD--
PpFAAH4     453 EMRLA-H-FVTIGSE-----CCTSLGVDYRES-GLKA-----------------SGAD--
PpFAAH5     403 AMRLA-H-YITIGSE-----CSAALGVQYENV-GQKV-----------------SGGD--
PpFAAH6     374 LGLPA-Y-YVLATSE-----ASSNLARYDGVRYGPRAHGEEVMSMYANSRAL-GFGPEVK
PpFAAH7     372 LGLPA-Y-YVLATSE-----ASSNLARYDGVRYGPRAHGEEVMSMYGNSRAQ-GFGSEVK
PpFAAH8     582 FI----M-NVTMGVD-----VLSHFDNWQRAG--LDV-------GYEDQT---AWPVELR
PpFAAH9     289 ELQKE-ISDSNLGSLALLRTAMQILQRWEFK-----L----------------NHEEWLT
RnFAAH      366 NNIPYALEVLSAGG------LFSDGGRSF----LQNFKGDFVDPCLGDLILILRLPSWFK
consensus   661          .. .  .....  .  .      .

AtFAAH      469 ------TRTSFA--IFRSFSASDYI-AAQCLRRRLMEYHLNI----------FKDVDVIV
PpFAAH1     464 ------VQGTLG--LFKEFTALDYL-SSQKIRRRAMYFYMQA----------FESVDVIV
PpFAAH2     544 ------VRVGFS--IYESFNSREFI-AAQQMRFRQMHYHNEI----------FKRADIII
PpFAAH3     488 ------VRSTIS--IFQGFSNREFV-TAQRLRSRCMQHHMEV----------FKEADFIV
PpFAAH4     486 ------VRVTSS--IYGSFNNREFI-GAQRMRFRQMHYHNEI----------FKKANIII
PpFAAH5     436 ------VRSTIT--IFQCFNNREYV-TAPRIRSRCMQHHMDI----------FEKADFIV
PpFAAH6     426 R---RILMGTYA--LSAGYYDAFYK-KAQQVRTVIQQDFKNA----------LEEVDLLI
PpFAAH7     424 R---RILMGTYA--LSAGYYDAFYK-KAQQVRTVIQQDFKNA----------LQEVDLLI
PpFAAH8     620 --------------RARLISAVDYV-QAQRARGLLATEVRHV--------IESQKVDAFI
PpFAAH9     327 TVKPDLAPALAARTKLALETSSNLVPLLQKIKDETRYAISEL----------LKNDSLLV
RnFAAH      416 RLLSLLLKPLFP--RLAAFLNSMRPRSAEKL--WKLQHEIEMYRQSVIAQWKAMNLDVLL
consensus   721       ..    ..    .        . ..    .    ..             .....

AtFAAH      510 TPTTGMTAPVIPPDALKNGETNIQVTTDLMRFVLAANLLGFPAISVPVGYD---------
PpFAAH1     505 TPTTGTTAPVISSAALTVGESDLTTVGNLMRFIIAPNFLGLPAISVPVGHD---------
PpFAAH2     585 TPTTGATAPLLRRNAENCGELDYGLGAKLMRFLIAGNFLGLPAISVPIGHD---------
PpFAAH3     529 TPTTACTAPPIRDDAEQYGELDYQHGGKLMRFIIAGNLLGLPAITLPVGYD---------
PpFAAH4     527 SPTTGATAPLIRRAAEKCGELDYVVGAKLMRYQIAGNFLGLPALSVPVGHD---------
PpFAAH5     477 TPTTACTASPVREVAEKYGELDYQNGG-------------LPAISIPVGYD---------
PpFAAH6     470 SPVTPTAAYKIGE---KVNDPLAMYVGDLMT--VNVNLAGLPALVVPCGLA---------
PpFAAH7     468 SPVAPTPAYKIDE---KMNDPLAMYVGDLMT--VNVNLAGLPALVVPCGLA---------
PpFAAH8     657 ------------------------GNSTDWERVCVGNLVGMPVIVIPTGFKKIN------
PpFAAH9     377 MPTVPDIPPKLNTKAEAL----EVFRNKTLDLICVAGMSSCCQVTMPA------------
RnFAAH      472 TPMLGPALDLN--------TPG--RATGAISYTVLYNCLDFPAGVVPVTTVTAEDDAQME
consensus   781 ..  .                       . . .  . .  ..  .*.

AtFAAH      561 -----------------K---EGLPIGLQIMGRPWAEATVLGLAAAVEELA-----PVTK
PpFAAH1     556 -----------------S---RGLPIGLQLIGRPWQEATLFRVAAAFEEVCT----PLRK
PpFAAH2     636 -----------------E---DGLPIGLQLIGRPWSEATLLHVAAVIEVCTL----ILHT
PpFAAH3     580 -----------------A---KGLPVGLQLVGKPWSEAALLRVAVAFEKICA----PQH-
PpFAAH4     578 -----------------A---DGLPIGMQLIGRPWSEATLLQVAAVIERLCS----PFQR
PpFAAH5     515 -----------------D---KGLPIGLQLIGRPWSEATLLRMAIAFEKICA----PQL-
PpFAAH6     516 -----------------KGGTSGLPVGLQMIGPAFREESILRPGHIFEQTLP----TSFS
PpFAAH7     514 -----------------KGGTCGLPVGLQMIGPAFGEDNLLRLGHIFEQTLP----TCFS
PpFAAH8     687 -----------------GTRRCTTVQTGIYAAPYQDGTVLALAMAYQAVTS----HHLQ
PpFAAH9     421 -----------------GNHDGVPMAVSLLARQGSDRFLLDTVLAIYSTVQEEDKVAAD
RnFAAH      522 LYKGYFGDIWDIILKKAMKNSVGLPVAVQCVALPWQEELCLRFMREVEQLMT----PQKQ
consensus   841                       ..... .. .. .   ..     . .  .....
```

```
AtFAAH     596 KPAIFYDILNTN-------------------------------------------------
PpFAAH1    592 RATTFYNILK---------------------------------------------------
PpFAAH2    672 SR-----------------------------------------------------------
PpFAAH3    615 RPKVYYDLLN---------------------------------------------------
PpFAAH4    614 RPEVLYDLLSTP-------NK----------------------------------------
PpFAAH5    550 QPEVYFDLLK----------------------------------------------------
PpFAAH6    555 APTLVS--------------------------------------------------------
PpFAAH7    553 APTLLGKQ------------------------------------------------------
PpFAAH8    725 RPPVDN--LGPD-------EDSS--------LKQFF-------------------------
PpFAAH9    463 QPSIVSDGNSAAAELAKEKGNAAFKEKDYKKAVGFYTDAIRLNGNNATYYNNRAMAYLQL
RnFAAH     578 PS-------------------------------------------------------------
consensus  901           ..............................................

AtFAAH         ------------------------------------------------------------
PpFAAH1        ------------------------------------------------------------
PpFAAH2        ------------------------------------------------------------
PpFAAH3        ------------------------------------------------------------
PpFAAH4        ------------------------------------------------------------
PpFAAH5        ------------------------------------------------------------
PpFAAH6        ------------------------------------------------------------
PpFAAH7        ------------------------------------------------------------
PpFAAH8        ------------------------------------------------------------
PpFAAH9    523 CSFSEAESDCTKALNLDKRSVKAYLRRGTAREFLGYYKEADEDFRQALIFEPTNKTASEA
RnFAAH         ------------------------------------------------------------
consensus  961 ............................................................

AtFAAH         ----------
PpFAAH1        ----------
PpFAAH2        ----------
PpFAAH3        ----------
PpFAAH4        ----------
PpFAAH5        ----------
PpFAAH6        ----------
PpFAAH7        ----------
PpFAAH8        ----------
PpFAAH9    583 LSRLKKLLYG
RnFAAH         ----------
consensus 1021 ..........
```

Figure S2. Alignment of full-length sequences of nine PpFAAH paralogs with AtFAAH and RtFAAH. Shortened or extended N and C termini along with conserved amidase signature region among the proteins can be compared. The symbols: asterisk, dot and gap for the consensus sequence indicate identical, same class and different class of residues at the same position, respectively. Red, green, and black colors also represent the same order of consensus symbols in terms of conserved residues

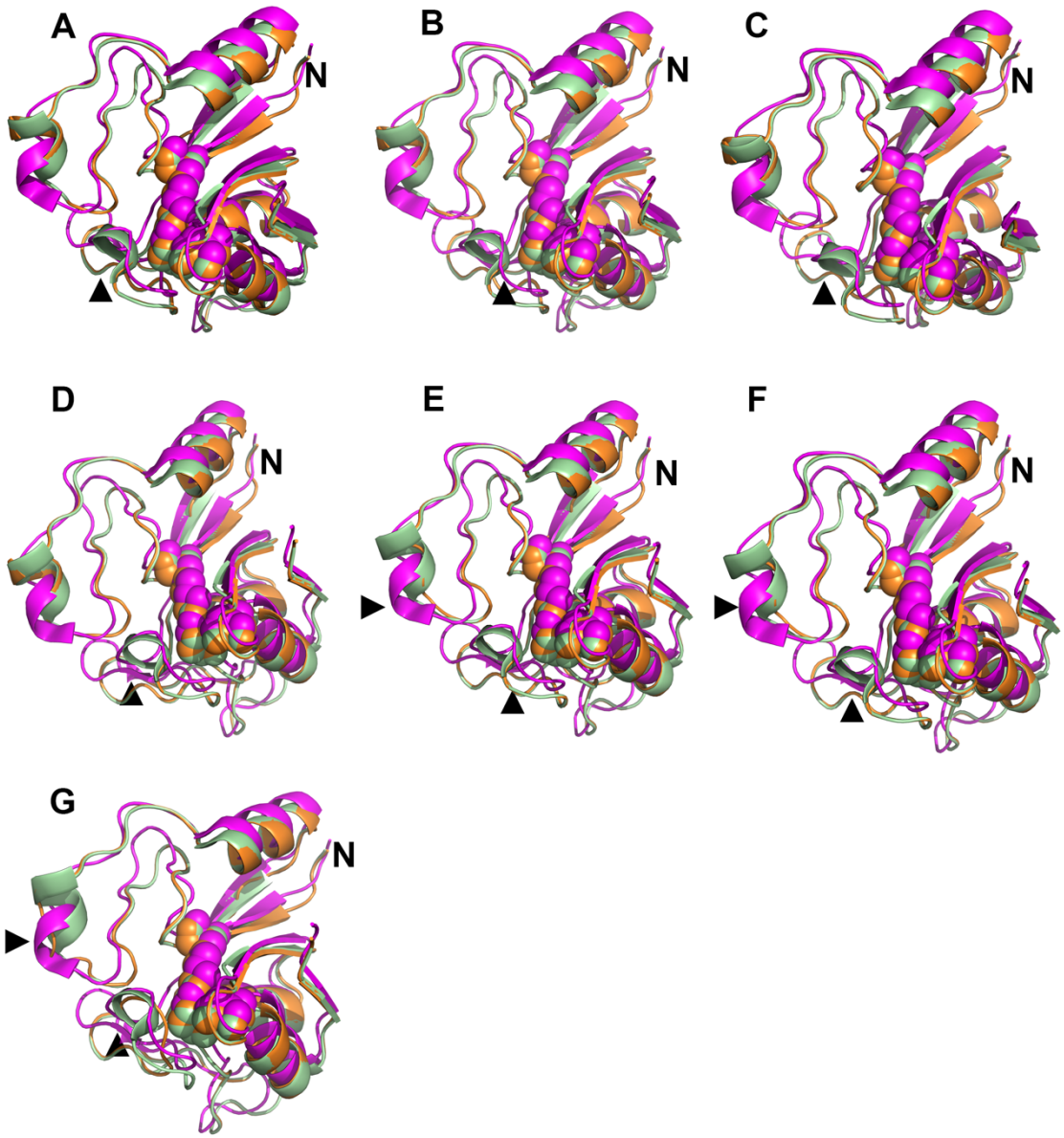


Figure S3. Structural alignment of amidase signature (AS) region. Amidase signature region of AtFAAH (pale green), rFAAH (magenta) and PpFAAH (orange) were done using Chimera 5 software. A – G represents PpFAAH2 to PpFAAH8 aligned with both At and RtFAAH. Catalytic triad (Lys-Ser-Ser) are shown as sphere. N represents the N-terminus region of AS sequence. Arrows points to differences in alignment of helices and loops

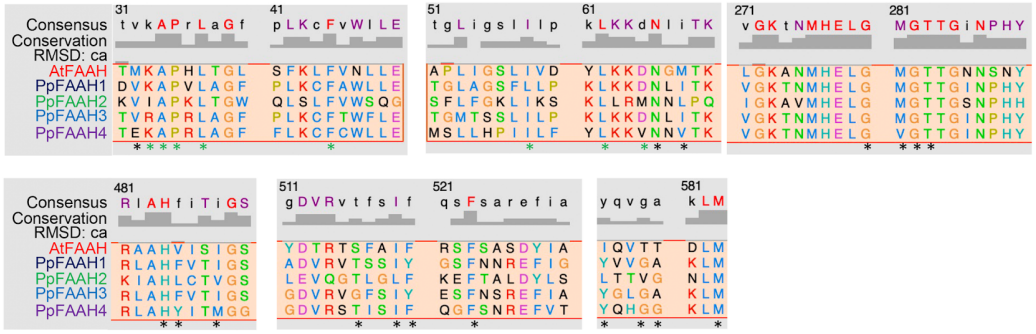


Figure S4. Structural alignment of membrane binding cap (MBC) and membrane access channel (MAC). Predicted structures of PpFAAH1 to PpFAAH4 were aligned with AtFAAH to determine MBC and MAC of PpFAAH. Green and black asterisks at the bottom of the sequence represents the important residues that make the MBC and MAC, respectively. Shadow height shown above the sequences indicates the conservation of residues, whereas the numbers indicate the consensus alignment of the residues

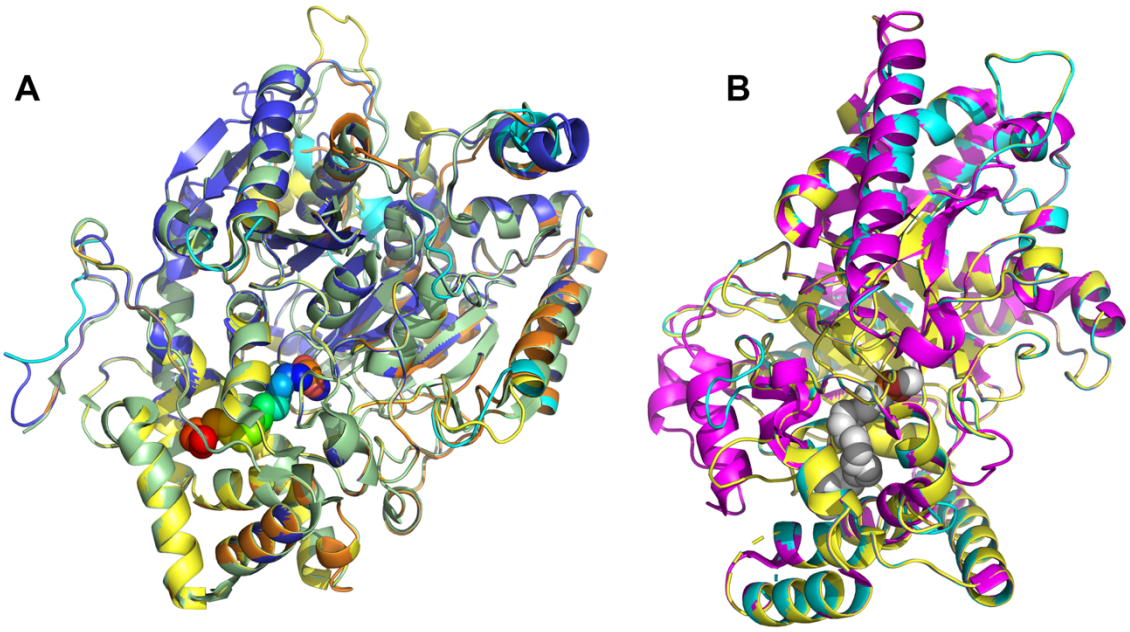


Figure S5. Substrate docking of PpFAAH. A). Overlaid structural alignment of PpFAAH1 to PpFAAH4 with At FAAH1 as a template (pale green) and B). PpFAAH6 and PpFAAH7 with Rt substrate MAFP. Protein structures are shown as cartoon with the substrate as sphere

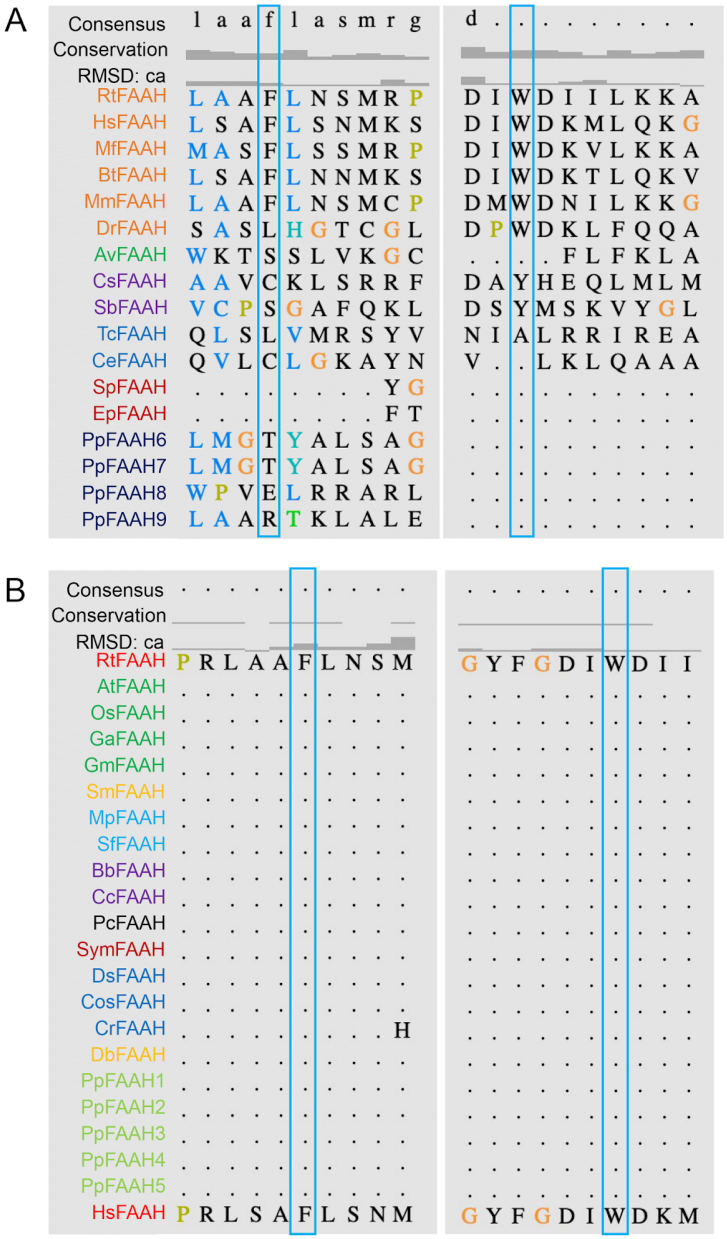


Figure S6. Sequence alignment of dynamic paddle residues. Predicted structural sequence alignment of template RtFAAH (1mt5) with FAAH from Kingdom A). Animalia and PpFAAH6 to PpFAAH9, and B). Plantae, Fungi, Chromista, Protozoa and PpFAAH1 to PpFAAH5 using Chimera 5 to determine the potential dynamic paddle. Blue boxes in the alignment shown the residues that potentially can make the dynamic paddle. Shadow height shown above the sequences indicates the conservation of residues, whereas the numbers indicate the consensus alignment of the residues. For description of protein names, see Table S2

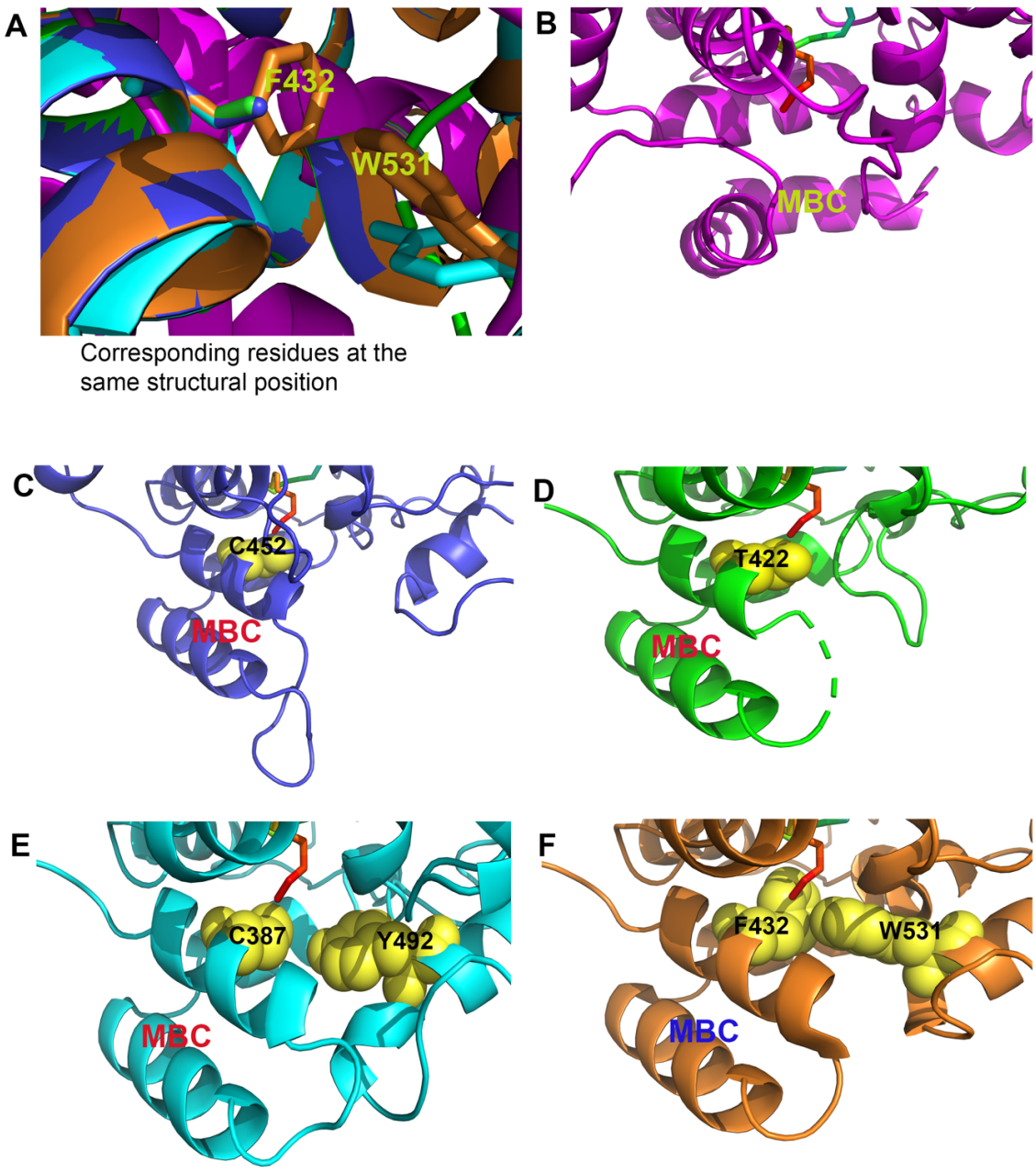


Figure S7. Comparison of dynamic paddle region of FAAH. A). Overlay of secondary structures of dynamic paddle region with F432 and W531 residues of RtFAAH with EpFAAH, CeFAAH, CsFAAH and PpFAAH6. Corresponding residues for the orthologs at the position of F432 and W531 are indicated by red arrows; Close-up images of probable dynamic paddle region of B). EpFAAH, C). CeFAAH, D). PpFAAH6, E). CsFAAH, F). RtFAAH. Images B to E also show the substrate MAFP (sticks, rainbow in color), residues of dynamic paddle region (sphere, yellow in color), and the two α-helices that are predicted to make MBC

Table S1. Details of PpFAAH paralog sequences along with AtFAAH, RtFAAH and HsFAAH

Name	Gene ID	Chromosome	Protein		Isoelectric point	Motif prediction (AA)	% Identity with		
			Length (AA)	Size (Kda)			AtFAAH	RtFAAH	HsFAAH1
AtFAAH	AT5G64440 (TAIR)	5	607	66.08	6.18	Amidase (198-321)	100	34	33
PpFAAH1	Pp3c23_9920V3.11	23	601	65.9	6.84	Amidase (195–318)	47	30	28
PpFAAH2	Pp3c3_2980V3.1	3	673	73.5	5.93	Amidase (269–392)	46	33	34
PpFAAH3	Pp3c7_18330V3.6	7	624	68.5	6.35	Amidase (213–336)	45	35	35
PpFAAH4	Pp3c26_13660V3.1	26	627	68.06	6.92	Amidase (211–334)	45	33	35
PpFAAH5	Pp3c11_92003.2	11	559	60.78	6.57	Amidase (161–284)	44	31	31
PpFAAH6	Pp3c27_1950V3.5	27	560	59.4	6.77	Amidase (135–257)	30	35	33
PpFAAH7	Pp3c16_16670V3.1	16	560	58.99	6.22	Amidase (133–255)	31	33	33
PpFAAH8	Pp3c4_17250V3.2	4	743	80.45	6.84	Amidase (364–483), Fascilin	26	28	28
PpFAAH9	Pp3c21_15890V3.1	21	592	64.35	7.96	Amidase (63–187), Tetratricopeptide	39	32	32
RtFAAH	100911581 (NCBI)	5	579	63.35	8.49	Amidase (135-257)	34	100	82
HsFAAH	2166 (NCBI)	1	579	63.07	7.82	Amidase (135-257)	31	82	100

Table S2. Percent identity matrix for PpFAAH paralogs, along with AtFAAH and RtFAAH

%*	AtFAAH1	PpFAAH1	PpFAAH2	PpFAAH3	PpFAAH4	PpFAAH5	PpFAAH6	PpFAAH7	PpFAAH8	PpFAAH9	RtFAAH
AtFAAH1	100.0	46.4	44.4	44.3	45.0	45.3	26.3	26.9	24.5	22.6	22.5
PpFAAH1	46.4	100.0	43.5	44.2	43.7	46.3	29.6	29.8	21.8	21.4	24.3
PpFAAH2	44.4	43.5	100.0	59.5	56.6	71.6	27.7	26.9	22.4	23.9	22.2
PpFAAH3	44.3	44.2	59.5	100.0	71.8	61.1	28.0	28.2	19.2	23.8	22.0
PpFAAH4	45.0	43.7	56.6	71.8	100.0	58.6	28.2	28.7	20.3	23.8	21.7
PpFAAH5	45.3	46.3	71.6	61.1	58.6	100.0	28.9	29.5	22.0	23.3	23.5
PpFAAH6	26.3	29.6	27.7	28.0	28.2	28.9	100.0	85.6	23.0	20.6	24.7
PpFAAH7	26.9	29.8	26.9	28.2	28.7	29.5	85.6	100.0	22.7	22.0	23.9
PpFAAH8	24.5	21.8	22.4	19.2	20.3	22.0	23.0	22.7	100.0	20.1	21.5
PpFAAH9	22.6	21.4	23.9	23.8	23.8	23.3	20.6	22.0	20.1	100.0	18.3
RtFAAH	22.5	24.3	22.2	22.0	21.7	23.5	24.7	23.9	21.5	18.3	100.0

* Created by Clustal2.1

Table S3. Details of the organisms of which FAAH was used for phylogenetic and 'Dynamic Paddle' analyses

Kingdom	Phylum/Division	Scientific Name	Code	Protein ID*
Animalia	Chordata	*Rattus norvegicus*	RtFAAH	NP_077046.1
Animalia	Chordata	*Homo sapiens*	HsFAAH	NP_001432.2
Animalia	Chordata	*Marmota flaviventris*	MfFAAF	XP_027793438.1
Animalia	Chordata	*Bos Taurus*	BtFAAH	XP_024845397.1
Animalia	Chordata	*Mus musculus*	MmFAAH	NP_034303.3
Animalia	Chordata	*Danio rerio*	DrFAAH	NP_001103295.1
Animalia	Arthropoda	*Armadillidium vulgare*	AvFAAH	RXG52942.1
Animalia	Platyhelminthes	*Clonorchis sinensis*	CsFAAH	RJW72397.1
Animalia	Platyhelminthes	*Schistosoma bovis*	SbFAAH	RTG91481.1
Animalia	Nematoda	*Toxocara canis*	TcFAAH	KHN72390 .1
Animalia	Nematoda	*Caenorhabditis elegans*	CeFAAH	Q17449
Animalia	Cnidaria	*Stylophora pistillata*	SpFAAH	PFX22943.1
Animalia	Cnidaria	*Exaiptasia pallida*	EpFAAH	KXJ22769.1
Plantae	Angiosperms	*Arabidopsis thaliana*	AtFAAH	AT5G64440.1
Plantae	Angiosperms	*Oryza sativa*	OsFAAH	XP_015633439.1
Plantae	Angiosperms	*Gossypium arboreum*	GaFAAH	XP_012462809.1
Plantae	Angiosperms	*Glycine max*	GmFAAH	XP_003545751.1
Plantae	Lycopodiophyta	*Selaginella moellendorffii*	SmFAAH	EFJ12565.1
Plantae	Bryophyta	*Marchantia polymorpha*	MpFAAH	PTQ37667.1
Plantae	Bryophyta	*Sphagnum fallax*	SfFAAH	Sphfalx0000s0679.2
Fungi	Ascomycota	*Beauveria bassiana*	BbFAAH	PMB70458.1
Fungi	Ascomycota	*Cladophialophora carrionii*	CcFAAH	OCT49863.1
Chromista	Heterokontophyta	*Phytophthora cactorum*	PcFAAH	RAW37580.1
Chromista	Domain: Eukaryota	*Symbiodinium microadriaticum*	SymFAAH	OLP90858.1
Chromista	Algae, Chlorophyta	*Dunaliella salina*	DsFAAH	Dusal.0428s00008.1
Chromista	Algae, Chlorophyta	*Coccomyxa subellipsoidea*	CosFAAH	XP_005650878.1
Chromista	Algae, Chlorophyta	*Chlamydomonas reinhardtii*	CrFAAH	PNW86966.1
Protozoa	Amoebozoa	*Dictyostelium discoideum*	DdFAAH	XP_643382.1
Plantae	Angiosperms	*Arabidopsis thaliana*	AtAmidase	At1G08980.1
Animalia	Chordata	*Homo sapiens*	HsAmidase	NP_777572

*Protein sequences were obtained from either NCBI, Uniport or Phytozome 12 database

Table S4. Details of predicted secondary structure of PpFAAH orthologs with AtFAAH and RtFAAH as templates

Protein	Template 6DII (AtFAAH)							Template 1MT5 (RtFAAH)						
	Aligned	Coverage (%)	Residues	# of α-helixes	# of β-sheets	Q score	RMSD	Aligned	Coverage (%)	Residues	# of α-helixes	# of β-sheets	Q score	RMSD
AtFAAH1	605	100	(1-605)	23	17	0.99	0.27	464	74	(100-591)	20	11	0.67	0.65
PpFAAH1	596	99	(4-601)	23	10	0.94	0.54	433	72	(127-582)	18	11	0.70	0.50
PpFAAH2	573	90	(83-666)	25	8	0.90	0.42	436	65	(133-600)	18	11	0.71	0.51
PpFAAH3	603	96	(11-624)	23	10	0.94	0.46	447	72	(145-598)	16	13	0.70	0.46
PpFAAH4	596	95	(17-623)	24	8	0.92	0.46	435	69	(199-665)	19	11	0.70	0.46
PpFAAH5	548	98	(1-559)	21	8	0.85	0.45	444	79	(81-545)	18	10	0.70	0.62
PpFAAH6	516	92	(2-558)	22	9	0.77	0.62	477	85	(52-539)	21	11	0.83	0.45
PpFAAH7	482	86	(46-556)	18	8	0.73	0.73	447	80	(41-541)	22	11	0.83	0.47
PpFAAH8	425	57	(293-728)	16	8	0.62	0.68	426	57	(291-718)	18	11	0.70	0.61
PpFAAH9	442	75	(6-454)	17	8	0.62	0.83	434	73	(4-446)	16	11	0.75	0.56
RtFAAH	440	80	(78-578)	18	8	0.75	0.38	537	100	(38-573)	22	11	1.00	0.00

Table S5. Predicted dimerization residues of PpFAAH1 to PpFAAH4, relative to AtFAAH

AtFAAH*	PpFAAH1	PpFAAH2	PpFAAH3	PpFAAH4
Gln5	Asn4	-	Pro15	-
Arg66	Arg65	Val133	Met75	Leu76
Thr68	Ile67	Ser135	Thr77	Leu78
Phe76	Tyr75	Tyr143	Tyr85	Tyr86
Asp225	Val222	Val296	Ala240	Val238
Thr454	Ala450	Gly529	Gly473	Gly471
Pro455	Gly451	Met530	Val474	Val472
Phe479	Phe474	Phe554	Phe498	Phe496
Ala481	Ala476	Ser556	Asn500	Asn498

*Substitution by same (green) or different class (red) of residue relative to AtFAAH

Table S6. Summary of predicted model for membrane binding cap

Protein	Terminus (AA position)	# of hydrophobic residues	# of helixes	Predicted model
AtFAAH1	N (27-60)	21/34	α1 and α2	Teeth on a comb
PpFAAH1	N (L21-P61)	24/41	"	"
PpFAAH2	N (A94-L127)	19/37	"	"
PpFAAH3	N (L39-L61)	19/23	"	TM and membrane integrated
PpFAAH4	N (A37-I71)	24/31	"	"
PpFAAH5	*Not available*			
PpFAAH6	C (A413-L435)	14/23	α18 and α19	Teeth on a comb
PpFAAH7	C (V422-F441)	14/20	"	"
PpFAAH8	C (A602-V616)	-	α12	"
PpFAAH9	C (V258-L271)	-	α9	"
RtFAAH	C (404-433)	23/34	α18 and α19	"

Table S7. Comparison of the residues at the entrance of ligand binding pocket

AtFAAH1*	PpFAAH1	PpFAAH2	PpFAAH3	PpFAAH4
Ala27	Ala26	Ala94	Ala36	Ala37
Pro28	Pro27	Pro95	Pro37	Pro38
Leu30	Leu29	Leu97	Leu38	Leu40
Phe38	Phe37	Phe108	Phe47	Phe48
Ile51	Ile50	Ile118	Ile60	Leu61
Leu55	Leu45	Leu122	Leu64	Leu65
Lys26	Ile25	Arg93	Lys35	Lys36
Asp58	Met57	Asp125	Val67	Asp68

*Substitution by same (green) or different class (red) of residue relative to AtFAAH

Table S8. Residues of substrate binding pockets in PpFAAH, relative to AtFAAH

AtFAAH1*	PpFAAH1	PpFAAH2	PpFAAH3	PpFAAH4	PpFAAH5
M25	V24	V92	E34	V35	-
A27	A26	A94	A36	A37	-
L55	L54	L122	L64	L65	M12
N59	N58	N126	N68	N69	N16
M61	L60	I126	V70	I71	I18
K205	K202	K276	K220	K218	K168
G255	G252	G326	G270	G268	G218
M256	M253	M327	V271	M269	A219
G257	G254	G328	G272	G270	G220
T258	T255	T329	T273	T271	T221
S281	S278	S352	S296	S294	S244
S305	S302	S376	S320	S318	S268
H441	H437	H516	H460	H458	H408
V442	L438	F517	Y461	F459	-
I445	V441	I520	M464	I462	I412
S472	T467	G547	T491	T489	T439
I475	L470	I550	I494	I492	-
F476	F471	Y554	F495	Y493	F443
F479	F474	F554	F495	F496	F446
I532	L527	Y607	Y551	Y549	-
T535	V530	G610	G554	G552	-
T536	G531	A661	G555	A553	-
M539	M534	M624	M558	M556	K493

*Substitution by same (green) or different class (red) of residue relative to AtFAAH

Table S9. Residues of substrate binding pockets in PpFAAH, relative to RtFAAH

RtFAAH*	PpFAAH6	PpFAAH7	PpFAAH8	PpFAAH9
K142	K140	K142	K371	K70
M191	G189	G191	A420	A121
L192	M190	M192	W421	Y122
S193	G191	G193	-	S123
F194	S192	S194	-	I124
G216	G214	G216	G442	G146
S217	S215	S217	S443	S147
T236	S234	S236	T462	T166
D237	D235	D237	E462	D167
I238	T236	T238	T464	T168
G239	G237	G239	V465	A169
G240	G238	G240	G466	G170
S241	S239	S241	S467	S171
Y335	G337	A339	-	Q269
L372	L374	L376	A580	R306
E373	P375	P377	E581	T307
S376	Y378	Y380	M584	Q310
A377	-	V381	N585	I311
L380	A381	-	M588	-
F381	T382	Y491	G589	-
L404	E405	E407	-	-
R428	-	-	A631	A335
A431	G429	G431	V616	A338
F432	T430	T432	E617	R339
T488	Y489	Y491	-	N399
G489	V490	V492	-	K400
I491	D492	D494	W662	L402
V495	V496	V498	C666	C406
W531	-	-	-	-

*Substitution by same (green) or different class (red) of residue relative to AtFAAH

CHAPTER 3. DIFFERENTIAL GENE EXPRESSION IN RESPONSE TO AN ENDOCANNABINOID IN *PHYSCOMITRELLA PATENS*

Imdadul Haq[1], Suhas Shinde[1], Michael Wang[2] and Aruna Kilaru[1]

[1]Department of Biological Sciences and Biomedical Sciences, East Tennessee State University, Johnson City, TN, 37614, USA

[2]BGI Genomics Co. Ltd., Yantian District, Shenzhen, 518083, China

kilaru@etsu.edu

Differential gene expression in response to an endocannabinoid in *Physcomitrella patens*

Introduction

Anandamide-mediated endocannabinoid signaling, and its physiological significance are well understood in mammalian systems. The role of anandamide or arachidonylethanolamide (AEA), a 20-carbon long polyunsaturated *N*-acylethanolamine (NAE) in early land plants that was discovered recently remains unresolved. In higher plants and animals, the metabolism of various NAEs is highly conserved with diverse biological functions.

