

Functional Analysis of HAP2 Reveals Insights into a Gamete Fusion Mechanism

By

Jennifer J. Forcina

B.S., University of New Hampshire, 2007

M.S., University of New Hampshire, 2009

Thesis

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

in the Division of Biology and Medicine:

Molecular Biology, Cell Biology, and Biochemistry Graduate Program at Brown

University

PROVIDENCE, RHODE ISLAND

May 2018

Copyright 2018 © by Jennifer Forcina

This dissertation by Jennifer J. Forcina is accepted in its present form by the
Molecular Biology, Cell Biology, and Biochemistry Graduate Program as satisfying
the dissertation requirement for the degree of Doctor of Philosophy.

Date _____

Dr. Mark A. Johnson, Advisor

Recommended to the Graduate Council

Date _____

Dr. Alison DeLong, Reader

Date _____

Dr. Gary Wessel, Reader

Date _____

Dr. Nicolas Fawzi, Reader

Date _____

Dr. Margaret Kielian, Outside Reader

Approved by the Graduate Council

Date _____

Dr. Andrew G. Campbell, Dean of the Graduate School

Curriculum Vitae

Jennifer Forcina

EDUCATION

- 2012-2018 **Brown University**, Providence, RI
 Ph.D. Molecular Biology, Cell Biology, and Biochemistry (expected May 2018)
 Dissertation: Functional Analysis of HAP2 Reveals Insights into a Gamete Fusion Mechanism
- 2007-2009 **University of New Hampshire**, Durham, NH
 M.S. Animal Science (December 2009)
 Thesis: Genetic Over-Expression of Cytokeratin 18 in Bovine Luteal Cells: Effects on Intermediate Filament Formation and Fas Ligand-Induced Apoptosis
- 2003-2007 **University of New Hampshire**, Durham, NH
 B.S. Medical Laboratory Science (May 2007)
 Summa Cum Laude

RESEARCH EXPERIENCE

- 2013-2018 **Brown University**, Providence, RI
 Graduate Student
 Ph.D. Thesis; Advisor: Dr. Mark Johnson
 Project: Mechanisms of HAP2-mediated gamete fusion in Arabidopsis
- Generation and characterization of transgenic plants
 - Transient transformation of plant material for protein analyses
 - Molecular cloning and site-directed mutagenesis
 - Genetic and genomic analyses
 - Fluorescence and confocal microscopy
 - Writing and preparation of grants, manuscript, posters, and oral presentations
 - Supervising and mentoring of undergraduate and rotating graduate students
- 07/2010-08/2012 **Cell Signaling Technology**, Danvers, MA
 Product Scientist I
- Antibody affinity purification and quality control, including Western Blot, Immunoprecipitation, and protein phosphorylation analysis
 - Mammalian cell culture and protein lysate preparation
 - Written and oral communication with customers in technical support of antibody products

- 10/2009-05/2010 **PPD Dermatology, Waltham, MA**
Research Specialist
- Development, optimization, and use of cell-based assays for screening of compounds for lead selection
 - Mammalian cell culture and high-throughput qPCR analysis
 - Role in the process improvement group
 - Use of Oracle software and GraphPad Prism for data management and analysis
- 8/2007-12/2009 **University of New Hampshire, Durham, NH**
M. S. Graduate student; Advisor: Dr. David Townson
Project: Characterization and manipulation of bovine luteal cells
- Primary bovine luteal cell culture, including tissue dissociation
 - Adenovirus transduction of primary bovine luteal cells
 - Immunofluorescence, flow cytometry, immunohistochemistry
 - Supervising and mentoring of undergraduate students
- 2/2006-5/2007 **University of New Hampshire, Durham, NH**
Undergraduate student; Advisor: Dr. Frank Rodgers
Project: Investigations into pro-/anti-apoptotic proteins in neutrophils
- Isolation of neutrophils from whole blood
 - Analysis of apoptosis-related proteins in neutrophils by Western Blot

RELEVANT SKILLS

- Molecular cloning and site-directed mutagenesis of plasmids
- Generation and analysis of transgenic plants
- DNA and RNA Isolation, PCR, RT-PCR
- Microscopy, including fluorescence and confocal of both live and fixed samples
- Genomic library preparation
- Mammalian cell culture, including primary cell culture, maintenance of adherent and suspension cell lines, and adenovirus transduction of primary bovine luteal cells
- Western blot, immunoprecipitation, flow cytometry, immunofluorescence, immunohistochemistry, and high-throughput qPCR
- Antibody affinity purification and quality control testing
- Protein expression and purification from bacteria using affinity and size exclusion chromatography
- Proficient with Mac OS, Windows, Adobe Illustrator, R studio, ImageJ, Microsoft Excel, Microsoft Powerpoint, Pymol, Python, Zen software (Zeiss)

TEACHING EXPERIENCE

- 2015-2016 **Certificate IV: The Teaching Consultant Program**
Harriet W. Sheridan Center for Teaching and Learning
Brown University, Providence, RI

- 07/2015 **Cold Spring Harbor Laboratory**, Cold Spring Harbor, NY
Teaching Assistant for Frontiers and Techniques in Plant Science
Instructors: Dr. Mark Johnson (Brown University), Dr. Shin Han Shiu
(Michigan State University), Dr. Marja Timmermans (CSHL)
- 2014-2015 **Certificate I: Sheridan Teaching Seminar – Reflective Teaching**
Harriet W. Sheridan Center for Teaching and Learning
Brown University, Providence, RI
- Spring 2014 **Brown University**, Providence, RI
Graduate Teaching Assistant for Plant Organism
Professors: Dr. Mark Johnson, Dr. Alison DeLong
- Spring 2008 **University of New Hampshire**, Durham, NH
Graduate Teaching Assistant for General Biology II
Supervisor: Dr. Carl Vaughan
- Fall 2007 **University of New Hampshire**, Durham, NH
Graduate Teaching Assistant for Cell Culture
Professor: Dr. Paul Tsang

OTHER PROFESSIONAL EXPERIENCE

- April 2017 Attended Pollen Network Imaging Workshop, University of Maryland
- 2014-2017 MCB Graduate Program First-Year Student Mentor
- 2014-2016 Sheridan Center Graduate Student Liaison for MCB Graduate Program
- 2014-2015 Graduate Student Member of MCB Graduate Program Admissions
Committee
- May 2014 Attended a Workshop in Next-Generation Sequencing hosted by Pollen
Research Collaborative Network, Charlotte, NC

HONORS AND AWARDS

- Brown University MCB Graduate Program Retreat Excellent Poster Award, August 2017
- American Society for Plant Biology Travel Grant Award, June 2017
- Plant Reproduction Registration Fee Award from Pollen RCN, March 2016
- Brown University MCB NIH Institutional T32 Training Grant Appointment, July 2014-June 2015
- Brown University MCB NIH Institutional T32 Training Grant Appointment, July 2013-June 2014
- Brown University MCB Graduate Program Retreat Excellent Poster Award, August 2013
- UNH COLSA Graduate Summer Fellowship, Summer 2009
- UNH COLSA Graduate Summer Fellowship, Summer 2008
- UNH Graduate TA Summer Fellowship, 2008
- Summa Cum Laude / Presidential Scholar, May 2007
- Undergraduate Research Opportunities Program (UROP) Award, September 2006-May 2007
- Martha Hopkins Scholarship, May 2006
- UNH Presidential Scholarship, September 2003-May 2007

PUBLICATIONS

J. Fedry*, **J. Forcina***, P. Legrand, G. Pehau-Arnaudet, A. Haouz, M. Johnson, F. A. Rey, and T. Krey. "HAP2 drives gamete fusion via insertion of highly divergent fusion loops into female gametes." Submitted. *co-first authors

A. Duncan*, **J. Forcina***, A. Birt, and D. Townson. "Estrous cycle-dependent changes of Fas expression in the bovine corpus luteum: influence of keratin 8/18 intermediate filaments and cytokines." *Reproductive Biology and Endocrinology* 10:90, 2012. PMID: 23113883
*co-first authors

SELECT PRESENTATIONS

"The membrane interaction surface of HAP2(GCS1), the ancient gamete fusion protein."
Gordon Research Conference: Fertilization & Activation of Development. Holderness School, Holderness, NH, July 2017.

Oral presentation selected from poster presentations.

"Arabidopsis HAP2(GCS1) Possesses a Putative Fusion Loop That is Essential for its Function in Fertilization."

Poster presented at Plant Biology in Honolulu, HI, June 2017.

"Double fertilization goes viral: HAP2-mediated gamete membrane fusion."

Oral presentation at Plants in New England (PINE), Whitehead Institute, Cambridge, MA, May 2017.

"Gamete fusion requires a domain in HAP2(GCS1) with predicted structural similarity to viral fusion proteins."

Oral presentation at EMBO Workshop on "Membrane fusion in health and disease", Paris, France. June 2016.

"Is HAP2(GCS1) the only DUO1-regulated gene required for gamete fusion?"

Poster presented at the 24th International Congress on Sexual Plant Reproduction, Tucson, AZ. March 2016.

"Defining the Function of HAP2(GCS1) in Arabidopsis."

Poster presented at the Gordon Research Conference on Fertilization and Activation of Development, Holderness, NH. July 2015.

"A genetic approach to identify novel components of gamete fusion mechanisms in Arabidopsis."

Poster Presented at 23rd International Congress on Sexual Plant Reproduction, Porto, Portugal. July 2014.

"Genetic Over-Expression of Cytokeratin 18 in Bovine Luteal Cells: Effects on Intermediate Filament Formation and FasL-Induced Death."

Poster presented at 42nd Annual Meeting: The Society for the Study of Reproduction, Pittsburgh, PA. July 2009.

Acknowledgements

I would like to thank my thesis advisor, Dr. Mark Johnson, for every way he has helped me become a better student, mentor, researcher, and colleague. His enthusiasm for science is impressive, and I have enjoyed all of our scientific conversations and musings. I appreciate his patience with my many questions, thoughts, and often random ideas. Mark constantly challenges me to be better in every way I can, and for that, I am entirely grateful. I have had many amazing and invaluable experiences as a result of joining his laboratory, and I cannot thank Mark enough for supporting my interests and my training beyond anything I expected.

I would like to thank my committee members for all their support, advice, and wisdom:

I would like to thank Dr. Nicolas Fawzi for taking a chance by being on my committee. I have benefitted greatly. He has helped me think differently about my project in many ways, and he has had a positive effect on both me and my thesis.

I would like to thank Dr. Gary Wessel for always being both supportive and helpful. Gary was my first rotation advisor, and I think it was an excellent start to my experiences at Brown University. I thank Gary for always bringing both positive energy and new perspectives to my committee meetings. Gary constantly encourages new experiences, new approaches, and putting yourself out there, and I think everyone needs someone like that in graduate school.

I thank Dr. Alison DeLong for all of her support, guidance, patience, and humor during this process. I thank Alison for sharing in the ability to laugh at ourselves. I appreciate all of our scientific and non-scientific conversations. I have learned a lot from Alison, and she has helped me be a better scientist and person.

I would like to thank all the past and current members of the Johnson Laboratory for all of their support and positive influences along the way. I would especially like to thank Dr. Kristin Beale for her mentorship and friendship. It has been wonderful to have someone who understands and is always willing to listen. I thank Dr. Alex Leydon for being an amazing mentor in the laboratory and for all the laughs, silliness, and seriousness over the years. Leydon helped me become a better scientist and provided encouragement and chaos in appropriate doses. I thank all of the wonderful undergraduate students I have been lucky enough to work with. I would like to specifically acknowledge Samyukta Mallick for all her help during her time working in the laboratory. I appreciate her positive attitude, enthusiasm, and work ethic.

I would like to thank our collaborators on the HAP2 structure-function manuscript, specifically Dr. Juliette Fedry, Dr. Thomas Krey, and Dr. Felix Rey. It has been an amazing and educational experience, and I am so grateful for this collaboration.

I want to thank everyone in the Arabidopsis group at Brown for their support, feedback, and willingness to hear all my practice presentations.

I would like to thank Geoff Williams for his help with microscopy and for answering all my questions.

I would like to thank the MCB graduate program, specifically Elaine Butler and Ashley Walker, for all their help with anything and everything I needed in graduate school.

Finally, I would like to thank my husband, Chris Connors, for having exceptional patience all during my time in graduate school and especially during my thesis writing. He has endured many scientific musings (or rants) and provided a lot of valuable insights into my thesis.

Table of Contents

Title Page	i
Copyright Page	ii
Signature Page	iii
Curriculum Vitae	iv
Acknowledgements	viii
Table of Contents	x
List of Figures and Tables	xiv

Chapter 1: Introduction: HAP2(GCS1) is a gamete fusion protein

that functions within an array of sexual systems	1
I. HAP2 mediates gamete fusion in a variety of species	1
Gamete fusion is essential to sexual reproduction	1
HAP2 expression is restricted to gametes and found in diverse sexual organisms	1
HAP2 is required for fertilization	3
HAP2 functions after gamete activation and adhesion	4
HAP2 is a gamete fusion protein	5
Hypotheses for HAP2-mediated gamete fusion can be derived from class II fusion proteins	8
II. Considerations for regulation of HAP2 activity in Arabidopsis fertilization	10
Arabidopsis pollen tubes deliver sperm to the site of fusion within the ovule	10
Hypothesis for regulation of HAP2 as a class II viral-like fusion protein	11
Concluding Remarks	12

Perspectives: What is the mechanism of HAP2-mediated gamete fusion in Arabidopsis?	12
References	14

Chapter 2: HAP2 drives gamete fusion via insertion of highly divergent fusion

loops into female gametes	18
Summary	19
Introduction	20
Results and Discussion	24
Distinct patterns of insertions and deletions in the tip of HAP2 domain II	24
The core structure of domain II is conserved across eukaryotes	27
An evolutionarily plastic domain II tip, exposing an amphipathic helix in Arabidopsis	30
The HAP2 ectodomains insert into liposomes as trimers via the domain II tip	32
α F amphipathicity is required for <i>AtHAP2</i> <i>in vitro</i> liposome insertion and <i>in vivo</i> function	33
The conserved E ¹¹⁷ -R ¹⁶³ salt-bridge and the helical character of α F are essential for Arabidopsis gamete fusion	38
The Rice α F is able to mediate fusion within the context of Arabidopsis HAP2	39
Concluding Remarks	41
Acknowledgements	43
References	45
Methods	49
Supplementary Figures	56
Supplementary Tables	64

Chapter 3: Expression of HAP2 in the female gametes as an opportunity to	
 assess models of HAP2 sufficiency	70
Summary	71
Introduction	72
Results	76
Discussion	83
Methods	85
Acknowledgements	89
Supplementary Figures	90
Supplementary Tables	91
References	94
 Chapter 4: Assessment of pH as a trigger for HAP2-mediated gamete fusion	 98
Summary	99
Introduction	100
Results	103
Discussion	115
Methods	118
Acknowledgements	123
Supplemental Figures	124
Supplemental Tables	131
References	134
 Chapter 5: HAP2 Fusion Helix: Conservation in Flowering Plants	
 and Expression in Sperm	138

Summary	139
Introduction	140
Results	142
Discussion	152
Methods	154
Acknowledgements	157
Supplementary Figures	158
Supplementary Tables	163
References	165

Chapter 6: Discussion: Arabidopsis HAP2 fuses gametes via an amphipathic

helix and a mechanism reminiscent of viral fusion proteins 168

Arabidopsis HAP2 presents an amphipathic helix at its membrane	
interaction surface	168
The HAP2 fusion helix is conserved in flowering plants	170
Sub-cellular localization of HAP2 may be highly regulated	171
Mutations in the fusion helix do not affect HAP2-YFP localization in sperm	
within pollen tubes	174
The environment at the site of gamete fusion contains factors that	
regulate HAP2 function	175
Perspectives: What is the pre-fusion form of HAP2?	177
How is HAP2 regulated?	179
Concluding Remarks	180
References	181

List of Figures and Tables

Chapter 1

Figure 1. Functional localization and characterization of HAP2 from a variety of species	2
Figure 2. Schematic views of the gamete fusion process in <i>Chlamydomonas</i> and <i>Arabidopsis</i>	7
Figure 3. Structure-based schematic of HAP2	9
Figure 4. Hypothetic model for HAP2 function as a viral-like fusion protein	10

Chapter 2

Figure 1. The HAP2 membrane interacting region has diverged across eukaryotic lineages	23
Figure 2. Structures of <i>Arabidopsis thaliana</i> and <i>Trypanosoma cruzi</i> HAP2	28
Figure 3. The <i>A. thaliana</i> HAP2 amphipathic fusion helix α F	32
Figure 4. <i>A. thaliana</i> HAP2e trimers insert into liposomes via a non-polar domain II apex	35
Figure 5. <i>Arabidopsis</i> gamete fusion requires presentation of hydrophobic residues within the amphipathic HAP2 α F helix	36
Figure 6. Variability in the HAP2 membrane interaction surface across eukaryotic phyla	40
Figure S1: Occurrence of the defining HAP2 pfam domain (10699) in eukaryotic genomes	56
Figure S2: Characterization of <i>At</i> HAP2e and <i>Tc</i> HAP2e and crystals of a protease-resistant <i>Tc</i> HAP2esub fragment	58
Figure S3: Comparison of trimeric class II fusion protein structures	60

Figure S4: <i>T. cruzi</i> HAP2e trimers insert into liposomes using a hydrophobic tip	61
Figure S5: Amphipathic helix prediction in loop 2 of plant and algal HAP2	62
Figure S6: Analysis of HAP2variant:YFP accumulation in sperm cells	63
Table S1: Accession numbers of amino acid sequences used in this study	64
Table S2: Data collection, phasing and refinement statistics	66
Table S3: Transgenic lines used to analyze function of HAP2:YFP variants	67
Table S4: Primers used for generation of HAP2:YFP variants and for genotyping and sequencing transformed plants	68

Chapter 3

Figure 1. Overview of an Arabidopsis ovule highlighting the location and shape of the egg cell	74
Figure 2. HAP2 expression in the egg cell	76
Figure 3. Ovules from EC1.1:HAP2-GFP/DD65:HAP2-mCherry plants	78
Figure 4. HAP2-GFP expression in synergid cells	82
Table 1. Transmission data for pEC1.1:HAP2-GFP/pDD65:HAP2:mCherry transformed plants pollinated with pollen from <i>hap2-1/+</i> plants	77
Table 2. Sex-specific crosses for the first generation of plants transformed with pEC1.1:HAP2-GFP/pDD65:HAP2:mCherry	79
Table 3. Transmission analysis of HAP2-YFP variants crossed to wild-type pistils or pEC1.1:HAP2-GFP/pDD65:HAP2-mCherry pistils	81
Figure S1. HAP2 expression in the egg cell after gamete fusion	90
Tables S1. Transmission data for primary transformed plants expressing HAP2-GFP in the egg cell	91
Tables S2. Transmission data for primary transformed plants expressing HAP2-GFP in the egg cell pollinated with <i>hap2-1/+</i> pollen	92

Table S3. Primers used in this study.	93
---------------------------------------	----

Chapter 4

Figure 1. Calibration of pHluorin in Arabidopsis pollen tubes	105
Figure 2. pH of pollen tubes grown both in SIV and in vivo for ~7 hours after pollination	106
Figure 3. pH at the site of fusion	107
Figure 4. Expression of pHluorin in synergid cells	109
Figure 5. Calibration of pHluorin in synergid cells	110
Figure 6. Expression of pHuji in synergid cells	111
Figure 7. Expression of HAP2 Domain III in synergid cells	113
Figure 8. Expression of a HAP2 Domain III with an extended C-terminus in synergid cells	114
Figure S1. Schematic demonstrating blocking mechanism of exogenous domain III of class II viral fusion proteins	124
Figure S2. Calibration trials for three independent experiments	125
Figure S3. Comparing pH of pollen tubes observed inside and outside ovules in the SIV assay	126
Figure S4. Expression of pHluorin in pollen tube with sperm nucleus marked with RFP	127
Figure S5. Comparing pHluorin ratio for the degenerating synergid versus the persistent or untargeted synergid	128
Figure S6. Arabidopsis HAP2 ectodomain monomer	129
Figure S7. Ovules expressing HAP2-DIII-GFP 16 hours after pollination with a sperm nuclei marker	130
Table S1. Male-specific transmission results for three primary transformed	

plants expressing HAP2-DIII:GFP in synergid cells	131
Table S2. Female-specific transmission results for three primary transformed plants expressing HAP2-DIII:GFP in synergid cells	132
Table S3. Female-specific transmission results for three primary transformed plants expressing HAP2-DIII-extended:GFP in synergid cells	133
 Chapter 5	
Figure 1. Genetic transmission assay results for HAP2 mutants	143
Figure 2. Expression of a sperm membrane marker with different HAP2 fusion helix variants	147
Figure 3. Expression of a sperm membrane marker with a double cysteine mutant in HAP2	148
Figure 4. Expression of a sperm ER marker	151
Figure S1. Arabidopsis HAP2 fusion helix at the tip of domain II	158
Figure S2. Predicted helix for inserted Tc.loop2 construct	159
Figure S3. Pearson's correlation coefficient for different HAP2 variants with the sperm membrane marker	160
Figure S4. HAP2 double cysteine mutant C312A/C319A is expected to disrupt two disulfide bonds	161
Figure S5. Examples of highly expressing HAP2 genomic fusion protein in sperm	162
Table S1. HAP2-YFP variant plants used for the genetic transmission analysis	163
Table S2. Primers used for site-directed mutagenesis of HAP2-YFP	164
 Chapter 6	
Figure 1. Structures of TBEV E and Arabidopsis HAP2 (predicted and solved)	169
Figure 2. Models for potential HAP2 fusion mechanisms	173

Figure 3. Models for potential regulation of the HAP2-mediated gamete fusion mechanism in Arabidopsis	176
Figure 4. Schematics representing potential pre-fusion models for HAP2	178

Chapter 1: Introduction: HAP2(GCS1) is a gamete fusion protein that functions within an array of sexual systems

I. HAP2 mediates gamete fusion in a variety of species.

Gamete fusion is essential to sexual reproduction.

Sexually dimorphic gametes are distinctly different from each other and are produced in species that include, but are not limited to, those that produce sperm and egg cells. Some organisms produce isogametes that are similar in their morphology (Lehtonen et al., 2016). Reproduction in both dimorphic (anisogamous) and isogamous species involves the merging of two gametes and their genetic material in a process called gamete fusion, or syngamy, which results in a zygote (Dresselhaus et al., 2016; Lehtonen et al., 2016). Gamete fusion requires the membranes from two separate gametes to fuse before nuclear fusion can occur (Clark et al., 2006; Dresselhaus et al., 2016; Lehtonen et al., 2016). One common mechanism by which cell membrane fusion occurs is initiation by a fusion protein, or fusogen (Oren-Suissa and Podbilewicz, 2007; Wong and Johnson, 2010). HAP2 is the first identified gamete fusion protein (Fedry et al., 2017; Valansi et al., 2017) and provides the opportunity to define the mechanism of gamete fusion in a variety of species (Liu et al., 2008; Steele and Dana, 2009; Wong and Johnson, 2010).

HAP2 expression is restricted to gametes and found in diverse sexual organisms.

HAP2 was initially discovered in species from the Plantae kingdom, specifically *Arabidopsis thaliana* (Johnson et al., 2004; Mori et al., 2006; von Besser et al., 2006) and *Lilium longiflorum* (GCS1, (Mori et al., 2006)). HAP2 was then quickly identified and detected in several other species, including *Chlamydomonas* (Liu et al., 2008; Mori et al., 2006), *Plasmodium falciparum* and *berghei* (Hirai et al., 2008; Liu et al., 2008; Mori et al.,

2006), and *Physarum polycephalum* (Mori et al., 2006). The conserved pfam10699 domain, or HAP2/GCS1 domain, has enabled identification of HAP2 proteins in different species in all other eukaryotic kingdoms except fungi ((Hirai et al., 2008; Liu et al., 2008; Steele and Dana, 2009; Wong and Johnson, 2010); Figure 1A).

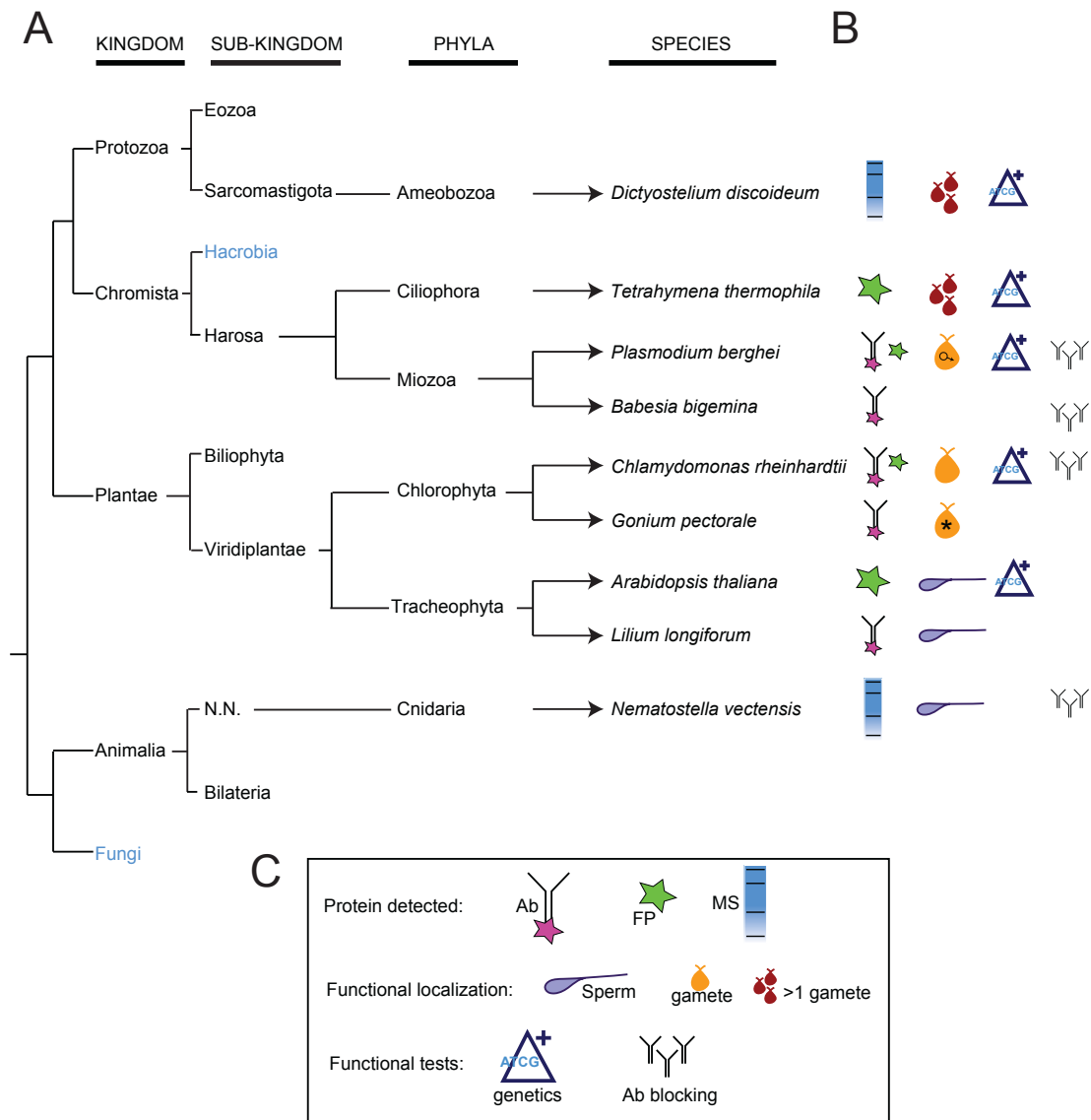


Figure 1. Functional localization and characterization of HAP2 from a variety of species. **A.** Species containing predicted HAP2 genes has been found in 4 out of the 5 eukaryotic kingdoms (shown in black), while absent in fungi (shown in blue). Within kingdoms that contain representatives with HAP2, all sub-kingdoms contain HAP2 representatives, except Hacrobia (shown in blue) as indicated by the presence of the pfam

10699 domain ((Fedry*, In preparation); Chapter 2). Only phyla with HAP2 proteins that have been experimentally tested are listed. **B.** Protein detection methods are indicated for each species (LC-MS/MS, fusion protein generation, antibody detection, or a combination of antibody detection to detect fusion proteins) with symbols explained in the legend (C). Results from localization studies are indicated for functionally relevant results. Finally, demonstration of function during fertilization is indicated, using either genetic and complementation assays or antibody blocking tests or both. For *Plasmodium berghei*, HAP2 has been detected in the male gamete, as indicated by the male symbol inside of the gamete cartoon (Liu et al., 2008). For *Gonium pectorale*, HAP2 has been detected in both gametes but is only present in one gamete after activation, as indicated by an asterisk inside the gamete cartoon (Kawai-Toyooka et al., 2014); therefore, expression in only one gamete is required for fertilization (Kawai-Toyooka et al., 2014).