The signaling pathway of anandamide is unknown in plants, partly because higher plants and well-studied plant model organisms do not synthesize endocannabinoids. Only early land plants were reported to synthesize endocannabinoids and their precursor arachidonic acid (AA)[1]. Typically, animals require linoleic acid (18:2) from external sources to be able to synthesize polyunsaturated fatty acids, including AA[2,3]. In addition to serving as a precursor of anandamide, AA is also an important component of phospholipid membrane of brain and skeletal muscle, mediator of cell signaling, blood clotting, and stress response[4]. Unlike higher plants, early land plants such as, mosses, hornworts, lycophytes, and liverworts synthesize high amounts of endogenous AA[4]; its function however, remains unclear. Recent studies showed that a tight regulation of AA is important in plant signaling and plays a role in biotic stress[5]. The ABA, unlike anandamide and its precursor, is more abundant throughout plant kingdom and is one of the most studied signaling molecules that is involved in diverse physiological pathways. In seed plants, ABA regulates growth and development, such as, seed dormancy, maturation and germination, seedling growth, stomatal regulation and flowering and senescence[6]; it is also a mediator of biotic and abiotic stresses[7,8]. For biotic stress, depending on the mode of pathogen entry and their infection, ABA could promote either resistance or susceptibility[7,8].

With emerging advancements on genome sequencing, RNA-seq has become a very important technique to identify global and specific molecular responses associated with growth, development, and biotic and abiotic stresses etc., in both plant and animal model systems. In plants, substantial experiments have been performed in order to identify specific gene expression pattern related to developmental stages, various biotic/abiotic stressor and their cross-relationship with phytohormones. Transcriptome analyses of ABA-inducible stress related gene expression in several model plant organisms showed an extensive and common involvement of

more than 50% of ABA responsive genes[9]. For example, in Arabidopsis, 63%, 54% and 10% genes related to drought, salinity and cold stresses, respectively are ABA-inducible[10]. In *P. patens*, in agreement with higher plants, a significant number of ABA-inducible genes are also stress-responsive[11]. Additionally, transcriptome studies also revealed differentially expressed genes among major developmental stages, protonema, gametophyte and sporophyte of *P. patens*[12].

Genomic studies in the last decade have established the significance of noncoding genome that was previously categorized as junk DNA, in directing numerous regulatory processes. In fact, the regulatory elements of noncoding genome, such as, cis- or trans-acting enhancers and promoters function in chromatin reprogramming and post-transcriptional regulation during RNA processing to dictate biological, physiological and developmental processes[13]. A high percentage of the genome is noncoding and generates long noncoding RNAs (lncRNAs), which are typically longer that 200nt with low protein-coding potential. For instance, 68% of the human genome is responsible for lncRNAs, of which 80% of them remain unannotated. Expression patterns are however, often correlated between mRNA and lncRNA, suggesting co-regulation of certain gene expression networks[14]. Understanding the function of lncRNA in mammals has led to development of mechanisms to treat breast[15], prostrate[16] and pancreatic[17] cancers, diabetes[18] and Alzheimer's[19]. In plants, a regulatory role for lncRNA in response to biotic and abiotic stress has emerged. In Arabidopsis, expression of thousands of natural antisense transcript (NATs)-lncRNAs were noted in response to stress and are required for cognate coding expression of sense genes[20]. Hundreds of novel and known lncRNAs were also identified in rice in response to major abiotic stresses such as, cold, heat, drought and salt[21]. In *Brassica rapa*, over 4000 lncRNA were identified as heat-responsive and in common association with hormones like salicylic acid and brassinosteroids[22].

Studies suggest that genome of an organism might encode for as many lncRNAs as mRNAs[14,23]. However, according to latest plant lncRNA database, CANTATAdb2.0[24], only 1498 lncRNA genes were identified in *P. patens*, which is a very low number compared to over 30 thousand protein coding genes. Overall, the understanding of lncRNA involvement in early plants is poorly understood. In this study, we aimed to investigate the temporal expression of both lncRNA and mRNA in response to AEA, AA and ABA. The overall goal was to identify

key components for AEA-mediated endocannabinoid signaling in *P. patens* and its association with the precursor molecule, AA and stress hormone ABA.

Materials and Methods

Plant material and treatment

The protonema of *P. patens* were grown in BCDAT medium (0.5 M $Ca(NO_3)_2$ $4H_2O$, 4.5 mM $FeSO_4$ $7H_2O$, 0.1 M $MgSO_4$ $7H_2O$, 1.84 mM KH_2PO_4, 1 M KNO_3, 4.5 mM $FeSO_4$ $7H_2O$, 500 mM Ammonium tartrate, 50 mM $CaCl_2$) with alternative TES (0.22 mM $CuSO_4$ $5H_2O$, 10 mM H_3BO_3, 0.23 mM $CoCl_2$ $6H_2O$, 0.1 mM Na_2MoO_4 $2H_2O$, 0.19 mM $ZnSO_4$ $7H_2O$, 2 mM $MnCl_2$ $4H_2O$, 0.17 mM KI) under 24 h light condition in a growth chamber (PERCIVAL;CU22L) for 10 days. After 10 days of growth, protonema were then treated with 50 µM AEA or AA and 10 µM ABA. Solvent DMSO (0.05%) was used as the control. Three biological replicates for each treatment at each time point were maintained and tissues were collected at 0h, 1h, 12h and 24h post-treatment. Harvested samples were squeeze dried and snap frozen in liquid N_2 before storing at -80°C.

RNA extraction and sequencing

RNA was isolated from stored samples of protonema tissues using Qiagen Plant RNA extraction Kit. Library preparation and sequencing were performed by BGI at their laboratory facility; biotin-labeled Ribo-Zero rRNA Removal Kit was used to remove rRNA from total RNA; TruSeq® Stranded Kit was used for first strand cDNA synthesis and libraries were sequenced on a BGISEQ-500 instrument. A total of 40 RNA-seq libraries were constructed. A comprehensive workflow of bioinformatic analysis was performed (Fig. 3.1A).

Data filtering and de novo transcriptome assembly

To obtain clean data, the short reads mapping software SOAPnuke[25] was used. Reads were mapped against ribosomal database, which allows five mismatches at the most to remove the reads. Further, process was followed to remove N bases more than 10%, and adapter, low quality and duplicate reads. After filtering, FASTQ format reads were used for downstream analyses.

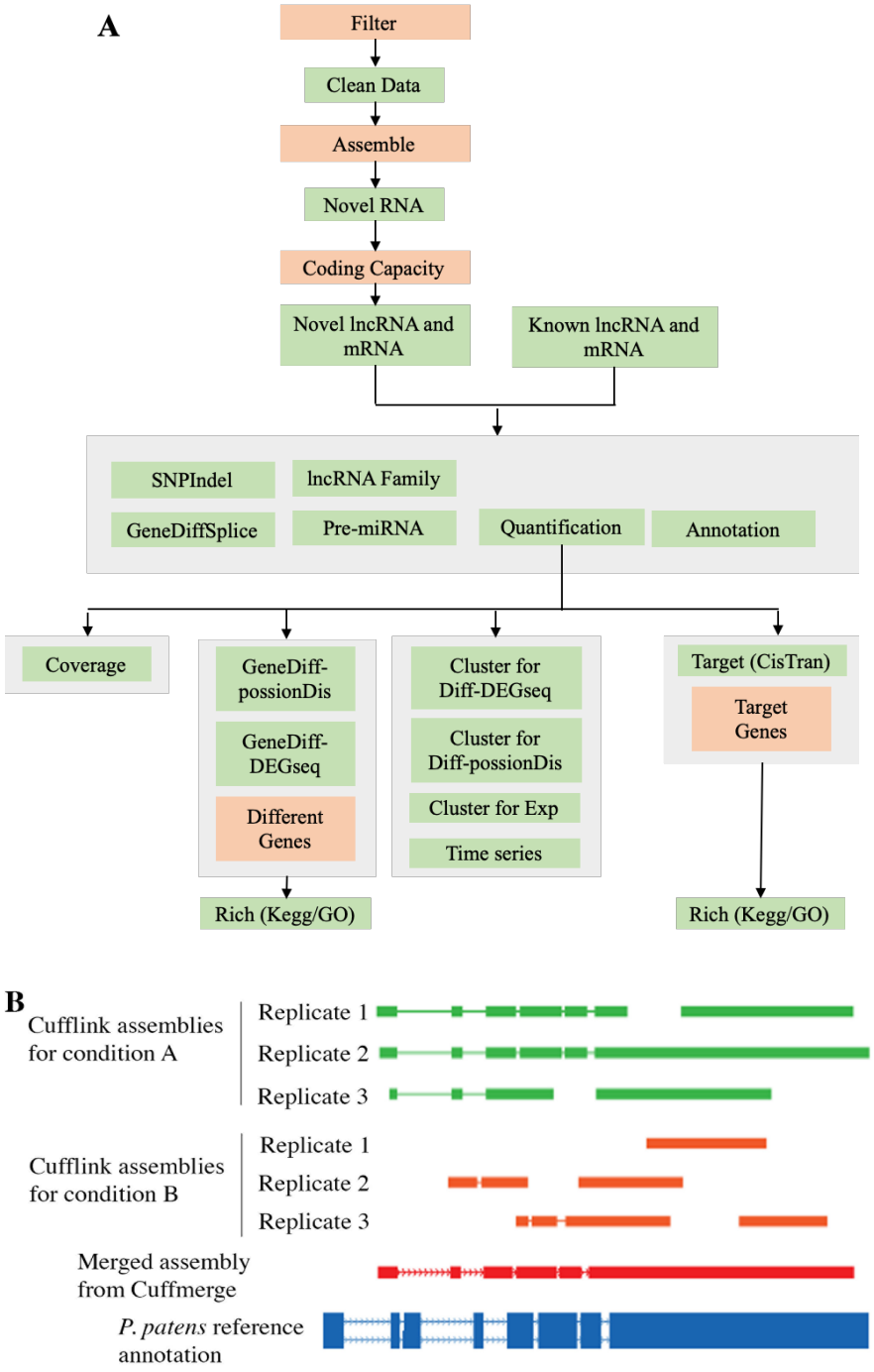


Figure 3.1. Bioinformatic analysis pipeline. A). Complete RNA-seq analysis workflow from data filtering to enrichment analysis. B). Merging transcripts from three replicates using Cuffmerge assembly tool. Combined transcripts were then annotated against *P. patens* reference genome

For assembly, filtered data were first mapped to the reference genome using HISAT2[26] software and then assembled using StringTie[27] and Cufflinks[28]. Cuffcompare, one of Cufflinks

tools was used to compare the transcripts to obtain the positional relationship with known mRNA and lncRNA. The parameters that used for this assembly were FPKM>= 0.5, Coverage > 1, and Length > 200. As for lncRNA based on NONCODE database, five categories; i, j, u, x, o of transcripts were most in our reads, which were kept. Cuffmerge, other tools of the Cufflinks, which used to merge multiple assembly. Since full assembly of low expressed genes are sometimes difficult due to insufficient depth of sequencing, we used Cuffmerge to assemble three replicates to merge the results of the assembly (Fig. 3.1B). The combined transcripts were used for subsequent analysis.

Prediction of coding ability

To distinguish mRNA from lncRNA, coding ability was predicted using three softwares and a database. The database that used was Pfam[29], which has a large collection of protein families represented by multiple sequence alignments and hidden markov models. Three softwares- CPC[30], CNCI[31] and txCdsPredict were used. Threshold that was set up for CPC and CNCI was 0; transcripts value less than 0 was lncRNA and higher than 0 was mRNA. For txCdsPredict, the threshold was 500, less and greater than 500 was lncRNA and mRNA, respectively.

Quantitative, gene coverage, differential expression and cluster analysis

In order to map the clean reads to the reference sequence and count the coverage, Bowtie2[32] software was used. To further calculate the expression value of genes and transcripts, RSEM (130) software was used. To analyze differential gene expression between different treatment groups DEGseq[33] software was used with fold change>=2.00 and FDR<=0.001 for filtering conditions of genes with significant differences. For cluster analysis, ComplexHeatmap[34] software was used.

SNPs and indel analysis

The GATK[35] software was used to determine the SNPs and indel from the mapping data of the reference genome. Briefly, raw RNA-seq reads were mapped to the reference genome of *P. patens* using HISAT2[26] software, followed by using Picard to mark duplicates. Next step was to use Split'N'Trim, in which reads were split with N into multiple supplementary alignments and reassign the mapping qualities for good alignments in order to match DNA conventions. Then BaseRecalibrator was used for each sample to detect and correct for patterns of systematic errors in the base quality scores. Finally, HaplotypeCaller was used to call variants to identify

SNPs and indels as raw variants. Raw variants of each sample were then filtered for filtered variants with SNPs and Indels. Files were then stored in VCF format for further analysis.

Enrichment analysis and data visualization

Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were performed for enrichment based on the p-value (0.01) and calculated FDR (false discovery rate). The cut off FDR value was set <=0.01 as significant enrichment. For further analysis of some of the raw RNA-seq data, and final visualization, R and R-studio were used. The R packages used for visualization were base, circlize[36], ComplexHeatmap[34], datasets, data.table, dplyr[37], ggplot2[38], graphics, grid, stats and plyr[39].

Results and Discussion

The occurrence of anandamide and its precursor AA, being limited to early land plants is intriguing. Like in mammals, it is not clear if AEA in bryophytes functions through an endocannabinoid receptor and imparts a protective role, among others. We carried out a comprehensive RNA-seq analyses in *P. patens* to gain a global perspective of the role of anandamide and identify any overlapping role with its precursor AA and the most common stress-responsive hormone, ABA. The importance of ABA in numerous physiological aspects, including stress tolerance has been well established in plants. Furthermore, NAEs in higher plants were shown to function through ABA-dependent and independent pathways. However, since the occurrence of AEA and its precursor AA are only limited to lower organisms, their physiological role and association with ABA remain unclear. As such, we examined the temporal changes in the expression of lncRNA and mRNA of *P. patens*, in response to exogenous AEA, ABA and AA treatment.

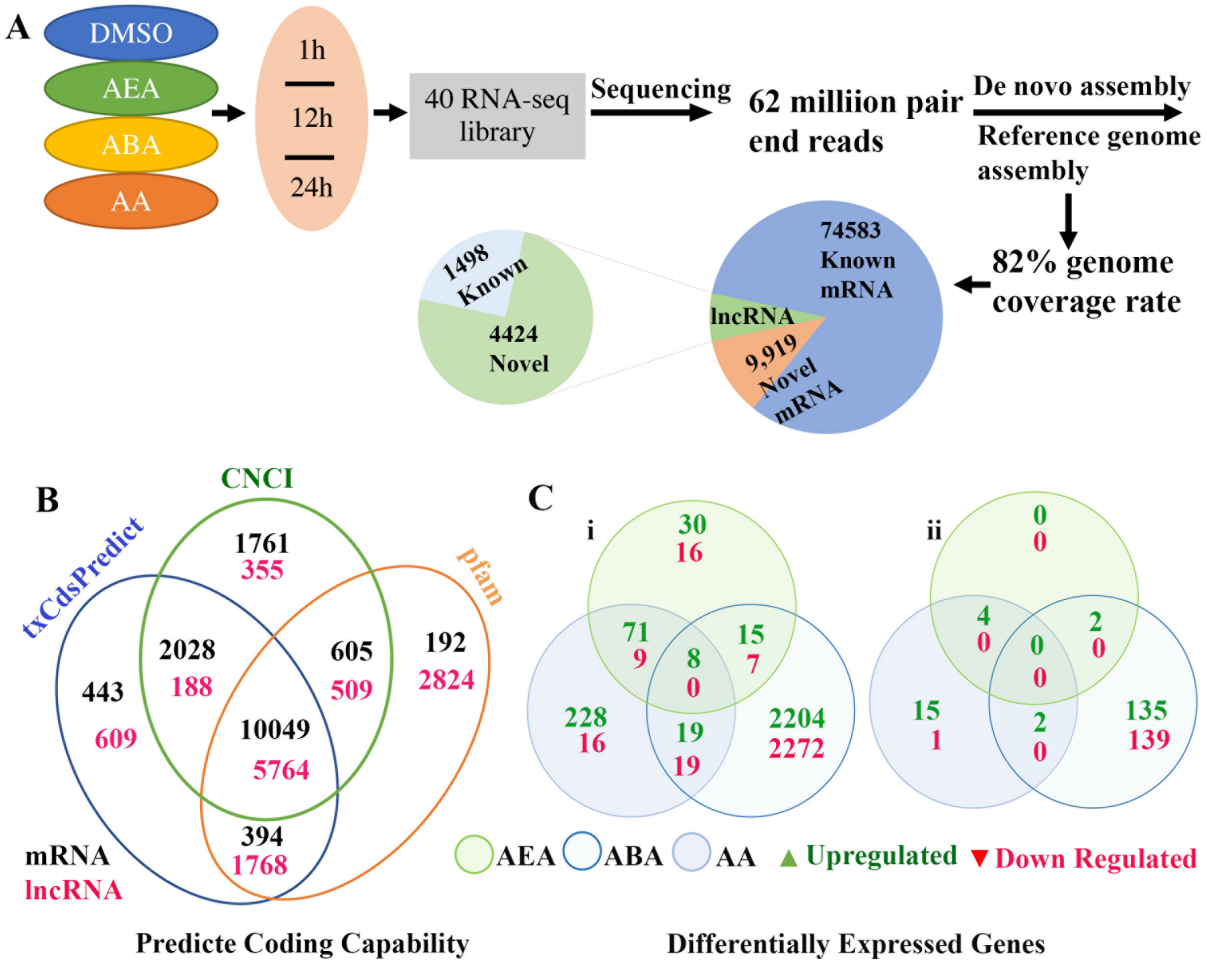


Figure 3.2. Overview of transcriptome analysis. A). Protonema of *P. patens* were treated with exogenous AEA, AA and ABA for 1, 12 and 24h. With three replicates for each sample, a total 40 RNA-seq libraries were synthesized. Reads were assembled with both *de novo* and reference genome of *P. patens*. Pie chart show the identified known and novel transcripts. B). Coding capacity was predicted for mRNA (black) and lncRNA (red) using CNCI and txCdsPredict software and mapped to pfam database; C). Number of common mRNA (i) and lncRNA (ii) that were always up- or down-regulated throughout the treatments

Transcriptome analysis identified previously unreported lncRNAs in P. patens

A *de novo* transcriptome assembly of *P. patens* was carried out using the 40 RNA-seq libraries that were generated in this study (Fig. 3.2A). In addition to the reference library built, the moss genome and non-redundant plant databases, including that of Arabidopsis were used for annotation of the transcriptome. Even though *P. patens* was the first non-seed organism to be

whole genome sequenced[40], is still not fully annotated. The *de-novo* assembly of deep sequencing data allowed us to identify 9,919 mRNA and 4,424 lncRNA that were previously unreported (Fig. 3.2A). We annotated identified lncRNA to determine novel lncRNA using CANTATAdb2.0 as plant database with known lncRNA. According to the database, there are 1,498 lncRNA have been identified in *P. patnes*[24], here we are reporting 4,424 novel lncRNA in *P. patens* (Fig. 3.2A).

After multiple steps of data filtering and assembly, three software (CPC[30], CNCI[31], and txCdsPredict) were used to map the transcripts of *P. patens* to pfam database (pfam[29]) to predict their coding capacity (Fig. 3.2B). While CPC was not useful, we followed a rigorous selection and only those mRNA or lncRNA that were identified by both CNCI and txCdsPredict and mapped to pfam database were chosen for further analyses. Furthermore, Pearson correlation coefficient analysis indicated significant correlation between the technical replicates. Exon number distribution revealed that most lncRNA transcripts were with only one exon, whereas most mRNA transcripts were with 10+ exons (Fig. 3.3A). The number of mNRA transcripts gradually lowered from exon number two to ten (Fig. 3.3A). Distribution of RNA length for majority of mRNA and lncRNA was between 0 to 2000 nucleotides (nt), while a few were over 10,000 and 5000 nt, respectively (Fig. 3.3B). A highest number of coding and non-coding genes were detected with only one transcript, over 75,000 and 5000 for mRNA and lcnRNA, respectively (Fig. 3.3C).

The identification and annotation of SNPs and Indels revealed their occurrence 2000 bases upstream (Up2k) and downstream (Down2k) of the genes and also within introns, exons, and intergenic regions (Fig. 3.4 A&B). These insertions, deletions and polymorphisms in the genomic regions can affect the pre- and post-transcriptional regulation of genes. For instance, upstream of a gene is used for pre-transcriptional regulation and a mutation in those regions can cause dysregulation of the gene; on the other hand, mutations in exon or intron can cause a post-transcription dysfunction such as immuration of mRNA or splicing. Among all the treatments, highest number of SNPs and Indels were observed in the exon regions and lowest in the intergenic regions (Fig. 3.4 A&B). However, the number of detectable SNPs and Indels varied with the type and duration of the treatment. For instance, highest number of Indels were detected with AEA treatment at 1h in regions of Up2K and exon (Fig. 3.4Bi), while highest number of

SNPs were detected in exon at 1h with ABA treatment (Fig. 3.4Aiii). Lowest number of SNPs and Indels were detected in intergenic regions with all treatments (Fig. 3.4 A&B).