Several studies have demonstrated that HAP2 is gamete specific (Kawai-Toyooka et al., 2014; Liu et al., 2008; Mori et al., 2006; von Besser et al., 2006). Due to its broad conservation in four out of the five eukaryotic kingdoms (Baldauf et al., 2000), HAP2 is expressed not only in sperm (Ebchuqin et al., 2014; Mori et al., 2006; von Besser et al., 2006) but also in isogametes from multiple species, including the seven-sex ciliate *Tetrahymena thermophila* (Cole et al., 2014), the slime mold *Dictyostelium discoideum* (Okamoto et al., 2016), and the algae *Gonium pectorale* ((Kawai-Toyooka et al., 2014); Figure 1B). In organisms that use sperm and eggs for fertilization, HAP2 protein has been detected in sperm ((Ebchuqin et al., 2014; Mori et al., 2006; von Besser et al., 2006); Figure 1B). In other species, HAP2 protein may be expressed in one mating type (Liu et al., 2008) or in more than one mating type ((Cole et al., 2014; Kawai-Toyooka et al., 2014; Okamoto et al., 2016); Figure 1B).

HAP2 is required for fertilization.

HAP2 is required to mediate successful fertilization in a variety of species, which has been demonstrated using a combination of genetic, cell biology, and biochemical approaches ((Wong and Johnson, 2010); Figure 1B). Genetic analysis has demonstrated that HAP2 is required in the male or minus gamete for fertilization in *Arabidopsis* (Johnson et al., 2004; Mori et al., 2006; von Besser et al., 2006), *Chlamydomonas* (Liu et al., 2008),

and *P. berghei* (Hirai et al., 2008), while being dispensable in the female or plus gametes (Hirai et al., 2008; Johnson et al., 2004; Liu et al., 2008; Mori et al., 2006). Additionally, antibody blocking experiments have demonstrated a requirement of HAP2 for successful fertilization in the parasites *P. berghei* and *P. falciparum* (Angrisano et al., 2017; Blagborough and Sinden, 2009) and *Babesia bigemina* (Camacho-Nuez et al., 2017), as well as the sea anemone *Nematostella vectensis* (Ebchuqin et al., 2014). In some isogamous species, neither HAP2 expression nor function is restricted to one gamete type (Cole et al., 2014; Okamoto et al., 2016). In Tetrahymena, all seven sexes express HAP2, and while fusion is significantly reduced when HAP2 is present on only one gamete, it still occurs (Cole et al., 2014), suggesting that HAP2 can function from only one gamete. Dictyostelium differentially expresses two isoforms of HAP2 within its three gamete types, but HAP2 is only required in two of the gamete types for fertilization (Okamoto et al., 2016). Together, results from these studies suggest that while HAP2 may be present in more than one gamete, its expression in one gamete is sufficient for fusion.

HAP2 functions after gamete activation and adhesion.

Regulation of HAP2 expression has been linked to gamete activation in several species (Kawai-Toyooka et al., 2014; Liu et al., 2008; Mori et al., 2006; Sprunck et al., 2012). Chlamydomonas produces isogametes of minus and plus mating types, and only after a mating pair adhere via their flagella are the gametes activated and HAP2 expressed on the minus gamete ((Liu et al., 2008); Figure 2A). In *G. pectorale*, gamete activation causes HAP2 degradation in one gamete, although it was initially expressed in both, so that only one gamete expresses HAP2 during fertilization (Kawai-Toyooka et al., 2014). In Arabidopsis, the egg cell secretes egg cell-1 proteins (EC1) that increase HAP2 protein localization in the sperm membrane in vitro, and the proposed role for EC1 is to activate sperm by regulating HAP2 function (Sprunck et al., 2012).

One of the challenges in identifying gamete fusion proteins has been determining the step in the fertilization process during which the protein functions (Dresselhaus et al., 2016; Klinovska et al., 2014; Oren-Suissa and Podbilewicz, 2007; Wright and Bianchi, 2016). Particularly difficult to parse out is whether a protein acts in gamete attachment (or adhesion) or in gamete fusion (Dresselhaus et al., 2016; Klinovska et al., 2014; Oren-Suissa and Podbilewicz, 2007; Wright and Bianchi, 2016). Recently, studies of the mammalian candidate fusion proteins of IZUMO1 and CD9, along with JUNO (the receptor for IZUMO1) have demonstrated they function in tight membrane adherence and are not sufficient for fusion (Bianchi et al., 2014; Chalbi et al., 2014; Wright and Bianchi, 2016). However, the possibility remains that these proteins may regulate or participate in a protein complex that may function during fusion (Aydin et al., 2016; Inoue, 2017; Inoue et al., 2015; Nishimura et al., 2016). Proteins required for gamete attachment have been identified in both *Chlamydomonas* and *Arabidopsis*. In *Chlamydomonas*, FUS1 is an attachment protein expressed on the minus gamete ((Misamore et al., 2003); Figure 2B). *Arabidopsis* GEX2 is expressed on the sperm and functions in attachment to the egg cell and central cell ((Mori et al., 2014); Figure 2F). A receptor has not been reported on the female side for FUS1 or GEX2. In agreement with its proposed role in fusion, transmission electron microscopy (TEM) of HAP2-deficient gametes in *Chlamydomonas* and *Tetrahymena* demonstrated a clear requirement for HAP2 after gamete attachment in these two species (Cole et al., 2014; Liu et al., 2008).

HAP2 is a gamete fusion protein.

A fusogen is not only required for the membrane merger event but must be sufficient to induce fusion in cells that would not normally fuse (Aguilar et al., 2013; Oren-Suissa and Podbilewicz, 2007; Wong and Johnson, 2010). Viral proteins are among some of the most well studied fusion proteins and are sufficient for membrane fusion when

expressed on one (the initiating) membrane, acting through a unilateral mechanism (Oren-Suissa and Podbilewicz, 2007; White et al., 2008; Wong and Johnson, 2010). SNARE proteins involved in vesicle fusion mechanisms also have been well studied and appear to use a distinctly different process, mediating fusion through a bilateral mechanism (Jahn and Scheller, 2006; Oren-Suissa and Podbilewicz, 2007; Wong and Johnson, 2010). Finally, cell-cell fusion events occur at distinct times during development in different tissues in a variety of species, including *Caenorhabditis elegans* and *Drosophila melanogaster* (Aguilar et al., 2013).

Recent structure-function studies of HAP2 from *Chlamydomonas*, *Tetrahymena*, and *Arabidopsis* have demonstrated that HAP2 is a fusion protein (Fedry et al., 2017; Pinello et al., 2017; Valansi et al., 2017). The x-ray crystal structure for *Chlamydomonas* HAP2 showed a trimer structure of its N-terminal extracellular region (ectodomain), analogous to the post-fusion form of class II fusion proteins ((Fedry et al., 2017); Figure 3A), including viral and somatic members of the category (Bressanelli et al., 2004; Gibbons et al., 2004; Modis et al., 2004; Perez-Vargas et al., 2014). *Chlamydomonas* HAP2 can interact with artificial membranes, and mutations in proposed fusion loops affect insertion into liposomes in vitro and fusion in vivo (Fedry et al., 2017). However, the fusion loop region was not resolved in the *Chlamydomonas* HAP2 structure (Fedry et al., 2017). A peptide corresponding to the membrane interaction region, including a predicted fusion loop, from *Tetrahymena* HAP2 was able to interact directly with lipids (Pinello et al., 2017). Lastly, *Arabidopsis* HAP2 was sufficient to drive membrane fusion when expressed ectopically on both cells in a mammalian cell culture system (Valansi et al., 2017).

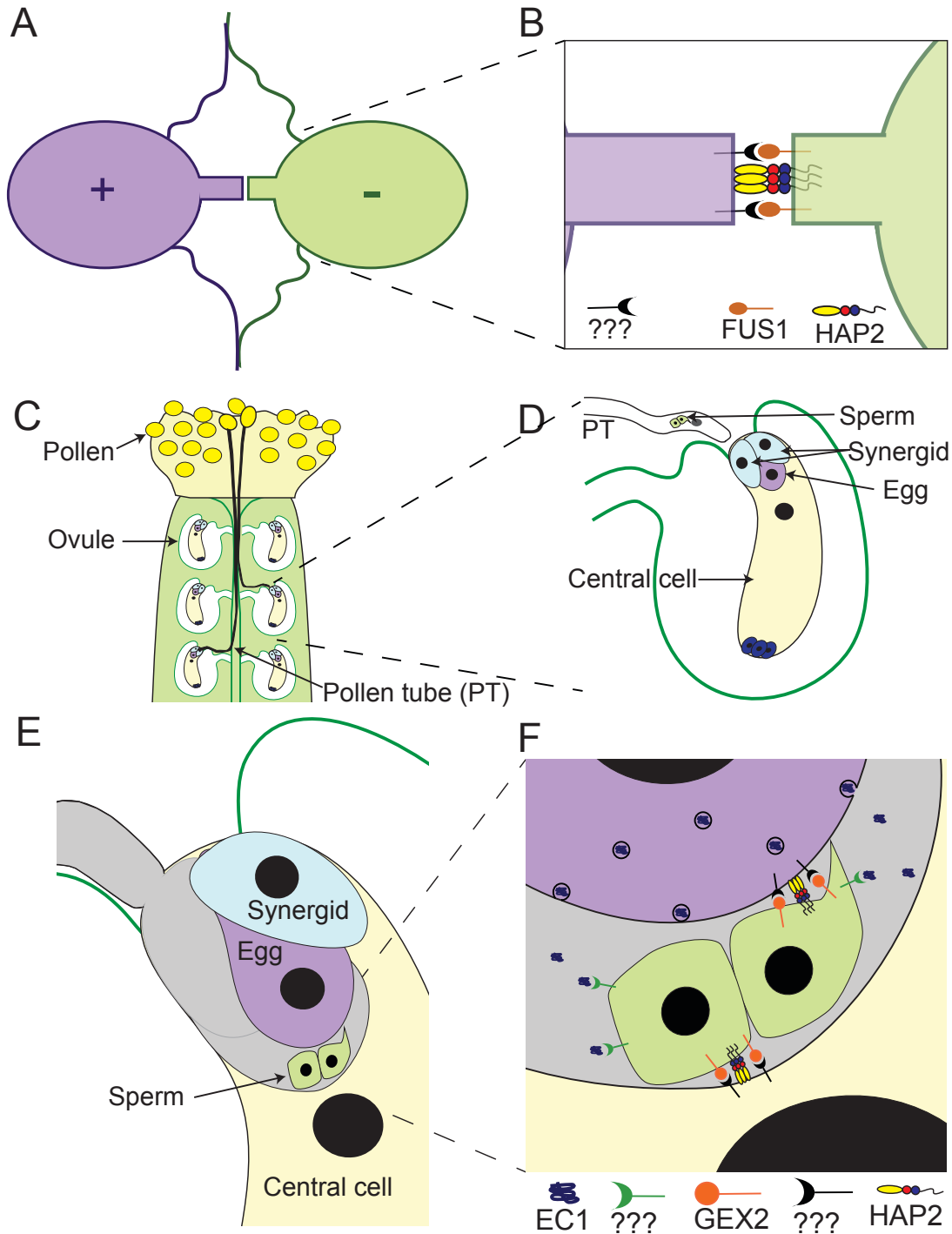


Figure 2. Schematic views of the gamete fusion process in *Chlamydomonas* and *Arabidopsis*. A. During fertilization, *Chlamydomonas* gametes adhere to each other via their flagella (Pan et al., 2003). B. Schematic of the gamete fusion site in *Chlamydomonas* after gamete activation and adhesion. FUS1 is shown attached to an unknown receptor. C. A schematic view of a cross-section through an *Arabidopsis* flower, which has been pollinated and pollen tubes have travelled through the transmitted tract and targeted ovules. D. Representation of a pollen tube approaching an ovule, with the sperm attached

to the pollen tube nucleus, and the seven cells of the female gametophyte, including the egg and central cell, shown (Drews and Yadegari, 2002). E. Arabidopsis ovule post-pollen tube burst but pre-membrane fusion. Pollen tube burst is shown in grey and overlaps with degenerated synergid area (also grey). F. Zoom in of the gamete fusion environment inside the ovule. GEX2 is shown binding to an unknown receptor. EC1 is shown being secreted from the egg cell and binding to an unknown receptor.

Hypotheses for HAP2-mediated gamete fusion can be derived from class II fusion proteins.

The ectodomains from Class II fusion proteins have a distinct three domain structure (Lescar et al., 2001; Rey et al., 1995), and a hydrophobic fusion loop is presented at the tip of domain II ((Bressanelli et al., 2004; Gibbons et al., 2004; Modis et al., 2004; Rey et al., 1995); Figure 3B) and interacts directly with the target membrane (Huang et al., 2010). Below the fusion loop, a conserved lipid binding pocket provides the interactions required for a tight fitting between the outer leaflet of the target membrane and the viral fusion protein tip (Guardado-Calvo et al., 2017). The class II viral fusion loop is also well constrained and does not change its structure during the activation and insertion process, which differs from that of class I fusion peptides which only adopt their final structure when interacting with the target membrane (Kielian and Rey, 2006; White et al., 2008). Trimer structure and membrane interaction analyses support the hypothesis (Fedry et al., 2017; Pinello et al., 2017) that HAP2 participates in a membrane fusion mechanism similar to that used by viruses which involves initiation through insertion of 'pre-hairpins' and eventual folding back of these 'hairpin' structures to pull the two opposing membranes closer together and enable fusion ((Apellaniz et al., 2014; Harrison, 2005; Harrison, 2015; Kielian and Rey, 2006; White et al., 2008); Figure 4).

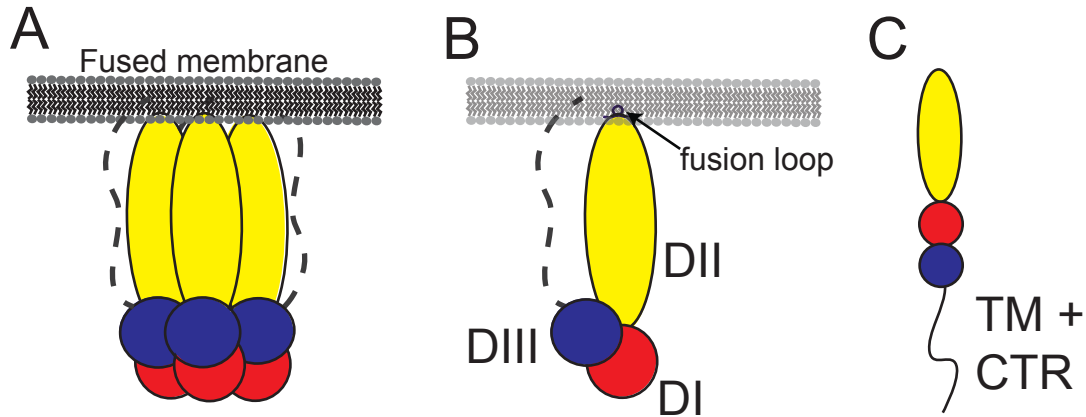


Figure 3. Structure-based schematic of HAP2. The structure of the post-fusion form of *Chlamydomonas* HAP2 (Fedry et al., 2017) provides the opportunity for informed schematics of HAP2, based on the models for other class II viral fusion proteins. A. Schematic representation of the HAP2 ectodomain structure. B. A schematic isolating a view of a monomer of class II fusion proteins to highlight the domains and the localization of the fusion loop in viral fusion structures. The dashed lines in A and B represent the stem region of the protein and the transmembrane domain which are not resolved in the structure. For *Chlamydomonas* HAP2, there was also no resolution of the predicted fusion loops at the tip of domain II. C. A structure-guided model of the domains of HAP2 as they may appear in an extended conformation, based upon class II fusion protein models, with the stem region, transmembrane domain (TM), and cytoplasmic region (CTR) now shown as a solid black line.

A somatic cell fusion protein, epidermal fusion factor-1 (EFF-1), that functions during development in *C. elegans* (Mohler et al., 2002; Shemer et al., 2004) is also structurally homologous to class II viral fusion proteins (Perez-Vargas et al., 2014). However, instead of a hydrophobic loop, EFF-1 has a short acidic patch in place of a hydrophobic fusion loop (Perez-Vargas et al., 2014). The proposed mechanism for EFF-1 is to oligomerize in *trans* across the two membranes rather than to insert a hairpin structure directly into the opposing membrane (Perez-Vargas et al., 2014). An alternate proposed hypothesis to the viral-fusion mechanism is that HAP2 may require a different class II-like protein on the female gamete (Valansi et al., 2017). A third hypothesis involves a cooperative model in which HAP2 functions unilaterally but fusion levels are enhanced when HAP2 is present on both membranes, potentially through interactions with itself in

trans, as hypothesized in *Tetrahymena* (Pinello et al., 2017), which may serve as a good model system to test cooperativity of HAP2.

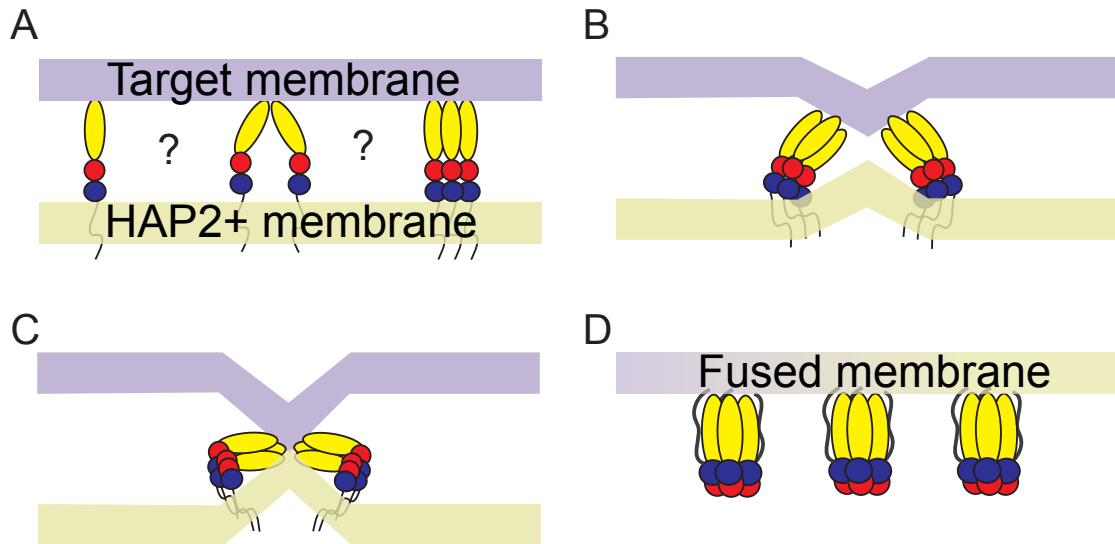


Figure 4. Model for HAP2 function as a viral-like fusion protein. This model is based upon the models of membrane fusion for viral proteins (Apellaniz et al., 2014; Harrison, 2005; Harrison, 2015; Kielian and Rey, 2006; Stiasny et al., 2011; White et al., 2008), with emphasis on the structural homology to class II viral fusion proteins (Fedry et al., 2017). A. The pre-fusion form of HAP2 is unknown, but models have presented the initial insertion of class II proteins in all three organizations shown (Apellaniz et al., 2014; Harrison, 2005; Harrison, 2015; Kielian and Rey, 2006; Stiasny et al., 2011; White et al., 2008). B. A step in the fusion process suggesting some distortion of plasma membrane shape or tension. C. Schematic representing the potential fold-back or pre-hairpin stage. D. Schematic representation of the post-fusion form of HAP2 as a trimer homologous to class II viral fusion proteins (Fedry et al., 2017).

II. Considerations for regulation of HAP2 activity in *Arabidopsis* fertilization.

Arabidopsis pollen tubes deliver sperm to the site of fusion within the ovule.

In double fertilization in flowering plants, the sperm are contained within the pollen tube, and the pollen tube delivers a pair of sperm, which are connected to each other (McCue et al., 2011), to the ovule ((Faure et al., 2002; Hamamura et al., 2011); Figure 2C-

D) in response to attractant proteins secreted from the synergid cells (Okuda et al., 2009; Takeuchi and Higashiyama, 2012). Pollen tube arrival is coordinated with degeneration of one of the synergid cells (Faure et al., 2002; Leydon et al., 2015). When the pollen tube bursts, it releases the sperm to the site of fusion, which is the area next to the two female gametes: the egg cell and the central cell ((Faure et al., 2002; Hamamura et al., 2011); Figure 1E-F). The sperm remain at the site of fusion for ~7-8 minutes before membrane fusion occurs (Hamamura et al., 2011). The area where the sperm reside is composed of the contents of the burst pollen tube and the degenerated synergid cell, creating a complex and poorly understood gamete fusion environment.

Hypothesis for regulation of HAP2 as a class II viral-like fusion protein.

Viral fusion proteins require a trigger to transition to their fusion competent state (White et al., 2008), and class II viral fusion proteins specifically require an acidic pH (White et al., 2008). A requirement for the decrease in pH has been demonstrated in liposome insertion assays (Acuna et al., 2015; DuBois et al., 2013; Gibbons et al., 2003; Stiasny et al., 2002). A conformational change in class II fusion proteins occurs at specific low pH levels, resulting in exposure of the fusion loop. Interestingly, *Chlamydomonas* HAP2 protein was able to insert into artificial membranes at a neutral pH (Fedry et al., 2017) and *Arabidopsis* HAP2 was able to bilaterally fuse mammalian cells without an acidic environment (Valansi et al., 2017), raising the question of whether HAP2 requires a pH trigger.

In *Arabidopsis*, EC1 proteins are proposed to activate the sperm after pollen tube burst, resulting in increased HAP2 at the sperm membrane (Sprunck et al., 2012). While this may partially explain why the two sperm do not fuse in the pollen tube, it is unclear how much HAP2 is already present at the sperm membrane as it travels through the pollen tube. The amount of HAP2 required at the sperm surface to mediate fusion is also

unknown. Furthermore, other modes of regulation, including a trigger for a potential HAP2 conformational change, may be occurring while the sperm are waiting at the site of fusion.

Concluding Remarks

HAP2 has been identified as the first, and only currently known, gamete fusion protein (Fedry et al., 2017; Valansi et al., 2017). The broad distribution of HAP2 provides the opportunity to study gamete fusion in a variety of species (Liu et al., 2008; Wong and Johnson, 2010), including agriculturally important plants and parasites that infect humans and animals. The opportunity now exists to test structure-based hypotheses regarding HAP2 function and regulation (Fedry et al., 2017). However, the membrane interaction surface of HAP2 needs to be defined. Since the membrane interaction region contains the proposed fusion loops and has the observed ability to interact with target membranes (Fedry et al., 2017), defining the structure of this region and its variability across species will be essential to our understanding of HAP2-mediated gamete fusion.

Perspectives: What is the mechanism of HAP2-mediated gamete fusion in Arabidopsis?

The goal of my thesis work has been to define how HAP2 functions in mediating gamete fusion in Arabidopsis. I have approached this in several ways. First, I have characterized the membrane interacting region of Arabidopsis HAP2 in a collaborative structure-function study described in our manuscript (Chapter 2). I have further investigated the functional conservation at the membrane interacting region within the flowering plant lineage (Chapters 2 and 5). I have investigated the endogenous and ectopic expression of HAP2 in gametophytic cells (Chapter 3). I have investigated the possibility of a pH trigger for Arabidopsis HAP2 by testing the pH environment surrounding the sperm both in the pollen tube and after sperm delivery to the female gametes (Chapter

4). Insights from my thesis work will continue to guide further experiments into HAP2 function and regulation in flowering plants.

References

- Acuna, R., Bignon, E.A., Mancini, R., Lozach, P.Y., and Tischler, N.D. (2015). Acidification triggers Andes hantavirus membrane fusion and rearrangement of Gc into a stable post-fusion homotrimer. *J Gen Virol* 96:3192-3197.
- Aguilar, P.S., Baylies, M.K., Fleissner, A., Helming, L., Inoue, N., Podbilewicz, B., Wang, H., and Wong, M. (2013). Genetic basis of cell-cell fusion mechanisms. *Trends Genet* 29:427-437.
- Angrisano, F., Sala, K.A., Da, D.F., Liu, Y., Pei, J., Grishin, N.V., Snell, W.J., and Blagborough, A.M. (2017). Targeting the Conserved Fusion Loop of HAP2 Inhibits the Transmission of *Plasmodium berghei* and *falciparum*. *Cell Rep* 21:2868-2878.
- Apellaniz, B., Huarte, N., Largo, E., and Nieva, J.L. (2014). The three lives of viral fusion peptides. *Chem Phys Lipids* 181:40-55.
- Aydin, H., Sultana, A., Li, S., Thavalingam, A., and Lee, J.E. (2016). Molecular architecture of the human sperm IZUMO1 and egg JUNO fertilization complex. *Nature* 534:562-565.
- Baldauf, S.L., Roger, A.J., Wenk-Siefert, I., and Doolittle, W.F. (2000). A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* 290:972-977.
- Bianchi, E., Doe, B., Goulding, D., and Wright, G.J. (2014). Juno is the egg Izumo receptor and is essential for mammalian fertilization. *Nature* 508:483-487.
- Blagborough, A.M., and Sinden, R.E. (2009). *Plasmodium berghei* HAP2 induces strong malaria transmission-blocking immunity in vivo and in vitro. *Vaccine*.
- Bressanelli, S., Stiasny, K., Allison, S.L., Stura, E.A., Duquerroy, S., Lescar, J., Heinz, F.X., and Rey, F.A. (2004). Structure of a flavivirus envelope glycoprotein in its low-pH-induced membrane fusion conformation. *The EMBO journal* 23:728-738.
- Camacho-Nuez, M., Hernandez-Silva, D.J., Castaneda-Ortiz, E.J., Paredes-Martinez, M.E., Rocha-Martinez, M.K., Alvarez-Sanchez, M.E., Mercado-Curiel, R.F., Aguilar-Tipacamu, G., and Mosqueda, J. (2017). Hap2, a novel gene in *Babesia bigemina* is expressed in tick stages, and specific antibodies block zygote formation. *Parasit Vectors* 10:568.
- Chalbi, M., Barraud-Lange, V., Ravaux, B., Howan, K., Rodriguez, N., Soule, P., Ndzoudi, A., Boucheix, C., Rubinstein, E., Wolf, J.P., et al. (2014). Binding of sperm protein Izumo1 and its egg receptor Juno drives Cd9 accumulation in the intercellular contact area prior to fusion during mammalian fertilization. *Development (Cambridge, England)* 141:3732-3739.
- Clark, N.L., Aagaard, J.E., and Swanson, W.J. (2006). Evolution of reproductive proteins from animals and plants. *Reproduction* 131:11-22.
- Cole, E.S., Cassidy-Hanley, D., Fricke Pinello, J., Zeng, H., Hsueh, M., Kolbin, D., Ozzello, C., Giddings, T., Jr., Winey, M., and Clark, T.G. (2014). Function of the male-gamete-specific fusion protein HAP2 in a seven-sexed ciliate. *Curr Biol* 24:2168-2173.
- Dresselhaus, T., Sprunck, S., and Wessel, G.M. (2016). Fertilization Mechanisms in Flowering Plants. *Curr Biol* 26:R125-139.
- Drews, G.N., and Yadegari, R. (2002). Development and function of the angiosperm female gametophyte. *Annual Review of Genetics* 36:99-124.
- DuBois, R.M., Vaney, M.C., Tortorici, M.A., Kurdi, R.A., Barba-Spaeth, G., Krey, T., and Rey, F.A. (2013). Functional and evolutionary insight from the crystal structure of rubella virus protein E1. *Nature* 493:552-556.
- Ebchuqin, E., Yokota, N., Yamada, L., Yasuoka, Y., Akasaka, M., Arakawa, M., Deguchi, R., Mori, T., and Sawada, H. (2014). Evidence for participation of GCS1 in fertilization of the starlet sea anemone *Nematostella vectensis*: implication of a

- common mechanism of sperm-egg fusion in plants and animals. *Biochem Biophys Res Commun* 451:522-528.
- Faure, J.E., Rotman, N., Fortune, P., and Dumas, C. (2002). Fertilization in *Arabidopsis thaliana* wild type: developmental stages and time course. *Plant J* 30:481-488.
- Fedry, J., Liu, Y., Pehau-Arnaudet, G., Pei, J., Li, W., Tortorici, M.A., Traincard, F., Meola, A., Bricogne, G., Grishin, N.V., et al. (2017). The Ancient Gamete Fusogen HAP2 Is a Eukaryotic Class II Fusion Protein. *Cell* 168:904-915 e910.
- Fedry*, J., Forcina*, J., Legrand, P., Pehau-Arnaudet, G., Haouz, A., Johnson, M.A., Rey, F.A., and Krey, T. (In preparation). HAP2 drives gamete fusion via insertion of highly divergent fusion loops into female gametes.
- Gibbons, D.L., Erk, I., Reilly, B., Navaza, J., Kielian, M., Rey, F.A., and Lepault, J. (2003). Visualization of the target-membrane-inserted fusion protein of Semliki Forest virus by combined electron microscopy and crystallography. *Cell* 114:573-583.
- Gibbons, D.L., Vaney, M.C., Roussel, A., Vigouroux, A., Reilly, B., Lepault, J., Kielian, M., and Rey, F.A. (2004). Conformational change and protein-protein interactions of the fusion protein of Semliki Forest virus. *Nature* 427:320-325.
- Guardado-Calvo, P., Atkovska, K., Jeffers, S.A., Grau, N., Backovic, M., Perez-Vargas, J., de Boer, S.M., Tortorici, M.A., Pehau-Arnaudet, G., Lepault, J., et al. (2017). A glycerophospholipid-specific pocket in the RVFV class II fusion protein drives target membrane insertion. *Science* 358:663-667.
- Hamamura, Y., Saito, C., Awai, C., Kurihara, D., Miyawaki, A., Nakagawa, T., Kanaoka, M.M., Sasaki, N., Nakano, A., Berger, F., et al. (2011). Live-Cell Imaging Reveals the Dynamics of Two Sperm Cells during Double Fertilization in *Arabidopsis thaliana*. *Curr Biol* 21:497-502.
- Harrison, S.C. (2005). Mechanism of membrane fusion by viral envelope proteins. *Adv Virus Res* 64:231-261.
- Harrison, S.C. (2015). Viral membrane fusion. *Virology* 479-480:498-507.
- Hirai, M., Arai, M., Mori, T., Miyagishima, S.Y., Kawai, S., Kita, K., Kuroiwa, T., Terenius, O., and Matsuoka, H. (2008). Male fertility of malaria parasites is determined by GCS1, a plant-type reproduction factor. *Curr Biol* 18:607-613.
- Huang, C.Y., Butrapet, S., Moss, K.J., Childers, T., Erb, S.M., Calvert, A.E., Silengo, S.J., Kinney, R.M., Blair, C.D., and Roehrig, J.T. (2010). The dengue virus type 2 envelope protein fusion peptide is essential for membrane fusion. *Virology* 396:305-315.
- Inoue, N. (2017). Novel insights into the molecular mechanism of sperm-egg fusion via IZUMO1. *J Plant Res* 130:475-478.
- Inoue, N., Hagihara, Y., Wright, D., Suzuki, T., and Wada, I. (2015). Oocyte-triggered dimerization of sperm IZUMO1 promotes sperm-egg fusion in mice. *Nat Commun* 6:8858.
- Jahn, R., and Scheller, R.H. (2006). SNAREs - engines for membrane fusion. *Nat Rev Mol Cell Biol* 7:631-643.
- Johnson, M.A., von Besser, K., Zhou, Q., Smith, E., Aux, G., Patton, D., Levin, J.Z., and Preuss, D. (2004). *Arabidopsis hapless* mutations define essential gametophytic functions. *Genetics* 168:971-982.
- Kawai-Toyooka, H., Mori, T., Hamaji, T., Suzuki, M., Olson, B.J., Uemura, T., Ueda, T., Nakano, A., Toyoda, A., Fujiyama, A., et al. (2014). Sex-specific posttranslational regulation of the gamete fusogen GCS1 in the isogamous volvocine alga *Gonium pectorale*. *Eukaryot Cell* 13:648-656.
- Kielian, M., and Rey, F.A. (2006). Virus membrane-fusion proteins: more than one way to make a hairpin. *Nat Rev Microbiol* 4:67-76.

- Klinovska, K., Sebkova, N., and Dvorakova-Hortova, K. (2014). Sperm-egg fusion: a molecular enigma of mammalian reproduction. *Int J Mol Sci* 15:10652-10668.
- Lehtonen, J., Kokko, H., and Parker, G.A. (2016). What do isogamous organisms teach us about sex and the two sexes? *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 371.
- Lescar, J., Roussel, A., Wien, M.W., Navaza, J., Fuller, S.D., Wengler, G., Wengler, G., and Rey, F.A. (2001). The Fusion glycoprotein shell of Semliki Forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH. *Cell* 105:137-148.
- Leydon, A.R., Tsukamoto, T., Dunatunga, D., Qin, Y., Johnson, M.A., and Palanivelu, R. (2015). Pollen Tube Discharge Completes the Process of Synergid Degeneration That Is Initiated by Pollen Tube-Synergid Interaction in Arabidopsis. *Plant Physiol* 169:485-496.
- Liu, Y., Tewari, R., Ning, J., Blagborough, A.M., Garbom, S., Pei, J., Grishin, N.V., Steele, R.E., Sinden, R.E., Snell, W.J., et al. (2008). The conserved plant sterility gene HAP2 functions after attachment of fusogenic membranes in Chlamydomonas and Plasmodium gametes. *Genes Dev* 22:1051-1068.
- McCue, A.D., Cresti, M., Feijo, J.A., and Slotkin, R.K. (2011). Cytoplasmic connection of sperm cells to the pollen vegetative cell nucleus: potential roles of the male germ unit revisited. *J Exp Bot* 62:1621-1631.
- Misamore, M.J., Gupta, S., and Snell, W.J. (2003). The Chlamydomonas Fus1 protein is present on the mating type plus fusion organelle and required for a critical membrane adhesion event during fusion with minus gametes. *Mol Biol Cell* 14:2530-2542.
- Modis, Y., Ogata, S., Clements, D., and Harrison, S.C. (2004). Structure of the dengue virus envelope protein after membrane fusion. *Nature* 427:313-319.
- Mohler, W.A., Shemer, G., del Campo, J.J., Valansi, C., Opoku-Serebuoh, E., Scranton, V., Assaf, N., White, J.G., and Podbilewicz, B. (2002). The type I membrane protein EFF-1 is essential for developmental cell fusion. *Developmental cell* 2:355-362.
- Mori, T., Igawa, T., Tamiya, G., Miyagishima, S.Y., and Berger, F. (2014). Gamete attachment requires GEX2 for successful fertilization in Arabidopsis. *Curr Biol* 24:170-175.
- Mori, T., Kuroiwa, H., Higashiyama, T., and Kuroiwa, T. (2006). GENERATIVE CELL SPECIFIC 1 is essential for angiosperm fertilization. *Nat Cell Biol* 8:64-71.
- Nishimura, K., Han, L., Bianchi, E., Wright, G.J., de Sanctis, D., and Jovine, L. (2016). The structure of sperm Izumo1 reveals unexpected similarities with Plasmodium invasion proteins. *Curr Biol* 26:R661-662.
- Okamoto, M., Yamada, L., Fujisaki, Y., Bloomfield, G., Yoshida, K., Kuwayama, H., Sawada, H., Mori, T., and Urushihara, H. (2016). Two HAP2-GCS1 homologs responsible for gamete interactions in the cellular slime mold with multiple mating types: Implication for common mechanisms of sexual reproduction shared by plants and protozoa and for male-female differentiation. *Dev Biol* 415:6-13.
- Okuda, S., Tsutsui, H., Shiina, K., Sprunck, S., Takeuchi, H., Yui, R., Kasahara, R.D., Hamamura, Y., Mizukami, A., Susaki, D., et al. (2009). Defensin-like polypeptide LUREs are pollen tube attractants secreted from synergid cells. *Nature* 458:357-361.
- Oren-Suissa, M., and Podbilewicz, B. (2007). Cell fusion during development. *Trends Cell Biol* 17:537-546.

- Pan, J., Misamore, M.J., Wang, Q., and Snell, W.J. (2003). Protein transport and signal transduction during fertilization in *Chlamydomonas*. *Traffic* (Copenhagen, Denmark) 4:452-459.
- Perez-Vargas, J., Krey, T., Valansi, C., Avinoam, O., Haouz, A., Jamin, M., Raveh-Barak, H., Podbilewicz, B., and Rey, F.A. (2014). Structural basis of eukaryotic cell-cell fusion. *Cell* 157:407-419.
- Pinello, J.F., Lai, A.L., Millet, J.K., Cassidy-Hanley, D., Freed, J.H., and Clark, T.G. (2017). Structure-Function Studies Link Class II Viral Fusogens with the Ancestral Gamete Fusion Protein HAP2. *Curr Biol* 27:651-660.
- Rey, F.A., Heinz, F.X., Mandl, C., Kunz, C., and Harrison, S.C. (1995). The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution. *Nature* 375:291-298.
- Shemer, G., Suissa, M., Kolotuev, I., Nguyen, K.C., Hall, D.H., and Podbilewicz, B. (2004). EFF-1 is sufficient to initiate and execute tissue-specific cell fusion in *C. elegans*. *Curr Biol* 14:1587-1591.
- Sprunck, S., Rademacher, S., Vogler, F., Gheyselinck, J., Grossniklaus, U., and Dresselhaus, T. (2012). Egg cell-secreted EC1 triggers sperm cell activation during double fertilization. *Science* 338:1093-1097.
- Steele, R.E., and Dana, C.E. (2009). Evolutionary history of the HAP2/GCS1 gene and sexual reproduction in metazoans. *PloS one* 4:e7680.
- Stiasny, K., Allison, S.L., Schlich, J., and Heinz, F.X. (2002). Membrane interactions of the tick-borne encephalitis virus fusion protein E at low pH. *Journal of virology* 76:3784-3790.
- Stiasny, K., Fritz, R., Pangerl, K., and Heinz, F.X. (2011). Molecular mechanisms of flavivirus membrane fusion. *Amino Acids* 41:1159-1163.
- Takeuchi, H., and Higashiyama, T. (2012). A species-specific cluster of defensin-like genes encodes diffusible pollen tube attractants in *Arabidopsis*. *PLoS Biol* 10:e1001449.
- Valansi, C., Moi, D., Leikina, E., Matveev, E., Grana, M., Chernomordik, L.V., Romero, H., Aguilar, P.S., and Podbilewicz, B. (2017). *Arabidopsis* HAP2/GCS1 is a gamete fusion protein homologous to somatic and viral fusogens. *J Cell Biol* 216:571-581.
- von Besser, K., Frank, A.C., Johnson, M.A., and Preuss, D. (2006). *Arabidopsis* HAP2(GCS1) is a sperm-specific gene required for pollen tube guidance and fertilization. *Development* (Cambridge, England) 133:4761-4769.
- White, J.M., Delos, S.E., Brecher, M., and Schornberg, K. (2008). Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme. *Crit Rev Biochem Mol Biol* 43:189-219.
- Wong, J.L., and Johnson, M.A. (2010). Is HAP2-GCS1 an ancestral gamete fusogen? *Trends Cell Biol* 20:134-141.
- Wright, G.J., and Bianchi, E. (2016). The challenges involved in elucidating the molecular basis of sperm-egg recognition in mammals and approaches to overcome them. *Cell Tissue Res* 363:227-235.

Chapter 2: HAP2 drives gamete fusion via insertion of highly divergent fusion loops into female gametes

Juliette Fedry^{1,2,3,4,10}, Jennifer Forcina^{5,10}, Pierre Legrand⁶, Gérard Péhau-Arnaudet⁷, Ahmed Haouz⁸, Mark Johnson^{5*}, Felix A. Rey^{1,2,11*}, Thomas Krey^{1,2,4,9*}

¹Unité de Virologie Structurale, Institut Pasteur, 25-28 Rue du Docteur Roux, 75724 Paris, France; ²CNRS UMR 3569, 25-28 Rue du Docteur Roux, 75724 Paris, France ; ³Université Paris Descartes Sorbonne Paris Cité, Institut Pasteur, Rue du Docteur Roux, 75015 Paris, France; ⁴Institute of Virology, Hannover Medical School, 30625 Hannover, Germany; ⁵Brown University, Department of Molecular Biology, Cell Biology, and Biochemistry, 60 Olive St, Providence, RI, USA 02912; ⁶Synchrotron SOLEIL, L'Orme des Merisiers, 91192 Gif-sur-Yvette, France; ⁷Ultrapole, CNRS UMR 3528, Institut Pasteur, 25-28 rue du Docteur Roux, 75724 Paris, France; ⁸Protéopôle, CNRS UMR 3528, Institut Pasteur, 25-28 rue du Docteur Roux, 75724 Paris, France; ⁹German Center for Infection Research (DZIF), 30625 Hannover, Germany

¹⁰Co-first author

¹¹Lead Contact

*Corresponding author

This manuscript was previously submitted for publication consideration. This manuscript is currently under revision for submission to a different journal. I helped draft and revise this manuscript. I contributed to the analysis in Figure 1. I did the analysis for Figure 5 and generated all of Figure 5, except 5E, with help from Dr. Mark Johnson and input from the other authors. I also contributed all of Figure S6 and generated Figure S1.

Summary

Gamete fusion is central to the reproduction of most eukaryotic species, yet the underlying mechanisms remain elusive. We recently showed that the sperm-restricted transmembrane protein HAP2 is a eukaryotic class II fusion protein, although the reported crystal structure did not resolve its membrane-interaction region. Here, we report a striking diversity in this critical region and provide insight into the mode of target membrane insertion using combined structural, biochemical, genetic and bioinformatic analyses. We reveal a compact loop-helix-loop motif in *T. cruzi* and an unprecedented amphipathic helix in *A. thaliana* HAP2. The diversity of HAP2 membrane interacting motifs observed so far is likely representative of adaptation to myriad eukaryotic sexual systems. We show that HAP2 uses a virus-like unilateral mechanism, requiring the insertion of non-polar residues into the opposing membrane, and bring new insight into its evolutionary path, providing a structural framework for identification of HAP2 orthologs in additional eukaryotic phyla.

Introduction

Fusion of the plasma membranes of two opposing gametes is the turning point in the life cycle of sexual organisms, bringing unique parental nuclei together into a common zygotic cytoplasm. Despite the central importance of this event, very little is known about the mechanisms responsible for merging gamete plasma membranes. The membrane fusion reaction is an exothermic process, but the required dehydration of the lipid head groups to allow direct contact between two lipid bilayers imposes a high kinetic barrier (Leikin et al., 1993). In other contexts - such as the fusion of enveloped viruses with host cells for entry, fusion of intracellular vesicles, or fusion of somatic cells - specialized effector proteins have been identified that catalyze the membrane fusion reaction by overcoming the kinetic barrier. Such proteins can function in two distinct manners. Viral fusion proteins as well as syncytins - which are endogenous-retrovirus-derived cellular fusion proteins (Blaise et al, 2003; Mi et al, 2000) - are initially anchored in only one of the two membranes to be fused and bridge the two membranes by insertion of a non-polar segment of the polypeptide chain (the “fusion loop”) into the target membrane (Harrison, 2015). In contrast, SNAREs (Südhof and Rothman, 2009), atlastins (Bian et al., 2011) and EFF-1 (Podbilewicz et al., 2006) have subunits anchored in both membranes and function by *trans*-oligomerization. In general, the fusion proteins deliver the mechanical force required for membrane merger through an exothermic conformational change following a specific trigger (White et al., 2008). Until very recently, there was no fusion effector protein identified for gamete merger during fertilization.

HAP2-GCS1 (HAP2 hereafter; Johnson et al., 2004; Mori et al., 2006), has been identified by limited amino acid sequence similarity (occurrence of pfam10699, Figure 1A) in genomes belonging to the major eukaryotic taxa except fungi (Liu et al., 2008). Although HAP2 orthologs have been found widely in plants and protists, positive detection in animals is scattered: current bioinformatic analyses have identified orthologs in

cnidarians, arthropods, annelids, and molluscs, but not in some well-studied clades like nematodes, echinoderms, and vertebrates (Figure S1). HAP2 is a single-pass transmembrane protein expressed only in gametes in all species tested (Cole et al., 2014; Ebchuqin et al., 2014; Liu et al., 2008; Okamoto et al., 2016; Steele and Dana, 2009; von Besser et al., 2006). Genetic analyses have demonstrated that it is necessary for gamete fusion in the flowering plant *Arabidopsis thaliana* (Mori et al., 2006; von Besser et al., 2006), the green alga *Chlamydomonas reinhardtii* (Liu et al., 2008), the malaria parasite *Plasmodium berghei* (Hirai et al., 2008; Liu et al., 2008), the ciliate *Tetrahymena thermophila* (Cole et al., 2014), and the slime mold *Dictyostelium discoideum* (Okamoto et al., 2016). Recently, the crystal structure of the HAP2 ectodomain from *C. reinhardtii* (CrHAP2; Fedry et al., 2017) revealed that HAP2 is homologous to the class II fusion proteins, which include viral (e.g., the fusion proteins of flaviviruses, alphaviruses, rubella virus, and several bunyaviruses (Guardado-Calvo and Rey, 2017; Modis, 2014)) and somatic cell fusion family (FF) proteins such as AFF-1 and EFF-1 (Avinoam et al., 2011; Mohler et al., 2002). Together, these data strongly support the notion that HAP2 is an ancestral gamete fusion protein.

Class II fusion proteins fold into an articulated rod made of three β -sheet-rich domains, with a central domain I flanked by domains II and III (see linear diagram in Figure 1A) (Dessau and Modis, 2013; Lescar et al., 2001; Rey et al., 1995). Domain III connects to the C-terminal transmembrane anchor via a flexible linker, while domain II bears the functional fusion loops at the opposite end. The tip of domain II is made of a core three-stranded antiparallel β -sheet (the *bdc* sheet) with intervening *bc* and *cd* inter-strand connections projecting at the distal end, which in the viral proteins inserts into the target membrane to initiate membrane fusion (Bressanelli et al., 2004; DuBois et al., 2013; Gibbons et al., 2004; Guardado-Calvo et al., 2016; Halldorsson et al., 2016; Modis et al., 2004). The low pH-triggered fusogenic conformational change of viral class II fusion

proteins leads first to an extended, trimeric intermediate conformation bridging the two membranes, with the target membrane-inserted domain II tip and the viral-membrane-anchored C-terminus at opposite ends of the protein (Liao et al., 2010). In contrast, the domain II tip in the somatic FF proteins lacks a hydrophobic fusion loop, does not insert into the target membrane, and the protein was proposed to undergo trimerization *in trans* to bridge the two membranes (Pérez-Vargas et al., 2014). In both viral and FF proteins, each subunit of the transient extended intermediate trimer then bends in half, projecting domain III toward the domain II tip. The resulting “hairpin” conformation draws the two membranes together by pulling the TM anchors to the trimer tip region in the post-fusion conformation. In the case of the FF proteins, as the TM segments of the trans-cellular trimer are anchored in the two opposing cells, the conformational change does not require the tip of domain II to insert into the opposite membrane. In spite of their clear structural homology, the members of the FF family of somatic fusion proteins have thus evolved a distinct mechanism of action from viral class II fusion proteins, as reflected by their requirement to be present at the surface of both cells to drive fusion (Podbilewicz et al., 2006).

It is currently not clear whether HAP2 functions by direct insertion into target membranes. A peptide corresponding to the predicted membrane-interacting segment connecting β -strands *c* and *d* (the “*cd* connection”) of *T. thermophila* HAP2 was shown to interact with lipid bilayers, as expected for a “fusion loop” (Pinello et al., 2017). The recombinant CrHAP2 ectodomain was also shown to bind liposomes in assays similar to those used to establish that class II viral fusion proteins function via host membrane insertion (Gibbons et al., 2003; Stiasny et al., 2002). But contradicting results were also presented: the heterologous expression of *Arabidopsis thaliana* HAP2 (AtHAP2) in mammalian cells was recently reported to induce cell- cell fusion only when HAP2 was expressed in both cells to be fused (Valansi et al., 2017). This study also indicated that

cell fusion could also take place if EFF-1 was expressed on the opposing cell, suggesting a fusion mechanism involving HAP2 homo- or hetero-oligomerization with EFF-1 *in trans* (Valansi et al., 2017), therefore raising mechanistic questions about the HAP2-driven membrane fusion process. Furthermore, although the functional studies indicated that the CrHAP2 soluble ectodomain inserted via the tip of domain II into liposomes, this region was disordered in the reported CrHAP2 crystal structure (Fedry et al., 2017), so the conformational and mechanistic significance of this putative critical membrane interacting region was not addressed. Together, the above results raised the possibility that HAP2 from different organisms may use different mechanisms.

Here, we observed that the amino acid sequence at the tip of HAP2 domain II – corresponding to the fusion loops in viral class II proteins - is highly variable across eukaryotic species, with the alignment showing a distinctive pattern of insertions and deletions interspersed with strictly conserved core structural elements. Two groups of HAP2 sequences exhibited the most contrasting patterns in the *cd* connection: the flowering plants and the kinetoplastid protists. We found that the recombinant ectodomains of AtHAP2 (AtHAP2e) and *T. cruzi* HAP2 (TcHAP2e), representative of these two groups, indeed insert into membranes as trimers via the tip of domain II as shown previously for the viral proteins and for CrHAP2. The crystal structures of AtHAP2e and TcHAP2 domain II revealed a striking diversity of structural elements that project from a conserved core *bdc* β -sheet toward the HAP2 apex. The kinetoplastid TcHAP2 had a compact domain II tip stabilized by an extra disulfide bond in the *cd* connection, while the flowering plant AtHAP2 displayed an unprecedented amphipathic fusion helix in this segment, poised to insert flat into the target membrane. Further functional studies demonstrated that the fusion helix is critical for liposome insertion of the soluble AtHAP2e *in vitro*, and that mutations in the fusion helix that abolish membrane insertion *in vitro* also block fertilization *in vivo*. This correlation provides strong support for a model in which

HAP2 functions like a viral fusion protein, via direct insertion into the target membrane to induce gamete fusion.

Results and Discussion

Distinct patterns of insertions and deletions in the tip of HAP2 domain II.

The antiparallel *bdc* β -sheet at the domain II tip is organized such that β -strands *b* and *c* point toward the target membrane at either side of the central β -strand *d*, which points away (Figure 1C). The alpha- and flavivirus fusion proteins were shown to insert apolar residues from the *cd* connection into target membranes for fusion (reviewed in Kielian and Rey, 2006). More recently, the bunyavirus class II fusion proteins were shown to insert apolar residues from both, the *cd* and *bc* connections (Guardado-Calvo et al., 2016; Halldorsson et al., 2016). In the reported *CrHAP2* X-ray structure, the polypeptide segment of the domain II tip that is positioned to most prominently project toward the target membrane is the *cd* connection, but most of it was not resolved in the crystals. Therefore, we focused here our analysis on this region of HAP2. We used a log-expectation algorithm (Edgar, 2004) to align representative HAP2 amino acid sequences from the diversity of eukaryotic phyla (Table S1). We identified seven clearly conserved structural anchors within the domain II tip: the three β -strands of the core *bdc* β -sheet, three cysteine residues in the connecting loops involved in three conserved disulfide bonds, and an invariant arginine in the *cd* connection known to be essential for *CrHAP2* function (Fedry et al., 2017). Remarkably, the sequences between these anchors were highly variable and defined six insertion/deletion hotspots in the *bc* and *cd* inter-strand connections (Figure 1B).

The first three identified HAP2 hotspots span the *bc* connection (Figures 1B and C). Hotspot 1 is immediately downstream of β -strand *b* and has the potential to contact the target membrane via its N-terminal, *b* strand-proximal side (Figure 1C). Residues from insertions in hotspot 1 (Figure 1B and D) - like the 19 inserted residues observed in *Toxoplasma gondii* HAP2 (NCBI accession KYF43852.1), which include aromatic and other non-polar residues - are therefore likely to play a role in target membrane insertion. Residues from hotspots 2 and 3, further downstream in the *bc* connection, are less likely to interact with the membrane, except perhaps in HAP2 orthologs like the one of the alga *Volvox carteri* (NCBI accession EFJ46133.1), which has a 17 residue insertion in hotspot 2 (Figure 1B) that could reach the membrane depending on the three-dimensional organization of the *bc* connection. The remaining three hotspots are in the *cd* connection; the least variable (hotspot 5) corresponds to the short $\alpha 0$ helix separating the two proposed fusion loops of *CrHAP2* (Fedry et al., 2017; Figure 1B) and is anchored by the conserved arginine and the second cysteine of disulfide 5 (Figure 1B). It varies in length from four to seven residues (Figure 1D) and a number of orthologs feature α -helix breaking residues (glycine or proline), suggesting that the *CrHAP2* $\alpha 0$ helix is not conserved. Hotspots 4 and 6 are in the region corresponding to the putative *CrHAP2* fusion loops 1 and 2, respectively (Figure 1B). Both are quite variable in length; hotspot 4 has a range of 80 residues: from a 14 amino acid deletion in the flowering plant *Oryza sativa* (rice) to a 66 amino acid insertion in the animal *Tribolium castaneum* (flour beetle). The kinetoplastid protists (e.g., *T. cruzi*) have an extra pair of cysteines in this region, predicted to form an additional disulfide bond. The region of HAP2 predicted to interact with the target membrane has therefore diverged considerably during the course of eukaryotic evolution.