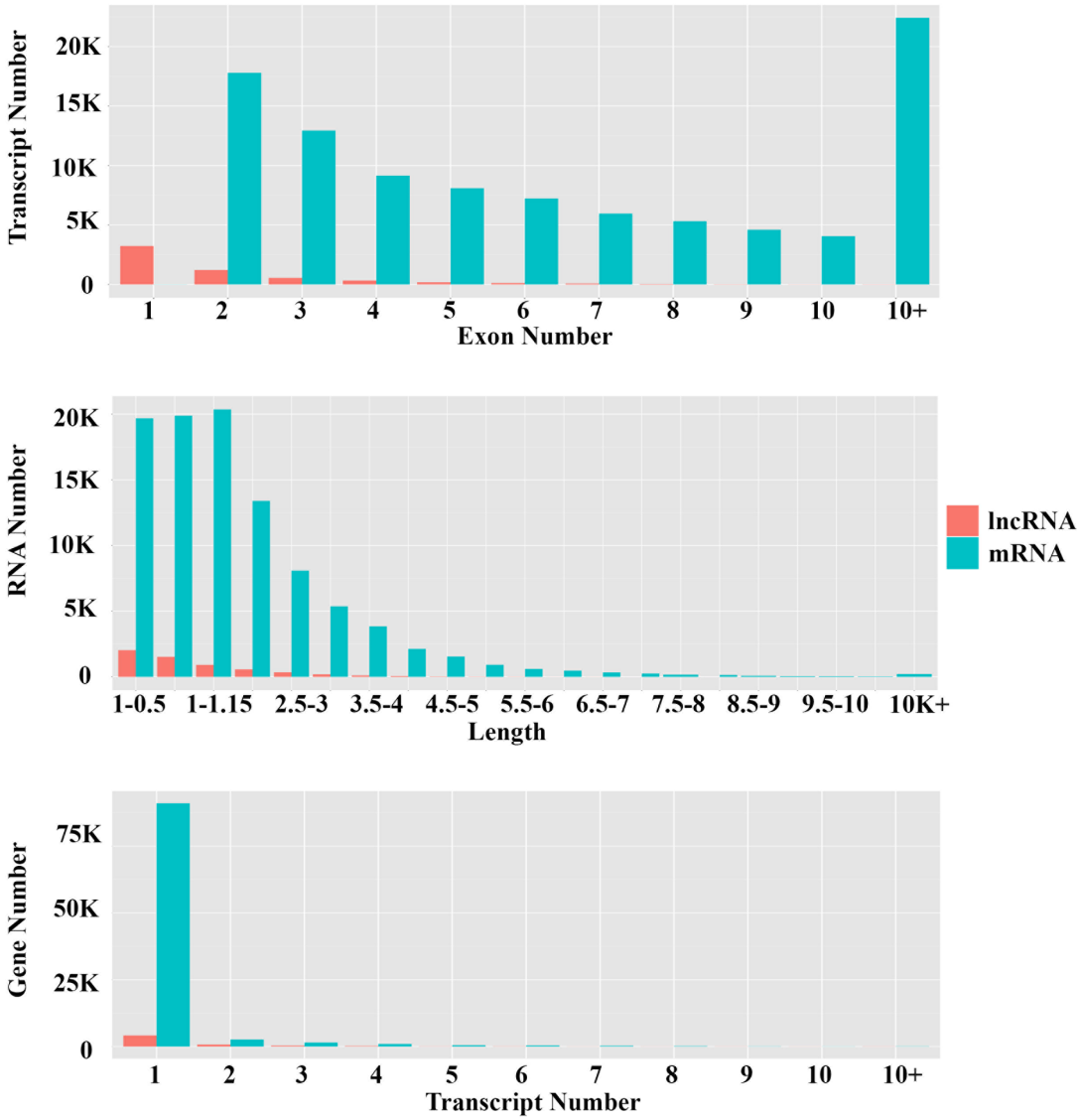


Figure 3.3. Distribution of transcripts. Distribution of mRNA (blue) and lncRNA (red) transcripts by their A). exon number B). read length and C). gene number. Highest number of genes were identified with one transcript

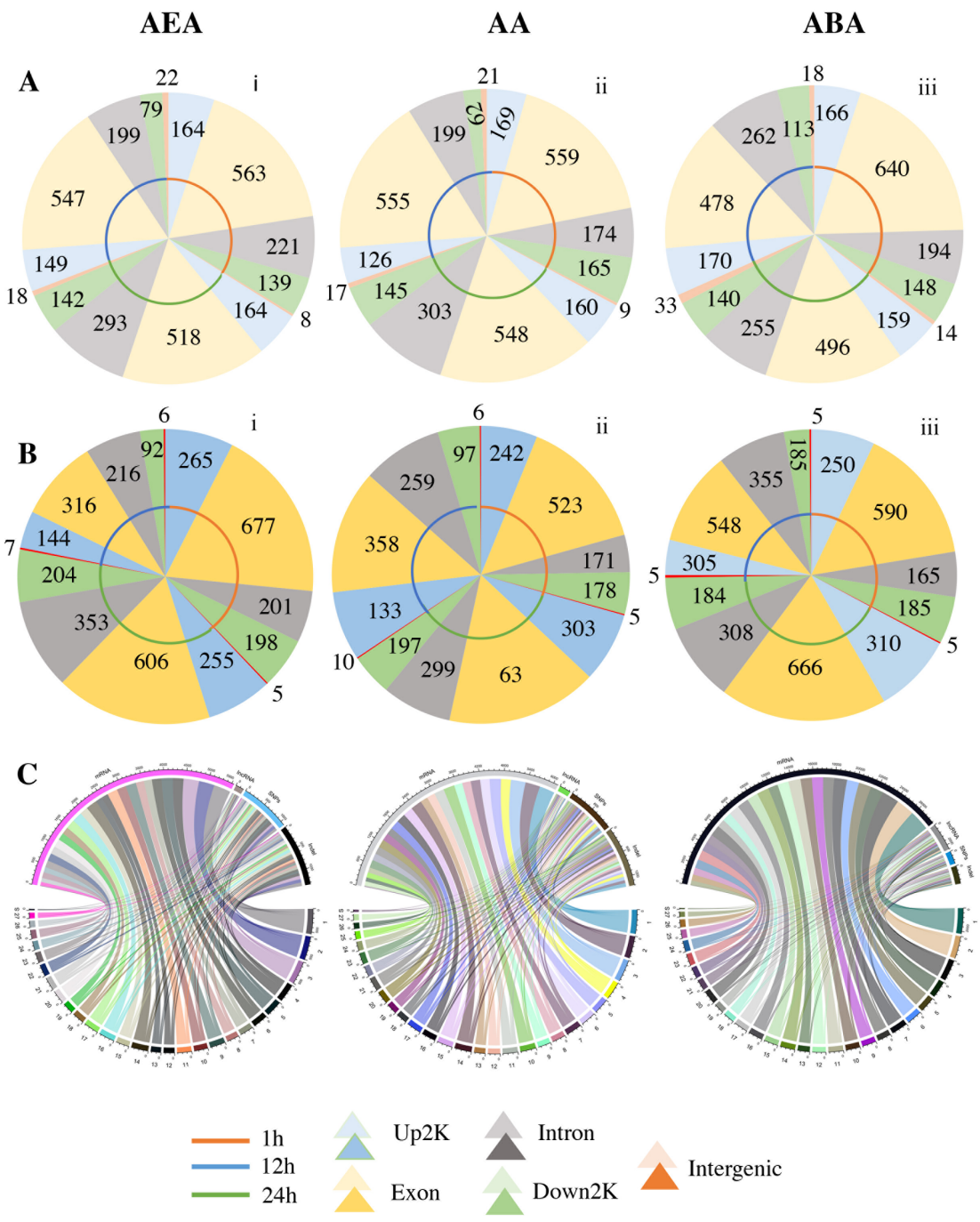


Figure 3.4. Identification of SNPs and Indels. Number of SNPs (A) and Indels (B) identified with exogenous AEA, AA and ABA treatments at 1h, 12h and 24h in exons, introns, intergenic, and 2K up and down stream of the coding region. C). Chromosomal mapping of the total expressed mRNA and lncRNA, and the identified SNPs and Indels

Analysis of chromosomal distribution of mRNA and lncRNA transcripts indicated that the highest number of genes were associated with Chr01 and lowest in Chr27. This distribution also corresponds to the size of the chromosome and in incongruence with previous gene distribution studies of *P. patens*[41] (Fig. 3.4C). However, the highest number of DEGs with various treatments were associated with Chr1, 2 and 3 (Fig. 3.4C). We also observed that even though the size of the chromosomes is similar from 5 to 16, the number of DEGs with AEA treatment were relatively higher in Chr11 and 14, AA in Chr10, 14 and 20 and ABA in Chr7, 14 and 20 (Fig. 3.4C). The discovery of these structural variations in genomic regions of *P. patens* provides a rich resource for further studies to understand the variation in gene expression and associated functional role.

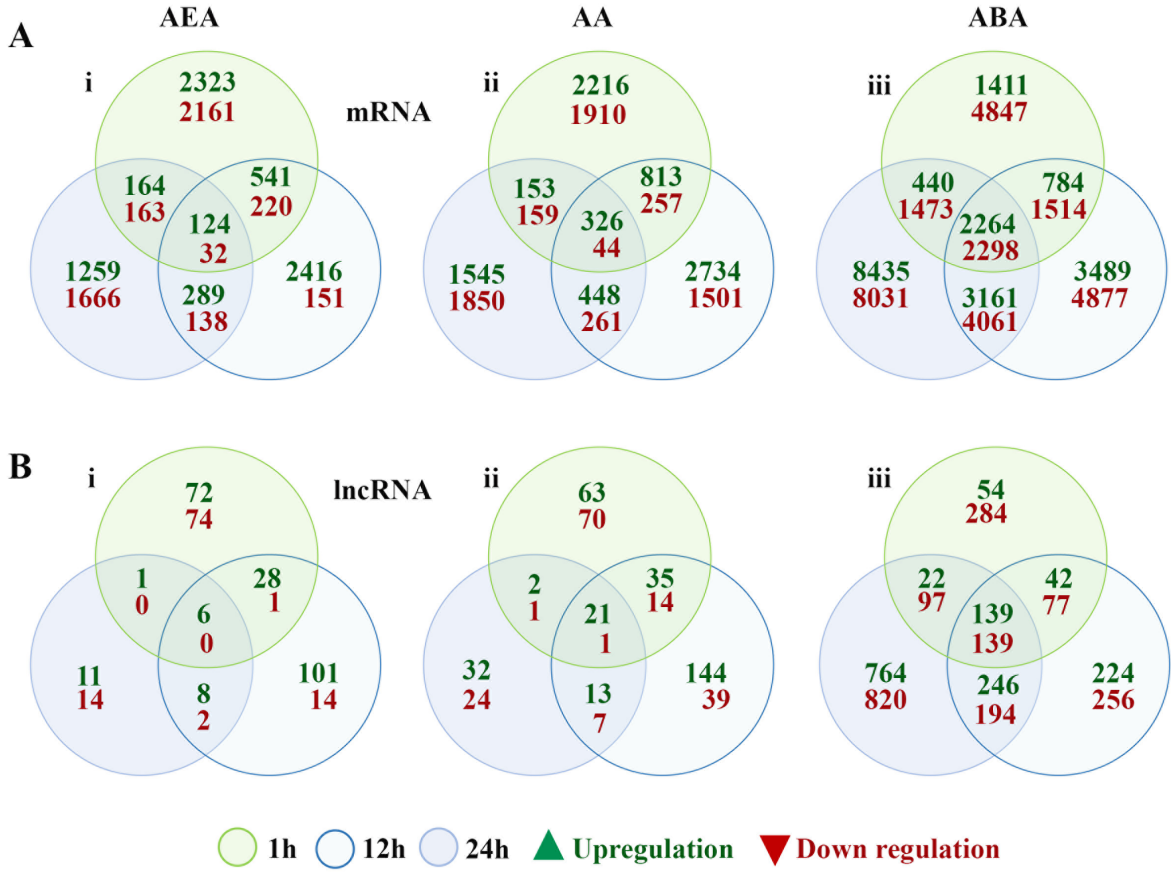


Figure 3.5. DEGs of mRNA and lncRNA. Number of up (green) and down (red) regulated mRNA (A) and lncRNA (B) with exogenous treatments of AEA, AA and ABA for 1h, 12h and 24h

Physcomitrella responds differentially to exogenous AEA, AA and ABA

The number of differentially expressed mRNA and lncRNA showed a significant variation between the treatments and duration. With AEA treatment, DEGs were higher at 1h (5728) and 12h (4111), which reduced significantly by 24h (3835) (Fig. 3.5Ai). With AA treatment, DEGs being higher at 1h (5878) and 12h (6348) and lower at 24h (4788) (Fig. 3.5Aii). The ABA treatment resulted in DEGs that reflect a gradient response to time with lowest number of up or down-regulated genes at 1h and the highest at 24h (Fig. 3.5Aiii); and in general, there were more down-regulated DEGs than the upregulated.

Among the three treatments, the strongest response was observed with ABA with highest up and downregulated DEGs at 24h. In contrast to ABA treatment, in both AEA and AA treatments, upregulated DEGs were higher compared to the down regulated (Fig. 3.5A). With AEA treatment, 139 known and 11 novel DEGs were common between the time points, of which 32 genes were down regulated. Moreover, within the AEA treatment, the highest number of DEGs were shared between 1h and 12h treatments. With AA treatment, commonly expressed DEGs over time were higher than what was noted for AEA treatment and the share of the upregulated DEGs were between 1h and 12h was higher, while down-regulated DEGs were more common between 12h and 24h but not very different from what was shared between 1h and 12h (Fig. 3.5Aii). On the other hand, with ABA treatment, 11.3% and 8.5% genes were common over time that were always up or down regulated, respectively. As the total number of known down regulated genes were higher than upregulated genes with ABA, a high number of common down regulated genes were detected between three time points (Fig. 3.5Aiii). For both known and novel genes, the number of down regulated genes were higher compared to upregulated genes (Fig. 3.5A). Interestingly, the majority of the genes that were up or down regulated were unique to the type and duration of treatment. Moreover, the fact that only a small number of DEGs were common between all the three time points for each treatment and between the treatments (Fig. 3.2C), suggests that the molecular response of *P. patens* protonema is unique to both time and treatment.

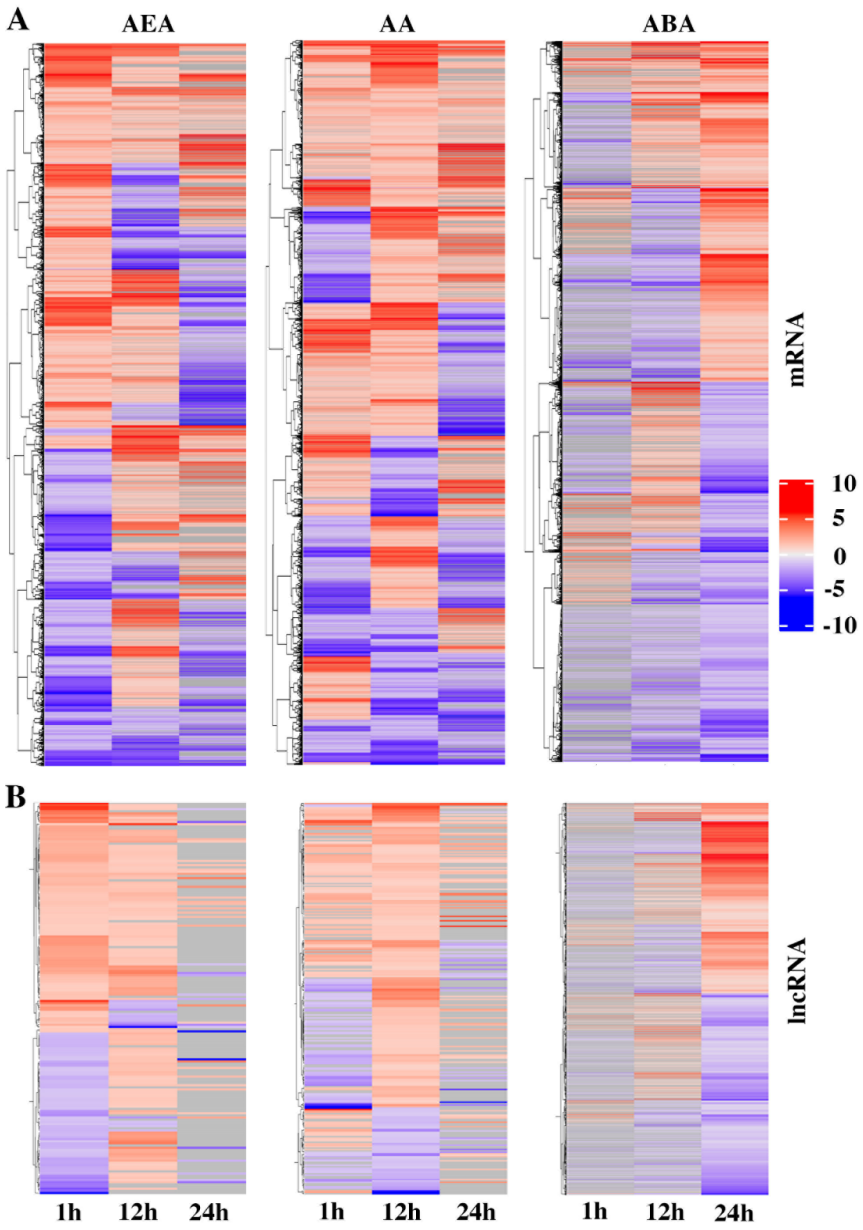


Figure 3.6. Heatmap of differentially expressed genes. Heatmap showing differentially expressed mRNA (A) and lncRNA (B) with exogenous treatments of AEA, AA and ABA at 1h, 12h and 24h. Log2 ratio range was between +10 to -10. Heatmaps were drawn using ComplexHeatmap program in R

By the generation of heatmaps, we have also analyzed if a gene that was up or down regulated at 1h has changed the expression pattern over time. As such, in all treatments about 4 to 8% of the genes altered their expression pattern, which typically was higher from 1h to 12h than 1h to 24h, for both AA and AEA treatments, whereas the converse was observed with ABA

(Fig. 3.6), similar to the representation by Venn diagrams (Fig. 3.5). Number of DEGs can be varied with biotic and abiotic perturbations and as well as the developmental stages of *P. patens*[12]. In higher plants like Arabidopsis, about 1-10% of the genome is ABA-regulated and responds to environmental stresses such as drought, cold and salinity, of which 25-50% of the genes were commonly expressed between the stresses[6]. We identified that about 15% of the genome responds to ABA treatment in *P. patens*, of which about 60% was down regulated. In a previous study however, chloronemata tissue of *P. patens* when treated with 10^{-5} M ABA for 1h showed a higher number of genes that were upregulated than downregulated, and a significant number of genes overlapped with desiccation treatment[11]. The variation between the two studies is likely due to their difference in developmental stage; transcriptome analysis of 34 different developmental stages of *P. patens* revealed that a significant number of genes are unique to different stages[12]. A cluster of genes remained highly up or down regulated with all treatments over time (Fig. 3.6).

Although the number of lncRNA that were differentially expressed was much lower than the mRNA, their temporal profiles for each treatment for both up and down regulated genes were mostly similar to the mRNA expression pattern (Fig. 3.6B). With AEA treatment, there were 332 lncRNA that were differentially expressed, of which 284 are novel. The number of lncRNA that responded to AA treatment were slightly higher than AEA, whereas with ABA treatment they significantly higher (Fig. 3.5B). For all treatments, the number of novel lncRNA that were detected was high since there were only 1422 lncRNA that were previously identified in *P. patens* genome. Like with mRNA expression, a higher number of genes were upregulated with AEA and AA treatments, and down regulated with ABA treatment (Fig. 3.5B, 3.6B).

Also, a similar trend was found in terms of number of DEGs at different time points; highest number of lncRNA differentially expressed with AEA was at 1h, however, number of upregulated lncRNA was higher at 12h (Fig. 3.5B, 3.6B). There were only a few differentially expressed lncRNA that were common between each treatment and the duration. With AEA, only six common lncRNA genes were upregulated at all time points but no down regulated genes were identified (Fig. 3.5Bi). Common lncRNA genes were higher between 1h and 12h, compared to other time points (Fig. 3.5B). On the other hand, with AA and ABA treatments, the common lncRNA genes between times were slightly higher (Fig. 3.5B). Both AA and ABA

showed similarity with AEA in terms of higher number of common genes between 1h and 12h (Fig. 3.6B). Analysis of altered expression of the same lncRNA between times surprisingly showed that there were none or very few same lncRNA that changed their expression over time (Fig. 3.6B). In general, the expression of lncRNAs is relatively low and are known to exhibit more tissue, cell, developmental and/or disease-specific profiles compared to mRNA[23]. The specific lncRNA response over time during AEA/AA/ABA treatment in *P. patens* suggests its temporal dependency. Moreover, the higher number of DEGs in both mRNA and lncRNA with AEA at 1h indicates that AEA has a short-term effect, while AA and ABA have prolonged effect on *P. patens* protonema.

AEA showed distinct molecular response that involves key physiological processes compared to AA and ABA

Gene ontology (GO) enrichment analysis was performed to evaluate the common and distinct responses of *P. patens* to AEA, AA and ABA in major physiological processes, such as involvement in biological process, cellular component and molecular function. The GO term for only significant DEGs (p-value<0.01) that were common at a temporal level and were either up- or down-regulated for each treatment were identified (Fig. 3.7). Most of the upregulated genes were common among the three treatments and represented cellular and metabolic process, biological regulation and response to stimulus and localization, whereas a low number of genes were involved in carbon/nitrogen utilization, detoxification, cell proliferation, and biological adhesion. The highly expressed genes that mostly involved in different biological processes are localized in the cell, cell part, membrane, organelle and membrane part, while less involved in cell junction, nucleoid, and symplast. However, the number of genes involved varied with treatments and duration. Moreover, for AEA or AA treatments, a higher number of upregulated genes were involved, whereas, with ABA, a significantly higher number of genes were down regulated in the physiological processes. A similar trend was observed when common genes between times were analyzed across treatments (Fig. 3.7).

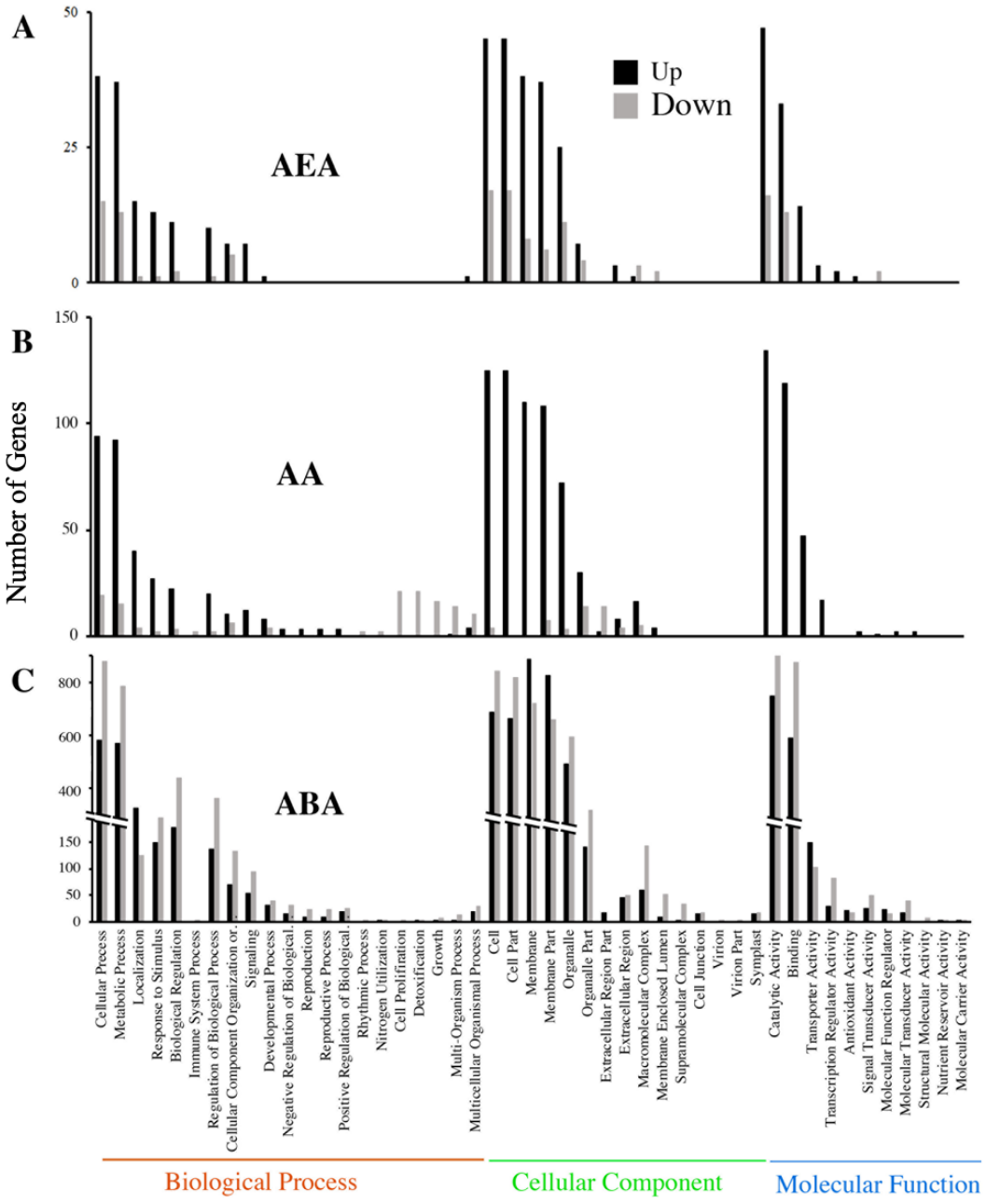


Figure 3.7. Gene ontology (GO) analysis. The GO analysis of common genes over time that were differentially expressed with exogenous AEA (A), AA (B), and ABA (C) treatment. Number of up and down regulated genes are involved in numerous biological processes in different cellular compartments and associated with diverse molecular functions

The GO term for most DEGs with AEA treatment was associated with biological processes. Among the upregulated genes, most notable associations were transmembrane

transport, protein domain specific binding, glutathione/sulfur/cellular modified amino acid/L-serine metabolic process, metallopeptidase activity, serine O-acetyltransferase activity and glutathione transferase activity (Fig. 3.8). Whereas, chloroplast and/or fission, phosphatidylinositol phosphate kinase and mitochondrial matrix-associated genes were downregulated (Fig. 3.8). On the other hand, with AA and ABA treatment, a higher number of genes were significantly up or down regulated at all time points compared to AEA. Genes that were always significantly upregulated were associated to amino acid/organic acid/carboxylic acid/anion transport and transmembrane transport, C22 sterol desaturase activity, ATPase activity-coupled to transmembrane movement of substances, sterol desaturate activity, alternate oxidase activity, ubiquinol: oxygen oxidoreductase activity, sulfite reductase activity, vacuolar membrane and integral/intrinsic component of membrane. While significantly downregulated genes associated with cellular component disassembly, defense response to bacteria, aging, leaf senescence, plant organ senescence, acyl desaturase activity, oxidoreductase activity and carboxyl-O-methyltransferase activity (Fig. 3.9).

The GO term analysis with AEA treatment showed differential expression with duration and the genes are likely associated with diverse physiological processes. For instance, at 1h the genes involved in cellular or metabolic processes were mostly upregulated but down regulated at 24h. At 1h, changes in biological processes associated with signal transduction, ion transport and organelle relocation or localization were noticed (Fig. 3.8). At 12h highly expressed genes were related to intracellular signal transduction, movement of a cell or subcellular component, response to stimulus and metabolic process (Fig. 3.8). At 24h, some of the highly expressed genes are related to cellular processes, response to stimulus and regulation of the metabolic process (Fig. 3.8). Overall, significant differences in expression within AEA treatment was time-dependent. Highly expressed genes were mostly associated with the biological process at 1h and 12h, which downregulated but at 24h.