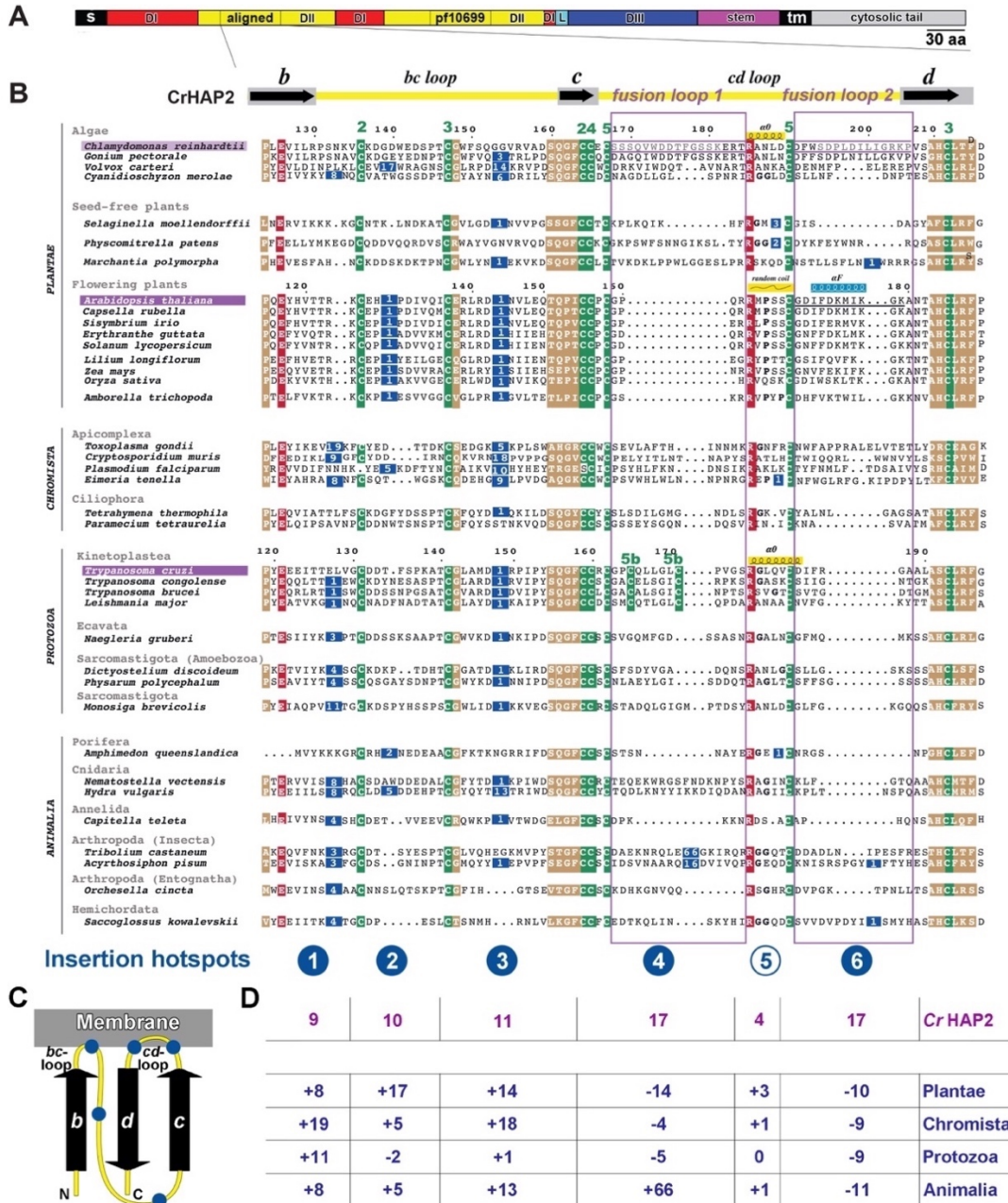


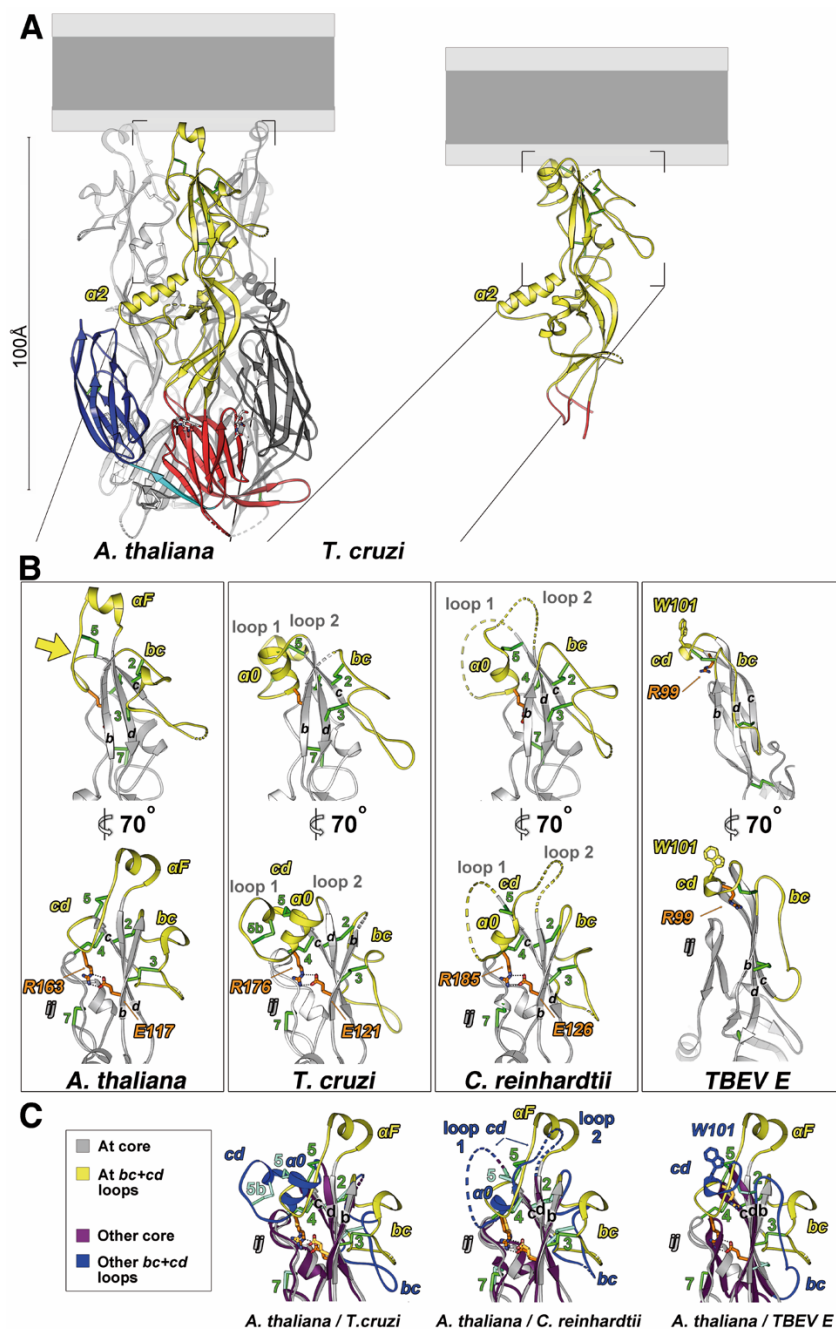
Figure 1. The HAP2 membrane interacting region has diverged across eukaryotic lineages. **A.** Linear diagram of the 705 amino acid *AtHAP2* sequence indicating the positions of the signal peptide (S), the polypeptide segments corresponding to domains I, II, and III (DI-III), the segment corresponding to the tip of domain II (the *bdc* β -sheet, aligned in **B**), the pfam HAP2-GCS1 domain (pf10699), the linker (L) between DI and DIII, the stem that connects the ectodomain to the transmembrane domain (tm), and the cytosolic tail. **B.** The *bdc* β -sheet region of a HAP2 amino acid sequence alignment obtained with MUSCLE (Edgar, 2004) using representative members from four out of the five eukaryotic kingdoms (named at left; Ruggiero et al., 2015); the secondary structure elements identified in the CrHAP2 crystal structure (Fedry et al., 2017) are indicated above the alignment as black arrows in grey boxes and the intervening loops in yellow (*bc* and

cd) (Fedry et al., 2017). The positions of putative CrHAP2 fusion loops 1 and 2 are indicated with purple text and boxes. The CrHAP2 $\alpha 0$ helix, separating loops 1 and 2 is indicated with a helical symbol immediately above the sequence. Highlighted in dark purple are *A. thaliana* and *T. cruzi* sequences, studied further in this work, and in pink the *C. reinhardtii* sequence. Conserved and semi-conserved residues are in white font on a red or beige background, respectively. A green background on the alignment highlights cysteines involved in disulfide bonds, most of which strictly conserved, numbered above the alignment according to the CrHAP2 structure. For clarity, insertions with respect to CrHAP2 are omitted but their length is indicated on a dark blue background; the major insertion hotspots in the membrane interaction surface are numbered below the alignment in blue circles. Potential helix-breaking residues (glycine or proline) in the *C. reinhardtii* $\alpha 0$ helix region of other species are indicated in bold characters. A random coil above the *A. thaliana* sequence highlights the absence of the $\alpha 0$ helix and a helical symbol on a cyan background indicates the position of the *A. thaliana* fusion helix αF (Figures 2 and 3). A black underline marks the *A. thaliana* segment deleted in the *At* $\Delta\alpha F$ construct used in genetic assays (Figure 5). **C.** Schematic diagram of the *bdc* β -sheet and the connecting loops at the tip of domain II relative to the target membrane. Blue circles indicate the position of insertion hotspots. **D.** The top line lists the amino acid lengths in between conserved framing residues in CrHAP2. Subsequent lines list the lengths of the longest amino acid insertions (+) or deletions (-) in the corresponding hotspot of the main phyla at each insertion hotspot relative to CrHAP2. See also Figure S1 and Table S1.

The core structure of domain II is conserved across eukaryotes.

HAP2 sequences from the flowering plants and from the kinetoplastids were among those displaying the most contrasting insertion/deletion patterns in the *cd* connection (hotspots 4 and 6, Figure 1B). We thus produced the recombinant ectodomains of the flowering plant *A. thaliana* HAP2 (*At*HAP2e, aa 25-494) and of the kinetoplastid *T. cruzi* (*Tc*HAP2e, aa 26-516) in *Drosophila* S2 cells and purified them from the cell supernatant (see Methods). Size exclusion chromatography (SEC) and multi-angle static light scattering (MALS) revealed that both proteins behaved as monomers in solution (Figure S2). *At*HAP2e crystallized in the $P6_3$ hexagonal space-group and diffracted anisotropically to a mean nominal resolution of 2.75Å. *Tc*HAP2e only crystallized upon limited proteolysis using subtilisin (see Methods; *Tc*HAP2e). The resulting purified protease-resistant 40 kD fragment (termed *Tc*HAP2e_{sub}) crystallized in the hexagonal space group $P6_122$, and the best crystals diffracted to 3.1Å resolution (see Methods and

Figure S2; crystallographic statistics are listed in Table S2). We determined both structures by the molecular replacement method using a search model derived from the structure of CrHAP2 (PDB 5MF1).



of the fusion loops with respect to a membrane upon insertion. The *At*HAP2 subunit in the foreground and the *Tc*HAP2 protomer are colored according to the class II convention: red, yellow, and blue for domains I, II, and III, respectively; cyan for the domain I-III linker. Disordered regions are shown as dashed segments, disulfide bonds as green sticks. **B.** Close-up view of the tip region of *At*HAP2, *Tc*HAP2 and *Cr*HAP2 (PDB 5MF1; (Fedry et al., 2017)) and TBEV E protein (PDB 1URZ; (Bressanelli et al., 2004)). The conserved core *bdc* β -sheet is shown in grey and the connecting (variable) loops in yellow. Side chains of residues forming the conserved salt bridge are shown in orange, hydrogen bonds as black dashed lines and disulfides in green sticks and numbered. For comparison, in TBEV E a conserved arginine of the fusion loop is represented in the same way. It does not form a salt bridge but underpins the fusion loop by making multiple polar interactions with the main chain. A yellow arrow in the left panel points to the *At*HAP2 region corresponding to helix $\alpha 0$ in other HAP2 orthologs. **C.** Pairwise superposition of the *At*HAP2 tip region (maintaining the color scheme of B) with *Tc*HAP2, *Cr*HAP2 and TBEV E, (with *bdc* β -sheet purple and connecting loops in blue) highlighting the structurally conserved domain II scaffold. See also Figure S2 and Table S2.

The experimental electron density map of *At*HAP2 allowed the tracing of 446 out of 469 ectodomain residues (25-493; the N-terminus is generated by signalase cleavage between residues 24 and 25). The resulting atomic model of *At*HAP2 revealed a trimer in unequivocal post-fusion hairpin conformation (Figures 2A and S3), as observed previously for *Cr*HAP2 (Fedry et al., 2017). The crystals of *Tc*HAP2_{e_{sub}} consisted of domain II with a short extension into domain I (Figure 2A). In contrast to the crystals of *Cr*HAP2 reported previously, the electron density for the *cd* connection was clear in both the *At*HAP2 trimer and the *Tc*HAP2_{e_{sub}} crystals (Figure S2H), allowing us to unambiguously trace the polypeptide chain in this region (Figure 2A and 2B). Importantly, the *Tc*HAP2_{e_{sub}} domain II monomer displayed the same conformation as domain II from both *At* and *Cr*HAP2e trimers. This observation is in line with the absence of conformational changes within domain II during the transition from pre- to post-fusion forms in several viral class II fusion proteins (Bressanelli et al., 2004; Gibbons et al., 2004; Halldorsson et al., 2016; Modis et al., 2004). We were thus able to model the *Tc*HAP2 trimer by superposing three *Tc*HAP2 domains II on either the *Cr*HAP2 or *At*HAP2 trimer (Figure S3). The $\alpha 2$ helix, which contributes in part to pfam10699 - the 50 aa stretch used to identify most of the HAP2 sequences annotated thus far (Figure S1) - is nearly identical across HAP2 structures

(Figure 2A, Figure S3). This helix is a characteristic HAP2 motif, as in the viral proteins it corresponds to a very short α -helix (two turns at most), which is presented in varying orientations in the available X-ray structures (Figure S3).

Superposition of the core *bdc* β -sheets of the HAP2 structures revealed root mean square deviations (rmsd) $<0.73\text{\AA}$ for pairwise comparisons of 19 C α atoms, highlighting the conservation of the structural core of the domain II tip (Figures 2B and C). Superposition of the *bdc* β -sheets of *At*HAP2 and the flavivirus class II protein from the tick-borne encephalitis virus (TBEV) resulted in an rmsd $<1.7\text{\AA}$ for 19 C α atoms. The E¹²⁶-R¹⁸⁵ salt bridge observed in *Cr*HAP2 (Fedry et al., 2017; E¹¹⁷-R¹⁶³ in *At*HAP2 and E¹²¹-R¹⁷⁶ in *Tc*HAP2) is another strikingly conserved HAP2-specific feature, with the side chains falling on top of each other when the superposition is based on the *bdc* β -sheet (Figures 2B and C). Furthermore, alignment of the *bdc* β -sheets also brings into superposition the *ij* loop (which is included in the pfam10699 segment). The strong structural conservation of the salt bridge and its interactions with the *ij* loop across HAP2 diversity thus supports its proposed role in orienting the fusion loops for proper membrane insertion, and further validates the use of the salt bridge residues as anchors for the sequence alignment (*Cr*HAP2 R¹⁸⁵ and E¹²⁶ in the strand *b* anchor, Figure 1B).

An evolutionarily plastic domain II tip, exposing an amphipathic helix in *Arabidopsis*.

The above comparison indicates that the core structural features of HAP2 domain II underpinning the membrane interaction surface are conserved across eukaryotes. In contrast, the segments between strands *b* and *c* and between *c* and *d* could not be aligned and differed markedly at the structural level (Figures 1, 2B and 2C). The *Tc*HAP2 *cd* and *bc* connections are located roughly at the same level, poised to interact with the target

membrane, indicating that they both comprise the membrane insertion surface (Figure 2B, second panel). A surface representation of the modeled *TcHAP2* trimer suggests a relatively flat membrane interaction surface with hydrophobic patches in the *bc* (V¹²⁹) and the *cd* inter-strand connection (L¹⁶⁷/L¹⁶⁸ and I¹⁸³/F¹⁸⁴, Figure S4A). As in *CrHAP2*, the *cd* connection features the $\alpha 0$ helix with the conserved arginine at its N-terminal end, but which continues with one additional helical turn after disulfide 5 (Fig. 2B, compare second and third panels). Loop 1, upstream $\alpha 0$, is stabilized by an extra disulfide bond, (numbered 5b in Figures 1B and 2B), as anticipated from the amino acid sequence alignment (Figure 1B). Loop 2, downstream, is very short: a single turn connecting $\alpha 0$ to the *d* strand.

In contrast to *TcHAP2*, the *AtHAP2* *cd* connection lacks the central $\alpha 0$ helix between the conserved arginine and disulfide bond 5. There is only a short loop from β -strand *c* to the conserved arginine, and after disulfide 5 the polypeptide chain projects apically toward the membrane, featuring an amphipathic helix (αF , Figure 2) positioned to interact flat against a target membrane by inserting its non-polar surface (I¹⁷¹, F¹⁷², M¹⁷⁵, I¹⁷⁶; Figures 3 and 4). The *bc* inter-strand connection appears too far from the membrane to be part of the insertion motif (compare first and second panels in Fig. 2B). We found that similar amphipathic helices are predicted in the amino acid sequence of the corresponding segment (hotspot 6, Figure 1B) of all plant and algae HAP2 sequences, including *CrHAP2* (Figure S5, Heliquet, Gautier et al., 2008). This finding provides new insight into the mode of *CrHAP2* membrane interaction, as this region was disordered in the previously reported structure (Fedry et al., 2017). Our bioinformatic analyses also predicted an amphipathic helix in the hotspot 4 sequence of the cnidarian animal *Nematostella vectensis* (Figure 1B), suggesting that this mode of target membrane interaction may be more widespread among eukaryotes.

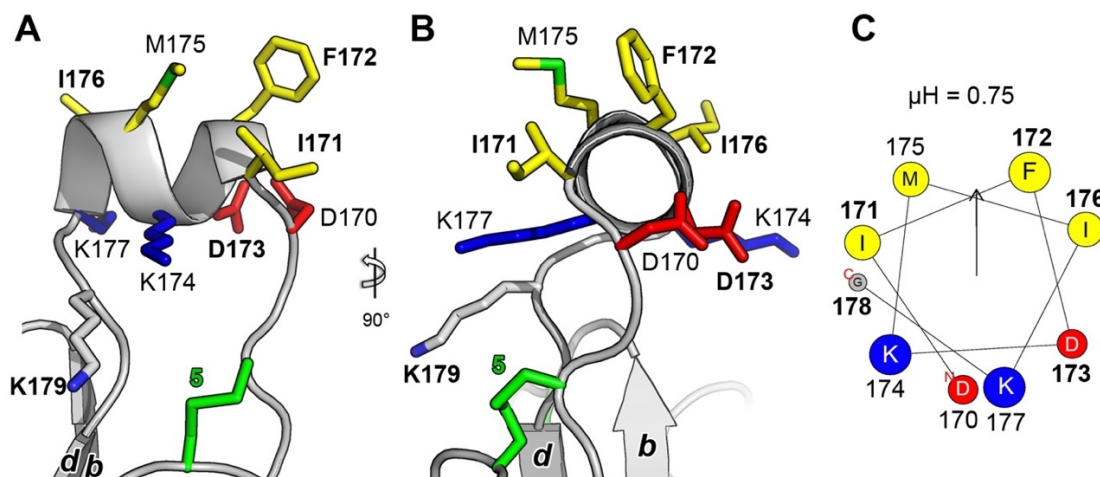


Figure 3. The *A. thaliana* HAP2 amphipathic fusion helix α F

A-B. Grey ribbon representation of the *At*HAP2 fusion helix α F at the tip of domain II. Side chains of apolar (yellow), negatively charged (red) and positively charged (blue) residues are represented as sticks, residues interrogated in subsequent experiments are labeled in bold. Green sticks indicate disulfide bonds, numbered as in Figure 1. α F is viewed laterally (**A**) and along the helical axis (**B**). **C.** Helical wheel projection of α F; circle size is proportional to side chain size (amino acids colored as in A and B). The hydrophobic moment of the amphipathic α helix (μ H) predicted by Heliquist (Gautier et al., 2008) is displayed as a vertical arrow and its value indicated above the diagram. See also Figure S5.

The HAP2 ectodomains insert into liposomes as trimers via the domain II tip.

We analyzed insertion into membranes of the *At*HAP2 ectodomain by mixing the monomeric protein (purified by size exclusion chromatography, Figure S2B) with liposomes of the composition defined in methods and analyzed liposome binding by ultracentrifugation in density gradients. The gradients were designed so that liposomes and associated proteins would float to the top fractions, while the free protein would sediment to the bottom. The soluble, wild type *At*HAP2e was indeed found in the top fractions, in contrast to mutants in the *cd* connection (Figure 4A), indicating that it binds to membranes and that the *cd* connection is involved in the interaction. Electron microscopy further showed that both *At*HAP2e (Figure 4B-D) and *Tc*HAP2e (Figure S4B-C) decorated the liposomes as projecting rods. Side views of the proteoliposome edge showed ~12nm long projections with a tapered end toward the membrane (Figure 4B and D, Figure S4B-

C), and top views showed a typical pattern of hexagonal packing of post-fusion trimeric ectodomains at the liposome surface (Figure 4C), consistent with the overall shape, orientation, and lateral packing of trimers observed earlier with the alphavirus and flavivirus class II fusion proteins (Gibbons et al., 2003; Stiasny et al., 2002). Importantly, in the experiments with the viral proteins, binding to liposomes was triggered by exposure to acid pH, which also led to concomitant trimerization of the protein into the post-fusion form. The size and shape of the projections observed on the HAP2e proteoliposomes are also compatible with a trimeric post-fusion form and not with monomers, indicating that the ectodomain underwent trimerization upon liposome binding (Figure 4B-D). As with CrHAP2, the crystallization trials were also set with monomeric HAP2e, and the conformational change in AtHAP2 to a post-fusion trimer must have also occurred in the crystallization drops prior to crystallization. The trigger of the transition from the HAP2 monomer to post-fusion trimer is currently not understood.

α F amphipathicity is required for AtHAP2 *in vitro* liposome insertion and *in vivo* function.

For a detailed functional analysis, we focused on the AtHAP2 tip because we had access to a genetic system (Figure 5) to complement biochemical analyses (Figure 4), and also because of the controversial results reported recently (Valansi et al., 2017) about the fusion mechanism of this protein, as mentioned above. To assess the function of HAP2 in flowering plant gamete fusion (Figure 5A-C), we utilized a genetic transmission assay (Figure 5D), which allowed us to determine whether HAP2 variants were able to restore gamete fusion to *hap2-2* null mutant sperm. In crosses between wild-type females and *hap2-2/HAP2* heterozygous males, the *hap2-2* allele is not inherited by progeny (Figure 5D) (von Besser et al., 2006; Wong and Johnson, 2010). However, if a single-locus

transgene carrying fully functional HAP2 is introduced into *hap2-2/HAP2* plants, transmission is restored and 33% of progeny will inherit *hap2-2* following a cross to a wild-type female (Figure 5D, AtHAP2 WT). This assay provided a quantitative readout for HAP2 function *in vivo* and facilitated dissection of the critical residues of the domain II tip (Figure 5D, Table S3). We first tested whether AtHAP2 α F is essential for function by deleting the helix along with two N- and C-terminal residues (box in Figure 5E, At $\Delta\alpha$ F). Importantly, deletion of α F rendered HAP2 non-functional; *hap2-2* transmission rates were zero (Figure 5F, At $\Delta\alpha$ F) even though expression of HAP2 $\Delta\alpha$ F:YFP in sperm was indistinguishable from the wild-type control (Figure 5G, Figure S6).

The AtHAP2 structure predicts that the non-polar residues (I¹⁷¹, F¹⁷², M¹⁷⁵, I¹⁷⁶) on the apical face of α F (Figure 2B, 4G) would be critical for insertion into female gamete plasma membranes. To address this hypothesis, we first tested whether these residues were important for association of AtHAP2e with liposomes (Figure 4A). When F¹⁷² was mutated to alanine, membrane association was decreased and when I¹⁷¹, F¹⁷², and I¹⁷⁶ were mutated to alanine (IFI>AAA triple mutant), membrane association was almost completely abrogated; the majority of AtHAP2e was found at the bottom of the sucrose gradient (Figure 4A) and was not associated with liposomes (Figure 4E, F). In *in vivo* analyses, mutating either I¹⁷¹ or I¹⁷⁶ to alanine slightly reduced HAP2 function in the gamete fusion assay (Figure 5F) and mutation of F¹⁷² to alanine more strongly reduced HAP2 function (Figure 5F). The IFI>AAA triple mutant completely eliminated HAP2 function *in vivo* (Figure 5F), in line with the abrogated membrane association observed *in vitro*. We conclude that these hydrophobic α F residues, which are not individually essential for function, provide multiple points for insertion of the HAP2 domain II tip into the egg and central cell plasma membrane during gamete fusion.

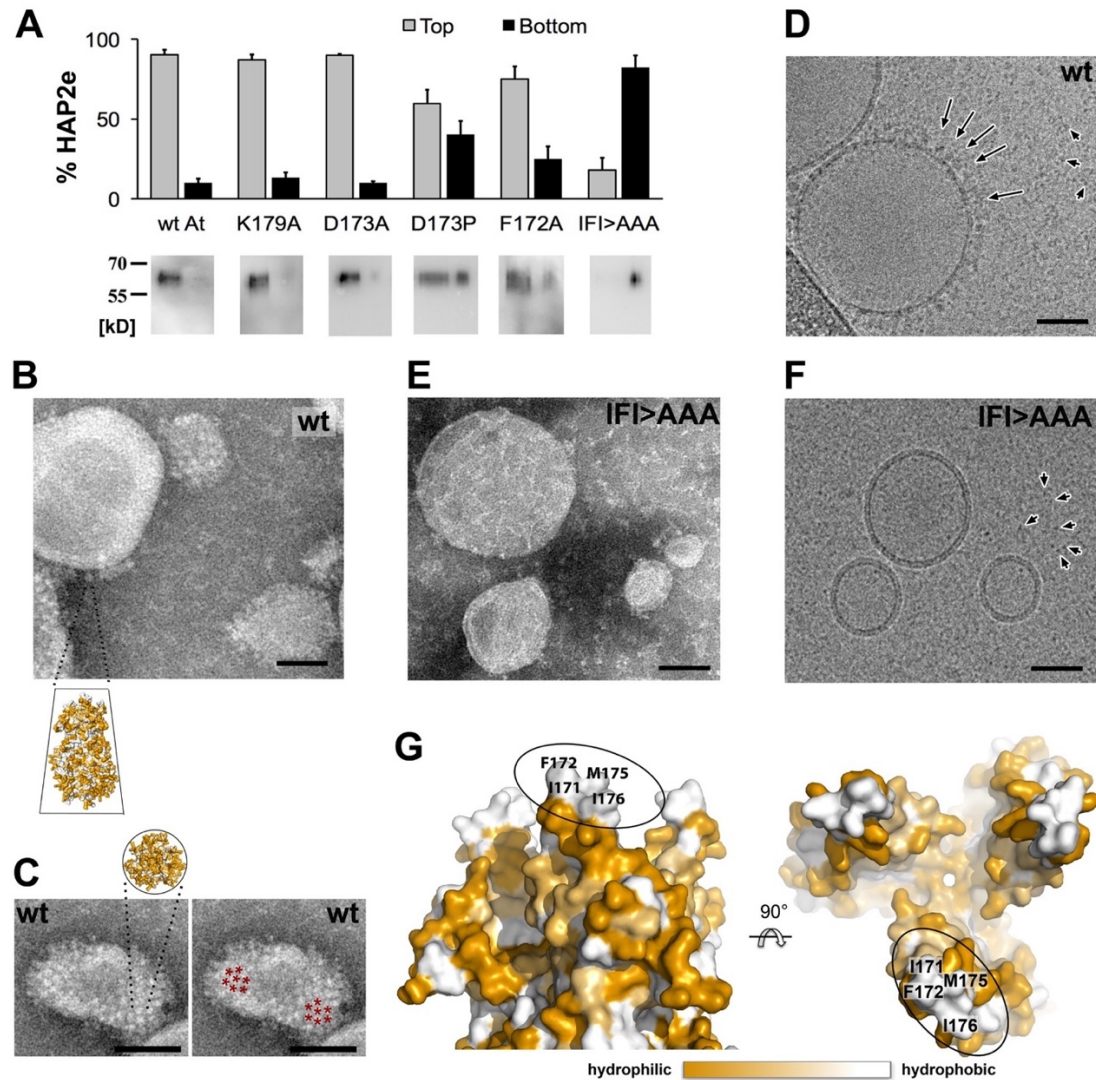


Figure 4. *A. thaliana* HAP2e trimers insert into liposomes via a non-polar domain II apex. A-F. Membrane insertion of the *At*HAP2 ectodomain. **A.** *At*HAP2e mutant membrane insertion was analyzed by differential sucrose gradient fractionation; the amount of HAP2e present in the top (grey) and bottom (black) fractions is shown as mean $\pm$ SD; corresponding immunoblots are shown below. Electron micrographs of liposomes incubated with wild type HAP2e (**B,C,D**) and the IFI>AAA mutant (**E,F**). Samples were analyzed by negative stain electron microscopy (EM) (**B,C,E**) or cryo-EM (**D,F**). Scale bar 50nm. Liposomes decorated with HAP2e display protein projections at the surface (**B,C,D**) contrasting with the smooth liposome surface in the presence of IFI>AAA mutant (**E,F**). The protein projections (some of which are indicated by long arrows in **D**) are rod-shaped with a tapered membrane-proximal end – an inset illustrates the fitting of a trimeric *At*HAP2e into one projection (**B**). The background is coated with proteins not bound to liposomes (short arrows in **D,F**). **C.** Top view of a liposome decorated with *At*HAP2e clustering in hexagonal arrays; an inset illustrates fitting of an *At*HAP2 trimer, red asterisks illustrate trimer clustering. **(G)** Surface representation of the membrane facing tip of the *At*HAP2 trimer viewed from the side (left panel) and the top (right panel). Surface residues

are colored according to hydrophilicity from dark orange (hydrophilic) to white (hydrophobic). The fusion loop region from one protomer is encircled and hydrophobic residues that are exposed at the tip are labeled. See also Figure S4.