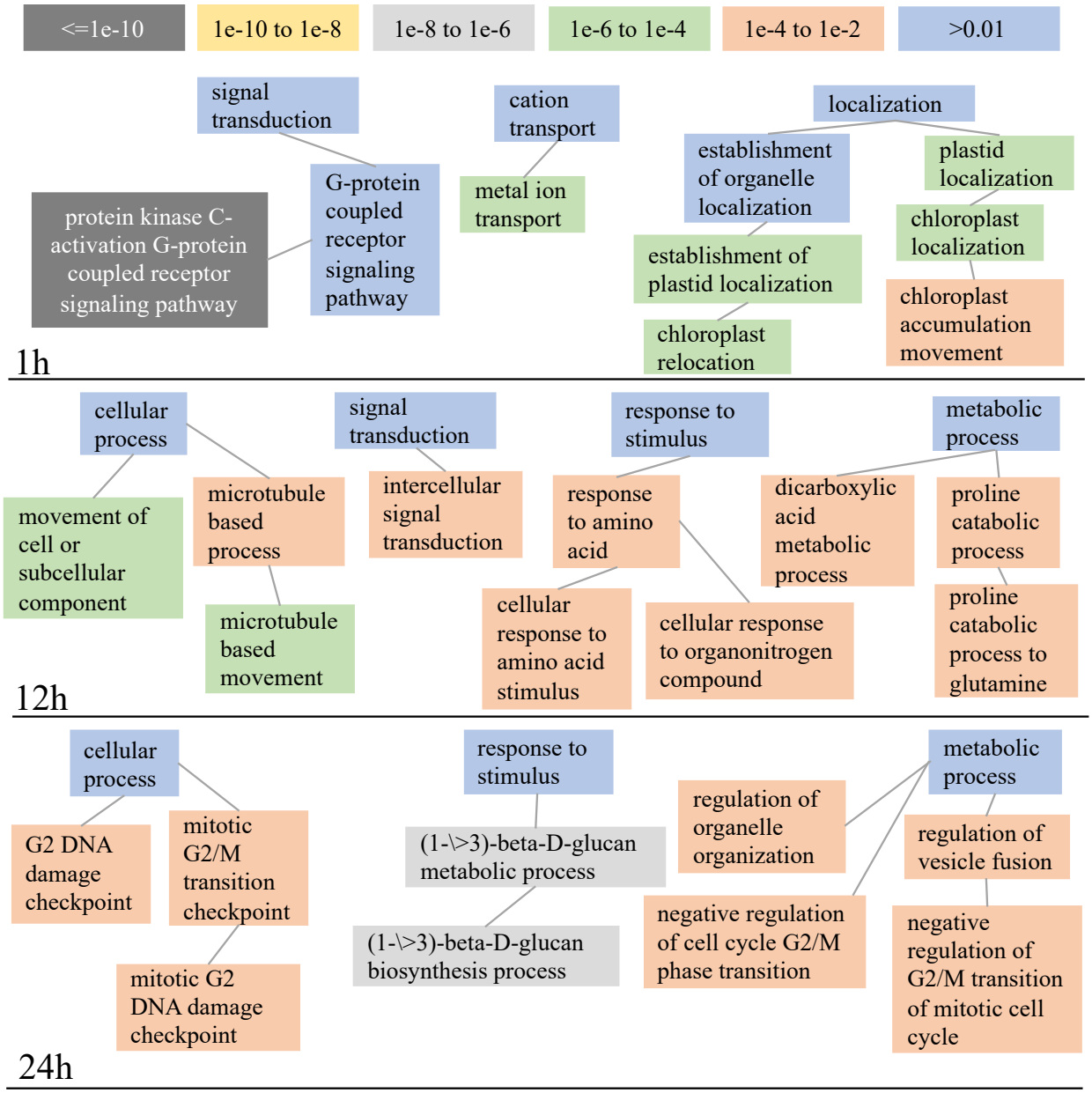


Figure 3.8. Gene ontology (GO) analysis of DEGs with AEA treatment. Highly expressed genes of significance at 1h, 12h and 24h, involved in different biological process are indicated. Color code was based on the p values

Interestingly, there were 17 genes ($p <= 10^{-10}$) that were highly expressed with AEA treatment that were related to signal transduction associated with activation of GPCR signaling pathway. This observation provides us with a first clue that AEA-mediated responses most likely involve a GPCR like receptor; however, a direct evidence for GPCR-like receptors and their

signaling in *P. patens* is yet to be demonstrated. In contrast, at 12h, highly expressed signal transduction related genes ($p = 10^{-4}$ to 10^{-2}) were only associated with intracellular signaling, while and no such genes were expressed at 24h (Fig. 3.8). This observation might implicate that a receptor-mediated AEA response is transient and likely attenuated in 24h. Moreover, over 70 highly expressed genes ($p = 10^{-6}$ to 10^{-4}) were related to chloroplast and plastid relocation and/or localization were detected at 1h, while gene expression associated with the movement of subcellular compartment continued at 12h but not detected at 24h (Fig. 3.8). In a cell, relocation of chloroplast occurs to optimize photosynthesis by moving away from the excessive illuminated area and protect photosynthetic machinery[42]. Furthermore, it is known that polyunsaturated NAEs can affect chloroplast biogenesis in higher plants[43]. Also, relocation and localization of chloroplast and/or plastid can reflect photosystem assembly or plastid reorganization, which is an indication of nutritional dependency of the cell or stress responses[42,44]. Additionally, expression of genes related to G2 DNA damage checkpoint, cellular response to UV-C and beta-D-glucan metabolic process at 24h might reflect cellular stress and thus, regulation of cell division[44]. Together, high expression of such classes of genes is an indication of cellular reorganization of organelles in the process of signal transduction to the downstream components and adjacent cells[45].

The GO analysis of DEGs in response to AA and ABA treatment showed a different pattern compared to AEA. With AA treatment at all the three time points, highly expressed genes were associated with metabolic process. At 1h, genes associated with a number of catabolic processes were significantly higher compared to at 12h or 24h (Fig. 3.9). Expression of some UV-C responsive and ubiquitin genes in response to AA and ABA treatment might imply checkpoints on gene regulation as well as higher protease activity inside cells[12]. Interestingly, neither AA nor ABA treatments induced the high expression of protein kinase C activation GPCR signaling proteins or organelle localization proteins, highlighting a distinct response to AEA.

Finally, GO analysis of commonly up and downregulated genes between treatments revealed that AEA and AA were mostly involved in upregulation of metabolic processes, biosynthetic process, transmembrane transport, protein domain specific binding, and down regulation of chloroplast/plastid fission and organization genes. On the other hand, between AEA

and ABA, while there were no down regulated genes that were common, a few genes were commonly upregulated. Although these upregulated genes were related to signaling, cell communication, and intracellular communication, the specific genes associated with these processes remained distinct between AEA and ABA.

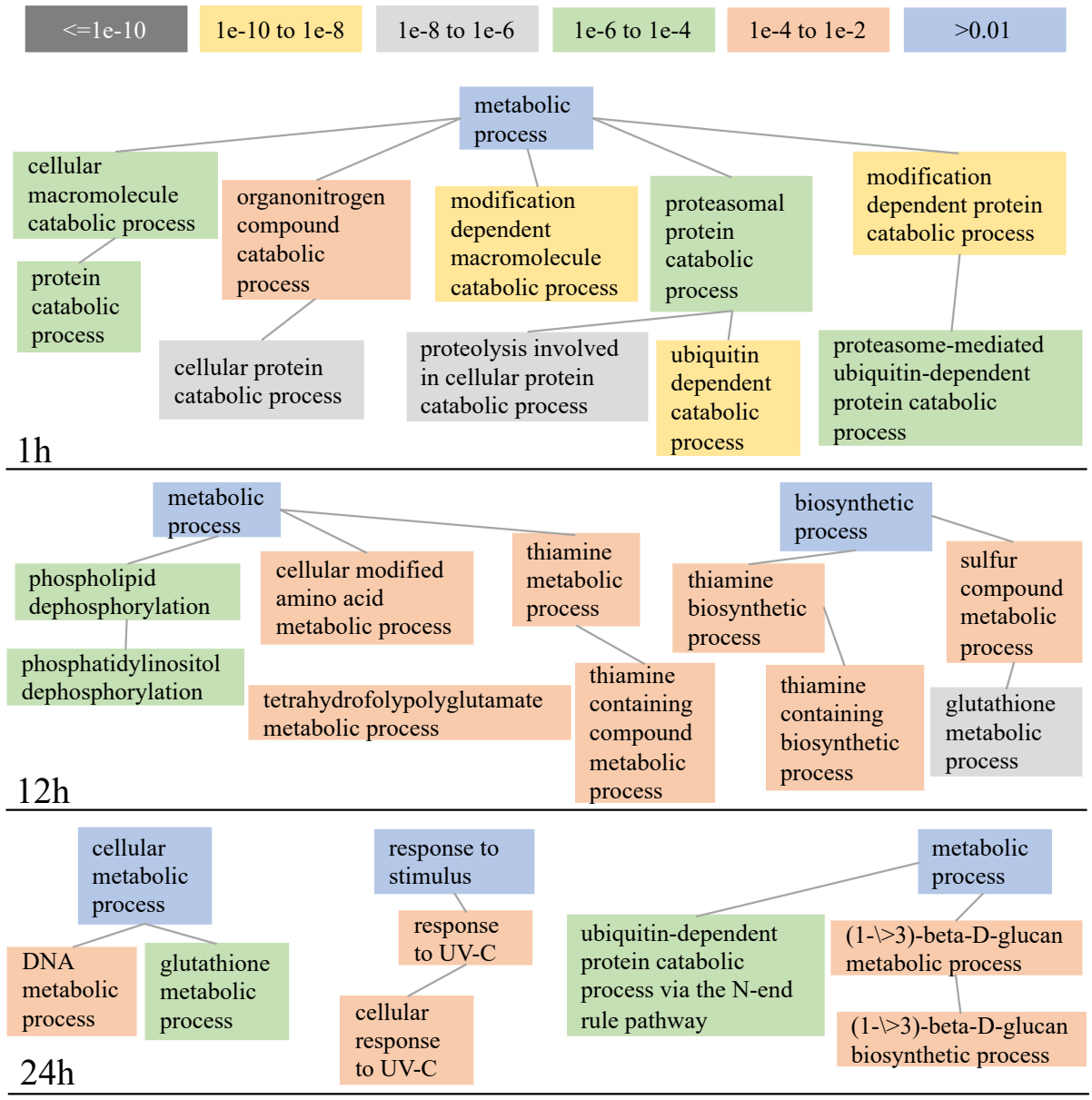


Figure 3.9. Gene ontology (GO) analysis of DEGs with AA treatment. Highly expressed genes of significance at 1h, 12h and 24h, involved in different biological process are indicated. Color code was based on the p values

Taken together, GO analysis showed that the exogenous AEA, AA or ABA affected biological processes in a distinct manner and most of these responses varied with treatment duration. A significant outcome of this study was the evidence that AEA is likely to be a receptor-bound signaling molecule and is capable of activating signal transduction.

AEA might possess a unique signaling pathway in P. patens

In plants, ABA is involved in seed maturation, seedling development, flowering, senescence, stomatal regulation, and a number of biotic and abiotic stresses, by inducing a number genes and participation in various signaling processes[6,8,46,47]. ABA-inducible genes were 15 to 40% common in response to stresses; highest with drought and lowest with cold stress[6]. Also, several common genes were induced in response to exogenous NAE 12:0 and ABA during Arabidopsis seedling development[48]. While over 200 genes were identified by a combination of molecular, biochemical and forward/reverse genetics studies, another 100 genes that were commonly induced or repressed by ABA were revealed from at least eight ABA-related transcriptome studies[6]. We analyzed the expression profiles of these well characterized genes, along with few additional DEGs, in order to identify an overlap in response to ABA and AEA. Furthermore, a potential signaling pathway for AEA was predicted by analyzing the expression of select receptors, secondary metabolite synthesizing genes, kinases/phosphatases and regulatory and metabolic pathway genes (Fig. 3.10). Analyzing these well annotated ABA-responsive genes against the transcriptome data of the current study revealed that a huge number of genes were exclusively expressed upon ABA treatment, but a few were common with AEA and AA treatments.

Receptors

Among the 24 receptors or receptor-like proteins associated with ABA in Arabidopsis[6], orthologs of only eight genes were differentially expressed with ABA in *P. patens*, of which three were highly upregulated (Fig. 3.11). Among them, GCR1 and GCR2 are GPCR-like receptors[49], which were down regulated with AEA treatment (Fig. 3.11). Furthermore, we identified a novel GPCR-like protein that did not respond to AA treatment but was upregulated with AEA and down regulated with ABA. Pyrabactin resistance (PYR)/PYL/RCAR (regulatory components of ABA receptor) family protein is well established as ABA receptor in Arabidopsis[8]. Out of 15 known receptors from this family protein, only one (PYL8) was upregulated and two others (PYL4 & PYL11) were down regulated with ABA; among these,

PYL4 was moderately upregulated with AEA. In *P. patens*, ABAR and A1E, which are known as ABA receptors in Arabidopsis were down regulated with ABA (Fig. 3.11). These data also suggest that response of receptors to AEA and ABA might be specific in *P. patens*.

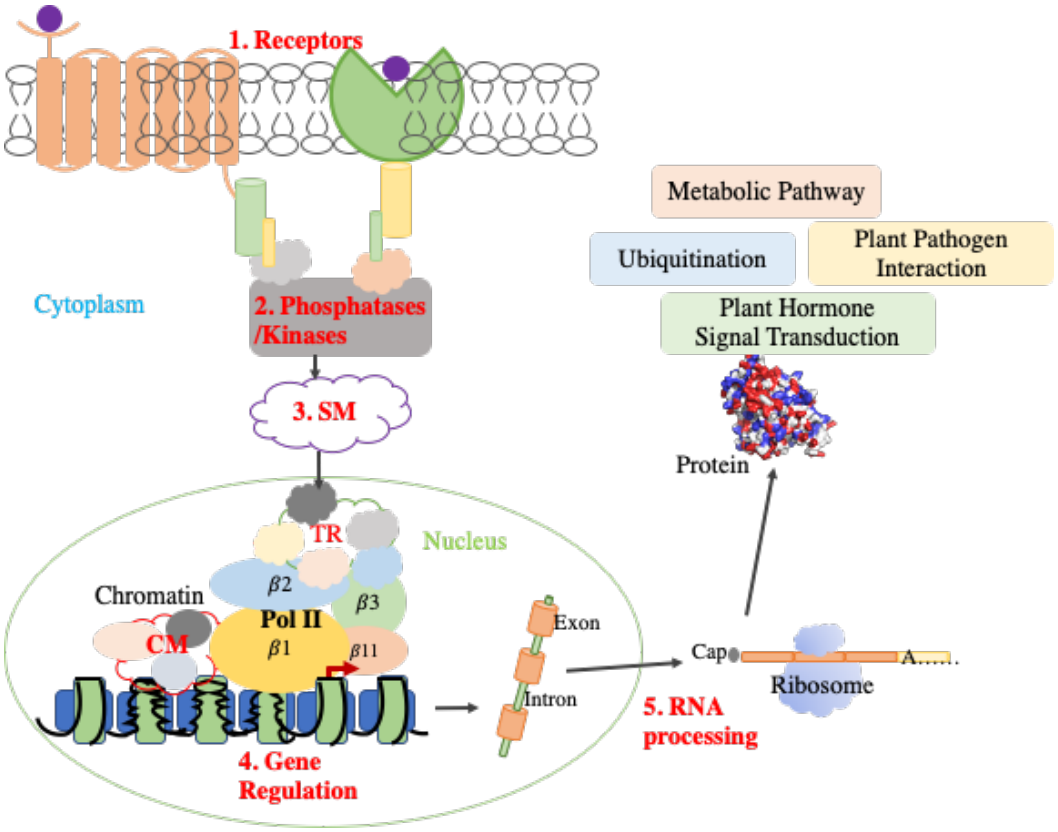


Figure 3.10. Schematic signaling pathway. The categories of signaling pathway include receptors, phosphatases/kinases, secondary metabolites (SM), gene regulation and RNA processing. Genes associated with transcription regulator (TR) and chromatin modifier (CM) are categorized as gene regulators.

Kinases or phosphatases

Out of 30 identified phosphatases and kinases in Arabidopsis[6] only 10 orthologs were differentially expressed in *P. patens* during the three treatments. Among the well characterized MAP kinases, MPK3 was highly upregulated by ABA but down regulated by AEA; however, a different ortholog of MPK3 was upregulated by AEA (Fig. 3.11). SNF1-related protein kinase (SnRK2) was highly upregulated by both ABA and AEA; an ortholog SnRK2 was moderately upregulated. Phosphatases like C-terminal domain phosphatase-like 3 (CPL3) was highly upregulated by all the treatments, whereas, HAB1 and HAB2 were moderately upregulated by

ABA and AEA. Among the secondary metabolite proteins associated with ABA signaling, only LPP2 and PLD were upregulated with ABA while four others were down regulated (Fig. 3.11). A NAD-dependent dehydrogenase, FLDH and RBOHF were highly upregulated with AEA, and rest of them were not differentially expressed (Fig. 3.11).

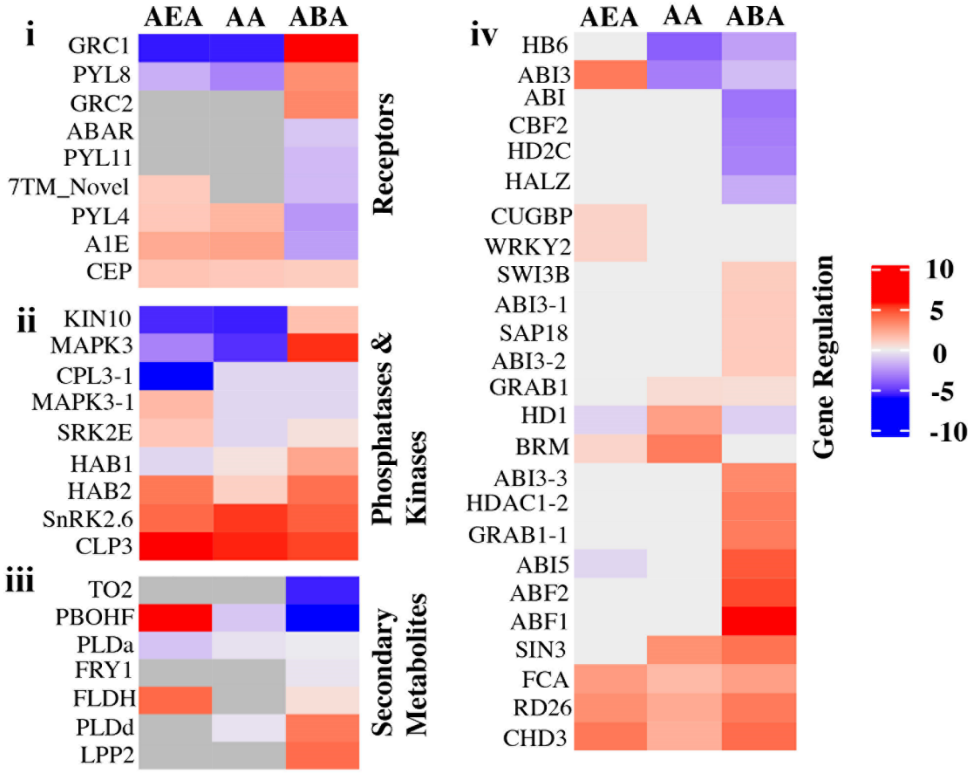


Figure 3.11. Heatmap of mRNA expression. Log2 expression of mRNA associated with receptors (i), phosphatases & kinases (ii), secondary metabolites (iii) and gene regulation (iv). Log2 ratio range was between +10 to -10. Heatmaps were drawn using ComplexHeatmap program in R

Gene regulatory components

Genes related to chromatin modification, transcription regulation and RNA processing and translation were categorized as gene regulators. Of the 71 genes that were recognized as ABA-associated[6], orthologs of 25 of them were differentially expressed in *P. patens* (Fig. 3.11). Most notable ABI family proteins were highly upregulated with ABA and down reregulated by AEA, similar to the response of Arabidopsis to NAE 12:0 treatment[50] (34) (Fig. 3.11). However, an another ortholog of ABI3 was upregulated by AEA and down regulated by ABA. Two chromatin modifier genes and one transcription regulatory gene were common between AEA and

ABA and highly upregulated. Also, a number of genes were only upregulated by ABA but were not differentially expressed by AEA or AA (Fig. 3.5d). The expression of transcription regulator, WRKY2, which is commonly associated with ABA and pathogen induced was however, not affected by ABA but upregulated with AEA (Fig. 3.5d). Among the nine catabolic enzymes of AEA[51], *PpFAAH1*, *PpFAAH2* and *PpFAAH8* highly upregulated by AEA, whereas *PpFAAH2* was also upregulated by ABA.

ABA metabolic genes

The genes that are involved in ABA synthesis, oxidation, conjugation and transport are categorized as ABA metabolic or transport genes. In Arabidopsis, 12 genes for synthesis, four for oxidation, 10 for conjugation and six for transport, a total of 32 genes previously characterized[6]. In *P. patens*, nine orthologs and paralogs of those genes were differentially expressed with ABA treatment; most of them were related to synthesis, one to oxidation and transport and none to ABA conjugation. Both oxidation and transport related genes, orthologs of Arabidopsis CYP707A3 and ABACG40 were down regulated. Out of seven orthologs of synthesis-related genes in Arabidopsis (AAO4, ABA4, ABA1, NCED3 and NCED9), four were down and three were up regulated. On the other hand, none of the genes related to ABA oxidation, conjugation or transport were differentially expressed with AA and AEA treatment, no. Only three genes, AAO4, ABA4 and ABA1 related to ABA synthesis were upregulated, in contrast to their response to ABA treatment. With exogenous ABA supply, it was not surprising that genes associated with ABA synthesis were down regulated, whereas their upregulation with both AA and AEA treatments suggests the need for ABA to modulate physiological processes.

Inducers and repressed genes with ABA

Upon ABA treatment, 68 and 47 genes were induced and repressed, respectively, and were in common with previous eight transcriptome studies in Arabidopsis[6,52,53]. Of these 36 and 17 orthologs and paralogs were induced and repressed, respectively in *P. patens*. Most notable orthologs that were induced are related to stress tolerance and transcription regulation, such as, COR47, RD20, RD26, ERD10, HAB1, benzodiazepine receptor-related and zinc finger family proteins. Among the genes that were repressed are scarecrow transcription factor family protein, nodule family protein, BAM2, ATPME3, LSH6, and ATEXPA1. On the other hand, with AA and AEA treatments, only three and five genes, respectively, were common with ABA inducible genes, of which, only RD26 and Zinc Finger transcription factors were upregulated, and the rest

were down regulated. In terms of ABA repressive genes, only three genes were common for both AA and AEA treatments, which were downregulated as with ABA treatment. These analyses reveal that while there may be some overlap in response to ABA, AA and AEA, they mostly induced and repressed different set of genes. In conclusion, ABA signaling pathway is somewhat conserved between angiosperms and bryophytes, and responds uniquely to ABA, AA and AEA in *P. patens*, suggesting a distinct physiological role for them.

lncRNAs identified as key regulatory components with AEA

The relation between the expression of lncRNAs and mRNA was established using Pearson's correlation coefficient and Spearman's coefficient methods. A cut off value of positive 0.6 was selected since the coefficient value close to positive one indicates better correlation between lncRNA and mRNA. In this study, a number of lncRNA shared a linear correlation with mRNA expression suggesting a regulatory role for them, which can be both *cis* and *trans* manner. For *cis* regulation, lncRNA can play a role by overlapping with 10 kb up and 20 kb downstream of the mRNA (Fig. 3.12). The expression of lncRNA and their mode of regulation of corresponding mRNA was unique between the three treatments. With AEA, highest number of lncRNA were observed at 1h, whereas, with AA and ABA at 12h and 24h, respectively (Fig. 3.12). Similar to mRNA and lncRNA DEGs, highest number of lncRNA associated with mRNA regulation were also observed with ABA treatment (Fig. 3.12). For all treatments, *cis* mRNA overlapping lncRNAs were most prominent. The mRNA and lncRNA correlated genes were categorized as described in previous section (Fig. 3.10).

Among the identified genes regulated by lncRNA, except for a few, their expression was unique with treatments. Corresponding lncRNAs for some of the main components of transcription and its regulation, such as RNA polymerase II subunit alpha (ropA), β1 (RPB1) and transcription initiation factor TFIID subunit 2 (TAF2) were identified with AEA treatment (Fig. 3.13). A 5013nt long *cis* overlapping lncRNA regulate the expression of both RPB1 and ropA. In general, lncRNA-mediated gene regulation studies are very limited in plants and as such, regulation of RNA pol II by lncRNA was not reported previously; however, a report suggested that animal RNA pol II subunit β2 is regulated by lncRNA[54]. In *P. patens*, we identified the locus of RBP1, ropA and its regulatory lncRNA in chromosome 13. The TAF2 is regulated by 2302nt long cis-Up10K lncRNA which was identified in locus of chromosome 21. Identification of regulatory components of transcription indicates the significant role AEA might play in gene

regulation. A number of gene regulatory lncRNAs were identified with ABA treatment as well; RBP1 was common between AEA and ABA. With AA, only lncRNA regulatory genes that were identified are PIF4 and MYC2 (Fig. 3.13). Other differentially expressed genes, specific to treatment are involved as transcription factor and in splicing, RNA transport, mismatch repair and protein processing; most of them were upregulated with the expression of corresponding upregulated lncRNA.

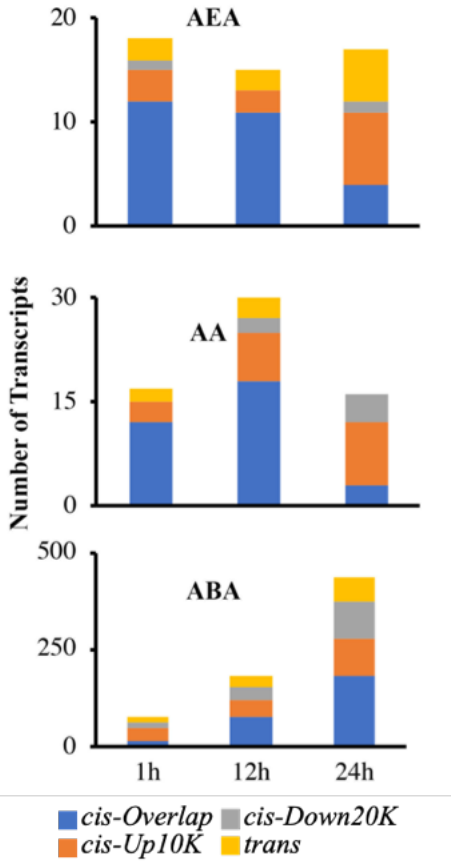


Figure 3.12. Distribution of lncRNA by their mode of mRNA regulation. A number of lncRNA expressions were associated with mRNA in *cis/tran* regulatory manner with AEA, AA and ABA treatment for 1h, 12h and 24h

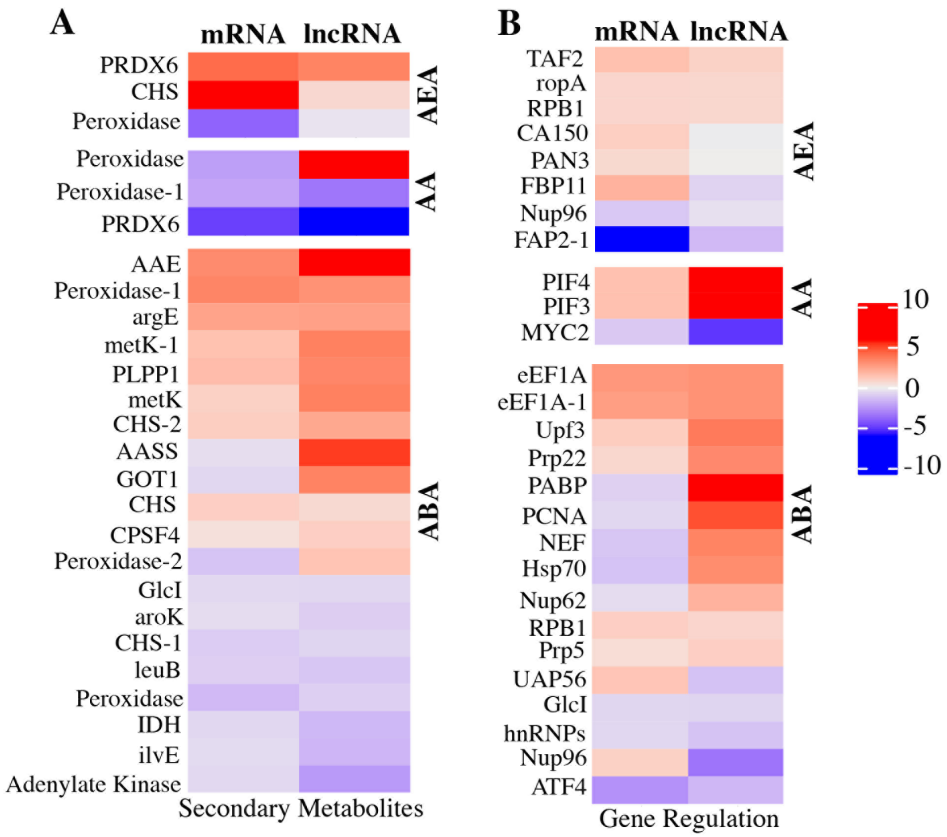


Figure 3.13. Heatmap of mRNA and lncRNA. Log2 expression of mRNA and corresponding lncRNA related to secondary metabolites (A) and gene regulation (B). Log2 ratio range was between +10 to -10. Heatmaps were drawn using ComplexHeatmap program in R

Some notable regulatory genes related to secondary metabolites, such as PABP, PCNA, NEF and Hsp70 were down regulated by upregulated lncRNA (Fig. 3.13). Whereas, UAP56, Nup98, FBP11 were upregulated with the down regulation of corresponding lncRNAs (Fig. 3.13). As for secondary metabolite synthesizing genes, only three genes were identified with AA and AEA treatment and two of them are identical, same sets of genes were also identified with ABA treatment, however, total 20 genes were differentially expressed with ABA.

One of the common genes between all treatments was differential expression of an ABC transporter subunit, also known as ATP-binding cassette (Fig. 3.13). With AEA, ABCB1 subunit upregulated with a down regulation of corresponding lncRNA, which is also true with ABA treatment. The mRNA and lncRNA of ABCB1 located in chromosome 13 and regulated as cis-mRNA-overlap manner. However, with AA, ABCG2 subunit was down regulated with

upregulation of corresponding lncRNA, in a cis-mRNA-up10k manner, and the locus of this gene is in chromosome 16. Further characterization of key lncRNAs and their regulatory role in mosses will provide evolutionary insights into regulation of metabolic pathways, hormone signaling and other molecular responses.

Conclusions

The transcriptome analysis with exogenous treatment of AEA, AA and ABA in temporal manner provided not only an overview of global molecular responses but also identified key molecules. These data will serve as a platform for future studies aimed at understanding the role of unique lipids in mosses and their association with diversified angiosperms. In this study, we provided a comprehensive deep sequencing transcriptome dataset that covered over 82% of *P. patens* reference genome that is currently available; additionally, with *de novo* assembly, we identified novel mRNA and lncRNA.

DEG analysis along with GO term association revealed that molecular responses and targets are unique when *P. patens* protonema was treated with AEA compared to AA or ABA. Upregulation of the genes related to putative GPCR signaling at 1h suggested that AEA might have a short-term effect, whereas ABA have an accumulative or continuous effect.

Analysis of gene expression in select categories showed the uniqueness of AEA signaling. Well known ABA receptors in other model organisms that were highly expressed with ABA in *P. patens* were not induced by AEA. Gene regulatory genes as well as secondary metabolites synthesizing genes also showed a pattern that was mostly treatment specific.

This study also identified number of novel lncRNA and their relationship in mRNA expression. Correlation between lncRNA and mRNA identified potential lncRNAs that could play regulatory role in AEA, AA or ABA mediated signaling.

Overall, transcriptome dataset generated in this study provides a global molecular understanding of a possible AEA-mediated endocannabinoid signaling with a parallel comparison with AA and ABA. Identification of key signaling molecules of AEA does indicate that early land plants like *P. patens* might possess novel AEA/endocannabinoid signaling pathway that is absent in higher plants. As such, this study also provides a substantial direction for further elucidation of molecular responses in early land plants.

REFERENCES

1. Gachet, M. S., Schubert, A., Calarco, S., Boccard, J. & Gertsch, J. Targeted metabolomics shows plasticity in the evolution of signaling lipids and uncovers old and new endocannabinoids in the plant kingdom. *Sci. Rep.* (2017). doi:10.1038/srep41177

2. Pandey, R., Mousawy, K., Nagarkatti, M. & Nagarkatti, P. Endocannabinoids and immune regulation. *Pharmacol. Res.* **60**, 85–92 (2009).

3. Blancaflor, E. B. *et al.* N-Acylethanolamines: Lipid metabolites with functions in plant growth and development. *Plant Journal* (2014). doi:10.1111/tpj.12427

4. Shanab, S. M. M., Hafez, R. M. & Fouad, A. S. A review on algae and plants as potential source of arachidonic acid. *J. Adv. Res.* **11**, 3–13 (2018).

5. Dedyukhina, E. G., Kamzolova, S. V. & Vainshtein, M. B. Arachidonic acid as an elicitor of the plant defense response to phytopathogens. *Chem. Biol. Technol. Agric.* **1**, 1–6 (2014).

6. Finkelstein, R. Abscisic Acid Synthesis and Response. *Arab. B.* **11**, e0166 (2013).

7. Ku, Y. S., Sintaha, M., Cheung, M. Y. & Lam, H. M. Plant hormone signaling crosstalks between biotic and abiotic stress responses. *International Journal of Molecular Sciences* **19**, (2018).

8. Lee, S. C. & Luan, S. ABA signal transduction at the crossroad of biotic and abiotic stress responses. *Plant, Cell and Environment* **35**, 53–60 (2012).

9. Thimm, O. Response of Arabidopsis to Iron Deficiency Stress as Revealed by Microarray Analysis. *Plant Physiol.* **127**, 1030–1043 (2001).

10. Seki, M. *et al.* Monitoring the expression pattern of around 7,000 Arabidopsis genes under ABA treatments using a full-length cDNA microarray. *Funct. Integr. Genomics* **2**, 282–291 (2002).

11. Stevenson, S. R. *et al.* Genetic Analysis of Physcomitrella patens Identifies ABSCISIC ACID NON-RESPONSIVE, a Regulator of ABA Responses Unique to Basal Land Plants and Required for Desiccation Tolerance. *Plant Cell* **28**, 1310–27 (2016).

12. Perroud, P. F. *et al.* The Physcomitrella patens gene atlas project: large-scale RNA-seq based expression data. *Plant J.* **95**, 168–182 (2018).

13. Kopp, F. & Mendell, J. T. Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell* **172**, 393–407 (2018).

14. Iyer, M. K. *et al.* The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* **47**, 199–208 (2015).

15. Zhang, T. *et al.* Long Non-Coding RNA and Breast Cancer. *Technology in cancer research & treatment* **18**, (2019).

16. Mitobe, Y., Takayama, K. ichi, Horie-Inoue, K. & Inoue, S. Prostate cancer-associated lncRNAs. *Cancer Letters* **418**, 159–166 (2018).

17. Gao, Y. *et al.* Comprehensive Characterization of Somatic Mutations Impacting lncRNA Expression for Pan-Cancer. (2019). doi:10.1016/j.omtn.2019.08.004

18. Li, Y. *et al.* Roles of Identified Long Noncoding RNA in Diabetic Nephropathy. *Journal of Diabetes Research* **2019**, (2019).

19. Tan, L., Yu, J. T., Hu, N. & Tan, L. Non-coding RNAs in Alzheimer's disease. *Molecular Neurobiology* **47**, 382–393 (2013).

20. Zhao, X. *et al.* Global identification of Arabidopsis lncRNAs reveals the regulation of MAF4 by a natural antisense RNA. *Nat. Commun.* **9**, 1–12 (2018).

21. Yuan, J. *et al.* Stress-responsive regulation of long non-coding RNA polyadenylation in Oryza sativa. *Plant J.* **93**, 814–827 (2018).

22. Wang, A. *et al.* Genome-wide analysis of long non-coding RNAs unveils the regulatory roles in the heat tolerance of Chinese cabbage (Brassica rapa ssp.chinensis). *Sci. Rep.* **9**, 1–14 (2019).

23. Hon, C.-C. *et al.* An atlas of human long non-coding RNAs with accurate 5′ ends. *Nature* **543**, 199–204 (2017).

24. Szcześniak, M. W., Rosikiewicz, W. & Makałowska, I. CANTATAdb: A Collection of Plant Long Non-Coding RNAs. *Plant Cell Physiol.* **57**, e8 (2016).

25. Chen, Y. *et al.* SOAPnuke: A MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* **7**, 1–6 (2018).

26. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–60 (2015).

27. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).

28. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).

29. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279-85 (2016).

30. Kong, L. *et al.* CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* **35**, W345-9 (2007).

31. Sun, L. *et al.* Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* **41**, e166 (2013).

32. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

33. Wang, L., Feng, Z., Wang, X., Wang, X. & Zhang, X. DEGseq: An R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* **26**, 136–

138 (2009).

34. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).

35. McKenna, A. *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

36. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. Circlize implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).

37. Wickham, H., François, R., Henry, L. & Müller, K. dplyr: A Grammar of Data Manipulation. (2018).

38. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (2016).

39. Wickham, H. The split-apply-combine strategy for data analysis. *J. Stat. Softw.* **40**, 1–29 (2011).

40. Rensing, S. A. *et al.* The Physcomitrella Genome Reveals Evolutionary Insights into the Conquest of Land by Plants. *Science (80-.).* **319**, 64–69 (2008).

41. Lang, D. *et al.* The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. *Plant J.* **93**, 515–533 (2018).

42. Jarillo, J. A. *et al.* Phototropin-related NPL1 controls chloroplast relocation induced by blue light. *Nature* **410**, 952–954 (2001).

43. Keereetaweep, J., Blancaflor, E. B., Hornung, E., Feussner, I. & Chapman, K. D. Ethanolamide oxylipins of linolenic acid can negatively regulate Arabidopsis seedling development. *Plant Cell* **25**, 3824–40 (2013).

44. Gaudet, P., Livstone, M. S., Lewis, S. E. & Thomas, P. D. Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. *Brief. Bioinform.* **12**, 449–462 (2011).

45. Krauss, G. *Biochemistry of Signal Transduction and Regulation.* (Wiley-VCH, 2014).

46. Fernando, V. C. D. & Schroeder, D. F. Role of ABA in Arabidopsis Salt, Drought, and Desiccation Tolerance. in *Abiotic and Biotic Stress in Plants - Recent Advances and Future Perspectives* (InTech, 2016). doi:10.5772/61957

47. Shu, K., Zhou, W., Chen, F., Luo, X. & Yang, W. Abscisic Acid and Gibberellins Antagonistically Mediate Plant Development and Abiotic Stress Responses. *Front. Plant Sci.* **9**, 416 (2018).

48. Teaster, N. D. *et al.* N-acylethanolamine metabolism interacts with abscisic acid signaling in Arabidopsis thaliana seedlings. *Plant Cell* **19**, 2454–2469 (2007).

49. Pandey, S. & Assmann, S. M. The Arabidopsis Putative G Protein-Coupled Receptor GCR1 Interacts with the G Protein Subunit GPA1 and Regulates Abscisic Acid Signaling. *Plant Cell* **16**, 1616–1632 (2004).

50. Teaster, N. D. *et al.* Overexpression of Fatty Acid Amide Hydrolase Induces Early

Flowering in Arabidopsis thaliana. *Front. Plant Sci.* (2012). doi:10.3389/fpls.2012.00032

51. Haq, I. & Kilaru, A. An endocannabinoid catabolic enzyme FAAH and its paralogs in an early land plant reveal evolutionary and functional relationship with eukaryotic orthologs. *Sci. Rep.* **10**, 3115 (2020).

52. Wang, R. S. *et al.* Common and unique elements of the ABA-regulated transcriptome of Arabidopsis guard cells. *BMC Genomics* **12**, (2011).

53. Choudhury, A. & Lahiri, A. Comparative analysis of abscisic acid-regulated transcriptomes in Arabidopsis. *Plant Biol.* **13**, 28–35 (2011).

54. Mee, D. J. M. du, Ivanov, M., Parker, J. P., Buratowski, S. & Marquardt, S. Efficient termination of nuclear lncRNA transcription promotes mitochondrial genome maintenance. *Elife* **7**, (2018).

CHAPTER 4. *IN VIVO* CHARACTERIZATION OF ENDOCANNABINOID SYSTEM MOLECULES

Imdadul Haq and Aruna Kilaru

Departments of Biological Sciences and Biomedical Sciences, East Tennessee State University, Johnson City, TN, 37614, USA

kilaru@etsu.edu

In vivo characterization of endocannabinoid system molecules

Introduction

The endocannabinoid system (ECS) is a well-known and studied signaling pathway in mammals. The name endocannabinoid is related to the psychotropic compound, D^9-tetrahydrocannabinol (THC) found in *Cannabis sativa.* The process of identification and characterization of the ECS components accelerated after the discovery of the receptors for THC, cannabinoid binding receptors 1 and 2 (CB1 and CB2)[1–4]. In mammals, the simplified version of ECS consists of ligands, a number of metabolic enzymes and receptors[5]. The two well-distinguished ligands are the lipids *N*-arachidonoylethanolamine (AEA/anandamide) and 2-arachidonoylglycerol (2-AG) from two distinct lipid families, *N*-acylethanolamines (NAEs) and 2-acylglycerols (2-AcGs), respectively[1], with CB1 and CB2 as the primary receptors[4,6]. The metabolic enzymes include *N*-acyl-phosphatidylethanolamine-hydrolyzing phospholipase D (NAPE-PLD), *sn*-1-specific diacylglycerol lipase (DGLa, DGLb), FAAH (fatty acid amide hydrolase), and monoacylglycerol lipase (MAGL/MGL)[7–11].

Both AEA and 2-AG are known as endocannabinoids because of their endogenous biosynthesis and specificity as ligands towards CB receptors. There are other molecules from the same family of NAE and 2-AcGs, which are not specific to CB receptors but rather use different G protein-coupled receptors (GPCRs), nuclear receptors, and ion channels, and play their part in numerous biological processes[12]. The NAE 20:4 was identified recently in bryophytes and lower organisms, but their physiological role has not yet been determined[13]. Targeted lipidomic analysis of various developmental stages of *P. patens* showed that NAE 20:4 levels were about 20% of the total NAEs in protonema and early and late gametophyte stages (data not published).

In both plants and animals, NAPE is known to be the precursor for NAE, however, identification and confirmation of key enzymes in plants is limited compared to animals. The catabolic reaction from NAPE to NAE occurs either in one step with the help of NAPE-PLD or two steps with PLC and PTN22, sPLA2 and Lyso PLD or ABHD4 and GDE1[14]. In plants, several PLD have been identified but NAPE specific PLD is not yet characterized. A mouse ABHD4 homolog was identified in Arabidopsis that previously was reported to be involved in starch metabolism, therefore it was considered as a poor candidate as NAPE synthase[15].

Significant research needs to be done to gain a better understanding of NAE synthesis in plants. On the other hand, the NAE catabolic enzyme fatty acid amide hydrolase (FAAH) is highly conserved in eukaryotes and has been studied extensively in both plants and animals[16]. In humans, two orthologs, FAAH1 and FAAH2 with specificity towards different NAEs were identified[11,17]. In Arabidopsis, AtFAAH1 is the most studied even though there are few other candidates. Knockout (KO) of *FAAH* resulted in increased NAEs, irrespective of the organisms, while the converse was true with the overexpressors (OE). In animals, FAAH is targeted for therapeutic treatments and its involvement is reported in anxiety, depression, obesity, diabetic, addiction to marijuana, tobacco or alcohol addiction and more[18–22]. In plants, overexpression of *FAAH* in Arabidopsis resulted in early flowering, and hypersensitivity to ABA and host/non-host pathogens, whereas KO lines showed hypersensitivity to NAE 12:0 during seedling germination[23–27]. In *P. patens*, we identified nine orthologs of FAAH of which PpFAAH1 – 4 showed similarity towards plant FAAH and PpFAAH6 – 9 towards animal FAAH[16]. Biochemical characterization of PpFAAH1 showed higher specificity to NAE 20:4 compared to NAE 16:0.

As endocannabinoid receptors, CB1 and CB2 are exclusive because of their tissue-specific expression and ligand specificity[28]. CB1 and CB2 are canonical GPCRs, which consist of seven transmembrane domains with downstream signal activation through G proteins[12]. In plants, receptors for NAEs are yet to be identified and characterized. Also, understanding of GPCRs in plants is limited. To date, there are four GPCR-like proteins in Arabidopsis that have been experimentally characterized; GCR1, GCR2, GTG1 and GTG2[29,30]. Out of the four, GCR2 was reported to be an ABA receptor, although later it was characterized as lanthionine synthetase[31]. Whereas both GTG1 and GTG2 were demonstrated as GPCR-type proteins; they showed intrinsic GTP-binding and GTPase activity[30]. Both of the proteins bind to ABA, and mutants of these genes exhibit ABA hypersensitivity. This indicated that GTG1 and GTG2 may be involved in ABA-mediated seed germination, root elongation, and flowering[31]. However, the bioinformatic analysis suggested that both proteins have nine transmembrane domains which is not in agreement with documented canonical GPCR[32].

Arabidopsis GCR1 is the only canonical GPCR characterized in plants, which was also supported by bioinformatic analysis[33]. However, GCR1 is not widely accepted by the scientific

community as a GPCR because there is no evidence for their ligands and interaction with Gα subunit protein (GPA1)[31]. Nevertheless, GCR1 was reported to be involved in a number of signaling and physiological aspects in Arabidopsis. Two independent groups reported different functions of GCR1; one group showed GCR1 as a cytokinin receptor but afterward discovered that an independent mutation was responsible for this cytokinin response[34]; the other group reported GCR1 to be involved in abolishing seed dormancy and decreasing flowering time[35]. Knockout lines of GCR1 were implicated in seed germination in response to gibberellins and brassinosteroids[34]. Pandey and Assmann (2004) reported that the GCR1 KO is hypersensitive to ABA that was involved root growth, stomatal response, and regulation of ABA-induced gene expression[29]. It also was hypersensitive to sphingosine-1-phosphate, a lipid metabolite and a transducer of the ABA signal upstream of GPA1, which indicates that GCR1 may act as a negative regulator of GPA1 facilitated ABA responses of guard cells[36,37].

The G-proteins in plants were first discovered in Arabidopsis; GPA1[35] was shown to be expressed during all stages of development except in mature seeds[38]. Since then, GPA1 has been studied extensively and reported to be involved in the development of external morphology as well as molecular physiology. For instance, *gpa1* mutants showed reduced cell division in hypocotyls and leaves but no alteration of root growth was observed[36]. Also, GPA1 affected cell division during seed germination, and was hypersensitive to ABA[38,39] and gibberellins but less sensitive to brassinosteroids[34]. In stomatal guard cells, *gpa1* mutants are ABA hypersensitive but exhibited less sensitivity in stomatal opening and inward K+ channel regulation[37]. Loss of GPA1 function in *Nicotiana plumbaginifolia* led to increased expression of NpGPA1 upon a naphthyl-acetic acid treatment, but reduced expression with ABA and salicylic acid treatment. Differential expression with gibberellins was not observed[40].

In *P. patens*, however, the canonical Gα subunit was reported to be missing; instead, an extra-large Gα (PpXGL) protein was identified and shown to be important to complete the life cycle[41]. In the same study, presence of two copies of Gβ and Gγ was also reported. Knockout lines of PpXLG or PpGβ2 resulted in slower gametophyte production with normal reproductive structures but sporophyte formation was omitted. PpXLG showed standard Gα subunit characteristics with GTP- binding and hydrolyzing ability[41].

Taken together, a global understanding of NAE signaling in plants is far from complete. In moss, endocannabinoids are identified and among the metabolic enzymes, only FAAH has been well-studied[16]. Other metabolic enzymes and receptors yet to be functionally characterized. Here, we attempted to identify and characterize some of the ECS components *in vivo* and *in vitro*.

Materials and Methods

Protein extraction and cellular fractionation

To determine the amidohydrolase activity, total proteins were extracted from protonema as well as early and late gametophyte stages as previously described[42]. Briefly, tissue samples were ground in liquid nitrogen and homogenized in a homogenization buffer (100 mM potassium phosphate, pH 7.2, 10 mM KCl, 1 mM EDTA and 400 mM sucrose). The homogenates were filtered using cheesecloth followed by centrifugation at 650g for 10 min at 4°C. The supernatant was centrifuged at 10,000g for 30 min at 4°C (Sorvall, SS 34 rotor). Soluble protein was collected as supernatant and used as total protein. The protein sample was then centrifuged at 150,000g for 60 min at 4°C in an ultracentrifuge (Sorvall Discovery 90). Microsomes were collected as a pellet and resuspended in homogenization buffer. Protein concentration was determined using Nanodrop (ND-1000).

Amidohydrolase activity

To determine the amidohydrolase activity for total protein and the microsomal fraction from protonema and gametophytes, 10 µg of extracted soluble protein samples were used. For substrate, 100 µM radiolabeled [1-^{14}C] NAE 20:4 was used. The protocol for enzyme assay, lipid extraction and product detection were as previously described[16].

Generation of constructs for PpFAAH1 knockout and overexpression

To create the KO lines using homologous recombination, a full-length transcript sequence including exons and introns of *PpFAAH1* was obtained from the Phytozome 12 website. The 5' and 3' flanking regions of *PpFAAH1* were amplified from the extracted genomic DNA of *P. patens*. The 5' flanking region contains 800bp, which includes 3 exons and 3 introns, whereas, the 3' flanking region contains 1100bp with 3 exons and 4 introns. Amplified PCR products were gel purified and cloned into a pMP1159 entry vector. Cloned vector was then linearized using SalI and NotI restriction enzymes and gel purified according to manufacturer's

instructions and used for polyethylene glycol (PEG) mediated protoplast transformation to generate KO mutants.

To generate the overexpression lines, full-length *PpFAAH1* was cloned into an entry vector as described before[16]. The confirmed *PpFAAH1* clone was sub-cultured in a destination vector, pBALGATE. The transformed plasmid was then linearized by SwaI and transformed using the PEG mediated transformation protocol described below.

PEG mediated transformation

Transformation of isolated protoplasts was performed as previously described[43]. Briefly, protoplasts were extracted from seven-day old *P. patens* grown in BCDAT medium (0.5 M $Ca(NO_3)_2$ $4H_2O$, 4.5 mM $FeSO_4$ $7H_2O$, 0.1 M $MgSO_4$ $7H_2O$, 1.84 mM KH_2PO_4, 1 M KNO_3, 4.5 mM $FeSO_4$ $7H_2O$, 500 mM Ammonium tartrate, 50 mM $CaCl_2$) with alternative TES (0.22 mM $CuSO_4$ $5H_2O$, 10 mM H_3BO_3, 0.23 mM $CoCl_2$ $6H_2O$, 0.1 mM Na_2MoO_4 $2H_2O$, 0.19 mM $ZnSO_4$ $7H_2O$, 2 mM $MnCl_2$ $4H_2O$, 0.17 mM KI). Tissues were transferred to a Petri dish with 8% mannitol and 2% driselase (Sigma D9515-25G) and incubated for 1h at room temperature with gentle shaking. The suspension was then filtered through 75 µM mesh (BD Falcon) followed by centrifugation at 250g for 5 min at 4 °C. The supernatant was discarded, and the protoplast pellet was resuspended in 8% mannitol followed by two repetitions of centrifugation and resuspension. Protoplasts were counted using a hemocytometer. For transformation, 10^6 protoplasts were resuspended in mannitol-Mg solution (0.4 M mannitol, 15 mM MgCl2, 4mM MES pH 5.7) and incubated for 20 min at room temperature. The knockout or overexpression construct containing DNA (15 µg) was added to the protoplasts with gentle swirl. Then 700 µl of PEG/Ca solution was added into each protoplast-DNA mixture, swirled and incubated for 30 min at room temperature. Subsequently, each mixture was diluted in W5 solution (154 mM NaCl, 125 mM CaCl2, 5 mM KCl, 2 mM MES pH 5.7) and centrifuged at 250g for 5 min. Supernatants were discarded and the pellet was resuspended in the PRM-T medium. One ml of resuspended protoplasts was then plated on PRM-B containing media overlaid with cellophane. Plates were maintained at 25°C in a growth chamber (PERCIVAL:CU22L) with 16h light and 8h dark cycle. After four days of growth, the cellophane with protoplast culture was transferred to plates containing hygromycin for selection of positive clones. Individual colonies were collected and sub-cultured for further analysis.

AEA and ABA treatment on FAAH overexpressor line 8 (OE8) mutant

The protonema of wild type and *FAAH* OEs were grown for 10 days on plates containing BCDAT growth media covered with cellophane. Individual colonies of 10-day old protonema were then transferred to 12 well plates containing BCDAT media with AA, AEA, ABA or DMSO. Four concentrations; 0, 1, 10 and 50 μM of AEA and ABA with three technical replicates were used in the experiment. Plain BCDT media or DMSO were used as negative controls for the experiment. Cultures were imaged on day 0 and every 3 days for the next 21 days using a digital camera (Canon; EOS 60D). The area of growth was determined from the images using ImageJ software. The experiment was repeated three times for biological replications and standard error was calculated.

Protein-protein interaction studies

The GST tagged PpFAAH1, expressed in *E. coli*[16] was immobilized in a glutathione fused agarose column. Total protein extractions from wild type (WT) and OE8 protonema cultured with and without exogenous ABA were incubated for 1h in PpFAAH1-glutathione agarose column. Columns were then handled by a protein mass spectrometry facility (MS Bioworks) and interacting proteins were identified using the following protocol. Each column was incubated in pre-heated 1.5X LDS buffer (106 mM Tris HCl, 141 mM Tris Base, 2% LDS, 10% Glycerol, 0.51 mM EDTA, 0.22 mM SERVA Blue, 0.175 mM Phenol Red, pH 8.5) for 15 min. Columns were centrifuged and the collected samples were separated by a MES buffer based SDS-PAGE using a 10% Bis-Tris NuPAGE (Invitrogen). Subsequently, in-gel trypsin digestion followed by quenching with formic acid was performed. Samples were then analyzed by nano LC-MS/MS using Waters NanoAcquity HPLC system (ThermoFisher Q Exactive). The peptides were loaded on a trapping column and eluted over a 75 μm analytical column at 350 nL/min. Both columns were packed with Luna C18 resin (Phenomenex). The MS was set in data-dependent mode. The resolution of Orbitrap MS was at 70,000 FWHM (full width at half maximum) and 17,500 FWHM for MS/MS. For the MS/MS selection, the fifteen most abundant ions were obtained. Each sample utilized an hour of instrument time. Mascot (Matrix Science software) was used for data processing, with the parameters of enzyme-Trypsin/P, Database-NCBI *Physcomitrella patens*, fixed modification, carbamidomethyl (C), mass value-monoisotopic, peptide mass tolerance-10 ppm, fragment mass tolerance- 0.02 Da, max missed cleavage-2. Scaffold (Proteome software) was used finally to parse the Mascot DAT files. Data

was validated and filtered to create non-redundant list of per sample. For data filtration, 1% protein and peptide false discovery rates (FDR) were used.

Cloning the PpGα1, PpGα2, GPCR like protein 1 (PpGLP1) and regulator of G protein signaling (PpRGS)

Full length *PpGLP1* (primers: FW-5'CACCATGATCGAAGGATTGTCGCCCGCA3', RV- 5'TTATGGGCCTTGATCAGCTTCCAG3') and *PpRGS* (primers: FW-5'CACCATGCCTGAGATATTGCCACATGGC3', RV-5'CTAAAATTCATGATGCCATAATTGC3') were PCR amplified and cloned into a p19 entry vector. Sequences for both *PpGα-1* and *PpGα-2* were synthesized commercially by IDT (Integrated DNA Technologies, Coralville, IA) and cloned into an entry vector, p19. Cloned vectors were confirmed by restriction digestion, colony PCR and sequencing. Confirmed clones were then subcloned in the destination vector, pDEST17, with a N-terminus 6xhis-tag.

Results and Discussion

In *P. patens*, we identified nine orthologs of fatty acid amide hydrolase, PpFAAH1-9, a catabolic enzyme of anandamide[16]. Further, we biochemically characterized PpFAAH1 and examined the structural details of the nine paralogs in relation to AtFAAH, RtFAAH and HsFAAH, to understand their functional, structural and evolutionary relationship[16]. To further our understanding, we examined *in vivo* amidohydrolase activity and gene expression at different developmental stages of *P. patens*, as well as generated mutants and identified interacting proteins of PpFAAH1 with or without ABA.

Table 4.1. In vivo amidohydrolase activity. Total and microsome protein of protonema and gametophyte were used to determine endogenous amidohydrolase activity against radiolabeled NAE 20:4. Total and specific activity were measured. Data are mean ± SE (n=3)

	Free Fatty Acid	
	Total Activity	Specific Activity
	nmol/min	nmol/min•mg
Total Protein of Protonema	19.89 ± 2.04	0.05
Microsomes of Protonema	6.81 ± 1.73	0.28
Total Protein of Gametophyte	49.19 ± 3.96	0.15
Microsomes of Gametophyte	33.12 ± 2.54	2.68

mRNA expression levels of PpFAAH paralogs varied with developmental stages

In vivo amidohydrolase activity was performed with total protein or microsomes from protonema and gametophyte tissues. The specific activity of the gametophyte microsomes and total protein using AEA as a substrate was higher when compared to the microsomal and total protein of protonema, respectively (Table 4.1). The total amidohydrolase activity of both total and microsomal protein from gametophyte tissue was two-fold and five-fold higher compared to the respective protein source of protonema (Table 4.1). In order to understand if the amidohydrolase activity at different developmental stages was related to *PpFAAH1* expression, we performed qPCR using the mRNA from the same developmental stage tissue samples. The results indicate that there was no significant difference in *PpFAAH1* expression in protonema and gametophyte tissues (Fig. 4.1A). These data suggest different possibilities; 1) PpFAAH1 could be the only active amidohydrolase enzyme in *P. patens* and its turnover rate from mRNA to enzyme must be higher in the gametophyte stage compared to protonema; 2) more than one PpFAAH are active and specific to NAE 20:4, and their expression was higher in gametophyte relative to protonema. To address these possibilities, we reviewed the expression data of nine FAAH paralogs, which were available from eFP browser. Although the expression data for protonemal stage was not available, it is clear that PpFAAH paralogs show tissue or developmental stage-specific mRNA expression patterns (Fig. 4.1B).