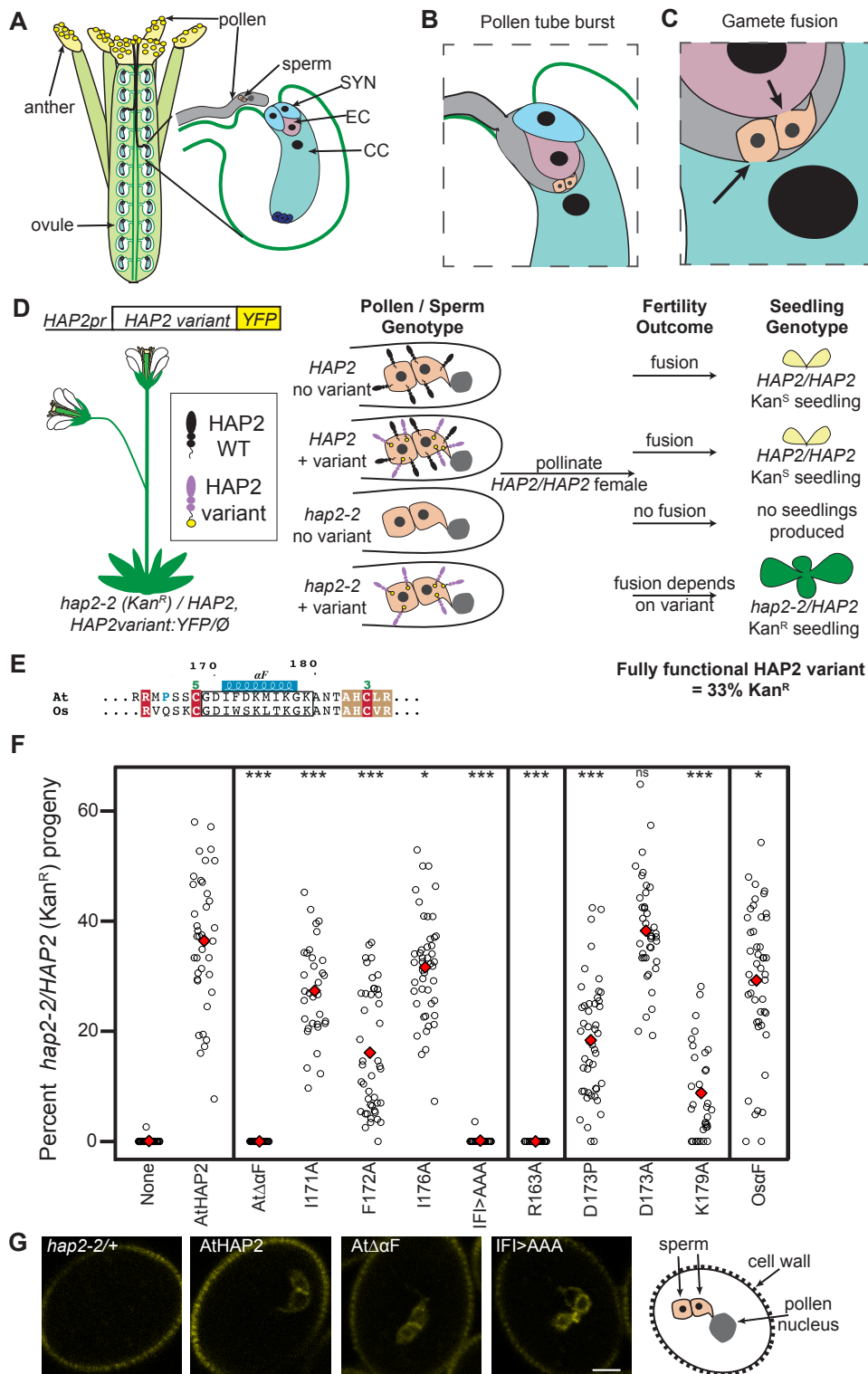


Figure 5. Arabidopsis gamete fusion requires presentation of hydrophobic residues within the amphipathic HAP2 α F helix. **A.** Schematic of male gamete delivery and double gamete fusion within an Arabidopsis flower. Four anthers are shown surrounding the pistil, which contains two columns of ovules; petals and sepals are not depicted. Two pollen tubes (black lines) are shown targeting ovules. The inset illustrates a pollen tube approaching a female gametophyte, which develops within the ovule. A pair of sperm cells and the pollen tube nucleus are shown within the pollen tube; synergid cells (SYN), egg cell (EC), central cell (CC), and three antipodal cells (dark blue) are shown within the female gametophyte. **B.** A schematic showing a pollen tube that has burst into a synergid, propelling the sperm to the site of fusion with the egg or central cell. **C.** Arrows highlight the fusion of one sperm with the egg and the other with the central cell. **D.** Assay for HAP2 function *in vivo* using transmission of *hap2-2* (T-DNA insertion carries kanamycin resistance, Kan^R). *hap2-2/+* plants carrying wild-type (WT) or variant HAP2:YFP transgenes were produced. These plants will generate four pollen/sperm genotypes. Endogenous HAP2 (black) or introduced HAP2:YFP construct (purple/yellow) are shown at the sperm surface. The fertility outcome for each sperm genotype is indicated; a functional copy of HAP2 (either endogenous or a functional HAP2:YFP variant) must be expressed by sperm to produce a viable seed. Only *hap2-2* sperm that carry a functional HAP2:YFP variant transgene will sire Kan^R progeny when outcrossed to *ms1* (*male sterile 1*) females. Expression of a fully functional HAP2:YFP variant will result in 33% Kan^R seedlings. **E.** The Arabidopsis (*At*) and Rice (*Os*) HAP2 membrane interaction surfaces including the α F helix are shown; the region chosen for mutagenesis is identified with a box. **F.** Quantitative analysis of effect of HAP2:YFP variants on *hap2-2* transmission. Each data point (circle) represents the percentage of progeny from a single cross-pollination that inherited the *hap2-2* allele; the mean is indicated with a red diamond. A minimum of three independent transgenic lines were analyzed for each HAP2:YFP variant (see Methods). HAP2:YFP variants tested are indicated on the x-axis: none, no transgene; others defined in text; p-values from the Wilcoxon rank sum test comparing each variant to *At*HAP2 are indicated: **p*<0.05, ***p*<0.01, ****p*<0.001. ns = not significant. **G.** HAP2:YFP is detectable in sperm cells of transgenic plants. Confocal micrographs show HAP2:YFP signal from the indicated variant and auto-fluorescence of the pollen cell wall. Scale bar, 5 μ m. A schematic depicts the arrangement of a pair of sperm cells within the pollen grain; the pollen nucleus and cell wall are depicted. See also Figure S6 and Tables S3 and S4.

Lysine 179 is situated just below α F (Figure 3) and is invariant among plant HAP2 sequences (Figure 1). To evaluate the roles of amino acids that are proximal to, but not part of, α F, we asked whether this residue was important for HAP2 function. Mutating K¹⁷⁹ to alanine strongly reduced the function of HAP2 in the gamete fusion assay (Figure 5F) but did not affect the ability of HAP2e to insert into liposomes *in vitro* (Figure 4A). This result suggests that K¹⁷⁹ is not required for membrane insertion, but rather for a subsequent stage of the gamete fusion reaction in plants. *At*HAP2 association with liposomes was enhanced when artificial membranes included the negatively charged

phospholipid DOPS (See Methods). This finding raises questions for future research on the plasma membrane composition of *Arabidopsis* female gamete plasma membranes and suggests that K¹⁷⁹ may play a role in driving the fusion reaction in this membrane environment.

The conserved E¹¹⁷-R¹⁶³ salt-bridge and the helical character of α F are essential for *Arabidopsis* gamete fusion.

We had proposed that the salt bridge between the invariant R¹⁸⁵ and E¹²⁶ in CrHAP2 (Figures 1B and 2B-C) constrains the domain II tip and helps present fusion loops to the target membrane (Fedry et al., 2017). However, *Tetrahymena thermophila* (Tt) mating was not eliminated when the corresponding arginine (R¹⁶⁴) was mutated in one of the two mating partners (Pinello et al., 2017), raising the question of whether the proposed function is maintained across species. We observed that *A. thaliana* gamete fusion was completely abolished upon mutation of the corresponding arginine residue to alanine (R¹⁶⁴A, Figure 5F). These results are consistent with the view that this salt-bridge is a critical feature of the AtHAP2 gamete fusion mechanism because it structures the domain II tip such that α F is presented correctly to interact with membranes. In TtHAP2, a positively charged lysine, potentially capable of forming the salt bridge in the absence of R¹⁶⁴, is located two residues downstream of the conserved arginine (Figure 1B). Indeed, while the corresponding CrHAP2 R¹⁸⁵A mutation eliminated function, an R¹⁸⁵K mutant restored fusion (Fedry et al., 2017), showing that a lysine can functionally substitute for the conserved arginine. Together, these data suggest that the required salt bridge may also be made by TtHAP2 K¹⁶⁶, thereby maintaining the orientation of the membrane interaction surface. Further experiments (e.g., an arginine/lysine double mutant in TtHAP2) would be necessary to evaluate this hypothesis.

To determine whether the helical nature of αF is critical for HAP2 function, we mutated D¹⁷³, which is in the middle of the hydrophilic side of αF , to proline, a mutation predicted to introduce a kink in the helix. We found that D¹⁷³P, but not D¹⁷³A (which is expected to maintain helical character), reduced HAP2 function *in vivo* (Figure 5F) and reduced the ability of HAP2e to associate with liposomes (Figure 4A). The fact that the D¹⁷³P mutant is capable of rescuing the *hap2-2* defect, even at lower levels, suggests that a kinked helix can also be functional. Of note, the predicted αF in some algal species contains a proline residue (Figure S5) - albeit at the center of a longer helix (3 turns vs 2 turns in *AtHAP2*), suggesting that a relaxed or kinked helical membrane interaction surface may have been selected to drive insertion of the domain II tip of HAP2 in some eukaryotic organisms.

The Rice αF is able to mediate fusion within the context of *Arabidopsis* HAP2.

To test whether other membrane interaction surfaces are functionally equivalent to *AtHAP2* αF in *Arabidopsis* gamete fusion, we replaced *At* αF with the corresponding amphipathic helix of rice (*Oryza sativa*, *Os*, Figure S5), a representative distantly related flowering plant species. This replacement resulted in a chimeric HAP2 variant with nearly identical function as wild type *AtHAP2* (*Os* αF , Figures 5E and 5F). While maintaining the amphipathic nature of the helix, *Os* αF shares an isoleucine at position 171 but substitutes a tryptophan at position 173 and a threonine at position 176 (Figure S5). The interchangeability of these residues within αF suggests plasticity in the mode of interaction between sperm-expressed HAP2 and its target membranes. Since *Arabidopsis* (dicot) and rice (monocot) represent two poles within flowering plant phylogeny, this result suggests the membrane interaction surface of HAP2 is interchangeable across ~150 million years of angiosperm evolution. In contrast, we previously showed that the entire rice HAP2 (with

Arabidopsis signal peptide, transmembrane domain, and C-terminal region) was not capable of rescuing Arabidopsis HAP2 function (Wong and Johnson, 2010).

Taken together, these results indicate that the Arabidopsis and rice membrane interaction surfaces are interchangeable, but that other regions of HAP2 function in a context-specific manner. Future work will exploit this finding to investigate how other domains of HAP2 could be involved in species-limited protein-protein interactions *in cis* with sperm surface proteins (e.g. the sperm adhesion protein GEX 2 (Mori et al., 2014), or *in trans* with egg or central cell-expressed proteins.

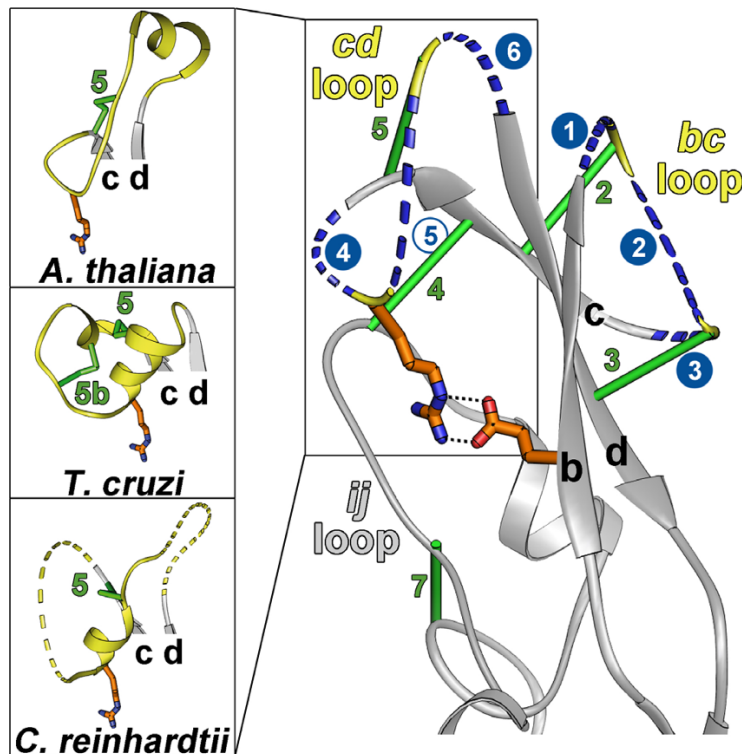


Figure 6. Variability in the HAP2 membrane interaction surface across eukaryotic phyla. Cartoon representation of the structurally conserved scaffold (grey) in the HAP2 tip region with variable loops in yellow. Blue dashed lines indicate the location of insertion/deletion hotspots in different HAP2 orthologs. These hotspots are numbered in blue circles matching those in Figure 1; disulfide bonds are shown as green sticks and numbered. The conserved salt bridge is represented as in Figure 2B. The three insets show the structures of the varied *bc* and *cd* connections of *At*, *Tc*, and *Cr*HAP2. See also Figure S3.

Concluding Remarks

The crystal structures of the tip of domain II of two distant HAP2 molecules reported here, together with *in vitro* and *in vivo* functional assays, provide strong support for HAP2 direct membrane insertion to drive gamete fusion, and furthermore define the functional structural elements critical for the interactions with the membrane of the female gamete. Such a unilateral fusion mechanism is in line with the observation that HAP2 is expressed in only one of two gametes in multiple plant, protozoan and animal species (Hirai et al., 2008; Kawai-Toyooka et al., 2014; Liu et al., 2008; Mori et al., 2006; Steele and Dana, 2009; von Besser et al., 2006) and that expression from only one gamete is sufficient for gamete fusion in all species tested thus far (Cole et al., 2014; Hirai et al., 2008; Liu et al., 2008; Mori et al., 2006; von Besser et al., 2006). The structural variability observed in the three HAP2 structures currently available (Figures 6 and S3), together with the highly variable amino acid sequence of the *bc* and *cd* connections among HAP2 orthologs (Figure 1B, hotspots) indicate that in spite of having a conserved overall structure, HAP2 has evolved multiple alternative modes of membrane interaction during eukaryotic evolution.

The recent finding that ectopically expressed *At*HAP2 in mammalian cells can cause syncytia formation only between HAP2 expressing cells (Valansi et al, 2017), led these authors to propose that, like FF proteins, HAP2 functions by *trans*-oligomerization without target membrane insertion. The strong correlation between the ability of the soluble *At*HAP2 ectodomain to insert into liposomes *in vitro* with its capacity to rescue defective *hap2-2* null gametes for fertilization *in vivo*, observed here, argue strongly against such a model. The incongruous sets of results reported by Valansi et al could be explained if HAP2 activity was enhanced above a critical threshold in the exotic membrane environment of the heterologous mammalian system when direct membrane insertion is occurring simultaneously on both opposing membranes, in keeping with the observation

that *T. thermophila* gamete fusion is enhanced when both mating partners express HAP2 (Cole et al., 2014; Pinello et al., 2017).

HAP2 orthologs have been identified by sequence similarity (pfam10699) in four out of the five eukaryotic kingdoms (Figures S1 and 1), suggesting that it was present in the last common ancestor to all eukaryotes and is a broadly used gamete fusion protein. It has not been identified, however, in some well-studied clades - e.g. fungi, nematodes, and vertebrates. While we cannot formally rule out the possibility of HAP2 replacement in some lineages by a fusion protein of different origin, its sporadic identification pattern among eukaryotes suggests that the most likely scenario is sequence divergence during evolution such that it is not recognized by current bioinformatic methods in many eukaryotic phyla. Such a situation is reminiscent of the viral class II proteins from different genera, which have no identifiable sequence motifs relating them and have only been identified through structural studies (Guardado-Calvo and Rey, 2017; Kielian and Rey, 2006). Yet the viral proteins clearly display the same structure (Figure S3) and employ the same fusion mechanism.

A possible evolutionary force that may have driven HAP2 divergence beyond detectability in some lineages is positive selection for sequence diversity to maintain barriers between species. Proteins mediating cellular interactions critical for fertilization are well known to diversify rapidly and reinforce interspecific fertility barriers (Swanson and Vacquier, 2002). Such interactions between HAP2 and partner proteins that regulate its fusion activity may have driven further divergence of the HAP2 sequence in organisms like yeast, *Caenorhabditis elegans*, mice, and humans. Our discovery of the significant structural conservation of HAP2 (e.g., the $\alpha 2$ helix), as shown in Fig. S3A – which is much greater than the one seen among viral class II proteins (compare the top and bottom rows in Fig. S3A) – now reveals new structural motifs to guide the search for orthologs in additional organisms. The linear alignment of the HAP2 sequences listed in Figure S1,

together with the three-dimensional alignment of the three HAP2 structures now available, provides information on residue pairs distant in the linear sequence but interacting with each other in the structure. The resulting co-variation patterns can be exploited to identify new signatures for the identification of HAP2 orthologs not detected by current methods, for instance in fungi or in vertebrates. In mice, sperm-expressed IZUMO (Inoue et al., 2005) and its egg-expressed receptor JUNO (Bianchi et al., 2014) are essential for gamete attachment (Aydin et al., 2016), the step preceding fusion, and are conserved in humans. It will be therefore very interesting now to use the insights into HAP2 structure-function provided here to identify proteins mediating gamete fusion in mammals.

Author contributions

J.Fe. and J.Fo. generated the aa sequence alignments. J.Fe., T.K. and F.A.R. conceived the in vitro experiments with the recombinant HAP2 ectodomain. J.Fe. made the constructs, expressed and crystallized the HAP2e proteins, optimized the crystals for diffraction with A.H., collected data and determined the HAP2 X-ray structure with P.L. J.Fe. also demonstrated insertion of HAP2e into liposomes, and G.P.A. performed electron microscopy on liposomes. J.Fo. and M.J. conceived the in vivo experiments. J.Fo. made the constructs and transgenic plants, and analyzed HAP2 function. T.K., F.A.R., and M.J. supervised experimental work. J.Fe., J. Fo, T.K., F.A.R., and M.J. wrote the manuscript with input from all other authors.

Acknowledgements

We thank W.J. Snell, members of the Snell, Rey, and Johnson groups and G. Obal for helpful discussions; Samyukta Mallick and the crystallization platform of the Pasteur Proteopole for technical help; staff of synchrotron beamlines PX-I at the Swiss Light

Source, Proxima-1 and -2 at SOLEIL for help during data collection; and M. Nilges and the Equipement d'excellence CACSICE for providing the Falcon II direct electron detector; G. Williams of Brown University's Leduc bioimaging facility; and A.-C. Lindner and J. Becker (Gulbenkian Institute) for annotation of *Physcomitrella patens* HAP2. F.A.R. acknowledges funding from the ERC Advanced grant project (340371) CelCelFus for this work, which also used general support from Institut Pasteur and CNRS. J.Fe. benefitted from an Allocation ministerielle pour l'Ecole Polytechnique AMX. M.J. and J.Fo. were supported by National Science Foundation (NSF) Grant IOS-1353798; with additional support from NSF Grant IOS-1540019 (M.J.) and National Institutes of Health Training Grant #T32-GM007601 (J.Fo.).

References

- Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., *et al.* (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301, 653-657.
- Avinoam, O., Fridman, K., Valansi, C., Abutbul, I., Zeev-Ben-Mordehai, T., Maurer, U.E., Sapir, A., Danino, D., Grunewald, K., White, J.M., *et al.* (2011). Conserved eukaryotic fusogens can fuse viral envelopes to cells. *Science* 332, 589-592.
- Aydin, H., Sultana, A., Li, S., Thavalingam, A., and Lee, J.E. (2016). Molecular architecture of the human sperm IZUMO1 and egg JUNO fertilization complex. *Nature* 534, 562-565.
- Backovic, M., and Krey, T. (2016). Stable *Drosophila* Cell Lines: An Alternative Approach to Exogenous Protein Expression. *Methods in molecular biology* (Clifton, NJ) 1350, 349-358.
- Baldauf, S.L., Roger, A.J., Wenk-Siefert, I., and Doolittle, W.F. (2000). A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* 290, 972-977.
- Bian, X., Klemm, R.W., Liu, T.Y., Zhang, M., Sun, S., Sui, X., Liu, X., Rapoport, T.A., and Hu, J. (2011). Structures of the atlastin GTPase provide insight into homotypic fusion of endoplasmic reticulum membranes. *Proceedings of the National Academy of Sciences* 108, 3976-3981.
- Bianchi, E., Doe, B., Goulding, D., and Wright, G.J. (2014). Juno is the egg Izumo receptor and is essential for mammalian fertilization. *Nature* 508, 483-487.
- Blaise, S., de Parseval, N., Benit, L., and Heidmann, T. (2003). Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. *Proc Natl Acad Sci U S A* 100, 13013-13018.
- Boavida, L.C., and McCormick, S. (2007). Temperature as a determinant factor for increased and reproducible in vitro pollen germination in *Arabidopsis thaliana*. *Plant J* 52, 570-582.
- Bressanelli, S., Stiasny, K., Allison, S.L., Stura, E.A., Duquerroy, S., Lescar, J., Heinz, F.X., and Rey, F.A. (2004). Structure of a flavivirus envelope glycoprotein in its low-pH-induced membrane fusion conformation. *The EMBO journal* 23, 728-738.
- Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, P., Roversi, P., Sharff, A., Smart, O., Vornrhein, C., *et al.* (2010). BUSTER version 2.9., Cambridge, United Kingdom, Global Phasing Ltd.
- Bunkoczi, G., and Read, R.J. (2011). Improvement of molecular-replacement models with Sculptor. *Acta Crystallogr D Biol Crystallogr* 67, 303-312.
- Celenza Jr., J.L., Grisafi, P.L., and Fink, G.R. (1995). A pathway for lateral root formation in *Arabidopsis thaliana*. *Genes & Development* 9, 2131-2142.
- Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16, 735-743.
- Cole, E.S., Cassidy-Hanley, D., Fricke Pinello, J., Zeng, H., Hsueh, M., Kolbin, D., Ozzello, C., Giddings, T., Winey, M., and Clark, T.G. (2014). Function of the Male-Gamete-Specific Fusion Protein HAP2 in a Seven-Sexed Ciliate. *Current biology : CB* 24, 2168-2173.
- Collaborative Computational Project (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50, 760-763.
- Dessau, M., and Modis, Y. (2013). Crystal structure of glycoprotein C from Rift Valley fever virus. *Proc Natl Acad Sci U S A* 110, 1696-1701.

- DuBois, R.M., Vaney, M.C., Tortorici, M.A., Kurdi, R.A., Barba-Spaeth, G., Krey, T., and Rey, F.A. (2013). Functional and evolutionary insight from the crystal structure of rubella virus protein E1. *Nature* 493, 552-556.
- Ebchuqin, E., Yokota, N., Yamada, L., Yasuoka, Y., Akasaka, M., Arakawa, M., Deguchi, R., Mori, T., and Sawada, H. (2014). Evidence for participation of GCS1 in fertilization of the starlet sea anemone *Nematostella vectensis*: implication of a common mechanism of sperm-egg fusion in plants and animals. *Biochemical and biophysical research communications* 451, 522-528.
- Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5, 113.
- Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66, 486-501.
- Evans, P. (2005). Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 62, 72-82.
- Fedry, J., Liu, Y., Pehau-Arnaudet, G., Pei, J., Li, W., Tortorici, M.A., Traincard, F., Meola, A., Bricogne, G., Grishin, N.V., Snell, W.J., Rey, F.A. and Krey, T. (2017). The Ancient Gamete Fusogen HAP2 Is a Eukaryotic Class II Fusion Protein. *Cell* 168, 904-915.e910.
- Gautier, R., Douguet, D., Antony, B., and Drin, G. (2008). HELIQUEST: a web server to screen sequences with specific alpha-helical properties. *Bioinformatics* 24, 2101-2102.
- Gibbons, D.L., Erk, I., Reilly, B., Navaza, J., Kielian, M., Rey, F.A., and Lepault, J. (2003). Visualization of the target-membrane-inserted fusion protein of Semliki Forest virus by combined electron microscopy and crystallography. *Cell* 114, 573-583.
- Gibbons, D.L., Vaney, M.C., Roussel, A., Vigouroux, A., Reilly, B., Lepault, J., Kielian, M., and Rey, F.A. (2004). Conformational change and protein-protein interactions of the fusion protein of Semliki Forest virus. *Nature* 427, 320-325.
- Guardado-Calvo, P., Bignon, E.A., Stettner, E., Jeffers, S.A., Perez-Vargas, J., Pehau-Arnaudet, G., Tortorici, M.A., Jestin, J.L., England, P., Tischler, N.D. and Rey, F.A. (2016). Mechanistic Insight into Bunyavirus-Induced Membrane Fusion from Structure-Function Analyses of the Hantavirus Envelope Glycoprotein Gc. *PLoS Pathog* 12, e1005813.
- Guardado-Calvo, P., and Rey, F.A. (2017). The Envelope Proteins of the Bunyavirales. *Adv Virus Res* 98, 83-118.
- Halldorsson, S., Behrens, A.J., Harlos, K., Huiskonen, J.T., Elliott, R.M., Crispin, M., Brennan, B., and Bowden, T.A. (2016). Structure of a phleboviral envelope glycoprotein reveals a consolidated model of membrane fusion. *Proc Natl Acad Sci U S A* 113, 7154-7159.
- Harrison, S.C. (2015). Viral membrane fusion. *Virology* 479-480, 498-507.
- Hellens, R.P., Edwards, E.A., Leyland, N.R., Bean, S., and Mullineaux, P.M. (2000a). pGreen: a versatile and flexible binary Ti vector for *Agrobacterium*-mediated plant transformation. *Plant Mol Biol* 42, 819-832.
- Hellens, R.P., Edwards, E.A., Leyland, N.R., Bean, S., and Mullineaux, P.M. (2000b). pGreen: a versatile and flexible binary Ti vector for *Agrobacterium*-mediated plant transformation. *Plant Molecular Biology* 42, 819-832.
- Hirai, M., Arai, M., Mori, T., Miyagishima, S.Y., Kawai, S., Kita, K., Kuroiwa, T., Terenius, O., and Matsuoka, H. (2008). Male fertility of malaria parasites is determined by GCS1, a plant-type reproduction factor. *Curr Biol* 18, 607-613.
- Inoue, N., Ikawa, M., Isotani, A., and Okabe, M. (2005). The immunoglobulin superfamily protein Izumo is required for sperm to fuse with eggs. *Nature* 434, 234-238.