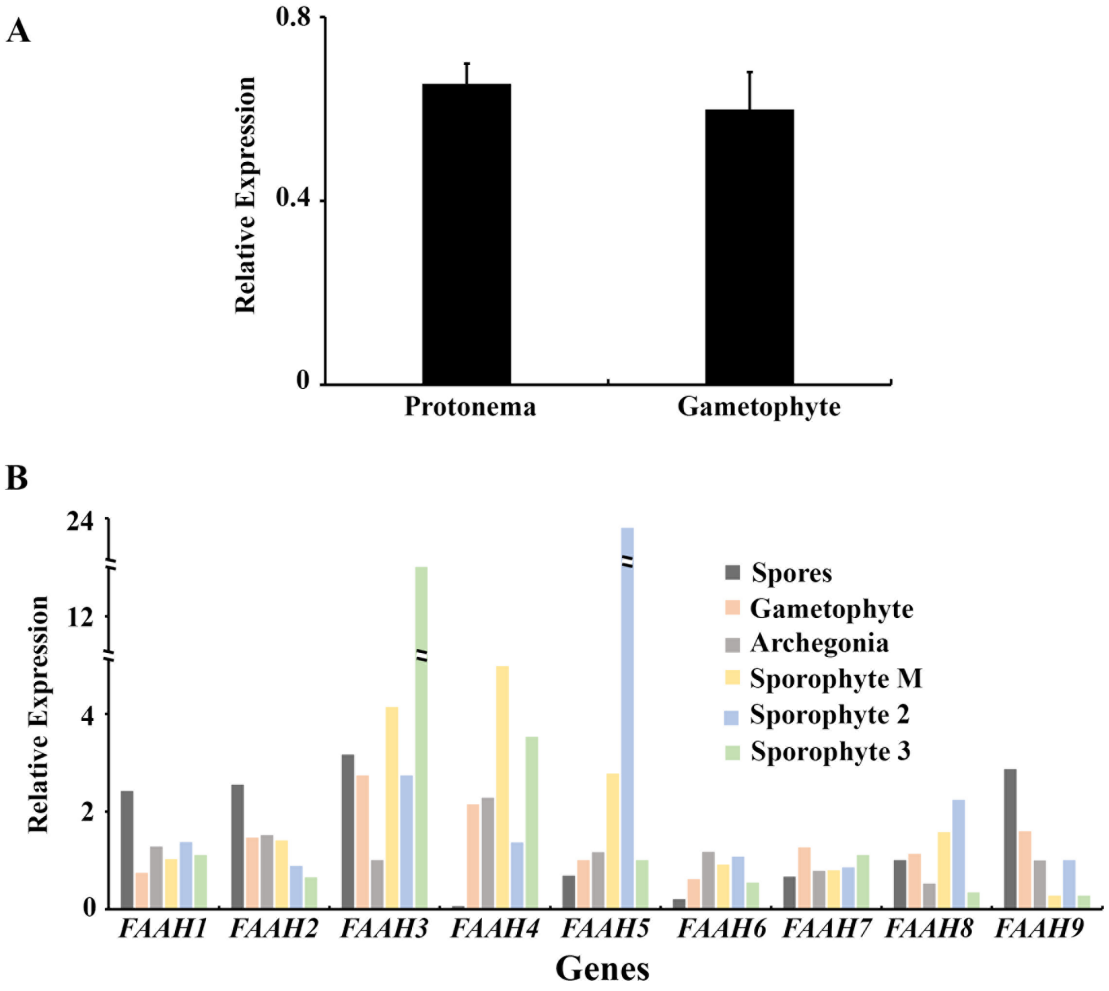


Figure 4.1. Expression of *PpFAAH* at developmental stages. A). Relative expression of *PpFAAH1* in protonema and gametophyte, which was calculated using a housekeeping gene *Actin5* as the base level. Data are mean ± SE (n=3). B). Absolute expression data of *PpFAAH1* to *PpFAAH9* in six developmental stages was obtained from the eFP browser

The eFP mRNA expression data of nine *PpFAAH* further revealed that the sporophyte or diploid stages together (M, 2 and 3) show much higher expression levels for most of the *PpFAAH* paralogs (Fig. 4.1B), relative to the haploid stages (spores, archegonia and gametophyte). Highest expression of *PpFAAH1* was observed in haploid spores, which declined with the development of gametophyte. All the *PpFAAH* paralogs were moderately expressed in archegonia, while *PpFAAH3* and *PpFAAH4* were among the paralogs that showed higher expression in the gametophyte (Fig. 4.1B). Interestingly, *PpFAAH4* was highly expressed in mature sporophyte and sporophyte 3 stages but was absent in spores (Fig. 4.1B). In general, the

expression of *PpFAAH6* and *PpFAAH7* remained low in all the developmental stages. Among all *FAAH* paralogs, *PpFAAH5* expression was highest in sporophyte stage 2 by 24-fold while the second highest expression was of *PpFAAH3* in sporophyte 3 with 12-fold higher expression (Fig. 4.1B). These varied levels of expression of *FAAH* paralogs in different developmental stages suggest that they might be involved in stage-specific functions. These data also suggest that amidohydrolase activity might be contributed by different PpFAAH paralogs depending on the developmental stage.

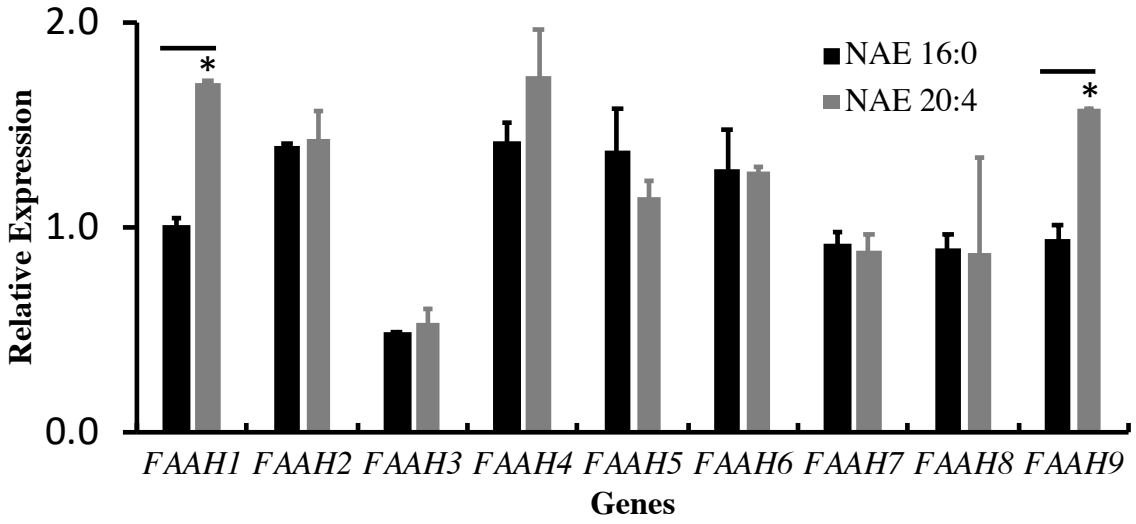


Figure 4.2. Expression of *PpFAAH* upon exogenous treatment. *PpFAAH1* to *PpFAAH9* expression in protonema cultured with or without exogenous NAE 16:0 (black) or NAE 20:4 (gray) was analyzed using qPCR. The asterisk (*) sign and line (-) on the top of the bar graph represent significant difference relative to control without inhibitor. Data are mean ± SE (n=3)

PpFAAH paralogs show differential response to saturated and unsaturated NAEs

Structural analyses of PpFAAH paralogs in *P. patens* suggested that they might be substrate-specific[16]. Among the two human FAAH, HsFAAH1 was specific to NAE 20:4 whereas HsFAAH2 preferred NAE 18:1[17]. To determine the substrate preference of the PpFAAH paralogs, protonema of *P. patens* were cultured in the presence of exogenous NAE 16:0 and NAE 20:4 for 6h. The mRNA extracted from these cultures was analyzed by qPCR to determine the relative expression levels of nine *PpFAAHs*. Protonema with NAE 20:4 treatment showed a significant ($p = 0.02$; student t-test) upregulation of *PpFAAH1*, *PpFAAH4*, and *PpFAAH9*, relative to NAE 16:0 treatment (Fig. 4.2). The expression *PpFAAH2, PpFAAH6, PpFAAH7* and

PpFAAH8 did not differ between the treatments. Relatively, higher expression of *PpFAAH5* was observed with the NAE 16:0 treatment, whereas *PpFAAH3* expression was lowest with both treatments (Fig. 4.2). These data from *in vitro* experiments suggest that PpFAAH are specific to saturated or unsaturated NAEs; however, this needs to be validated with *in vivo* studies. Considering the *in vivo* amidohydrolase activity with the expression data from eFP browser and exogenous NAE treatment, we could predict that more than one FAAH enzyme is likely active in *P. patens* with substrate specificity. However, additional studies are required to confirm their specificity and tissue-specific functional role.

Overexpression of PpFAAH1 showed exogenous NAE tolerance but inhibited developmental transition

To understand the implications of altered NAE levels, including anandamide, we have generated FAAH KO and OE mutants using homologous recombination. The KO lines are yet to be confirmed but we eight PpFAAH1 OE lines (OE1 to OE8) were confirmed by qPCR. The OE lines OE1 to OE7 showed two to three-fold higher expression compared to wild type, whereas the OE8 line showed almost six-fold higher expression (Fig. 4.3A). Mutant line OE8 was used for further *in vivo* characterization. Observation of growth on regular BCDAT medium of OE8 and WT showed an obvious phenotypic difference between them (Fig. 4.3B). The OE8 line did not make a developmental transition to a mature gametophyte whereas the other OE lines (OE1 to OE7) did not show such phenotype. In case of OE8 mutant line, generation of a leafy gametophyte structure is missing and instead showed continuous filamentous growth like protonema (Fig. 4.3B). We further evaluated the response of OE8 to the presence of exogenous AEA and ABA. The OE8 line showed tolerance to exogenous AEA even at a higher concentration (50 μM) and an enhanced growth at lower concentration (1 mM; Fig. 4.4B), relative to WT, which was inhibited in a dose-dependent manner (Fig. 4.4A). Although the growth response was similar between WT and OE8 by 21 days at lower concentrations of AEA, the OE8 continued to show better growth that WT at higher AEA concentration (Fig. 4.4C). In contrast, both WT and OE8 showed no tolerance to exogenous ABA (Fig. 4.5A and B) and their cumulative growth response up to 21 days remained comparable (Fig. 4.5C).

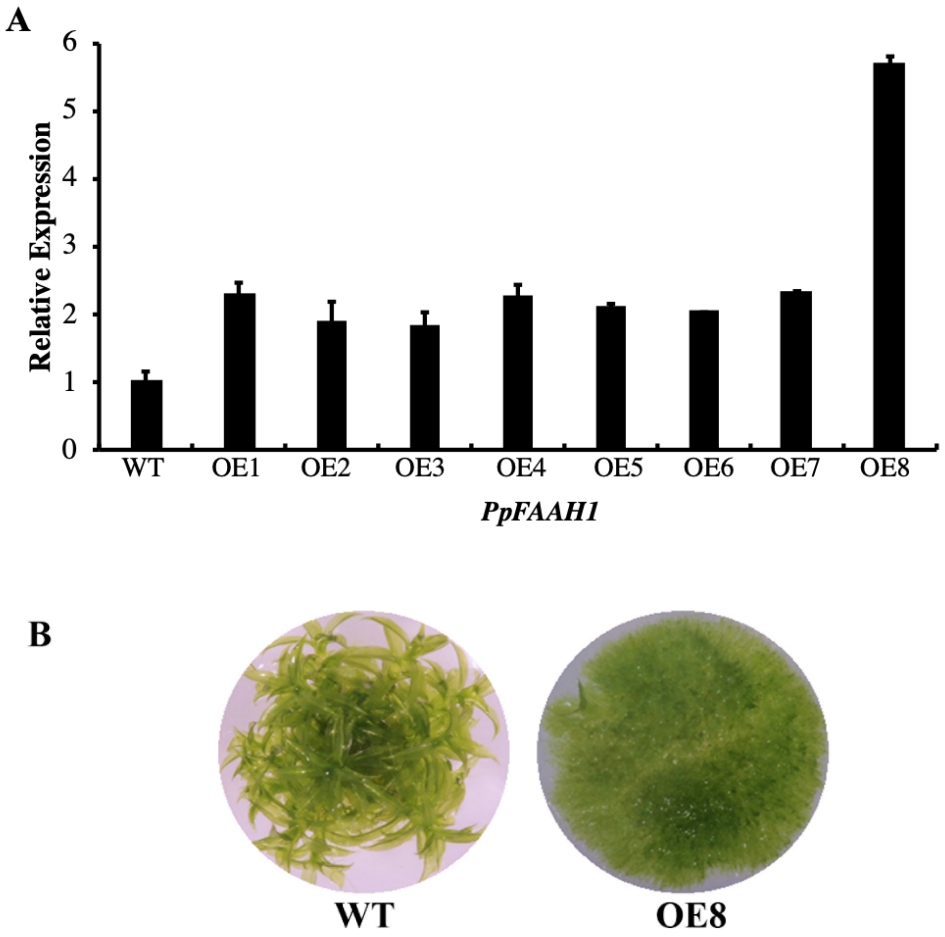


Figure 4.3. Over expression of *PpFAAH1*. A). Eight *PpFAAH1* overexpression lines (OE1 to OE8) were confirmed using qPCR. Relative expression was determined using Actin5 and WT *PpFAAH1*. Data represent mean ± SE (n=3). B). Phenotype of 21-day old WT and OE8 line

Interestingly, the exogenous AEA did not recover the phenotype of OE8, suggesting that the phenotype is likely associated with expression levels of *PpFAAH1* or other the levels of other NAE types, but necessarily associated with anandamide content. In Arabidopsis, FAAH OE showed enhanced growth and early flowering but hypersensitivity to biotic and abiotic stresses [23,26,27]. Domain deletion studies of AtFAAH further revealed that while enhanced growth and tolerance to NAE phenotype was associated with catalytic activity of FAAH, the hypersensitive response to stressors and ABA was independent of its enzyme activity, suggesting a possible dual role for AtFAAH[44]. Although the OE8 of *P. patens* did not show any sensitivity against ABA (Fig. 4.5), their inability to switch to gametophyte stage might likely be related to non-enzymatic role of *PpFAAH1*.

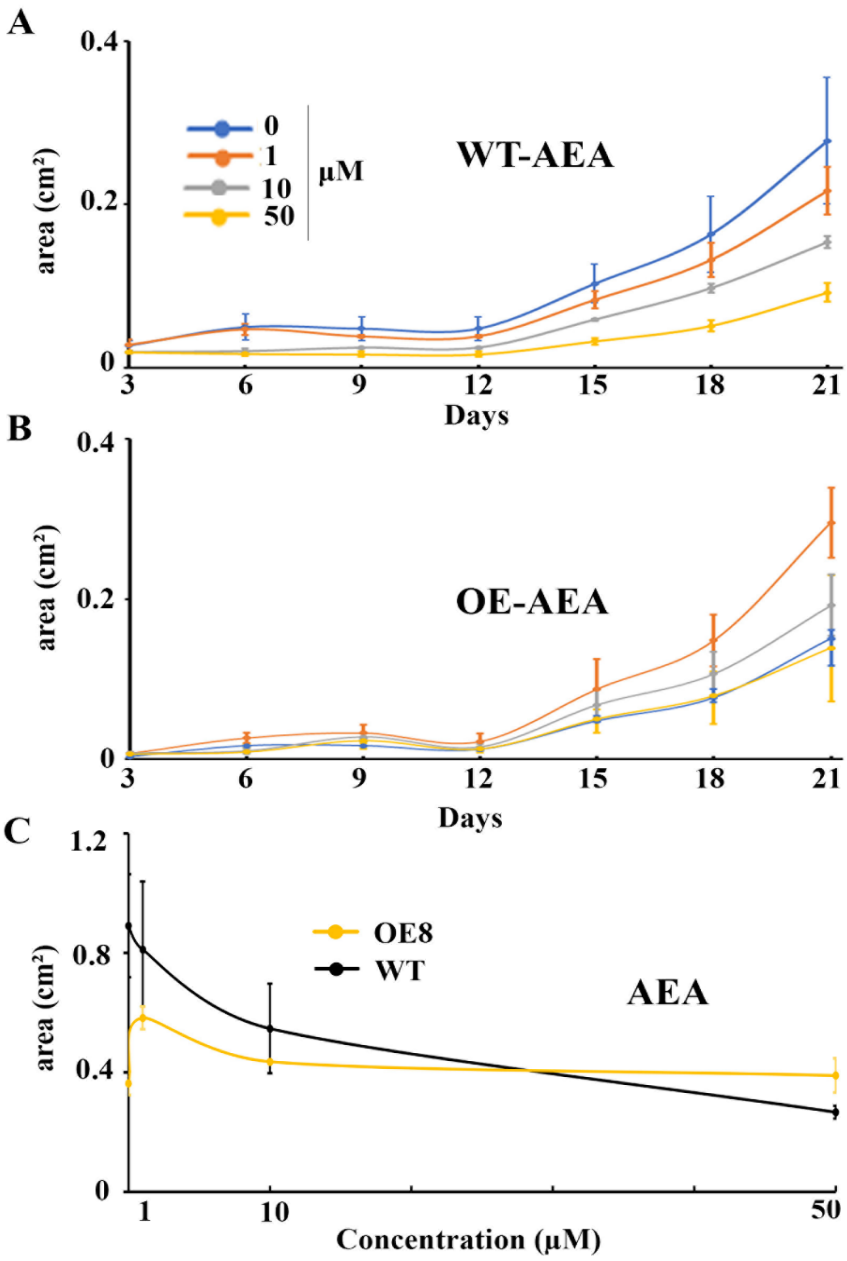


Figure 4.4. Growth assay of WT and OE in response to AEA treatment. Growth of WT (A) and OE (B) with AEA treatment for 21 days. C). Cumulative growth of WT and OE for 21 days at different concentration of AEA. Data represent mean ± SE (n=3)

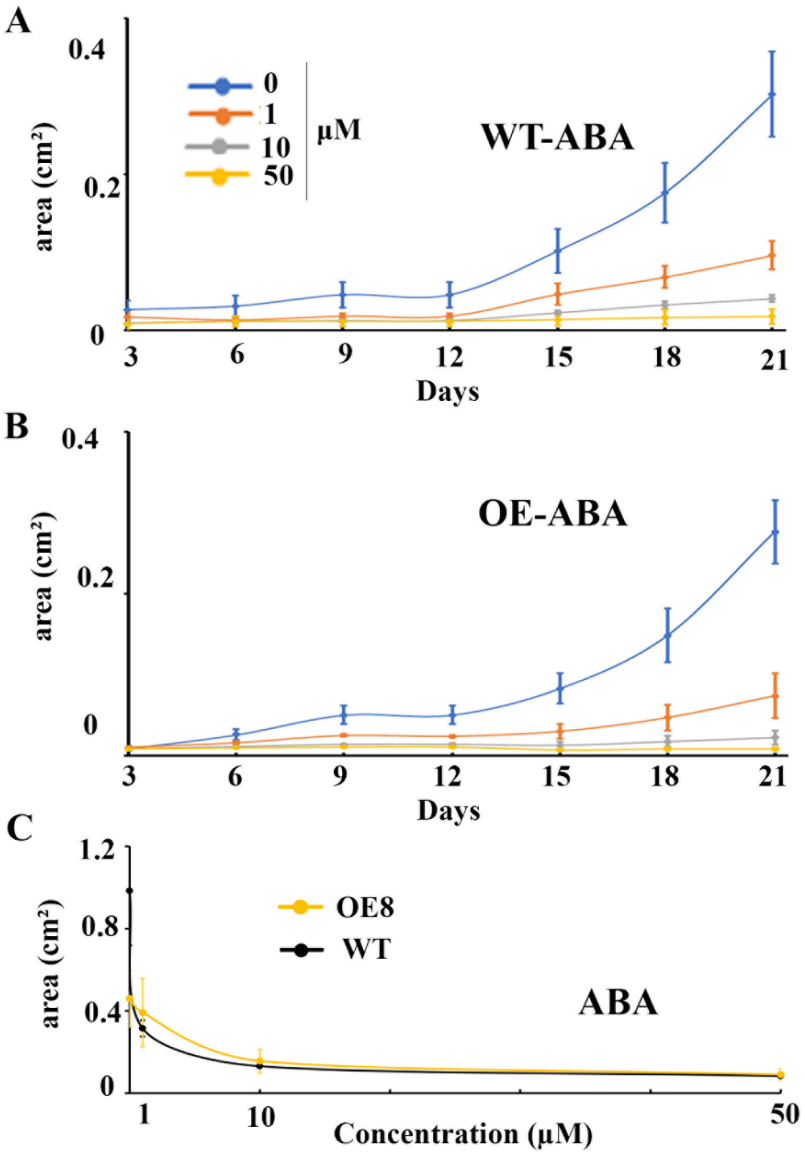


Figure 4.5. Growth assay of WT and OE in response to ABA treatment. Growth of WT (A) and OE (B) with ABA treatment for 21 days. C). Accumulative growth of WT and OE for 21 days at different concentration of ABA. Data represent mean ± SE (n=3)

Alternatively, the reduction of AEA level in OE8 due to the specificity of PpFAAH1, could impact AEA-mediated developmental pathway in an irreversible manner, as shown with PpXLG or PpGβ2 KO lines that could not form a sporophyte[41]. Furthermore, we could examine if the other *PpFAAH* OE lines with 2-3 fold higher expression levels of *PpFAAH1* but no phenotype show same levels of tolerance to AEA, as OE8; if such is the case, OE8 phenotype could be associated with a possible non-catalytic activity of PpFAAH1 and likely through

interaction with other proteins. Quantification of endogenous levels of NAEs in these various PpFAAH1 OE lines will also provide some mechanistic insights.

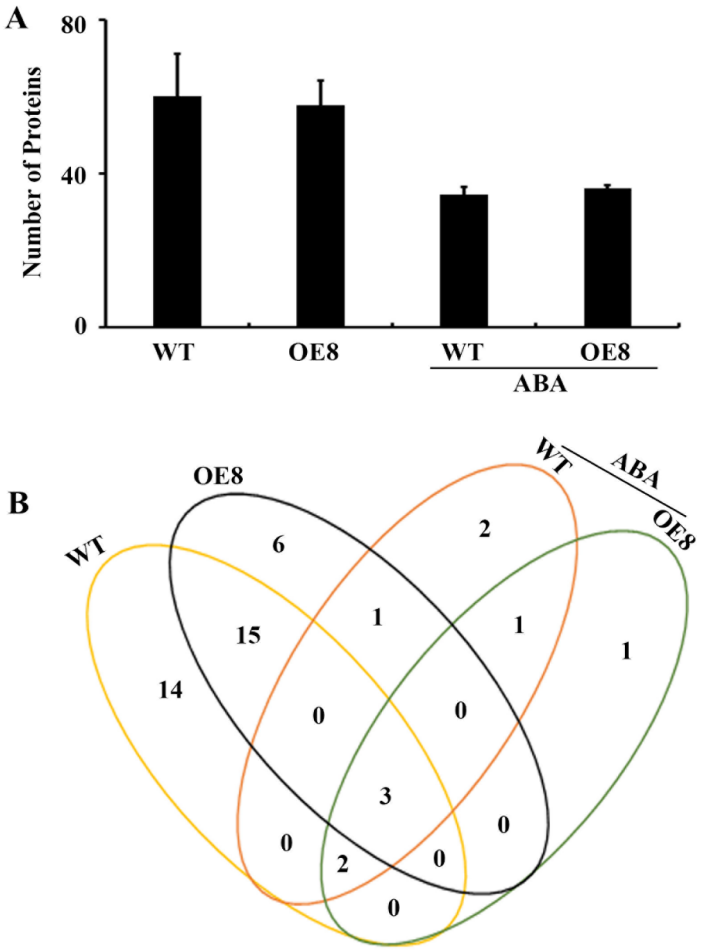


Figure 4.6. Identification of PpFAAH1 interacting proteins. A). Total number of interacting proteins were identified from protein samples of wild type (WT) and PpFAAH1 overexpressor (OE8) protonema cultured with or without ABA treatment for 24h. B). Venn diagram shows the number of interacting proteins identified for each treatment and genotype

Identification of interacting proteins of PpFAAH1

Interacting proteins of PpFAAH1 and their localization in organelles such as chloroplasts and mitochondria implies its involvement in energy related metabolic and developmental processes. To identify if such interactions play a role in developmental transition and growth, we carried out a pull-down assay followed by LC-MS/MS. To identify interacting partners of PpFAAH1, GST-tagged PpFAAH1 was used as a bait and total protein from WT and OE8 grown with and without exogenous ABA was used as prey. Interestingly, the highest number of unique

peptides and proteins identified with matching spectra were observed in both WT and OE8 protein samples from cultures that were not treated with ABA (Fig. 4.6A). The identified proteins from each sample were then analyzed by removing the proteins in the experimental samples that were also identified with the negative control. Further, proteins were selected as an interacting protein of PpFAAH1 only if the identified proteins were recognized by more than two log fold from the threshold. Subsequently, exclusive and common proteins between samples identified (Fig. 4.6B; Table 4.2). Following these analyses, a total of 34 proteins were identified with WT and 25 with OE8, of which 18 of them were common. Only nine and seven proteins were identified with ABA treated WT and OE8, respectively, of which six of them were common. Additionally, there were five common proteins in WT and three in OE8 that were grown with or without ABA treatment (Fig. 4.6B). The identification of different numbers and types of proteins among these four samples suggests plants under stress conditions possibly alter their interacting partners. To evaluate the possible hypothesis, we cross checked those proteins with our RNA-seq data that was obtained with exogenous treatment of ABA (Chapter 3).

Our RNA-seq data generated for protonema with 1h, 12h and 24h of exogenous ABA treatment were used to compared with the protein data. Of the 46 differentially expressed genes 35 of them correspond to the interacting proteins identified. Also, 13 out of 16 genes that were found in common between WT and OE8, but not with ABA treatment, were down regulated (Table 1). The proteins such as chlorophyll a-b binding protein, ADP ribosylation factor 2-like, and fructose-bisphosphate aldolase 1 are involved in energy producing pathways. Other proteins such as cell division protein FtsZ and heat shock protein 90-2 are growth and stress responsive and down regulated with ABA treatment. Therefore, down regulation but no interaction with PpFAAH1 with exogenous ABA treatment indicates the involvement of FAAH1 independent of ABA or ABA-mediated stress responses. On the other hand, elongation factor Tu and chloroplast-like protein, that were common between WT and OE with ABA treatment, showed upregulation in RNA-seq analysis. Additionally, among the 20 interacting proteins that were exclusive to WT or OE8, 16 of them were down regulated and 4 were not differentially expressed (Table 1). Identification of PpFAAH1 interacting proteins and validation with RNA-seq data confirmed that PpFAAH1 and other partners interact in a different manner depending on the bio-physiological conditions. One possibility could be without stress condition, FAAH

interaction is higher with other proteins but in stress conditions FAAH might be free to catabolize NAEs and the same class of molecules to help overcome the severe situation.

Most of the interacting proteins identified were organelle bound, mostly chloroplastic, and some mitochondrial or ribosomal; few were also cytosolic. This suggests that PpFAAH1, like mammalian FAAH is likely bound to or localized to an organelle. In mammals, FAAH is localized frequently in the intracellular compartment membrane of mitochondria and smooth endoplasmic reticulum (ER) and less often to the somatic plasma membrane[45]. Human FAAH1 was also found in lipid droplets (LDs)[46]. It has been demonstrated that endocannabinoid and its metabolic enzymes are found in the same organelles[45,47]. In mammals, anandamide is localized mostly in organelles such as ER, LDs, mitochondria, and lysosomes[45]. Thus far, there are no reports of anandamide or FAAH subcellular localization in plants. Identification of most of the interacting proteins of FAAH associated with chloroplast or mitochondria, suggests that the general location of anandamide might also be the same. Because of the hydrophobic nature of lipid or anandamide, FAAH is not expected to be a free-floating molecule in the cytoplasm[45]. In plants, chloroplasts are known to be the site of lipid biosynthesis[48], therefore localization of AEA and its catabolic enzyme in chloroplasts would be a reasonable hypothesis and requires additional confirmation.