- Johnson, M.A., von Besser, K., Zhou, Q., Smith, E., Aux, G., Patton, D., Levin, J.Z., and Preuss, D. (2004). *Arabidopsis* hapless mutations define essential gametophytic functions. *Genetics* 168, 971-982.
- Kabsch, W. (1988). Automatic indexing of rotation diffraction patterns. *J Appl Crystallogr*, 67-72.
- Kawai-Toyooka, H., Mori, T., Hamaji, T., Suzuki, M., Olson, B.J., Uemura, T., Ueda, T., Nakano, A., Toyoda, A., Fujiyama, A., *et al.* (2014). Sex-specific posttranslational regulation of the gamete fusogen GCS1 in the isogamous volvocine alga *Gonium pectorale*. *Eukaryot Cell* 13, 648-656.
- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10, 845-858.
- Kielian, M., and Rey, F.A. (2006). Virus membrane-fusion proteins: more than one way to make a hairpin. *Nature reviews* 4, 67-76.
- Koncz, C., and Schell, J. (1986). The Promoter of TI-DNA Gene 5 Controls the Tissue-Specific Expression of Chimeric Genes Carried by a Novel Type of *Agrobacterium* Binary Vector. *Molecular & General Genetics* 204, 383-396.
- Leikin, S., Parsegian, V.A., Rau, D.C., and Rand, R.P. (1993). Hydration forces. *Annu Rev Phys Chem* 44, 369-395.
- Lescar, J., Roussel, A., Wien, M.W., Navaza, J., Fuller, S.D., Wengler, G., Wengler, G., and Rey, F.A. (2001). The Fusion glycoprotein shell of Semliki Forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH. *Cell* 105, 137-148.
- Liao, M., Sanchez-San Martin, C., Zheng, A., and Kielian, M. (2010). In vitro reconstitution reveals key intermediate states of trimer formation by the dengue virus membrane fusion protein. *J Virol* 84, 5730-5740.
- Liu, Y., Tewari, R., Ning, J., Blagborough, A.M., Garbom, S., Pei, J., Grishin, N.V., Steele, R.E., Sinden, R.E., Snell, W.J., *et al.* (2008). The conserved plant sterility gene HAP2 functions after attachment of fusogenic membranes in *Chlamydomonas* and *Plasmodium* gametes. *Genes Dev* 22, 1051-1068.
- Mi, S, Lee, X., Li, X.P., Veldman, G.M., Finnerty, H., Racie, L., LaVallie, E., Tang, X.Y., Edouard, P., Howes, S., Keith, J.C., McCoy, J.M. (2000). Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis *Nature* 403, 785-789
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. *J Appl Crystallogr* 40, 658-674.
- Modis, Y. (2014). Relating structure to evolution in class II viral membrane fusion proteins. *Curr Opin Virol* 5, 34-41.
- Modis, Y., Ogata, S., Clements, D., and Harrison, S.C. (2004). Structure of the dengue virus envelope protein after membrane fusion. *Nature* 427, 313-319.
- Mohler, W.A., Shemer, G., del Campo, J.J., Valansi, C., Opoku-Serebuoh, E., Scranton, V., Assaf, N., White, J.G., and Podbilewicz, B. (2002). The type I membrane protein EFF-1 is essential for developmental cell fusion. *Dev Cell* 2, 355-362.
- Mori, T., Igawa, T., Tamiya, G., Miyagishima, S.Y., and Berger, F. (2014). Gamete attachment requires GEX2 for successful fertilization in *Arabidopsis*. *Curr Biol* 24, 170-175.
- Mori, T., Kuroiwa, H., Higashiyama, T., and Kuroiwa, T. (2006). GENERATIVE CELL SPECIFIC 1 is essential for angiosperm fertilization. *Nat Cell Biol* 8, 64-71.
- Okamoto, M., Yamada, L., Fujisaki, Y., Bloomfield, G., Yoshida, K., Kuwayama, H., Sawada, H., Mori, T., and Urushihara, H. (2016). Two HAP2-GCS1 homologs responsible

for gamete interactions in the cellular slime mold with multiple mating types: Implication for common mechanisms of sexual reproduction shared by plants and protozoa and for male-female differentiation. *Dev Biol* 415, 6-13.

Pérez-Vargas, J., Krey, T., Valansi, C., Avinoam, O., Haouz, A., Jamin, M., Raveh-Barak, H., Podbilewicz, B., and Rey, F.A. (2014). Structural basis of eukaryotic cell-cell fusion. *Cell* 157, 407-419.

Pinello, J.F., Lai, A.L., Millet, J.K., Cassidy-Hanley, D., Freed, J.H., and Clark, T.G. (2017). Structure-function studies link class II viral fusogens and the ancestral gamete fusion protein HAP2. *Curr Biol* 27, 651-660.

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

Rey, F.A., Heinz, F.X., Mandl, C., Kunz, C., and Harrison, S.C. (1995). The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution. *Nature* 375, 291-298.

RStudio Team (2016). RStudio: Integrated Development for R. RStudio (Boston, MA: RStudio, Inc.).

Ruggiero, M.A., Gordon, D.P., Orrell, T.M., Bailly, N., Bourgoin, T., Brusca, R.C., Cavalier-Smith, T., Guiry, M.D., and Kirk, P.M. (2015). A higher level classification of all living organisms. *PLoS One* 10, e0119248.

Steele, R.E., and Dana, C.E. (2009). Evolutionary history of the HAP2/GCS1 gene and sexual reproduction in metazoans. *PLoS ONE* 4, e7680.

Stiasny, K., Allison, S.L., Schalich, J., and Heinz, F.X. (2002). Membrane interactions of the tick-borne encephalitis virus fusion protein E at low pH. *J Virol* 76, 3784-3790.

Südhof, T.C., and Rothman, J.E. (2009). Membrane fusion: grappling with SNARE and SM proteins. *Science* (New York, NY) 323, 474-477.

Swanson, W.J., and Vacquier, V.D. (2002). The rapid evolution of reproductive proteins. *Nat Rev Genet* 3, 137-144.

Team, R.C. (2015). R: A language and environment for statistical computing. (Vienna, Austria: R Foundation for Statistical Computing).

Thorn, A., and Sheldrick, G.M. (2011). ANODE: anomalous and heavy-atom density calculation. *J Appl Crystallogr* 44, 1285-1287.

Valansi, C., Moi, D., Leikina, E., Matveev, E., Grana, M., Chernomordik, L.V., Romero, H., Aguilar, P.S., and Podbilewicz, B. (2017). Arabidopsis HAP2/GCS1 is a gamete fusion protein homologous to somatic and viral fusogens. *J Cell Biol* 216, 571-581.

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

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

Wilson, Z.A., Morroll, S.M., Dawson, J., Swarup, R., and Tighe, P.J. (2001). The Arabidopsis *MALE STERILITY1* (*MS1*) gene is a transcriptional regulator of male gametogenesis, with homology to the PHD-finger family of transcription factors. *Plant Journal* 28, 27-39.

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

Methods

CONTACT FOR REAGENT AND RESOURCE SHARING

Requests for resources and reagents should be directed to Felix A. Rey (felix.rey@pasteur.fr).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Plant growth conditions.

Seeds were stratified (at least 2 days at 4°C) on solid Murashige and Skoog (MS) medium (Sigma Aldrich, St. Louis, MO) supplemented with appropriate antibiotics and germinated at 22°C under constant light (Percival incubator). 7-14 days later, seedlings were transferred to sterile #2MIX potting media (www.fafard.com) with fertilizer (N:P:K, 15:5:15) and were grown at 20°C, 50-60% humidity, on a 16 hour light / 8 hour dark light cycle in growth chambers (Environmental growth chambers, Chagrin Falls, OH, USA). **S2 cell lines.**

Drosophila Schneider 2 cells (S2; ATCC CRL-1963) were cultured in Schneider's complete media (Thermo Scientific) before transfection and in Insect Xpress media after transfection (Lonza, Basel, Switzerland). Culturing and transfection of S2 cells has been described previously (Backovic and Krey, 2016).

METHOD DETAILS

Expression and purification of the soluble HAP2 ectodomain

Codon-optimized synthetic cDNA corresponding to a soluble C-terminally truncated version of the HAP2 ectodomain (HAP2e) comprising residues 25-494 from *Arabidopsis thaliana* and 26-516 from *Trypanosoma cruzi* were cloned into a modified Drosophila S2 expression vector described previously (Krey et al., 2010) and transfection was performed as reported earlier (Backovic and Krey, 2016). For large-scale productions,

cells were induced with 4 μM CdCl_2 at a density of approximately 7×10^6 cells/ml for 8 days, pelleted, and the soluble ectodomains were purified by affinity chromatography from the supernatant using a StrepTactin Superflow column followed by size exclusion chromatography using a Superdex200 column in 10mM TRIS pH8 100mM NaCl. Pure proteins were concentrated to approximately 8 and 4 mg/ml respectively.

Size Exclusion Chromatography with Multi-Angle Light Scattering

Purified HAP2e ectodomains were subjected to SEC using a Superdex 200 column (GE HealthCare) equilibrated with 10mM TRIS pH8 100mM NaCl. Separation was performed at 20°C with a flow rate of 0.5 ml min⁻¹. Online MALS detection was performed with a DAWN-HELEOS II detector (Wyatt Technology, Santa Barbara, CA, USA) using a laser emitting at 690 nm. Online differential refractive index measurement was performed with an Optilab T-rEX detector (Wyatt Technology). Data were analyzed, and weight-averaged molecular masses (M_w) and mass distributions (polydispersity) for each sample were calculated using the ASTRA software (Wyatt Technology).

Liposome co-flotation experiments

DOPE (1,2-dioleoyl-sn-glycero-3-phosphoethanolamine), DOPC (1,2-dioleoyl-sn-glycero-3-phosphocholine), DOPS (1,2-dioleoyl-sn-glycero-3-phospho-L-serine), cholesterol and sphingomyelin were purchased from Avanti Polar Lipids. Liposomes were freshly prepared in PBS by the freeze-thaw and extrusion method using molar ratios of 17% DOPE 16% DOPC 50% cholesterol 17% sphingomyelin. Alternatively, for co-flotation assays on wild type and mutant AtHAP2e proteins, the following alternative composition was used : 50% cholesterol, 30% DOPC, 20% DOPS. 0.07 μM purified HAP2e were mixed with 8 mM liposomes and incubated 1h at 21 °C in 100 μL PBS. Samples were then adjusted to a final concentration of 20% sucrose, overlayed with a 5-60% sucrose gradient

(in PBS) and centrifuged 1h at 4 °C at ~152.000 x g. Fractions from the top, middle and bottom of the gradient were analyzed by immunoblotting using a monoclonal anti strep tag antibody and the bands quantified using the GeneTools Syngene software. The percentage of HAP2e in either fraction was calculated as the ratio between HAP2e in individual fractions and total HAP2e (sum of HAP2e in top and bottom fractions).

Electron microscopy

Purified HAP2e (*A. thaliana* or *T. cruzi*) mixed with liposomes, was spotted on glow discharged carbon grids (CF300, EMS, USA), negatively stained with 2% phosphotungstic acid (PTA) PH 7.4, analyzed with a Tecnai G2 Bio-Twin electron microscope (FEI, USA) and imaged with an Eagle camera (FEI, USA). For cryo electron microscopy liposomes mixed with purified HAP2e were applied on a glow discharged Lacey Carbon grid (Agar Scientific, UK). Samples were plunge-frozen in liquid ethane using an automated system (Leica EMGP, Austria) and visualized on a Tecnai F20 electron microscope operating at a voltage of 200 kV. Image frames were recorded in low-dose mode on a Falcon II direct electron detector (FEI, USA).

Crystallization and structure determination

Crystals of HAP2e from *Arabidopsis thaliana* were grown at 293K using the hanging-drop vapor-diffusion method in drops containing 1 µL protein solution mixed with 1 µL reservoir solution containing 100 mM Sodium Citrate pH 4.5 and 200 mM zinc acetate. Diffraction quality rod-like crystals appeared after 1 week and were flash-frozen in mother liquor containing 30% (v/v) glycerol. The crystals diffracted anisotropically to 2.24Å along the c* axis but only to 3.7Å in orthogonal directions. Therefore, an ellipsoidal cut off was applied using the StarAniso server (STARANISO version 1.7.2 18-Apr-17 Ian J. Tickle, Global Phasing Ltd., Cambridge, UK ; <http://staraniso.globalphasing.org/cgi->

bin/staraniso.cgi) in order to remove noise and refinement was carried out to a nominal resolution of 2.75Å.

Limited proteolysis of HAP2e from *Trypanosoma cruzi* was carried out by adding subtilisin (dissolved in 10mM Tris pH8, 30mM NaCl at 10mg/mL) to a solution containing *T. cruzi* HAP2e at 16mg/mL in 10mM Tris pH8, 100mM NaCl at a 1:300 w:w ratio. After one hour incubation at room temperature, the protease was inactivated by addition of 1mM PMSF and the protease-resistant fragment accounting to ~60% of the digested protein was purified by SEC using a Superdex200 column (*TcHAP2e_{sub}*). Crystals of *TcHAP2e_{sub}* were grown at 293K using the hanging-drop vapor-diffusion method in drops containing 1 µL protein solution mixed with 1 µL reservoir solution containing 100mM CHES pH 9.0, 200mM NaCl, and 10%w/v PEG 8k. Diffraction quality rod-like crystals appeared after 7-10 days and were flash-frozen in mother liquor containing 25% (v/v) glycerol.

Data collection was carried out at the the ESRF (ID30A-3) and the Synchrotron Soleil (Proxima-1). Data were processed, scaled and reduced with XDS (Kabsch, 1988), Pointless (Evans, 2005) and programs from the CCP4 suite (Collaborative Computational Project, 1994).

To determine the structure of the protease-resistant fragment of *T. cruzi* HAP2, we initially used Sculptor (Bunkoczi and Read, 2011) to create a search model for molecular replacement based on the structure of a protomer of the *C. reinhardtii* HAP2 trimer (PDB 5MF1). The full monomer model was divided into 3 individual domains (domains I, II and III) and loops and exposed side chains were trimmed off. The structure was finally determined by the molecular replacement method using Phaser (McCoy et al., 2007) and an isolated domain II as search model. Phases were refined using the anomalous signal of a highly redundant Sulfur-SAD data set collected at a wavelength of 2.06641 Å on crystals of the native protein.

The full *C. reinhardtii* HAP2 monomer (PDB 5MF1) was used to create a search model with Sculptor (Bunkoczi and Read, 2011). The so-called "tip region" was deleted (*bc* and *cd* connections). The structure was determined by the molecular replacement method using Phaser (McCoy et al., 2007). Zinc atoms were localized in the density by calculating an anomalous difference map using ANODE (Thorn and Sheldrick, 2011).

For both orthologs, model building was performed using Coot (Emsley et al., 2010), and refinement was done using AutoBuster (Bricogne et al., 2010).

Generating hap2-2/HAP2 transgenic plants carrying HAP2:YFP variants.

Arabidopsis HAP2 coding sequence variants were generated using a transfer-DNA plasmid containing the native HAP2 promoter and coding sequence with a C-terminal YFP fusion (*pHAP2:HAP2cds:YFP*; PGL290; CDS, GenBank AAY51999.1; pGreen (Hellens et al., 2000a) backbone; Basta resistant seedlings). Mutations were made using the NEB Q5 Site-Directed Mutagenesis Kit (NEB #E0554S); primer sequences are provided in Table S4. Mutations were confirmed by sequencing (Genewiz, South Plainfield, NJ). *Agrobacterium tumefaciens*-mediated transformation (strain GV3101 (Koncz and Schell, 1986) with pSoup helper plasmid (Hellens et al., 2000b)) of *hap2-2/HAP2* (SALK_152706; Alonso et al., 2003; von Besser et al., 2006)) was performed by floral dip (Clough and Bent, 1998). *hap2-2/HAP2* seedlings were selected on MS plates supplemented with sucrose (5g/L) and kanamycin sulfate (50mg/L; Life Technologies/Thermo Fisher Scientific, Waltham, MA).

After seeds were collected from transformed plants, they were sterilized using a 50% bleach solution containing 0.02% Triton X-100 for 7 minutes, followed by washing 4 times with sterilized water. Seeds were re-suspended in sterilized 0.1% agarose and were plated on 15cm dishes containing MS media supplemented with Basta (25mg/L; Chem Service Inc., West Chester, PA; Oakwood Chemical, Estill, SC). After at least 2 days at

4°C, plates were moved to 22°C incubator with constant light. After 7-14 days, Basta resistant seedlings were transferred to soil made up with 1X fertilizer (4 plants / pot). For primary transformants (T₁), DNA was isolated using the leaf boil method (Celenza Jr. et al., 1995). T₁ plants were genotyped for the *hap2-2* allele using primer sets to detect both the Salk tDNA insertion (Lbal, Alfonso *et al.* 2003; hap2seqTR3, von Besser *et al.* 2006) and WT genomic HAP2 (hap2c2; hap2seqTR3; see primer table S4).

Genetic transmission assay for assessing transmission of hap2-2 allele.

Pollen from T₁ plants that were heterozygous for *hap2-2/+* were hand-pollinated onto *male sterile 1* (*ms1-1*) pistils (Wilson et al., 2001). Seeds from individual *ms1* crosses were gas-sterilized and plated on MS Kanamycin (50mg/L) plates supplemented with sucrose (5g/L). Seeds were cold-treated at 4°C for at least 2 days then transferred to 22°C incubator with constant light. *hap2-2/+* seeds were simultaneously plated on MS Kanamycin plates to ensure the drug resistance is working correctly. For analysis of *hap2-2* transmission, resistant versus sensitive seedlings were counted manually after 6-10 days at 22°C.

Mutations were confirmed in at least one T₁ line per transgene by sequencing of a PCR product. PCR products were generated using a EcoRI-HAP2-CVFP-F1 forward primer and the HAP2Ex7R reverse primer.

Analysis of HAP2:YFP variant expression.

Expression of HAP2:YFP in *pHAP2:HAP2variant:YFP* transgenic plants was initially screened using fluorescence microscopy. Pollen from T₁ plants was hydrated on slides in pollen growth media (Boavida and McCormick, 2007) and imaged using a Zeiss Axiovert 200M microscope (images not shown). Each transgenic line was analyzed using confocal microscopy (Figure 5G, Supplementary Figure S6) of the T₁ and/or T₂ generation.

Pollen were hydrated in pollen growth media and imaged using a Zeiss LSM 800 confocal microscope. All images were taken using excitation with the 488 laser at 10% (488 settings kept constant) and using the 40x water objective with a 2x zoom. Single optical slices were collected (974 x 974 pixels, 16-bit). All images were imported to ImageJ (FIJI version), and cropped to 300x300 pixels.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data are presented as mean $\pm$ SD unless otherwise indicated in figure legends and experimental repeats are indicated in figure legends.

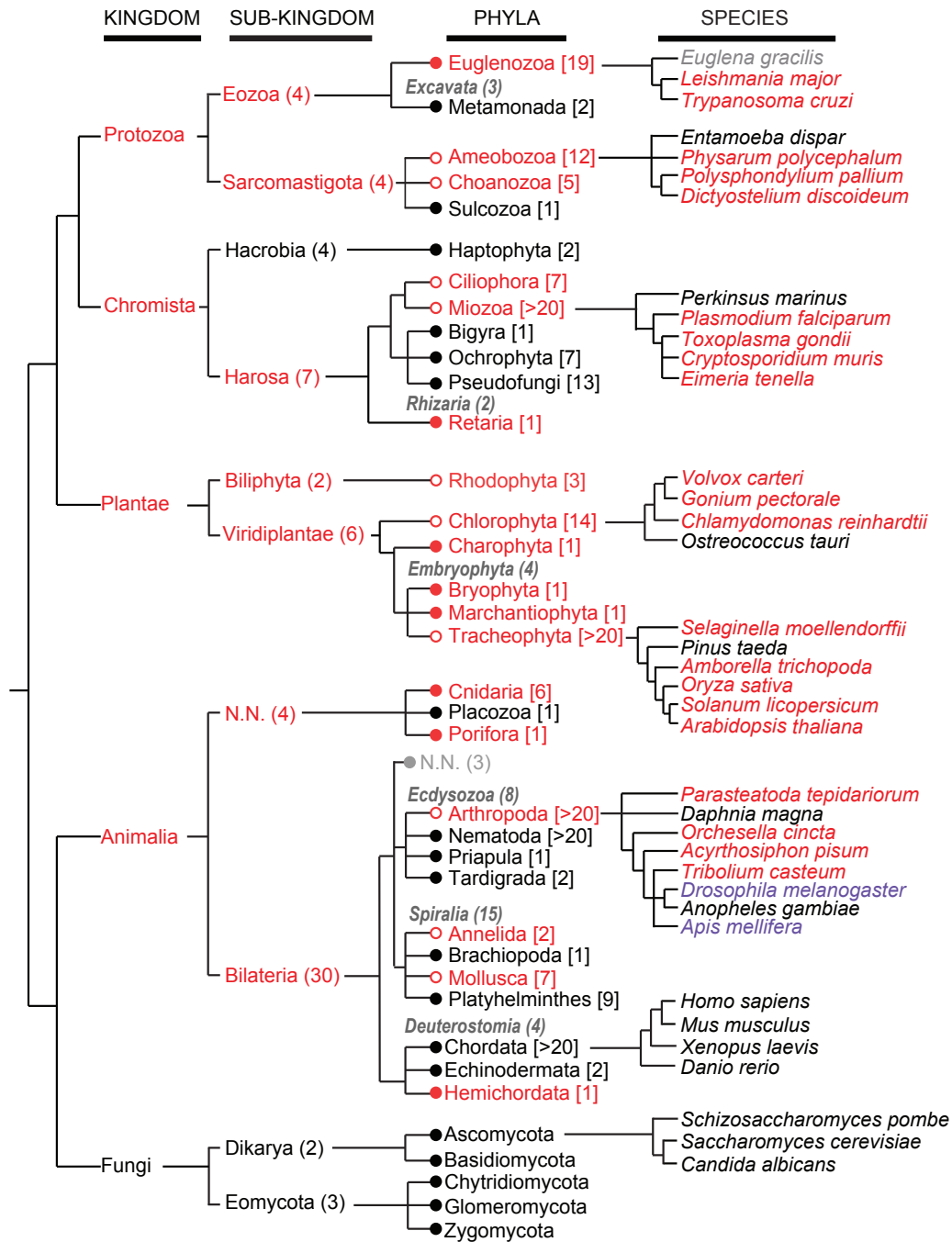
Statistical analysis of transmission of hap2-2 allele.

The following criteria were used to analyze HAP2:YFP variant transgenic plants: at least three independent transgenic lines were analyzed, at least three crosses were performed for each line, only crosses with at least 10 seeds were analyzed (Table S3). Transmission data did not follow a normal distribution for each transgene (confirmed by Anderson-Darling test for normality in R studio (RStudio Team, 2016), using R version 3.2.3 (Team, 2015)); therefore, a non-parametric statistical test was used. Each HAP2 mutant was compared to the WT transgene (*AtHAP2*) using a Wilcoxon rank sum test with continuity correction in R studio.

DATA AND SOFTWARE AVAILABILITY

The atomic coordinates and structure factors for two structures have been deposited in the protein database (PDB) under the accession numbers 5OW3 (*AtHAP2*) and 5OW4 (*TcHAP2*).

Supplementary Figures



Pfam10699 present Pfam10699 not found
 HAP2 identified, pfam10699 absent
 Insufficient annotated genome information
 ● Pfam10699 present in all available genomes
 ○ Pfam10699 absent in some genomes
 ● Pfam10699 absent in all available genomes

Figure S1: Occurrence of the defining HAP2 pfam domain (10699) in eukaryotic genomes. Related to Figure 1. A schematic phylogeny representing the five eukaryotic kingdoms, sub-kingdoms, and phyla for which whole-genome sequence information is available (classification system; Baldauf et al., 2000; Ruggiero et al., 2015). Selected groups of species are provided for some phyla. Red font indicates presence of pfam10699 (the HAP2-GCS1 domain). Filled red circles: all available genomes of the phylum contain pfam10699; open red circles: the phylum contains examples both of genomes that have or do not have pfam10699; filled black circles: phyla in which none of the available genomes contains pfam10699. HAP2 was identified in *D. melanogaster* and *A. mellifera*, but these sequences (and genomes) lack pfam10699 (green font). Parenthetical numbers indicate the total number phyla within the sub-kingdom. Only phyla with at least one whole-genome sequence with > 10,000 annotated proteins are listed. Bracketed numbers indicate the number of species from each phylum used in this analysis.

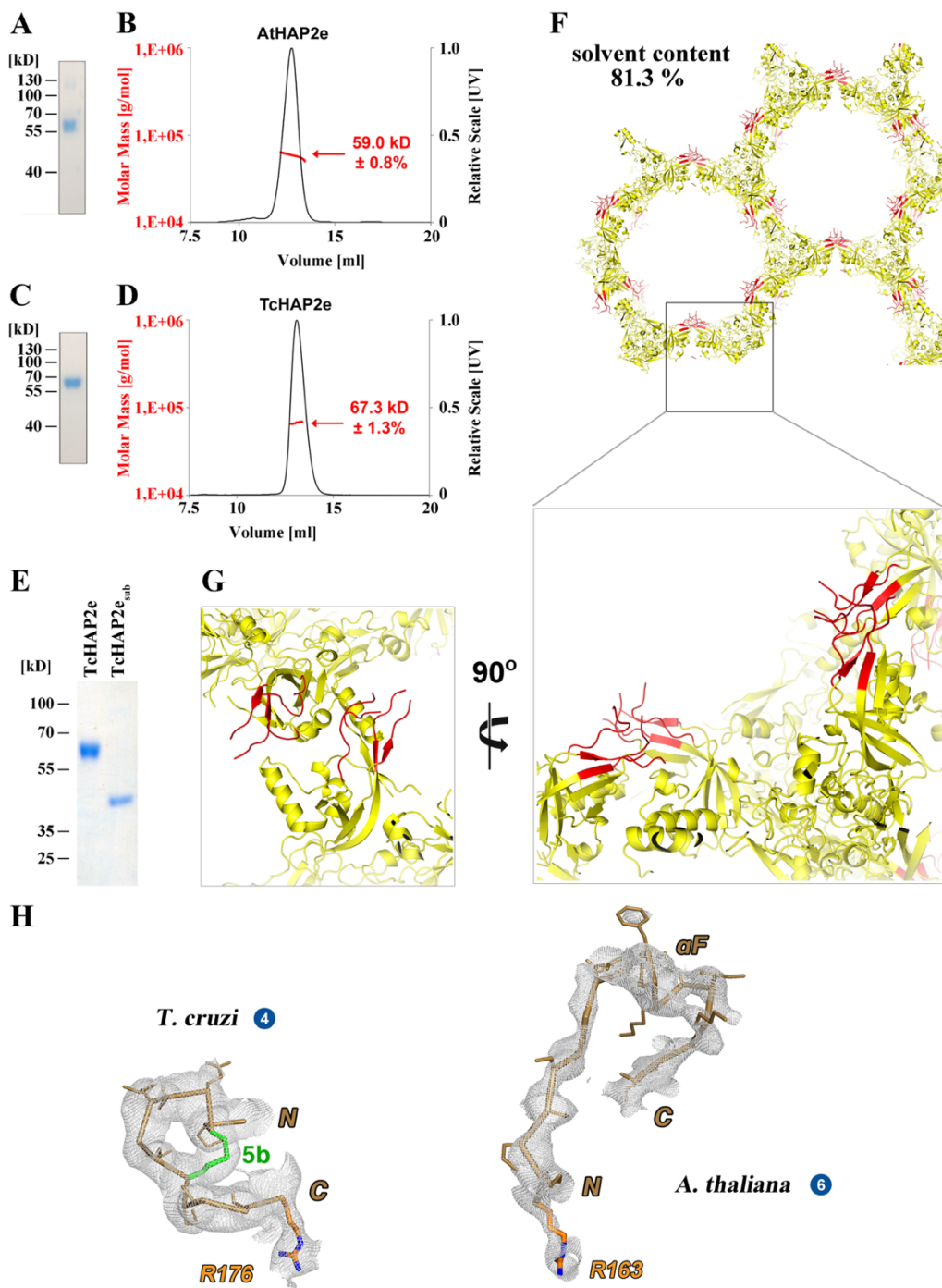


Figure S2: Characterization of AtHAP2e and TcHAP2e and crystals of a protease-resistant TcHAP2e_{sub} fragment. Related to Figure 2. (A) AtHAP2e and (B) TcHAP2e purity was assessed by Coomassie-blue-stained SDS-PAGE. SEC analysis of AtHAP2e (C) and (D) TcHAP2e on a Superdex 200 column in 10mM Tris pH8, 100mM NaCl revealed the presence of a single peak. SEC-MALS analysis indicated that this peak for both Tc and At HAP2e corresponds to a monomer. The y axis on the right gives the UV absorption at 280 nm, on a relative scale from 0 to 1 (plotted in black) while the y axis on the left gives the molecular mass by MALS (plotted and labeled in red). E. TcHAP2e and the purified protease-resistant fragment resulting from subtilisin cleavage (TcHAP2e_{sub}) were separated by SDS-PAGE stained with Coomassie blue. F-G. Overview (top panel) and close-up (bottom panel) of the TcHAP2e_{sub} crystal packing. Domain II (yellow) assembles as a rigid dodecameric ring, whereas domain I projects into solvent within this ring, explaining why domain I is only partially traceable in the experimental electron density. H. Composite omit electron density of part of the cd strand connection of TcHAP2 (left) and AtHAP2 (right) corresponding to hotspots 4 and 6, respectively, contoured at 0.9 σ , allowing for manual building of the model in this region. The side chain of the arginine involved in the conserved salt bridge is shown as stick, other side chains as lines.

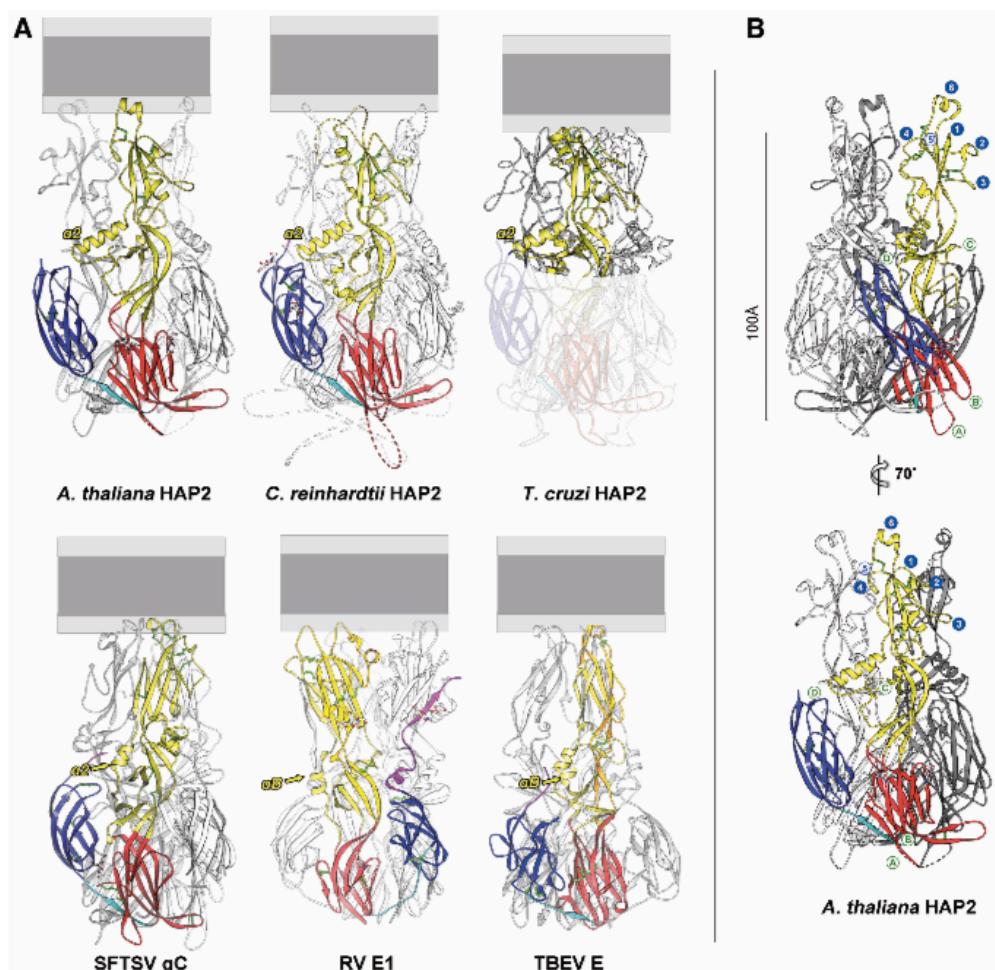


Figure S3: Comparison of trimeric class II fusion protein structures. Related to Figure 6. **A.** Comparison of the observed *Cr* and *At*HAP2 trimers with a putative *Tc*HAP2 trimer (top panel) and viral class II fusion protein trimers (bottom panel). Note the conserved overall organization of the trimers. All proteins are colored as in Figure 2. HAP2 structures are aligned vertically using the structurally conserved helix $\alpha 2$; the striking divergence of viral class II fusion proteins precluded an equivalent alignment. The model for the *T. cruzi* HAP2 ectodomain in its post-fusion conformation presented in (**A**), right panel, was generated using the PHYRE II server (Kelley et al., 2015); the resulting model was superposed on either the *At* or *Cr*HAP2 trimers and the experimentally determined structure of *T. cruzi* HAP2 domain II was then superposed and merged into this trimer. The two independent *Tc*HAP2 trimer models generated from either the *At*HAP2 or *Cr*HAP2 were structurally very similar with rmsd = 2.6Å for 711 common Ca atoms (237 per subunit). The PHYRE II-derived parts of the protein are shaded in white. **B.** View of the *At*HAP2 trimer with insertion hotspots in the *bc* and *cd* connections indicated with blue encircled numbers as in Figures 1 and 6. Green encircled letters highlight insertion points outside of the membrane interacting region, at the end of domain I opposite to the domain I/II interface. These loops on domain I interact with domain III in the pre-fusion form the viral class II proteins.

Trypanosoma cruzi

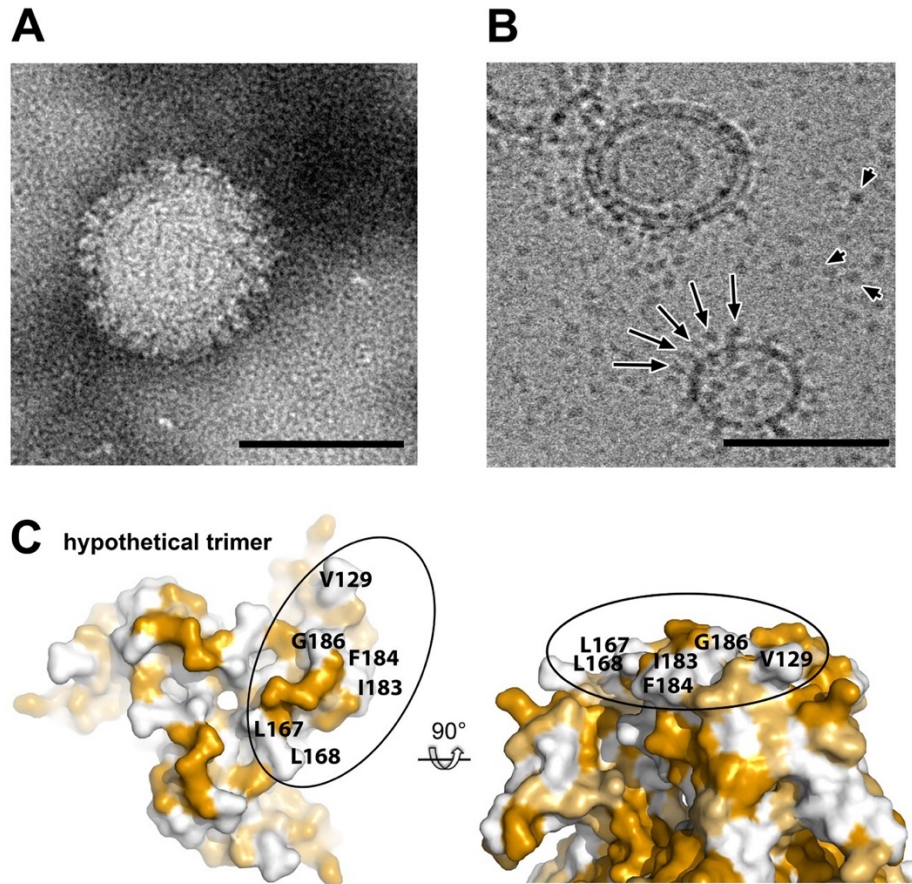


Figure S4: *T. cruzi* HAP2e trimers insert into liposomes using a hydrophobic tip. Related to Figure 4. **A.** Surface representation of the membrane facing tip of the putative *T. cruzi* HAP2 trimer obtained as described in the legend to Fig. S3 viewed from the membrane (left panel) and from the side (right panel). Surface residues are colored according to hydrophilicity from dark orange (hydrophilic) to white (hydrophobic). The fusion loop region from one protomer is encircled and hydrophobic residues that are exposed at the tip are labeled. Residue 129 belongs to the *bc* connection, the others to *cd*. **B-C.** Membrane insertion of *T. cruzi* HAP2e. Electron micrographs of liposomes incubated in presence of HAP2e were analyzed by negative staining (**B**) or cryo-EM (**C**). Scale bar 50nm. Liposomes decorated with HAP2e display protein projections at the surface. The protein projections (some of which are indicated by long arrows in **C**) are shaped as tapered rods similar to the ones of *A. thaliana* HAP2e (see Figure 4) and form lateral assemblies. The background is coated with proteins not bound to liposomes (short arrows in **C**).

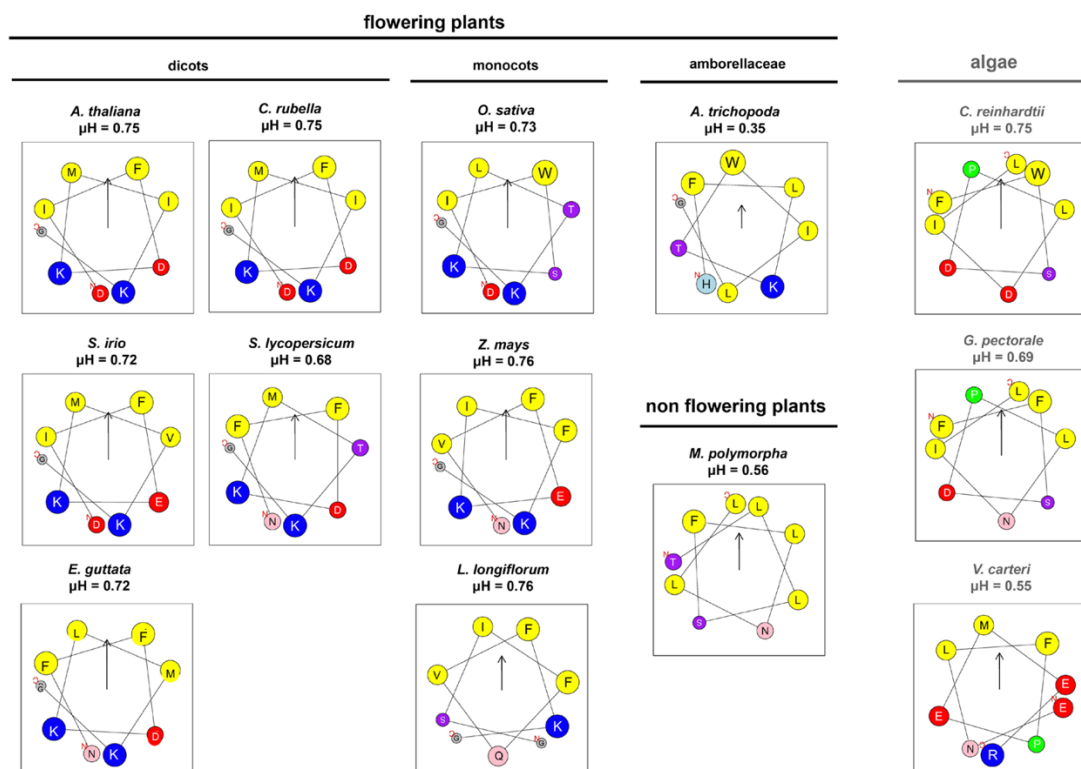


Figure S5: Amphipathic helix prediction in loop 2 of plant and algal HAP2. Related to Figure 3. Helical wheel projections of HAP2 loop 2 residues from indicated organisms. The circle size is proportional to the size of the displayed side chain. Non-polar residues are colored in yellow. Charged residues are colored red (negative charge) and blue (positive charge). Polar uncharged residues are purple and light pink. Glycines are displayed in grey, prolines in green. The hydrophobic moment values of the amphipathic alpha helices μH predicted by Heliquist is shown below the organism name, suggesting the formation of an amphipathic helix in the putative membrane-interacting region of plant and algal HAP2.

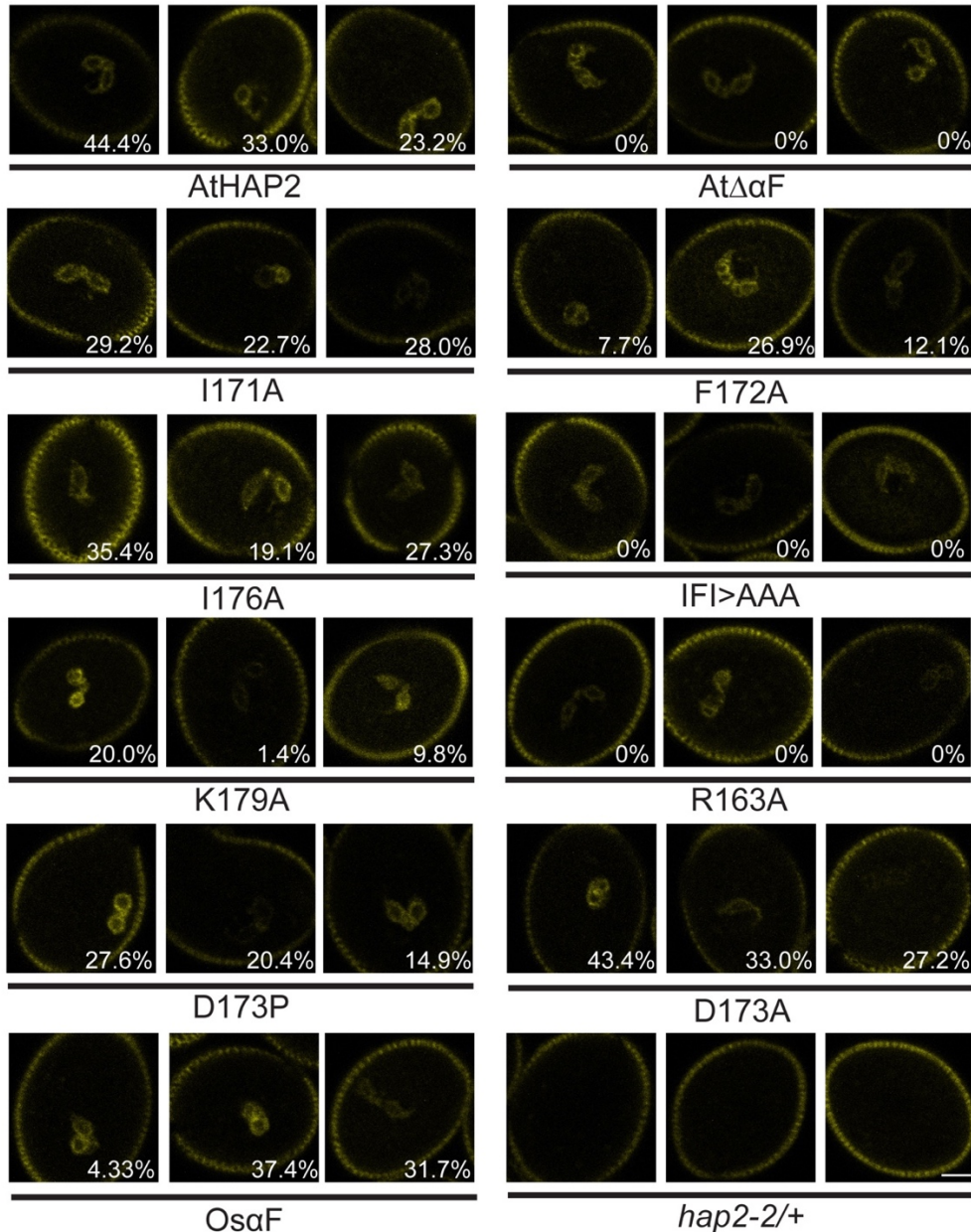


Figure S6: Analysis of HAP2variant:YFP accumulation in sperm cells. Related to Figure 5. Confocal micrographs of pollen grains from three independent transgenic lines expressing the indicated HAP2:YFP variant construct are shown. Each pollen grain shown contains a pair of sperm representing the maximum YFP intensity observed in the represented transgenic line. Signal from the pollen grain wall is due to autofluorescence and observed in the *hap2-2/+* control, which does not contain a HAP2: YFP construct. Supplemental Table 3 lists the number of independent transgenic lines that were produced for each HAP2:YFP variant and tested using our genetic transmission assay for HAP2 function (Figure 5). The average percent transmission of *hap2-2* (33% is expected for a fully functional single insertion-site line, Figure 5) for each line is indicated in the bottom right corner of each image. Scale bar, 5μm.

Table S1. Accession numbers of amino acid sequences used in this study. Related to Figure 1.

Organism	Protein	mRNA/cds	genomic
<i>Chlamydomonas reinhardtii</i>	A4GRC6.2 ¹	XM_001695841 .1	NW_001843718.1; FUSM
<i>Gonium pectorale</i>	BAO57179.1	AB915401.1	AB915402.1
<i>Volvox carteri</i>	EFJ46133.1	-----	GL378353.1
<i>Cyanidioschyzon merolae</i>	XP_005536505 .1	XM_005536448 .1	NC_010137.1; CYME_CMK076C
<i>Selaginella moellendorffii</i>	XP_002971451 .1	XM_002971405 .1 (partial)	NW_003314278.1; SELMODRAFT_451371; GCS1-2
<i>Physcomitrella patens</i>	-----	-----	Pp Chr. 25 10991182 – 11030780*
<i>Marchantia polymorpha</i>	BAV38815.1	LC172179.1	Mapoly0066s0032 ²
<i>Arabidopsis thaliana</i>	AAY51999.1	NM_117241.2	NC_003075.7; HAP2
<i>Capsella rubella</i>	XP_006289435 .1	XM_006289373 .1	NW_006238918.1; CARUB_v10002940mg
<i>Sisymbrium irio</i>	ADE20442.1	GU724984.1 (partial)	-----
<i>Erythranthe guttata</i>	XP_012846387 .1	XM_012990933 .1	NW_012193798.1; LOC105966372
<i>Solanum lycopersicum</i>	XP_019070618 .1	XM_019215073 .1	NC_015445.2; LOC101265357
<i>Lilium longiflorum</i>	BAE71142.1	AB206810.1	-----
<i>Zea mays</i>	NP_001307741 .1	NM_001320812 .1	NC_024460.2; LOC100274016
<i>Oryza sativa</i>	BAF16968.1	AK072871.1	LOC_Os05g18730 ²
<i>Amborella trichopoda</i>	XP_020531486 .1	XM_020675827 .1	NW_006500346.1; LOC18447729
<i>Toxoplasma gondii</i>	KYF43852.1	-----	AGQS02004567.1
<i>Cryptosporidium muris</i>	EEA05614.1	XM_002139927 .1	NW_002196571.1; CMU_026210
<i>Plasmodium falciparum</i>	XP_001347424 .1	PF3D7_101420 0 ³	KI926414.1

<i>Eimeria tenella</i>	BAM16295.1	AB723702.1	ETH_00017050/ETH_00017055 ^{4,5}
<i>Tetrahymena thermophila</i>	AIA57699.1	KJ629172.1	GG662480.1
<i>Paramecium tetraurelia</i>	CAK65033.1	XM_001432393.1 (partial)	NW_001799006.1; GSPATT00034511001
<i>Trypanosoma cruzi</i>	EAN93043.1	XM_809801.1 (partial)	NW_001849404.1; Tc00.1047053509105.4
<i>Trypanosoma congolense</i>	CCC94129.1	-----	HE575323.1
<i>Trypanosoma brucei</i>	EAN78468.1	-----	CM000208.1
<i>Leishmania major</i>	XP_003722443.1	XM_003722395.1 (partial)	NC_007284.2; LMJF_35_0460
<i>Naegleria gruberi</i>	EFC41606.1	XM_002674304.1	NW_003163287.1; NAEGRDRAFT_70404
<i>Dictyostelium discoideum</i>	BAS29571.1	LC075766.1	DDB_G0276069 ⁶
<i>Physarum polycephalum</i>	BAE71144.1	AB206812.1	-----
<i>Monosiga brevicolis</i>	EDQ88884.1	XM_001746445.1	NW_001865049.1; MONBRDRAFT_8819
<i>Amphimedon queenslandica</i>	XP_019853252.1	XM_019997693.1	NW_003546408.1; LOC109582771
<i>Nematostella vectensis</i>	EDO36432.1	XM_001628445.1 (partial)	NW_004172600.1; LOC101240634
<i>Hydra vulgaris</i>	XP_004211319.2	XM_004211271.2	NW_004172600.1; LOC101240634
<i>Capitella teleta</i>	ELU07639.1	-----	KB299712.1
<i>Tribolium castaneum</i>	EFA06462.1	XM_008196950.2	NC_007422.5; LOC103313508
<i>Acyrtosiphon pisum</i>	XP_016661643.1	XM_016806154.1	NW_003384018.1; LOC100569456
<i>Orchesella cincta</i>	ODN05384.1	-----	LJIJ01000024.1
<i>Saccoglossus kowalevskii</i>	XP_006821859.1	LOC102804021	NW_003142801; LOC102804021

¹UniProtKb/Swiss-Prot ID in NCBI. ²Phytozome ID. ³PlasmoDB ID. ⁴ToxoDB. ⁵Walker et al. 2015. ⁶DictyBase ID. *Hand annotation using RNA-seq data, personal communication with A.-C. Lindner and J. Becker, Plant Genomics Lab, Instituto Gulbenkian de Ciência

Table S2. Data collection, phasing and refinement statistics. Related to Figure 2.

	AtHAP2			TcHAP2		
Data collection						
Space group	$P 6_3$			$P 6_1 22$		
Cell dimensions						
a, b, c (Å)	77.26	77.26	219.53	150.10	150.10	123.80
α, β, γ (°)	90.0	90.0	120.0	90.0	90.0	120.0
Resolution (Å)	49.33-2.75 (2.82-2.75) ^a			49.12-3.1 (3.31-3.1)		
R_{meas}	0.256 (6.2)			0.163 (1.74)		
$I/\sigma(I)$	6.08 (0.37)			11.4 (1.3)		
$CC_{1/2}$	0.998 (0.36)			0.997 (0.41)		
Completeness (%)	99.7 (99.2)			97.7 (98.2)		
Redundancy	7.0 (6.9)			8.9 (8.6)		
Refinement						
Resolution (Å)	49.38-2.75 (3.07-2.75)			49.12-3.1 (3.31-3.1)		
No. reflections	10646			15430		
$R_{\text{work}} / R_{\text{free}}$	0.231 / 0.277			0.233 / 0.26		
No. atoms						
Protein	3511			1855		
Solvent	43			2		
B factors						
Protein	67.80			84.33		
R.m.s deviations						
Bond lengths (Å)	0.007			0.008		
Bond angles (°)	0.91			0.99		
Ramachandran plot [§]						
Favoured (%)	94.1			96.1		
Allowed (%)	5.9			3.9		
Outliers (%)	0			0		

One crystal was used to collect each of the diffraction data sets used to determine the crystal structure.

^a Values in parentheses are for highest-resolution shell.

^b After ellipsoidal truncation performed using the StarAniso server (Global Phasing Ltd.).

[§] Ramachandran statistics were calculated with MolProbity

Table S3. Transgenic lines used to analyze function of HAP2:YFP variants. Related to Figure 5.

Construct ID	HAP2 Variant	Lines (n)	Crosses (n)	Seedlings / Cross (avg.)	Progeny (n)
pgl294	AtHAP2	8	39	41.9	1635
pgl1070	AtΔαF	6	28	29.5	825
pgl1165	I171A	5	33	43.9	1448
pgl1061	F172A	7	44	34.4	1513
pgl1057	I176A	8	46	41.3	1900
pgl1315	IFI>AAA	6	26	24.4	634
pgl1371	R163A	4	24	30.0	721
pgl1510	D173P	8	47	46.0	2164
pgl1374	D173A	7	39	38.6	1507
pgl1056	K179A	7	33	36.2	1195
pgl1151	OsaF	9	46	34.6	1592
	<i>hap2-2</i> /HAP2	n/a	39	31.9	1244
Total numbers (n) of independent transgenic lines (Lines), cross-pollinations (Crosses), progeny tested (Progeny), and the average number of seedlings analyzed for each cross pollination (Seedlings/Cross) are given for each construct. ID, construct identification number.					

Table S4. Primers used for generation of HAP2:YFP variants and for genotyping and sequencing transformed plants. Related to Figure 5.