Table 4.2. List of interacting proteins of PpFAAH1. Proteins identified from WT and OE8 protonema with or without exogenous ABA treatment using LC/MS/MS. Expression of the corresponding gene with ABA treatment at 24h from RNA-seq (Chapter 3) data is included in the last column

Sample	**NCBI ID**	**Protein Name**	**RNA-Seq Expression with ABA**
WT	NP_904195.1	ATP synthase CF1 beta subunit (chloroplast)	No
	NP_904200.1	ribosomal protein S4 protein (chloroplast)	No
	NP_904216.1	ATP synthase CF1 alpha subunit (chloroplast)	No
	XP_024357001.1	glutamine synthetase, chloroplastic-like	No
	XP_024370033.1	thioredoxin H-type-like isoform X2	Up
	XP_024373747.1	phosphoglycerate kinase, cytosolic-like isoform X1	Down

	XP_024375438.1	transketolase, chloroplastic-like	Down
	XP_024379832.1	uncharacterized protein LOC112284343	Down
	XP_024385403.1	uncharacterized protein LOC112287035	Down
	XP_024389237.1	arginine--tRNA ligase, cytoplasmic-like	Down
	XP_024389614.1	auxin transport protein BIG-like	Down
	XP_024400532.1	malate dehydrogenase, chloroplastic-like	Down
	XP_024400995.1	glutathione S-transferase-like	Down
	XP_024402853.1	glyceraldehyde-3-phosphate dehydrogenase, cytosolic-like	Down
OE	XP_024361069.1	probable histidine kinase 6	Down
	XP_024366797.1	translation factor GUF1 homolog, chloroplastic	Down
	XP_024388771.1	oxygen-evolving enhancer protein 2, chloroplastic-like	Up
	XP_024390742.1	elongation factor G, chloroplastic-like	Down
	XP_024391009.1	6-phosphogluconate dehydrogenase, decarboxylating 1-like	Down
	XP_024395990.1	phosphoenolpyruvate carboxylase 2-like	Up
WT & OE	XP_024359341.1	fructose-bisphosphate aldolase 1, chloroplastic-like	Down
	XP_024360203.1	non-symbiotic hemoglobin	Down
	XP_024365396.1	tubulin alpha-1 chain	No
	XP_024367679.1	alanine--tRNA ligase-like	Down
	XP_024367961.1	elongation factor 2-like isoform X1	Up
	XP_024368037.1	eukaryotic initiation factor 4A-10-like	No
	XP_024371797.1	ribulose bisphosphate carboxylase small chain clone 512-like	Down
	XP_024373095.1	chlorophyll a-b binding protein, chloroplastic-like	No
	XP_024374287.1	ADP-ribosylation factor 2-like	Down
	XP_024374405.1	uncharacterized protein LOC112281765	Down
	XP_024380704.1	carbonic anhydrase 2-like	Down
	XP_024381523.1	cell division protein FtsZ homolog 2-1, chloroplastic-like	Down
	XP_024396882.1	heat shock protein 90-2	Down
	XP_024397067.1	linoleate 9S-lipoxygenase-like	Down
	XP_024403903.1	peroxisomal (S)-2-hydroxy-acid oxidase GLO5-like	Down
WT-ABA	XP_024366844.1	40S ribosomal protein S11-like	No
	XP_024397046.1	glutathione S-transferase F9-like	Down

OE-ABA	XP_024369099.1	cysteine desulfurase, mitochondrial-like	No
WT/OE-ABA	XP_024384950.1	elongation factor Tu, chloroplastic-like	Up
OE & WT-ABA	XP_024388849.1	uncharacterized protein LOC112288653	No
WT & WT/OE-ABA	XP_024370152.1	CLP protease regulatory subunit CLPX1, mitochondrial-like	No
	XP_024381159.1	carbamoyl-phosphate synthase large chain, chloroplastic-like	Down
WT, OE & WT/OE-ABA	XP_024357897.1	uncharacterized protein LOC112273402	Down
	XP_024379116.1	succinate-semialdehyde dehydrogenase, mitochondrial-like	No
	XP_024380963.1	MLP-like protein 423	Down

The identified interacting proteins of PpFAAH1 show functional diversity with most them involved in producing energy for the cell. Most notable were ATP synthase CF1 alpha or beta subunits, glutamate synthetase, malate dehydrogenase, glyceraldehyde-3-phosphate dehydrogenase, fructose-bisphosphate aldolase 1, and phosphoenolpyruvate carboxylase 2-like. Some of them are also related to cell morphogenesis, such as the tubulin alpha-1 chain, cell division protein FtsZ homolog 2-1, auxin transport protein BIG-like, transketolase, and oxygen-evolving enhancer protein 2. A few of them are related to transcription and translation processes such as elongation factor Tu, elongation factor G, 40S ribosomal protein S11-like, translation factor GUF1 homolog, ribosomal protein S4 proteins, and CLP regulatory subunit CLPX1. One of the common proteins that was identified with both WT and OE was linoleate 9S lipoxygenase-like, which is a metabolic enzyme capable of oxidizing polyunsaturated fatty acids and NAEs[49]. Animal FAAH was shown to interact with NAPE-PLD to regulate AEA level in neuronal cells [50]. Although functional plant NAPE-PLD is yet to be identified, interaction of PpFAAH1 with a lipoxygenase-like enzyme could also be a means to regulate the levels of these signaling lipids. Among the three proteins that were common for all treatments, one was unannotated, and the other two were succinate semialdehyde dehydrogenase and major latex protein (MLP)-like protein 423.

Identification and characterization of additional components of ECS

In addition to the ligand anandamide and its metabolic enzyme FAAH, we expect the occurrence of GPCRs and the associated G proteins and regulator of G proteins (RGS) as part of the endocannabinoid signaling system. Here we identified and partially characterized some of

these components. In mammals, the well-studied CB1 and CB2 receptors are GPCRs. The canonical GPCR proteins consist of seven transmembrane domains with the N-terminus as extracellular and the C-terminus intracellular[31]. The intracellular C-terminus interacts with the G-α subunit of G protein complex. Thus far, an extra-large Gα (XLG), Gβ, and Gγ were previously characterized in *P. patens* and were shown to be involved in growth and development[41]. However, canonical GPCR, Gα, and RGS were not reported in *P. patens*.

To identify GPCR candidates in *P. patens*, the proteome of *P. patens* was screened for all the proteins with seven transmembrane domains. The candidates were narrowed down systematically by using several transmembrane domain prediction programs. Finally, we used GPCR specific bioinformatic tools to confirm identification of potential candidates of GPCRs in *P. patens*. As a quality control for these bioinformatic tools, the proteome of human and Arabidopsis were analyzed in parallel to identify their known GPCRs. After systematic and rigorous bioinformatic analyses, we selected one GPCR like protein 1 (GLP1, accession number: Pp3c15_15980V3.1) for further characterization. To identify canonical Gα and RGS in *P. patens*, we used homologs of human Gα (NC_000009) and RGS (accession number: NP_002919) to search the Phytozome 12.0 database. The search resulted in identification of putative Gα and RGS (transcript name: Pp3c3_6370V3.2) orthologs in the *P. patens* genome. Further systematic analysis also identified a spliced form of Gα, therefore, we named Gα-1 and Gα-2 (transcript name: Pp3c2_8810V3.2).

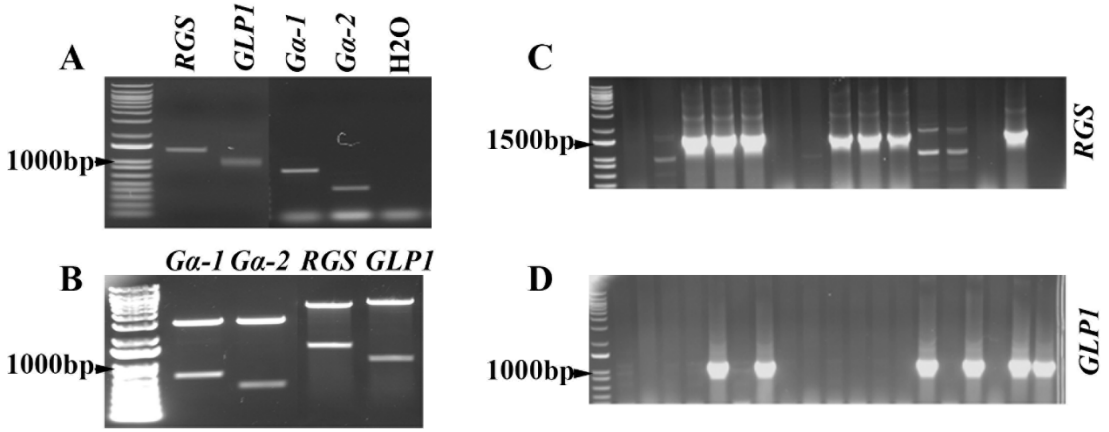


Figure 4.7. Cloning of G protein signaling components. A) Amplification of *RGS*, *GLP1*, *Gα-1* and *Gα-2* from the cDNA of *P. patens* protonema. Confirmation of insertion of the genes in the entry vector using restriction enzymes (B) and colony PCR (C and D)

Cloning of GPCR components

A total of four genes, identified as the putative components of G protein signaling were cloned using Gateway technology. The putative *PpGα1, PpGα2, PpGLP1* and *PpRGS* genes were amplified from extracted RNA of *P. patens* protonema and cloned in entry vector, p19 (Fig. 4.7). The insertion of genes into the entry vector were confirmed by colony PCR, restriction digestion and sequencing (Fig. 4.7 B-D). Confirmed clones were then subcloned into destination vectors, pDEST15 and pDEST17 with either a GST or His tag in the N-terminus of the vectors, respectively. Further expression and purification are required to determine the functional role of these putative genes in endocannabinoid signaling in *P. patens*.

Conclusions

Understanding the ECS in plants would be novel and could open a new avenue to explore the mechanisms for stress tolerance in plants, should the ECS be involved in stress responses. In *P. patens*, the existence of the ECS is evident from the identification of ligand (AEA) and its metabolic enzyme (FAAH) and partial characterization G protein components.

Specifically, we were able to show that more than one active FAAH are present in *P. patens* and these are differentially expressed during development and in response to the type of exogenously applied NAEs. The mRNA expression data also suggest that among the nine FAAH paralogs, some of them are likely specialized to respond in a developmental stage-specific and substrate specific manner. The higher *in vivo* amidohydrolase activity in microsomes of protonema and gametophyte suggests that FAAH are likely associated with organellar membranes. However, the amidohydrolase activity observed in protonema and gametophyte could be a result of more than one active FAAH paralog that responds to NAE 20:4. With PpFAAH1, PpFAAH5, and pPFAAH9 showing the highest mRNA expression in response to anandamide, we hypothesize that these might be the major contributors to the *in vivo* FAAH activity that was observed. The mRNA expression for these three genes was, however, low in the gametophyte stage, which suggest that these genes respond on demand. It would be worthwhile to quantify the mRNA expression of all the nine FAAH paralogs in protonema as these data were not available on the eFP browser.

Over expression of *PpFAAH1* affected developmental transition that was not rescued by exogenous anandamide, although it was tolerated. This phenotype was not observed with OE

lines that showed lower FAAH1 expression levels. These data together suggest that, while there was tolerance of OE8 to higher concentrations of anandamide, the developmental phenotype was perhaps associated with the non-catalytic activity of FAAH. It is possible the PpFAAH1 might play an additional role in *P. patens*, independently or in interaction with other proteins. The interacting proteins of PpFAAH1 identified in response to ABA treatment suggests an alternate mechanism for PpFAAH1 that might be independent of catalytic activity. Correlation between interacting proteins and transcriptome data in response to ABA treatment supported data from both analyses and also confirms the ability of PpFAAH1 to play a role in development and response to stress conditions.

Finally, we were able to identify a number of putative ECS components such as GLP1, RGS and Gα, which emphasizes the existence of ECS in *P. patens*. Further research is essential to identify a complete endocannabinoid signaling pathway and the relevance of its occurrence in early land plants.

REFERENCES

1. Munro, S., Thomas, K. L. & Abu-Shaar, M. Molecular characterization of a peripheral receptor for cannabinoids. *Nature* **365**, 61–65 (1993).

2. Bornheim, L. M., Everhart, E. T., Li, J. & Correia, M. A. Characterization of cannabidiol-mediated cytochrome P450 inactivation. *Biochem. Pharmacol.* **45**, 1323–1331 (1993).

3. Crawley, J. N. *et al. Anandamide, an endogenous ligand of the cannabinoid receptor, induces hypomotility and hypothermia in vivo in rodents*. *Pharmacology Biochemistry and Behavior* **46**, (1993).

4. Ameri, A. The effects of cannabinoids on the brain. *Prog. Neurobiol.* **58**, 315–48 (1999).

5. Matsuda, L. A., Lolait, S. J., Brownstein, M. J., Young, A. C. & Bonner, T. I. Structure of a cannabinoid receptor and functional expression of the cloned cDNA. *Nature* **346**, 561–564 (1990).

6. Howlett, A. C. Cannabinoid receptor signaling. *Handb. Exp. Pharmacol.* 53–79 (2005).

7. Bisogno, T. *et al.* Cloning of the first sn1-DAG lipases points to the spatial and temporal regulation of endocannabinoid signaling in the brain. *J. Cell Biol.* **163**, 463–468 (2003).

8. Dinh, T. P. *et al.* Brain monoglyceride lipase participating in endocannabinoid inactivation. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 10819–24 (2002).

9. Tripathy, S., Venables, B. J., Chapman, K. D., Gage, D. & Chapman, K. N - Acylethanolamines in Signal Transduction of Elicitor Perception. Attenuation of Alkalinization Response and Activation of Defense Gene Expression. *Plant Physiol.* **121**, 1299–1308 (1999).

10. Okamoto, Y., Morishita, J., Tsuboi, K., Tonai, T. & Ueda, N. Molecular characterization of a phospholipase D generating anandamide and its congeners. *J. Biol. Chem.* **279**, 5298–305 (2004).

11. Cravatt, B. F. *et al.* Molecular characterization of an enzyme that degrades neuromodulatory fatty-acid amides. *Nature* **384**, 83–7 (1996).

12. Mackie, K. Cannabinoid Receptors: Where They are and What They do. *J. Neuroendocrinol.* **20**, 10–14 (2008).

13. Gachet, M. S., Schubert, A., Calarco, S., Boccard, J. & Gertsch, J. Targeted metabolomics shows plasticity in the evolution of signaling lipids and uncovers old and new endocannabinoids in the plant kingdom. *Sci. Rep.* (2017). doi:10.1038/srep41177

14. Ueda, N., Tsuboi, K. & Uyama, T. Enzymological studies on the biosynthesis of N-acylethanolamines. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* **1801**, 1274–1285 (2010).

15. Jin, X.-H. *et al.* Discovery and characterization of a Ca2+-independent phosphatidylethanolamine N-acyltransferase generating the anandamide precursor and its congeners. *J. Biol. Chem.* **282**, 3614–23 (2007).

16. Haq, I. & Kilaru, A. An endocannabinoid catabolic enzyme FAAH and its paralogs in an early land plant reveal evolutionary and functional relationship with eukaryotic orthologs. *Sci. Rep.* **10**, 3115 (2020).

17. M-H Wei, O Toure, G M Glenn, M Pithukpakorn, L Neckers, C Stolle, P Choyke, *et al.* Novel mutations in FH and expansion of the spectrum of phenotypes expressed in families with hereditary leiomyomatosis and renal cell cancer. *J. Med. Genet.* (2006). doi:10.1136/jmg.2005.033506

18. Lazary, J., Eszlari, N., Juhasz, G. & Bagdy, G. Genetically reduced FAAH activity may be a risk for the development of anxiety and depression in persons with repetitive childhood trauma. *Eur. Neuropsychopharmacol.* **26**, 1020–1028 (2016).

19. Durand, E. *et al.* Evaluating the Association of FAAH Common Gene Variation with Childhood, Adult Severe Obesity and Type 2 Diabetes in the French Population. *Obes. Facts* **1**, 305–309 (2008).

20. Spagnolo, P. A. *et al. FAAH* Gene Variation Moderates Stress Response and Symptom Severity in Patients with Posttraumatic Stress Disorder and Comorbid Alcohol Dependence. *Alcohol. Clin. Exp. Res.* **40**, 2426–2434 (2016).

21. Blüher, M. *et al.* Dysregulation of the peripheral and adipose tissue endocannabinoid system in human abdominal obesity. *Diabetes* **55**, 3053–60 (2006).

22. Bühler, K.-M. *et al.* Risky alcohol consumption in young people is associated with the fatty acid amide hydrolase gene polymorphism C385A and affective rating of drug pictures. *Mol. Genet. Genomics* **289**, 279–289 (2014).

23. Kang, L. *et al.* Overexpression of a fatty acid amide hydrolase compromises innate immunity in Arabidopsis. *Plant J.* (2008). doi:10.1111/j.1365-313X.2008.03603.x

24. Teaster, N. D. *et al.* N-acylethanolamine metabolism interacts with abscisic acid signaling in Arabidopsis thaliana seedlings. *Plant Cell* **19**, 2454–2469 (2007).

25. Keereetaweep, J., Blancaflor, E. B., Hornung, E., Feussner, I. & Chapman, K. D. Ethanolamide oxylipins of linolenic acid can negatively regulate Arabidopsis seedling development. *Plant Cell* **25**, 3824–40 (2013).

26. Wang, Y.-S. *et al.* Manipulation of Arabidopsis fatty acid amide hydrolase expression modifies plant growth and sensitivity to N-acylethanolamines. *Proc. Natl. Acad. Sci.* (2006). doi:10.1073/pnas.0603571103

27. Teaster, N. D. *et al.* Overexpression of Fatty Acid Amide Hydrolase Induces Early Flowering in Arabidopsis thaliana. *Front. Plant Sci.* (2012). doi:10.3389/fpls.2012.00032

28. Katona, I. & Freund, T. F. Multiple Functions of Endocannabinoid Signaling in the Brain. *Annu. Rev. Neurosci.* (2012). doi:10.1146/annurev-neuro-062111-150420

29. Pandey, S. & Assmann, S. M. The Arabidopsis Putative G Protein-Coupled Receptor GCR1 Interacts with the G Protein Subunit GPA1 and Regulates Abscisic Acid Signaling. *Plant Cell* **16**, 1616–1632 (2004).

30. Pandey, S., Nelson, D. C. & Assmann, S. M. Two Novel GPCR-Type G Proteins Are Abscisic Acid Receptors in Arabidopsis. *Cell* **136**, 136–148 (2009).

31. Urano, D. & Jones, A. M. Heterotrimeric G Protein–Coupled Signaling in Plants. *Annu. Rev. Plant Biol.* **65**, 365–384 (2014).

32. Humphrey, T. V. & Botella, J. R. Re-evaluation of the cytokinin receptor role of the Arabidopsis gene GCR1. *J. Plant Physiol.* **158**, 645–653 (2001).

33. Colucci, G., Apone, F., Alyeshmerni, N., Chalmers, D. & Chrispeels, M. J. GCR1, the putative Arabidopsis G protein-coupled receptor gene is cell cycle-regulated, and its overexpression abolishes seed dormancy and shortens time to flowering. *Proc. Natl. Acad. Sci.* **99**, 4736–4741 (2002).

34. Steber, C. M. *et al.* A Role for Brassinosteroids in Germination in Arabidopsis. *PLANT Physiol.* **125**, 763–769 (2001).

35. Ma, H., Yanofsky, M. F. & Meyerowitz, E. M. Molecular cloning and characterization of GPA1, a G protein alpha subunit gene from Arabidopsis thaliana. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 3821–5 (1990).

36. Ullah, H. *et al.* Modulation of Cell Proliferation by Heterotrimeric G Protein in Arabidopsis. *Science (80-.).* **292**, 2066–2069 (2001).

37. Wang, X.-Q., Ullah, H., Jones, A. M. & Assmann, S. M. G Protein Regulation of Ion Channels and Abscisic Acid Signaling in Arabidopsis Guard Cells. *Science (80-.).* **292**, 2070–2072 (2001).

38. Ullah, H. *et al.* The beta-subunit of the Arabidopsis G protein negatively regulates auxin-induced cell division and affects multiple developmental processes. *Plant Cell* **15**, 393–409 (2003).

39. Lapik, Y. R. & Kaufman, L. S. The Arabidopsis cupin domain protein AtPirin1 interacts with the G protein alpha-subunit GPA1 and regulates seed germination and early seedling development. *Plant Cell* **15**, 1578–90 (2003).

40. Kaydamov, C., Tewes, A., Adler, K. & Manteuffel, R. Molecular characterization of cDNAs encoding G protein α and β subunits and study of their temporal and spatial expression patterns in Nicotiana plumbaginifolia Viv. *Biochim. Biophys. Acta - Gene Struct. Expr.* **1491**, 143–160 (2000).

41. Hackenberg, D., Perroud, P.-F., Quatrano, R. & Pandey, S. Sporophyte Formation and Life Cycle Completion in Moss Requires Heterotrimeric G-Proteins. *Plant Physiol.* **172**, 1154–1166 (2016).

42. Shrestha, R., Noordermeer, M. A., van der Stelt, M., Veldink, G. A. & Chapman, K. D. N-acylethanolamines are metabolized by lipoxygenase and amidohydrolase in competing pathways during cottonseed imbibition. *Plant Physiol.* **130**, 391–401 (2002).

43. Liu, Y.-C. & Vidali, L. Efficient Polyethylene Glycol (PEG) Mediated Transformation of the Moss Physcomitrella patens *J. Vis. Exp.* (2011). doi:10.3791/2560

44. Kim, S.-C. *et al.* Mutations in *Arabidopsis* Fatty Acid Amide Hydrolase Reveal That Catalytic Activity Influences Growth but Not Sensitivity to Abscisic Acid or Pathogens. *J. Biol. Chem.* **284**, 34065–34074 (2009).

45. Maccarrone, M. Metabolism of the endocannabinoid anandamide: Open questions after 25 years. *Frontiers in Molecular Neuroscience* **10**, (2017).

46. Oddi, S. *et al.* Evidence for the intracellular accumulation of anandamide in adiposomes. *Cell. Mol. Life Sci.* **65**, 840–850 (2008).

47. Gulyas, A. I. *et al.* Segregation of two endocannabinoid-hydrolyzing enzymes into pre- and postsynaptic compartments in the rat hippocampus, cerebellum and amygdala. *Eur. J. Neurosci.* **20**, 441–458 (2004).

48. Finkelstein, R. Abscisic Acid Synthesis and Response. *Arab. B.* **11**, e0166 (2013).

49. Kilaru, A. *et al.* Lipoxygenase-mediated Oxidation of Polyunsaturated *N* -Acylethanolamines in *Arabidopsis*. *J. Biol. Chem.* **286**, 15205–15214 (2011).

50. Si, P. *et al.* Association of polymorphisms of NAPE-PLD and FAAH genes with schizophrenia in Chinese Han population. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* **35**, 215–218 (2018).

CHAPTER 5. CONCLUSION AND FUTURE DIRECTIONS

Conclusions

This study was conducted to address a fundamental question "what is the role of unique endocannabinoid, anandamide that is not synthesized endogenously in higher plants but synthesized in bryophytes and lower organisms?". To answer this question, we aimed for identification and characterization molecules associated with endocannabinoid metabolism, and its hypothetical signaling pathway at a molecular level. Thus, we identified nine orthologs of anandamide catabolic enzyme, PpFAAH1-9 in *P. patens* and PpFAAH1was characterized biochemically. Using extensive bioinformatic analyses, we predicted an evolutionary relationship between the plant and animal FAAH.

Transcriptome analysis revealed that AEA have short-term effect, which could be receptor mediated. The GO terms and categorized gene analysis showed that signal transduction activated through AEA can affect pre and post transcriptional or translational regulation. Identified novel lncRNA specific to AEA, AA or ABA, which provides a new direction for exploring the underlying molecular signaling pathways.

In vivo amidohydrolase activity showed that the function of PpFAAH is tissue specific. Overexpression of PpFAAH1 provided evidence for requirement of NAEs and/or regulated FAAH expression in developmental transition of mosses. Interacting proteins of PpFAAH1, identified with or without ABA treatment showed significant differences. Additionally, a correlation of these interacting proteins with the transcriptome study validated the hypothesis that PpFAAH1 dissociates from its interacting partners under stress conditions, as reflected by ABA treatment. Identification of interacting proteins in subcellular compartments suggested that anandamide and its metabolic enzymes are the localized in membrane-bound organelles.

Our studies conclude with the evidence that anandamide is an essential molecule in growth and development in *P. patens*. This study provides an early evidence for endocannabinoid signaling and a unique role for anandamide that is distinct from that of ABA in an early land plant.

Future Directions

In the course of identifying anandamide catabolic enzyme in *P. patens*, we identified nine and characterized one of them. Characterization of the remaining FAAH paralogs is pertinent to gain detailed understanding of NAE catabolism in *P. patens*. Since the predicted structure of identified FAAH in moss suggested a close relation with both the plant and animal FAAH, characterizing the remaining PpFAAH will provide a broader understanding of their functional regulatory aspects. Also, tissue-specific expression of FAAH paralogs and their differential expression to NAE treatment suggest that these various FAAH might play a specific or redundant role in mosses. As such, additional studies are necessary to differentiate the roles of individual paralogs in growth, development and stress responses.

Our study also revealed that PpFAAH might play a regulatory role in addition to its catalytic function. The overexpression of *PpFAAH1* showed an inhibition of developmental transition that could not be rescued with exogenous anandamide. Further characterization of *PpFAAH1* knock outs and overexpressors by analyzing their lipidome and transcriptome will likely reveal the specific role of NAEs as well FAAH in growth and development of moss. Generation of multiple knockout of FAAH genes is also valuable to identify the role of redundancy.

We identified interacting proteins of PpFAAH1 using mass spectrometry, which are predicted to be localized in different cellular compartments and involved in multiple functions. Further confirmation of interacting proteins by yeast two-hybrid studies will validate the current data. Also, FAAH localization studies will provide better understanding of their cellular role and their ability to physically interact with other proteins and regulation FAAH/NAE-mediated pathways.

Transcriptome analysis with temporal exogenous treatments of AEA, AA and ABA identified number of key molecules that were uniquely responded to each treatment, such as receptors, kinases, secondary metabolite producing genes and gene regulatory genes. Characterizing these molecules will provide a better understanding of anandamide signaling pathway. Transcriptome data also identified key lncRNA as regulatory molecules; elucidating their regulatory function will offer a great deal of insight into gene regulation in plants.

We also have initiated heterologous expression of some key components of ECS. Further characterization of putative GPCR and G proteins and their interaction with anandamide is necessary for unequivocal conclusion of the existence of a novel endocannabinoid signaling pathway in plants.

REFERENCES

Ameri, Angela, Alwina Wilhelm, and Thomas Simmet. 1999. “Effects of the Endogeneous Cannabinoid, Anandamide, on Neuronal Activity in Rat Hippocampal Slices.” *British Journal of Pharmacology* 126(8):1831–39.

Anon. 2019. “One Thousand Plant Transcriptomes and the Phylogenomics of Green Plants.” *Nature* 574(7780):679–85.

Arif, M. Asif, Manuel Hiss, Marta Tomek, Hauke Busch, Rabea Meyberg, Stefanie Tintelnot, Ralf Reski, Stefan A. Rensing, and Wolfgang Frank. 2019. “ABA-Induced Vegetative Diaspore Formation in Physcomitrella Patens.” *Frontiers in Plant Science* 10:315.

Bachur, N. R., K. Masek, K. L. Melmon, and S. Udenfrind. 1965. “Fatty Acid Amides of Ethanolamine in Mammalian Tissues.” *The Journal of Biological Chemistry* 240:1019–24.

Beike, Anna K., Daniel Lang, Andreas D. Zimmer, Florian Wüst, Danika Trautmann, Gertrud Wiedemann, Peter Beyer, Eva L. Decker, and Ralf Reski. 2015. “Insights from the Cold Transcriptome of *Physcomitrella Patens* : Global Specialization Pattern of Conserved Transcriptional Regulators and Identification of Orphan Genes Involved in Cold Acclimation.” *New Phytologist* 205(2):869–81.

Berger, A., G. Crozier, T. Bisogno, P. Cavaliere, S. Innis, and V. Di Marzo. 2001. “Anandamide and Diet: Inclusion of Dietary Arachidonate and Docosahexaenoate Leads to Increased Brain Levels of the Corresponding N-Acylethanolamines in Piglets.” *Proceedings of the National Academy of Sciences* 98(11):6402–6.

Blancaflor, Elison B., Guichuan Hou, and Kent D. Chapman. 2003. “Elevated Levels of N-Lauroylethanolamine, an Endogenous Constituent of Desiccated Seeds, Disrupt Normal Root Development in Arabidopsis Thaliana Seedlings.” *Planta* 217(2):206–17.

Blancaflor, Elison B., Aruna Kilaru, Jantana Keereetaweep, Bibi Rafeiza Khan, Lionel Faure, and Kent D. Chapman. 2014. “N-Acylethanolamines: Lipid Metabolites with Functions in Plant Growth and Development.” *Plant Journal*.

Boger, Dale L., Haruhiko Sato, Aaron E. Lerner, Michael P. Hedrick, Robert A. Fecik, Hiroshi

Miyauchi, Gordon D. Wilkie, Bryce J. Austin, Matthew P. Patricelli, Benjamin F. Cravatt, and Julius Rebek. 2000. “Exceptionally Potent Inhibitors of Fatty Acid Amide Hydrolase: The Enzyme Responsible for Degradation of Endogenous Oleamide and Anandamide.”

Bornheim, Lester M., E. Thoma. Everhart, Jianmin Li, and M. Almir. Correia. 1993. “Characterization of Cannabidiol-Mediated Cytochrome P450 Inactivation.” *Biochemical Pharmacology* 45(6):1323–31.

Bracey, Michael H., Michael A. Hanson, Kim R. Masuda, Raymond C. Stevens, and Benjamin F. Cravatt. 2002. “Structural Adaptations in a Membrane Enzyme That Terminates Endocannabinoid Signaling.” *Science*.

Calignano, A., G. La Rana, A. Giuffrida, and D. Piomelli. 1998. “Control of Pain Initiation by Endogenous Cannabinoids.” *Nature* 394(6690):277–81.

Chapman, Kent D. 1998. “Phospholipase Activity during Plant Growth and Development and in Response to Environmental Stress.” *Trends in Plant Science* 3(11):419–26.

Chapman, Kent D., Barney Venables, Robert Markovic, Raymond W. Blair Jr RW, Bettinger, and Chris Bettinger. 1999. “N-Acylethanolamines in Seeds. Quantification Of Molecular Species and Their Degradation upon Imbibition.” *Plant Physiology* 120(4):1157–64.

Cotter, Matthew Q., Neal D. Teaster, Elison B. Blancaflor, and Kent D. Chapman. 2011. “N-Acylethanolamine (NAE) Inhibits Growthin Arabidopsis Thaliana Seedlings via ABI3-Dependent and -Independent Pathways.” *Plant Signaling and Behavior* 6(5):671–79.

Cove, David. 2005. “The Moss Physcomitrella Patens.”

Cove, David, Magdalena Bezanilla, Phillip Harries, and Ralph Quatrano. 2006. “Mosses as Model Systems for the Study of Metabolism and Development.”

Cravatt, B. F., K. Demarest, M. P. Patricelli, M. H. Bracey, D. K. Giang, B. R. Martin, and A. H. Lichtman. 2001. “Supersensitivity to Anandamide and Enhanced Endogenous Cannabinoid Signaling in Mice Lacking Fatty Acid Amide Hydrolase.” *Proceedings of the National Academy of Sciences of the United States of America* 98(16):9371–76.

Cravatt, B. F., O. Prospero-Garcia, G. Siuzdak, N. B. Gilula, S. J. Henriksen, D. L. Boger, and R. A. Lerner. 1995. “Chemical Characterization of a Family of Brain Lipids That Induce Sleep.” *Science (New York, N.Y.)* 268(5216):1506–9.

Cravatt, Benjamin F., Alan Saghatelian, Edward G. Hawkins, Angela B. Clement, Michael H. Bracey, and Aron H. Lichtman. 2004. “Functional Disassociation of the Central and Peripheral Fatty Acid Amide Signaling Systems.” *Proceedings of the National Academy of Sciences of the United States of America* 101(29):10821–26.

Crawley, Jacqueline N., Rebecca L. Corwin, John K. Robinson, Christian C. Felder, William A. Devane, and Julius Axelrod. 1993. *Anandamide, an Endogenous Ligand of the Cannabinoid Receptor, Induces Hypomotility and Hypothermia in Vivo in Rodents*. Vol. 46.

Cuming, Andrew C., Sung Hyun Cho, Yasuko Kamisugi, Helen Graham, and Ralph S. Quatrano. 2007. “Microarray Analysis of Transcriptional Responses to Abscisic Acid and Osmotic, Salt, and Drought Stress in the Moss, Physcomitrella Patens.” *New Phytologist* 176(2):275–87.

Devane, W. A., L. Hanus, A. Breuer, R. G. Pertwee, L. A. Stevenson, G. Griffin, D. Gibson, A. Mandelbaum, A. Etinger, and R. Mechoulam. 1992. “Isolation and Structure of a Brain Constituent That Binds to the Cannabinoid Receptor.” *Science (New York, N.Y.)* 258(5090):1946–49.

Eveland, Andrea L., Namiko Satoh-Nagasawa, Alexander Goldshmidt, Sandra Meyer, Mary Beatty, Hajime Sakai, Doreen Ware, and David Jackson. 2010. “Digital Gene Expression Signatures for Maize Development.” *Plant Physiology* 154(3):1024–39.

Faure, Lionel, Denis Coulon, Jeanny Laroche-Traineau, Marina Le Guedard, Jean-Marie Schmitter, Eric Testet, René Lessire, and Jean-Jacques Bessoule. 2009. “Discovery and Characterization of an *Arabidopsis Thaliana N* -Acylphosphatidylethanolamine Synthase.” *Journal of Biological Chemistry* 284(28):18734–41.

Fesenko, Igor, Regina Khazigaleeva, Ilya Kirov, Andrey Kniazev, Oksana Glushenko, Konstantin Babalyan, Georgij Arapidi, Tatyana Shashkova, Ivan Butenko, Victor Zgoda,

Ksenia Anufrieva, Anna Seredina, Anna Filippova, and Vadim Govorun. 2017. "Alternative Splicing Shapes Transcriptome but Not Proteome Diversity in Physcomitrella Patens." *Scientific Reports* 7(1):2698.

Frank, Wolfgang, Diah Ratnadewi, and Ralf Reski. 2005. "Physcomitrella Patens Is Highly Tolerant against Drought, Salt and Osmotic Stress." *Planta* 220(3):384–94.

Fu, Jin, Silvana Gaetani, Fariba Oveisi, Jesse Lo Verme, Antonia Serrano, Fernando Rodríguez De Fonseca, Anja Rosengarth, Hartmut Luecke, Barbara Di Giacomo, Giorgio Tarzia, and Daniele Piomelli. 2003. "Oleylethanolamide Regulates Feeding and Body Weight through Activation of the Nuclear Receptor PPAR-Alpha." *Nature* 425(6953):90–93.

Gachet, Mariá Salomé, Alexandra Schubert, Serafina Calarco, Julien Boccard, and Jürg Gertsch. 2017. "Targeted Metabolomics Shows Plasticity in the Evolution of Signaling Lipids and Uncovers Old and New Endocannabinoids in the Plant Kingdom." *Scientific Reports*.

Gao, Yue, Xin Li, Hui Zhi, Yunpeng Zhang, Peng Wang, Yanxia Wang, Shipeng Shang, Ying Fang, Weitao Shen, Shangwei Ning, Steven Xi Chen, and Xia Li. 2019. "Comprehensive Characterization of Somatic Mutations Impacting LncRNA Expression for Pan-Cancer."

Giang, Dan K., Benjamin F. Cravatt, and Richard Lerner. 1997. "Molecular Characterization of Human and Mouse Fatty Acid Amide Hydrolases." *Biochemistry* 94:2238–42.

Hampson, A. J., M. Grimaldi, J. Axelrod, and D. Wink. 1998. "Cannabidiol and 9-Tetrahydrocannabinol Are Neuroprotective Antioxidants." *Proceedings of the National Academy of Sciences* 95(14):8268–73.

Haq, Imdadul and Aruna Kilaru. 2020. "An Endocannabinoid Catabolic Enzyme FAAH and Its Paralogs in an Early Land Plant Reveal Evolutionary and Functional Relationship with Eukaryotic Orthologs." *Scientific Reports* 10(1):3115.

Heckman, D. S., D. M. Geiser, B. R. Eidell, R. L. Stauffer, N. L. Kardos, and S. B. Hedges. 2001. "Molecular Evidence for the Early Colonization of Land by Fungi and Plants." *Science* 293(5532):1129–33.

Horrocks, Lloyd A. and Young K. Yeo. 1999. “HEALTH BENEFITS OF DOCOSAHEXAENOIC ACID (DHA).” *Pharmacological Research* 40(3):211–25.

Howlett, A. C. 2005. “Cannabinoid Receptor Signaling.” *Handbook of Experimental Pharmacology* (168):53–79.

Jin, Xing-Hua, Yasuo Okamoto, Jun Morishita, Kazuhito Tsuboi, Takeharu Tonai, and Natsuo Ueda. 2007. “Discovery and Characterization of a Ca2+-Independent Phosphatidylethanolamine N-Acyltransferase Generating the Anandamide Precursor and Its Congeners.” *The Journal of Biological Chemistry* 282(6):3614–23.

Jones, S. and J. M. Thornton. 1996. “Principles of Protein-Protein Interactions.” *Proceedings of the National Academy of Sciences* 93(1):13–20.

Kaewsuwan, Songsri, Edgar B. Cahoon, Pierre-Francois Perroud, Chanpen Wiwat, Nathinee Panvisavas, Ralph S. Quatrano, David J. Cove, and Nuntavan Bunyapraphatsara. 2006. “Identification and Functional Characterization of the Moss Physcomitrella Patens Δ 5 -Desaturase Gene Involved in Arachidonic and Eicosapentaenoic Acids Biosynthesis.” *Journal of Biologicla Chemistyr*.

Kammerer, W. and D. J. Cove. 1996. “Genetic Analysis of the Effects of Re-Transformation of Transgenic Lines of the MossPhyscomitrella Patens.” *MGG Molecular & General Genetics* 250(3):380–82.

Kang, Li, Yuh Shuh Wang, Srinivasa Rao Uppalapati, Keri Wang, Yuhong Tang, Vatsala Vadapalli, Barney J. Venables, Kent D. Chapman, Elison B. Blancaflor, and Kirankumar S. Mysore. 2008. “Overexpression of a Fatty Acid Amide Hydrolase Compromises Innate Immunity in Arabidopsis.” *Plant Journal*.

Kathuria, Satish, Silvana Gaetani, Darren Fegley, Fernando Valiño, Andrea Duranti, Andrea Tontini, Marco Mor, Giorgio Tarzia, Giovanna La Rana, Antonio Calignano, Arcangela Giustino, Maria Tattoli, Maura Palmery, Vincenzo Cuomo, and Daniele Piomelli. 2003. “Modulation of Anxiety through Blockade of Anandamide Hydrolysis.” *Nature Medicine* 9(1):76–81.

Keereetaweep, Jantana, Elison B. Blancaflor, Ellen Hornung, Ivo Feussner, and Kent D.

Chapman. 2013. “Ethanolamide Oxylipins of Linolenic Acid Can Negatively Regulate Arabidopsis Seedling Development.” *The Plant Cell* 25(10):3824–40.

Kilaru, Aruna, Elison B. Blancaflor, Barney J. Venables, Swati Tripathy, Kirankumar S. Mysore, and Kent D. Chapman. 2007. “The N-Acylethanolamine-Mediated Regulatory Pathway in Plants.” *Chemistry & Biodiversity* 4(8):1933–55.

Kilaru, Aruna and Kent D. Chapman. 2012. “N-Acylated Phospholipid Metabolism and Seedling Growth.” *Plant Signaling & Behavior* 7(9):1200–1202.

Kilaru, Aruna, Cornelia Herrfurth, Jantana Keereetaweep, Ellen Hornung, Barney J. Venables, Ivo Feussner, and Kent D. Chapman. 2011. “Lipoxygenase-Mediated Oxidation of Polyunsaturated *N* -Acylethanolamines in *Arabidopsis*.” *Journal of Biological Chemistry* 286(17):15205–14.

Kim, Sang-Chul. 2010. “Functional Characterization of Plant Fatty Acid Amide Hydrolases.”

Kim, Sang-Chul, Li Kang, Satish Nagaraj, Elison B. Blancaflor, Kirankumar S. Mysore, and Kent D. Chapman. 2009. “Mutations in *Arabidopsis* Fatty Acid Amide Hydrolase Reveal That Catalytic Activity Influences Growth but Not Sensitivity to Abscisic Acid or Pathogens.” *Journal of Biological Chemistry* 284(49):34065–74.

Kroemer, K., R. Reski, and W. Frank. 2004. “Abiotic Stress Response in the Moss Physcomitrella Patens: Evidence for an Evolutionary Alteration in Signaling Pathways in Land Plants.” *Plant Cell Reports*.

Labahn, Jörg, Sebastian Neumann, Georg Büldt, Maria-Regina Kula, and Joachim Granzin. 2002. “An Alternative Mechanism for Amidase Signature Enzymes.” *Journal of Molecular Biology* 322(5):1053–64.

Lambert, Didier M., Severine Vandevoorde, Kent-Olov Jonsson, and Christopher J. Fowler. 2002. “The Palmitoylethanolamide Family: A New Class of Anti-Inflammatory Agents?” *Current Medicinal Chemistry* 9(6):663–74.

Leonti, Marco, Laura Casu, Stefan Raduner, Filippo Cottiglia, Costantino Floris, Karl-Heinz Altmann, and Jürg Gertsch. 2010. “Falcarinol Is a Covalent Cannabinoid CB1 Receptor Antagonist and Induces Pro-Allergic Effects in Skin.” *Biochemical Pharmacology*

79(12):1815–26.

Li, Yan, Keyang Xu, Kechen Xu, Sixiang Chen, Yifang Cao, and Huakui Zhan. 2019. “Roles of Identified Long Noncoding RNA in Diabetic Nephropathy.” *Journal of Diabetes Research* 2019.

Lichtman, Aron H., E. Gregory Hawkins, Graeme Griffin, and Benjamin F. Cravatt. 2002. “Pharmacological Activity of Fatty Acid Amides Is Regulated, but Not Mediated, by Fatty Acid Amide Hydrolase in Vivo.” *The Journal of Pharmacology and Experimental Therapeutics* 302(1):73–79.

Lister, Ryan, Ronan C. O’Malley, Julian Tonti-Filippini, Brian D. Gregory, Charles C. Berry, A. Harvey Millar, and Joseph R. Ecker. 2008. “Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis.” *Cell* 133(3):523–36.

Liu, Jun, Choonkyun Jung, Jun Xu, Huan Wang, Shulin Deng, Lucia Bernad, Catalina Arenas-Huertero, and Nam-Hai Chua. 2012. “Genome-Wide Analysis Uncovers Regulation of Long Intergenic Noncoding RNAs in *Arabidopsis*.” *The Plant Cell* 24(11):4333–45.

Long, Jonathan Z., Weiwei Li, Lamont Booker, James J. Burston, Steven G. Kinsey, Joel E. Schlosburg, Franciso J. Pavón, Antonia M. Serrano, Dana E. Selley, Loren H. Parsons, Aron H. Lichtman, and Benjamin F. Cravatt. 2009. “Selective Blockade of 2-Arachidonoylglycerol Hydrolysis Produces Cannabinoid Behavioral Effects.” *Nature Chemical Biology* 5(1):37–44.

Lucanic, Mark, Jason M. Held, Maithili C. Vantipalli, Ida M. Klang, Jill B. Graham, Bradford W. Gibson, Gordon J. Lithgow, and Matthew S. Gill. 2011. “N-Acylethanolamine Signalling Mediates the Effect of Diet on Lifespan in Caenorhabditis Elegans.” *Nature* 473(7346):226–29.

Di Marzo, Vincenzo, Nunzio Sepe, Luciano De Petrocellis, Alvin Berger, Gayle Crozier, Ester Fride, and Raphael Mechoulam. 1998. “Trick or Treat from Food Endocannabinoids?” *Nature* 396(6712):636–636.

Maurelli, S., T. Bisogno, L. De Petrocellis, A. Di Luccia, G. Marino, and V. Di Marzo. 1995. “Two Novel Classes of Neuroactive Fatty Acid Amides Are Substrates for Mouse

Neuroblastoma ‘Anandamide Amidohydrolase’.” *FEBS Letters* 377(1):82–86.

McKinney, Michele K. and Benjamin F. Cravatt. 2005. “Structure and Function of Fatty Acid Amidy Hydrolase.” *Annual Review of Biochemistry* 74:411–32.

Mitobe, Yuichi, Ken ichi Takayama, Kuniko Horie-Inoue, and Satoshi Inoue. 2018. “Prostate Cancer-Associated LncRNAs.” *Cancer Letters* 418:159–66.

Motes, Christy M., Priit Pechter, Cheol Min Yoo, Yuh-Shuh Wang, Kent D. Chapman, and Elison B. Blancaflor. 2005. “Differential Effects of Two Phospholipase D Inhibitors, 1-Butanol and N-Acylethanolamine, on in Vivo Cytoskeletal Organization and Arabidopsis Seedling Growth.” *Protoplasma* 226(3–4):109–23.

Natarajan, V., P. C. Schmid, P. V Reddy, and H. H. Schmid. 1984. “Catabolism of N-Acylethanolamine Phospholipids by Dog Brain Preparations.” *Journal of Neurochemistry* 42(6):1613–19.

Neu, Axel, Csaba Földy, and Ivan Soltesz. 2007. “Postsynaptic Origin of CB1-Dependent Tonic Inhibition of GABA Release at Cholecystokinin-Positive Basket Cell to Pyramidal Cell Synapses in the CA1 Region of the Rat Hippocampus.” *The Journal of Physiology* 578(1):233–47.

O’Donoghue, Martin-Timothy, Caspar Chater, Simon Wallace, Julie E. Gray, David J. Beerling, and Andrew J. Fleming. 2013. “Genome-Wide Transcriptomic Analysis of the Sporophyte of the Moss Physcomitrella Patens.” *Journal of Experimental Botany* 64(12):3567–81.

Okamoto, Yasuo, Jun Morishita, Kazuhito Tsuboi, Takeharu Tonai, and Natsuo Ueda. 2004. “Molecular Characterization of a Phospholipase D Generating Anandamide and Its Congeners.” *The Journal of Biological Chemistry* 279(7):5298–5305.

Patricelli, M. P. and B. F. Cravatt. 1999. “Fatty Acid Amide Hydrolase Competitively Degrades Bioactive Amides and Esters through a Nonconventional Catalytic Mechanism.” *Biochemistry* 38(43):14125–30.

Patricelli, M. P., H. A. Lashuel, D. K. Giang, J. W. Kelly, and B. F. Cravatt. 1998. “Comparative Characterization of a Wild Type and Transmembrane Domain-Deleted

Fatty Acid Amide Hydrolase: Identification of the Transmembrane Domain as a Site for Oligomerization." *Biochemistry* 37(43):15177–87.

Perroud, Pierre François, Fabian B. Haas, Manuel Hiss, Kristian K. Ullrich, Alessandro Alboresi, Mojgan Amirebrahimi, Kerrie Barry, Roberto Bassi, Sandrine Bonhomme, Haodong Chen, Juliet C. Coates, Tomomichi Fujita, Anouchka Guyon-Debast, Daniel Lang, Junyan Lin, Anna Lipzen, Fabien Nogué, Melvin J. Oliver, Inés Ponce de León, Ralph S. Quatrano, Catherine Rameau, Bernd Reiss, Ralf Reski, Mariana Ricca, Younousse Saidi, Ning Sun, Péter Szövényi, Avinash Sreedasyam, Jane Grimwood, Gary Stacey, Jeremy Schmutz, and Stefan A. Rensing. 2018. "The Physcomitrella Patens Gene Atlas Project: Large-Scale RNA-Seq Based Expression Data." *Plant Journal* 95(1):168–82.

Pertwee, Roger G. 2009. "Cannabinoid Pharmacology: The First 66 Years." *British Journal of Pharmacology* 147(S1):S163–71.

Rahman, Iffat Ara Sonia, Kazuhito Tsuboi, Toru Uyama, and Natsuo Ueda. 2014. "New Players in the Fatty Acyl Ethanolamide Metabolism." *Pharmacological Research* 86:1–10.

Ruibal, Cecilia, Alexandra Castro, Valentina Carballo, László Szabados, and Sabina Vidal. 2013. "Recovery from Heat, Salt and Osmotic Stress in Physcomitrella Patens Requires a Functional Small Heat Shock Protein PpHsp16.4." *BMC Plant Biology* 13(1):174.

Saghatelian, Alan, Michele K. McKinney, Michael Bandell, Ardem Patapoutian, and Benjamin F. Cravatt. 2006. "A FAAH-Regulated Class of N-Acyl Taurines That Activates TRP Ion Channels." *Biochemistry* 45(30):9007–15.

Sante, Richard. 2014. "Occurrence and Implications of the N-Acylethanolamine Metabolic Pathway in Physcomitrella Patens." *Electronic Theses and Dissertations*.

Schmid, H. H. 2000. "Pathways and Mechanisms of N-Acylethanolamine Biosynthesis: Can Anandamide Be Generated Selectively?" *Chemistry and Physics of Lipids* 108(1–2):71–87.

Shrestha, Rhidaya, Richard A. Dixon, and Kent D. Chapman. 2003. "Molecular Identification

of a Functional Homologue of the Mammalian Fatty Acid Amide Hydrolase in Arabidopsis Thaliana." *Journal of Biological Chemistry*.

Shrestha, Rhidaya, Minke A. Noordermeer, Marcelis van der Stelt, Gerrit A. Veldink, and Kent D. Chapman. 2002. "N-Acylethanolamines Are Metabolized by Lipoxygenase and Amidohydrolase in Competing Pathways during Cottonseed Imbibition." *Plant Physiology* 130(1):391–401.

Stelt, Mario, Susan H. Fox, Michael Hill, Alan R. Crossman, Stefania Petrosino, Vincenzo Di Marzo, and Jonathan M. Brotchie. 2005. "A Role for Endocannabinoids in the Generation of Parkinsonism and Levodopa-induced Dyskinesia in MPTP-lesioned Non-human Primate Models of Parkinson's Disease." *The FASEB Journal* 19(9):1140–42.

Stevenson, Sean R., Yasuko Kamisugi, Chi H. Trinh, Jeremy Schmutz, Jerry W. Jenkins, Jane Grimwood, Wellington Muchero, Gerald A. Tuskan, Stefan A. Rensing, Daniel Lang, Ralf Reski, Michael Melkonian, Carl J. Rothfels, Fay-Wei Li, Anders Larsson, Gane K. S. Wong, Thomas A. Edwards, and Andrew C. Cuming. 2016. "Genetic Analysis of Physcomitrella Patens Identifies ABSCISIC ACID NON-RESPONSIVE, a Regulator of ABA Responses Unique to Basal Land Plants and Required for Desiccation Tolerance." *The Plant Cell* 28(6):1310–27.

Sun, Yong-Xin, Kazuhito Tsuboi, Yasuo Okamoto, Takeharu Tonai, Makoto Murakami, Ichiro Kudo, and Natsuo Ueda. 2004. "Biosynthesis of Anandamide and N-Palmitoylethanolamine by Sequential Actions of Phospholipase A2 and Lysophospholipase D." *The Biochemical Journal* 380(Pt 3):749–56.

Tan, Lin, Jin Tai Yu, Nan Hu, and Lan Tan. 2013. "Non-Coding RNAs in Alzheimer's Disease." *Molecular Neurobiology* 47(1):382–93.

Tan, Shawn Lu Wen, Saakshi Chadha, Yansheng Liu, Evelina Gabasova, David Perera, Karim Ahmed, Stephanie Constantinou, Xavier Renaudin, MiYoung Lee, Ruedi Aebersold, and Ashok R. Venkitaraman. 2017. "A Class of Environmental and Endogenous Toxins Induces BRCA2 Haploinsufficiency and Genome Instability." *Cell* 169(6):1105-1118.e15.

Teaster, Neal D., Jantana Keereetaweep, Aruna Kilaru, Yuh-Shuh Wang, Yuhong Tang, Christopher N. Q. Tran, Brian G. Ayre, Kent D. Chapman, and Elison B. Blancaflor. 2012. "Overexpression of Fatty Acid Amide Hydrolase Induces Early Flowering in Arabidopsis Thaliana." *Frontiers in Plant Science*.

Teaster, Neal D., Christy M. Motes, Yuhong Tang, William C. Wiant, Matthew Q. Cotter, Yuh Shuh Wang, Aruna Kilaru, Barney J. Venables, Karl H. Hasenstein, Gabriel Gonzalez, Elison B. Blancaflor, and Kent D. Chapman. 2007. "N-Acylethanolamine Metabolism Interacts with Abscisic Acid Signaling in Arabidopsis Thaliana Seedlings." *Plant Cell* 19(8):2454–69.

Thomas, E. A., B. F. Cravatt, P. E. Danielson, N. B. Gilula, and J. G. Sutcliffe. 1997. "Fatty Acid Amide Hydrolase, the Degradative Enzyme for Anandamide and Oleamide, Has Selective Distribution in Neurons within the Rat Central Nervous System." *Journal of Neuroscience Research* 50(6):1047–52.

Uauy, R., P. Mena, B. Wegher, S. Nieto, and N. Salem. 2000. "Long Chain Polyunsaturated Fatty Acid Formation in Neonates: Effect of Gestational Age and Intrauterine Growth." *Pediatric Research* 47(1):127–35.

Ueda, Natsuo, Kazuhito Tsuboi, and Toru Uyama. 2010a. "Enzymological Studies on the Biosynthesis of N-Acylethanolamines." *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 1801(12):1274–85.

Ueda, Natsuo, Kazuhito Tsuboi, and Toru Uyama. 2010b. "N-Acylethanolamine Metabolism with Special Reference to N-Acylethanolamine-Hydrolyzing Acid Amidase (NAAA)." *Progress in Lipid Research* 49(4):299–315.

Venables, Barney J., Cheryl A. Waggoner, and Kent D. Chapman. 2005. "N-Acylethanolamines in Seeds of Selected Legumes." *Phytochemistry* 66(16):1913–18.

Wang, Haibin, Huirong Xie, Yong Guo, Hao Zhang, Toshifumi Takahashi, Philip J. Kingsley, Lawrence J. Marnett, Sanjoy K. Das, Benjamin F. Cravatt, and Sudhansu K. Dey. 2006. "Fatty Acid Amide Hydrolase Deficiency Limits Early Pregnancy Events." *The Journal of Clinical Investigation* 116(8):2122–31.

Wang, Meng Cheng, Zhen Ying Peng, Cui Ling Li, Fei Li, Chun Liu, and Guang Min Xia. 2008. “Proteomic Analysis on a High Salt Tolerance Introgression Strain of Triticum Aestivum/Thinopyrum Ponticum.” *Proteomics* 8(7):1470–89.

Wang, Xiaoqin and Yikun He. 2009. “Negative Effects of Desiccation on the Protein Sorting and Post-Translational Modification.” *Plant Signaling & Behavior* 4(5):435–37.

Wang, Y. S., R. Shrestha, A. Kilaru, W. Wiant, B. J. Venables, K. D. Chapman, and E. B. Blancaflor. 2006. “Manipulation of Arabidopsis Fatty Acid Amide Hydrolase Expression Modifies Plant Growth and Sensitivity to N-Acylethanolamines.” *Proceedings of the National Academy of Sciences*.

Wei, Binqing Q., Tarjei S. Mikkelsen, Michele K. McKinney, Eric S. Lander, and Benjamin F. Cravatt. 2006. “A Second Fatty Acid Amide Hydrolase with Variable Distribution among Placental Mammals.” *The Journal of Biological Chemistry* 281(48):36569–78.

Wolf, Luise, Luca Rizzini, Ralf Stracke, Roman Ulm, and Stefan A. Rensing. 2010. “The Molecular and Physiological Responses of Physcomitrella Patens to Ultraviolet-B Radiation.” *Plant Physiology* 153(3):1123–34.

Xiao, Lihong, Hui Wang, Ping Wan, Tingyun Kuang, and Yikun He. 2011. “Genome-Wide Transcriptome Analysis of Gametophyte Development in Physcomitrella Patens.” *BMC Plant Biology* 11(1):177.

Xiao, Lihong, Liechi Zhang, Ge Yang, Honglin Zhu, and Yikun He. 2012. “Transcriptome of Protoplasts Reprogrammed into Stem Cells in Physcomitrella Patens” edited by C. Gibas. *PLoS ONE* 7(4):e35961.

Yu, M., D. Ives, and C. S. Ramesha. 1997. “Synthesis of Prostaglandin E2 Ethanolamide from Anandamide by Cyclooxygenase-2.” *The Journal of Biological Chemistry* 272(34):21181–86.

Yuan, Jiapei, Jingrui Li, Yang Yang, Chang Tan, Yumin Zhu, Long Hu, Yijun Qi, and Zhi John Lu. 2018. “Stress-responsive Regulation of Long Non-coding RNA Polyadenylation in Oryza Sativa.” *The Plant Journal* 93(5):814–27.

Zhang, Guojie, Guangwu Guo, Xueda Hu, Yong Zhang, Qiye Li, Ruiqiang Li, Ruhong

Zhuang, Zhike Lu, Zengquan He, Xiaodong Fang, Li Chen, Wei Tian, Yong Tao, Karsten Kristiansen, Xiuqing Zhang, Songgang Li, Huanming Yang, Jian Wang, and Jun Wang. 2010. "Deep RNA Sequencing at Single Base-Pair Resolution Reveals High Complexity of the Rice Transcriptome." *Genome Research* 20(5):646–54.

Zhang, Tianzhu, Hui Hu, Ge Yan, Tangwei Wu, Shuiyi Liu, Weiqun Chen, Yong Ning, and Zhongxin Lu. 2019. "Long Non-Coding RNA and Breast Cancer." *Technology in Cancer Research & Treatment* 18.

Zhao, Xinyue, Jingrui Li, Bi Lian, Hanqing Gu, Yan Li, and Yijun Qi. 2018. "Global Identification of Arabidopsis LncRNAs Reveals the Regulation of MAF4 by a Natural Antisense RNA." *Nature Communications* 9(1):1–12.

Zoerner, Alexander A., Frank-Mathias Gutzki, Sandor Batkai, Marcus May, Christin Rakers, Stefan Engeli, Jens Jordan, and Dimitrios Tsikas. 2011. "Quantification of Endocannabinoids in Biological Systems by Chromatography and Mass Spectrometry: A Comprehensive Review from an Analytical and Biological Perspective." *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 1811(11):706–23.

VITA

MD IMDADUL HAQ

Education: PhD, Biomedical Sciences, East Tennessee State University, Johnson City, Tennessee, 2020

MS, Biology, East Tennessee State University, Johnson City, Tennessee, 2014

MS, Biotechnology, University of Malaya, Kuala Lumpur Malaysia, 2011

BS, Biology, University of Dhaka, Dhaka Bangladesh, 2007

Professional Experience: Teaching Assistant, East Tennessee State University, Johnson City Tennessee, 2015-2019

Graduate Assistant, East Tennessee State University, College of Arts and Sciences, 2015-2020

Bioinformatics Intensive, National Center for Genomic Resources, Santa Fe, New Mexico, 2019

Proteomics, Cold Spring Harbor Laboratory, New York, 2018

Publications: Haq I and Kilaru A (2020). An endocannabinoid catabolic enzyme FAAH and its paralogs in an early land plant reveal evolutionary and functional relationship with eukaryotic orthologs. Scientific Reports. 10:3115

Haq I, Thakuri B, Hobbs T, Davenport M, and Kumar D (2020). Tobacco SABP2-interacting protein SIP428 is a SIR2 type

deacetylase. Plant Physiology and Biochemistry (*In press*).

Smith, John J. (2010). "What Everyone Should Know."

Chigurupati P, Haq I, and Kumar D (2016). Tobacco methyl salicylate esterase mediates nonhost resistance. Current Plant Biology. 6, 48-55.

Hossain M, Barrow J, Shen Y, Haq I, and Bungert J (2015). Artificial Zinc Finger DNA Binding Domains: Versatile Tools for Genome Engineering and Modulation of Gene Expression. J. of Cell. Biochem. 116 (11) 2435- 2444.

Kumar D, Haq I, Chapagai D, Tripathi D, Donald D, Hossain M, and Devaiah S (2015) Salicylic acid Signaling: Current Perspective on Role in Plants. *Recent Advances in Phytochemistry*, Springer. 45, 115-136.

Honors and Awards:

Sigma XI Research Award, 2019

Best Poster, Appalachian Student Research Forum, 2017

Significant Presentation, Gordon Research Conference, 2019

Phytochemical Society of North America, 2017

iMOSS, 2018

American Society of Plant Biologists, 2017 & 2019

National Institute of Child Health & Human Development, 2018

National Center for Genomic Resources, 2018