Primer Name	Primer	Reason
hap2SeqTR3	CAATCAAACCTGCGCAGAAGGAAGC	<i>hap2-2</i> genotyping
Hap2C2	GGCCTCACTCGTTCTCAATTGGAG	<i>hap2-2</i> genotyping
Lba1	TGGTTCACGTAGTGGGCCATCG	<i>hap2-2</i> genotyping
HAP2Ex7R	GATTAACCTTTGAGAAAGTTATCATTTG	sequencing variants
ECORI_HAP2_CVFP_F1	gaattcCTTACTTACATTAGGGATGTACCT	sequencing variants
HAP2_FL_F172A_F1	GTGGAGATATTGCTGATAAGATG	F172A
HAP2_FL_F172A_R1	AAGAAGAAGGCATTCTTCGTT	F172A
HAP2_FL_K179A_F1	GATTAAAGGGGCGGCAAATACC	K179A
HAP2_FL_K179A_R1	ATCTTATCAAAAATATCTCCACAAGAAGAAGGC	K179A
HAP2_FL_I176A_F1	TGATAAGATGGCTAAAGGGAAG	I176A
HAP2_FL_I176A_R1	AAAATATCTCCACAAGAAGAAGGC	I176A
HAP2_del_pFL_F1	GCAAATACCGCGCATTGCCTTCG	Δ 169-179
HAP2_del_pFL_R1	ACAAGAAGAAGGCATTCTTCGTTGCGGTCC	Δ 169-179

HAP2_OspFL_R1	CTTACTCCAAATATCTCCACAAGAAGAAGGCAT	Os.FH replacement
HAP2_OspFL_F1	TTGACAAAAGGGAAGGCAAATACCGCGCA	Os.FH replacement
HAP2_I171A_triple_R1	AAGAAGGCATTCTTCGTTGCGGTC	I171A
HAP2_I171A_F2	CTTGTGGAGATGCTTTTGATAAGAT	I171A
HAP2_triple_build_R2	CTTATCAGCAGCATCTCCACAAGAAGAAGGCATTC	IFI>AAA
HAP2_triple_build_F2	ATGGCTAAAGGGAAGGCAAATACCGCGCATTG	IFI>AAA
R185A_AtHAP2_R1	CACATGGACAACAAATCGGCTGAGTC	R163A
R185A_AtHAP2_F1	GACCGCAACGAGCAATGCCTTCTTCT	R163A
HAP2_D173A_F1	GAGATATTTTTGCTAAGATGATTAAAGGGAAGGC	D173A
HAP2_D173A_R1	CACAAGAAGAAGGCATTCTTCGTTGC	D173A
HAP2_D173P_F1	GTGGAGATATTTTTCCCAAGATGATTAAAG	D173P
HAP2_D173P_R1	AAGAAGAAGGCATTCTTCGTT	D173P

Chapter 3: Expression of HAP2 in the female gametes as an opportunity to assess models of HAP2 sufficiency

Jennifer Forcina, Samyukta Mallick, Samantha Paul, Alexander Leydon, and Mark Johnson.

This chapter is currently unpublished. I performed all crosses and image analysis in this chapter. I prepared the figures and the manuscript with advice from Dr. Mark Johnson. Samyukta Mallick helped with the female HAP2 cloning. Samantha Paul and Alexander Leydon generated the egg cell membrane marker.

Summary

HAP2 is a gamete fusion protein and required for fertilization in a broad range of eukaryotic species. In all species tested, HAP2 is only required on one membrane to mediate fusion. However, the exact fusion mechanism for HAP2 is not known. Therefore, we wanted to develop assays to test models of HAP2 sufficiency in plant cells. We introduced HAP2 fusion proteins into cells of the female gametophyte to test whether HAP2 could function unilaterally, either by rescuing fertilization in fusion-defective sperm or by initiating ectopic fusion events within the ovule. HAP2-GFP expression was detected in the egg cell was not limited to the egg cell membrane. We were unable to detect HAP2 fusion protein expression in the central cell, and HAP2-GFP expressed in the synergid cell was not membrane-localized. We did not observe ectopic fusion phenotypes. Additionally, we did not observe rescue of fusion from female gametes containing our construct encoding HAP2-GFP in the egg cell and HAP2-mCherry in the central cell with HAP2-deficient, fusion defective sperm in our assay. We also tested a role for HAP2 in a cooperative model of fusion where HAP2 variants with differing levels of function were expressed in the sperm and HAP2 was introduced into the female gametes, but we did not detect increased fusion in our assay. Therefore, we did not observe evidence for a cooperative model of HAP2-mediated fusion in our assay. However, we did not rule out the potential for a cooperative HAP2 gamete fusion model, and additional experiments are needed to clearly determine the mechanism of increased fusion by HAP2 when it is present in more than one membrane.

Introduction

Membrane fusion proteins are required for viral fusion, vesicle fusion, and cell-cell fusion events, and they act directly in fusing membranes (Aguilar et al., 2013; Jahn and Scheller, 2006; Oren-Suissa and Podbilewicz, 2007; White et al., 2008; Wong and Johnson, 2010). A fusion protein must be sufficient to mediate ectopic fusion events (Aguilar et al., 2013; Oren-Suissa and Podbilewicz, 2007; Wong and Johnson, 2010). HAP2 has structural homology to class II fusion proteins ((Fedry et al., 2017; Fedry*, Submitted); Chapter 2), a group encompassing both viral fusion proteins (Gibbons et al., 2004; Lescar et al., 2001; Modis et al., 2004; Rey et al., 1995) and the somatic cell fusion protein, epithelial fusion failure -1 (EFF-1; (Perez-Vargas et al., 2014)). Class II viral fusion proteins induce fusion unilaterally by direct insertion into the target membrane of a hydrophobic fusion loop at its membrane interacting surface (Bressanelli et al., 2004; Gibbons et al., 2003; Gibbons et al., 2004; Huang et al., 2010; Kielian and Rey, 2006; Modis et al., 2004; Stiasny et al., 2002). Both a short *Tetrahymena thermophila* HAP2 peptide that includes the proposed fusion loop (Pinello et al., 2017) and the entire *Chlamydomonas reinhardtii* and *Arabidopsis thaliana* HAP2 ectodomains ((Fedry et al., 2017; Fedry*, Submitted); chapter 2) are able to interact directly with artificial membranes. Additionally, in vitro functional analysis of the membrane interacting surface of HAP2 identified hydrophobic amino acids that are essential for insertion into liposomes for *Chlamydomonas* and *Arabidopsis* ((Fedry et al., 2017; Fedry*, Submitted); chapter 2). The membrane interacting surface of *Chlamydomonas* HAP2 was not resolved in the structure (Fedry et al., 2017). *Arabidopsis* HAP2 has an amphipathic fusion helix, with a requirement for a hydrophobic patch at its membrane interaction surface, and we previously identified several mutants in the fusion helix that impair HAP2 function ((Fedry et al., 2017; Fedry*, Submitted); chapter 2).

Expressing Arabidopsis HAP2 in a mammalian cell fusion assay demonstrated sufficiency of HAP2 as a fusion protein (Valansi et al., 2017). However, HAP2 was required on both cells for fusion, which suggests the potential for a bilateral fusion mechanism (Valansi et al., 2017). HAP2 could also fuse, to similar levels, with a cell expressing EFF-1 on the opposing membrane, raising the possibility that EFF-1 and HAP2 could function in trans to initiate membrane fusion (Valansi et al., 2017). EFF-1 has a small acidic patch in the region corresponding to the viral fusion loop (Perez-Vargas et al., 2014), and during development in *Caenorhabditis elegans*, EFF-1 fuses cells through a homotypic bilateral fusion process (Mohler et al., 2002; Podbilewicz et al., 2006; Shemer et al., 2004). The fusion mechanism of EFF-1 is not known, but it has been proposed to form trimers in *trans* (Valansi et al., 2017), bringing the two membranes together in a method reminiscent of SNARE proteins (Jahn and Scheller, 2006). EFF-1 can mediate fusion in combination with another *C. elegans* development fusion family (FF) protein, anchor-cell fusion failure -1 (AFF-1; (Avinoam et al., 2011)). The ability of EFF-1 to mediate fusion with HAP2 could suggest a broad ability of some class II fusion proteins to mediate fusion through a heterotypic, bilateral fusion mechanism. An alternative hypothesis is that HAP2 can act both unilaterally and bilaterally in a proposed 'hybrid mechanism' (Pinello et al., 2017), or cooperative model. In *Tetrahymena thermophila*, HAP2 is expressed on all gametes but is sufficient for fusion when present on only one membrane (Cole et al., 2014). However, the rate of gamete fusion increases greatly when HAP2 is present on both fusing gametes (Cole et al., 2014). Therefore, one could propose that the bilateral action of HAP2 in *Tetrahymena* (Cole et al., 2014) and in the ectopic mammalian fusion assay (Valansi et al., 2017) could be explained by a 'cooperative' model where HAP2 can function in *trans* to mediate fusion, in addition to an ability to function unilaterally in membrane fusion (Pinello et al., 2017).

Arabidopsis HAP2 is only expressed in sperm and is required from the male for gamete fusion in vivo (Johnson et al., 2004; Mori et al., 2006; von Besser et al., 2006), supporting a unilateral mechanism. However, the possibility of an undiscovered class II fusion protein on the female side cannot be dismissed (Valansi et al., 2017). The genetic data, combined with structural and biochemical evidence, supports a model of direct membrane interaction but does not exclude the possibility of a ‘cooperative’ mechanism ((Fedry*, Submitted; Johnson et al., 2004; Mori et al., 2006; von Besser et al., 2006); chapter 2).

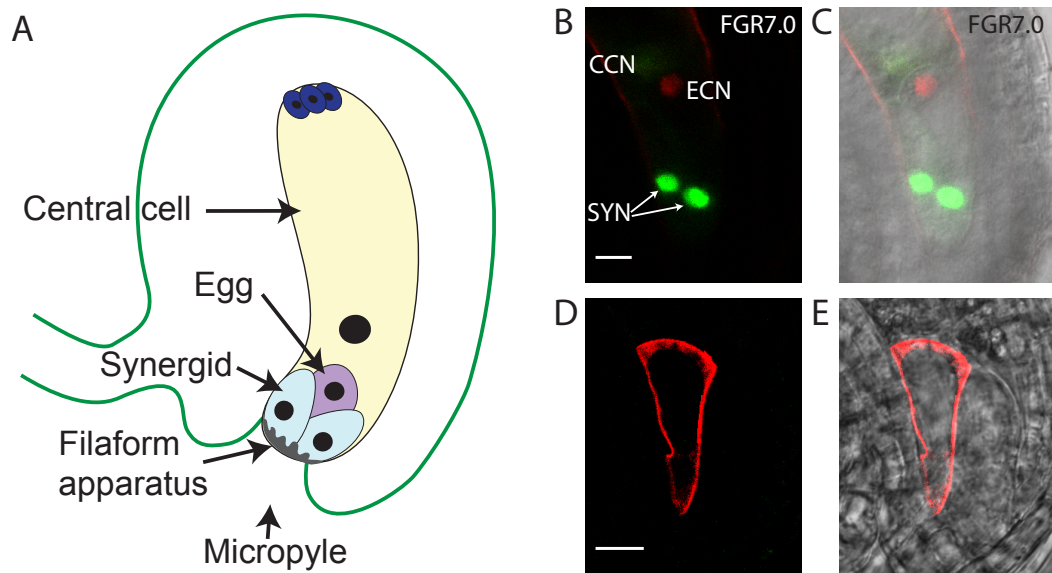


Figure 1. Overview of an Arabidopsis ovule highlighting the location and shape of the egg cell. **A.** A schematic of an Arabidopsis ovule with the micropylar end oriented toward the bottom of the page. The female gametes (egg and central cell) are indicated, as well as the synergid cells. The filiform apparatus at the micropylar end of the synergid cells is also indicated. **B-C.** Expression of FGR7.0 (Volz et al., 2013) in an ovule. FGR7.0 contains fluorescent markers labeling the synergid nuclei with GFP, the egg cell nucleus with RFP, and the central cell nucleus with YFP. The confocal image is shown as a merge of the RFP and GFP channels (**B**) and then as a merge of RFP, GFP, and DIC (differential interference contrast; **C**). CCN= central cell nucleus. ECN = egg cell nucleus. SYN = synergid cell nuclei. Scale bar is 10um. **D-E.** An ovule expressing SUC9-mCherry in the egg membrane. The ovule is from a plant that is expressing both SUC9-mCherry and HAP2-GFP. The GFP and RFP channels are merged in panel **D**, and the RFP and DIC channels are merged in **E**. Scale bar is 10um.

Arabidopsis sperm are not motile and instead they are transported to the female gametes by a pollen tube (Faure et al., 2002; Hamamura et al., 2011; McCue et al., 2011). The pollen tube is guided to the ovule by attractant molecules secreted by two synergid cells within the ovule (Takeuchi and Higashiyama, 2012). Synergid cells are located at the micropylar end of the ovule (Figure 1A), where the pollen tube will eventually enter (Dresselhaus et al., 2016). The two synergid cells, three antipodal cells, and 2 gametes (egg and central cell) comprise the female gametophyte ((Drews and Yadegari, 2002); Figure 1A-C), and the egg and central cell each must fuse with a sperm cell for successful seed development, or production of a viable offspring (Berger et al., 2008). After the pollen tube enters the ovule, one synergid degenerates and the pollen tube bursts, releasing the sperm to the site of fusion next to the egg and central cells (Faure et al., 2002; Hamamura et al., 2011). EC1 (egg cell-1) proteins are secreted from the egg cell as a consequence of pollen tube burst, and EC1 proteins are proposed to activate sperm by enhancing HAP2 accumulation at the sperm membrane (Sprunck et al., 2012). Sperm do not fuse immediately after they are delivered to the site of fusion, but instead experience a delay, or wait period, before fusion occurs (Hamamura et al., 2011).

To further investigate the sufficiency of HAP2 to mediate fusion, we introduced HAP2 into female gametes, providing us the opportunity to test whether HAP2 from the egg cell can rescue fusion with sperm that are deficient in HAP2 and therefore unable to mediate fusion (Johnson et al., 2004; Mori et al., 2006; von Besser et al., 2006). We also tested whether HAP2 expression in female gametes increases the level of fusion in partially defective HAP2 mutants ((Fedry*, Submitted); Chapter 2), which would support a cooperative fusion model of direct insertion supplemented with *trans*-oligomerization to enhance fusion events. We further tested whether we could express HAP2 in the synergid cells.

Results

HAP2 expression in the egg cell is not sufficient to rescue fusion.

Sperm that lack HAP2 are unable to fuse (Johnson et al., 2004; Mori et al., 2006; von Besser et al., 2006), so we wanted to test whether HAP2 expression in the egg cell would be sufficient to fuse the egg to fusion-defective sperm. We successfully expressed both an egg membrane marker (Figures 1D-E and 2F-H) and a HAP2-GFP fusion protein in the egg cell under the EC1.1 promoter (pEC1.1:HAP2-GFP; Figure 2). Some HAP2-GFP expression appears to be localizing to the membrane region of the egg cell, as indicated by an egg membrane marker (Figure 2E-H and Figure S1A). HAP2 expression is also detected in the egg cell after successful pollen tube burst and membrane fusion (Fig S1A). Expression of HAP2 in the egg cell did not appear to negatively affect fertilization rates (Table S1). Additionally, expression of HAP2 in the egg alone was unable to rescue the fertilization defect in *hap2* sperm (Tables S2).

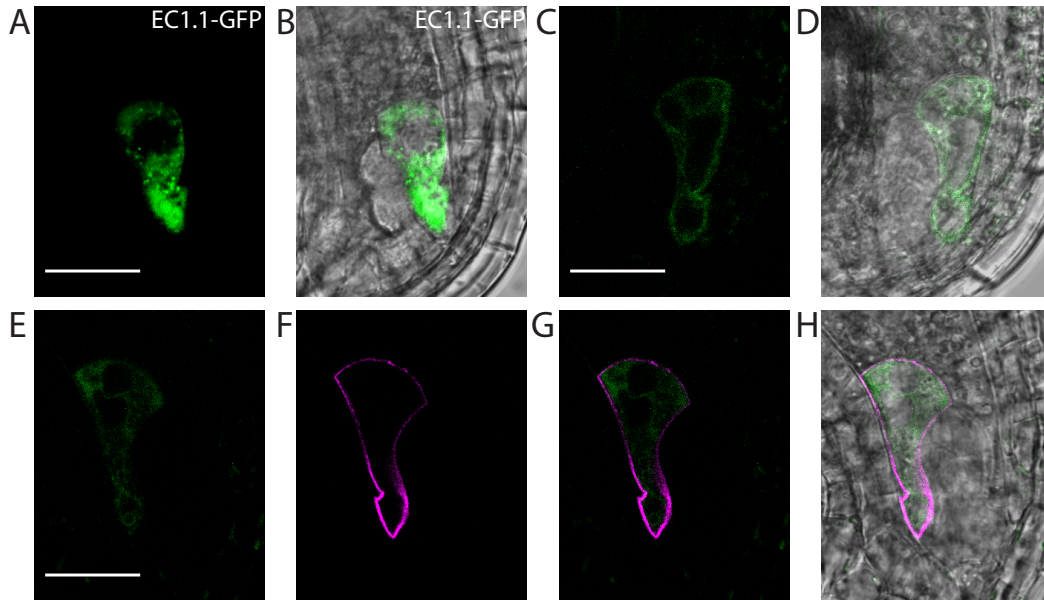


Figure 2. HAP2 expression in the egg cell. **A-B.** Confocal image of an ovule expressing EC1.1-GFP under the EC1.1 promoter (Sprunck et al., 2012) in the egg cell. The GFP channel alone is shown in panel **A**, and the GFP and DIC channels are merged in panel **B**. **C-D.** Confocal image of HAP2-GFP in the egg cell within an ovule from an emasculated

pistil. GFP channel is shown in panel **C**, and panel **D** is a merge of the GFP and DIC channels. **E-H**. Confocal image of an ovule expressing both HAP2-GFP and SUC9-mCherry under the EC1.1 promoter. The GFP channel is shown in **E**, while the RFP channel is shown in **F**. In panel **G**, the GFP and RFP channels are merged, while GFP, RFP, and DIC are merged in panel **H**. Scale bar is 20um.

We tried to express HAP2-GFP in the egg cell simultaneously with HAP2-mcherry in the central cell (EC1.1promoter:HAP2-GFP/DD65promoter:HAP2-mCherry plasmid; Figure 3). Again, HAP2-GFP expression could be detected readily in the egg cell, but HAP2-mCherry was not detected in the central cell (Figure 3). HAP2-mCherry expression may be present in the central cell but in low levels that are not detectable due to the size of the central cell and its localization within the embryo sac, which is highly auto-fluorescent. To test whether the choice of fusion protein had an effect on detection of HAP2 in the central cell, we also generated a construct to try to express HAP2-GFP in the central cell, but we did not observe HAP2-GFP expression in the central cell (data not shown). Crosses with plants expressing HAP2 in the egg cell and containing the transgene encoding HAP2 in the central cell did not rescue the *hap2* fusion defect (Table 1).

Table 1. Transmission data for EC1.1:HAP2-GFP/DD65:HAP2-mCherry transformed plants pollinated with pollen from *hap2-1/+* plants.

	T1 x <i>hap2-/+</i> data				
T1	# BastaS	# BastaR	% BastaR	n value	# crosses
<i>ms1</i>	195	0	0	195	5
1.1	27	0	0	27	2
1.4	42	0	0	42	2
2.1	21	0	0	21	2
2.2	16	0	0	16	2
<i>ms1</i> = <i>male-sterile 1</i> pistils. T1 indicates primary transformed line analyzed. Basta sensitivity (BastaS) and resistance are shown (BastaR). n = total number of seeds across the number of crosses listed.					

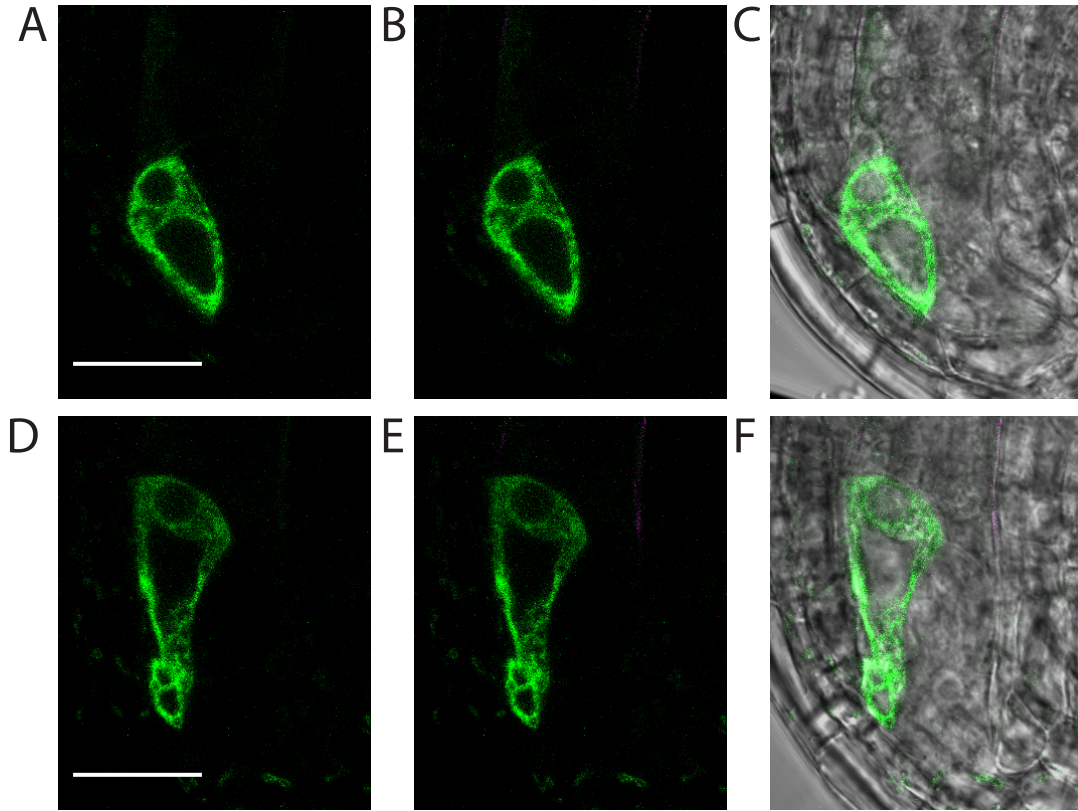


Figure 3. Ovules from EC1.1:HAP2-GFP/DD65:HAP2-mCherry plants. Confocal images of ovules containing the dual construct encoding HAP2-GFP in the egg cell under the EC1 promoter and HAP2-mCherry in the central cell under the DD65 promoter. **A-C.** One ovule expressing HAP2-GFP in the egg cell but HAP2-mCherry was not observed in the central cell. **D-F.** Another example of an ovule expressing HAP2-GFP in the egg cell but no observable HAP2-mCherry in the central cell. The GFP channel is shown in panels **A** and **D**. A merge of GFP and RFP is shown in panels **B** and **E**. Panels **C** and **F** merge GFP, RFP, and DIC channels. Scale bar is 20um.

We tested whether there was a negative effect on double fertilization and therefore seed production as a result of introducing our EC1.1:HAP2-GFP/DD65:HAP2-mCherry construct into female gametes by performing sex-specific cross-pollinations tests (Table 2). The male-specific transmission assay with *male-sterile-1* (*ms1*) females was used to assess transmission of the transgene through the sperm, where our construct should not be expressed, as a rough indicator of whether multiple insertions of the transgenes had occurred. In a single-copy line, the transmission of the transgene is expected to be 50%,

and therefore, we would expect to see 50% transmission of the antibiotic resistance associated with the transgene (Hygromycin). While we observed one plant (T1 = 2.2; Table 2) that had decreased transmission through the female, this was not consistently observed and therefore is more likely to be a result of the insertion site of the transgene rather than an effect of the expression of the transgene.

Table 2. Sex-specific crosses for the first generation of plants transformed with EC1.1:HAP2-GFP/DD65:HAP2-mCherry.

	ms1 x T1 data				T1 x Col-0 data				
	#	#	%		#	#	%		#
T1	HygS	HygR	HygR	n	HygS	HygR	HygR	n	crosses
1.1	105	97	48.02	202	15	24	61.54	39	2
1.2	42	55	56.70	97	24	25	51.02	49	1
1.4	79	91	53.53	170	17	13	43.33	30	1
2.1	41	101	71.13	142	23	39	62.90	62	3
2.2	105	92	46.70	197	19	6	24.00	25	2
2.3	68	94	58.02	162	n/a	n/a	n/a	n/a	n/a
3.1	91	90	49.72	181	n/a	n/a	n/a	n/a	n/a
Genetic expectation			50.00	50.00					
ms1 = male-sterile 1 female. ms1x T1 data is male specific. T1 indicates the primary transformant line analyzed. T1 x Col-0 is female specific. Hygromycin sensitivity (HygS) and resistance (HygR) are indicated. N= total number of seeds analyzed across total number of crosses.									

HAP2 expression from female cells did not improve fusion efficiency for partially functional HAP2 mutants.

While introducing HAP2 in the egg and central cell is not sufficient in our assay to mediate gamete fusion (Tables 1), we reasoned that HAP2 expression from the female side may improve fusion capabilities of partially functional HAP2 mutants ((Fedry*, Submitted); Chapter 2), which would provide support for a cooperative mode of fusion. We tested several mutants we published previously, including mutations affecting hydrophobic residues at the membrane interacting region of the fusion helix, specifically F¹⁷², I¹⁷¹, and I¹⁷⁶, and a lysine residue at the base of the fusion helix (K¹⁷⁹) that may be

important after initial membrane interaction ((Fedry*, Submitted); Chapter 2). We used pollen from plants expressing our HAP2 variants fused to YFP (HAP2-YFP variant; ((Fedry*, Submitted); Chapter 2)) and heterozygous for the *hap2-2* mutant allele (Alonso et al., 2003; von Besser et al., 2006) to pollinate pistils from plants that express HAP2-GFP in the egg cell and have the potential to express HAP2-mCherry in the central cell (Figure 1). We analyzed transmission of *hap2-2* through the male as a measure of gamete fusion ((Fedry*, Submitted; Mori et al., 2010; Wong et al., 2010); Chapter 2)). We saw no effect of HAP2 expression in female gametes on *hap2-2* transmission (Table 3), suggesting that in our assay there is no cooperativity between HAP2 on the egg and sperm cell that results in increased gamete fusion events leading to successful fertilization.

HAP2 expression in the synergid cell does not cause aberrant fusion events.

Although we did not observe rescue of sperm-egg fusion in our assay with HAP2 introduced into the egg and central cell (Table 1), we wanted to test whether expression of HAP2 in the synergid cells would result in ectopic fusion events with other cells within the ovule. For example, if HAP2 is expressed in the synergid cell membrane and is sufficient for fusion in this context, without a need for any other proteins or specific conditions, as previously observed in the ectopic mammalian cell fusion assay (Valansi et al., 2017), then HAP2 could cause aberrant cell fusion events between the synergid cells, potentially interfering with the function of the synergid cells to attract pollen tubes. Although it is possible that fusion of synergid cells may not interfere with their ability to secrete proteins that attract pollen tubes but rather with synergid degeneration after pollen tube arrival. Another possibility is that the synergid cells could fuse with the egg cell, perhaps interfering with the function of the egg cell, such as secretion of EC1 proteins or the ability to develop into an embryo. If high rates of ectopic fusion were to initiate from HAP2-expressing synergid cells, we would expect fertilization and possibly pollen tube

attraction to be affected. However, HAP2-GFP was not localized to the synergid cell membrane, which would preclude its ability to initiate plasma membrane fusion. In support of functional synergid cells, we observed successful pollen tube attraction and pollen tube burst in ovules expressing HAP2-GFP in their synergid cells (Figure 4).

Table 3. Transmission analysis of HAP2-YFP variants crossed to wildtype pistils or pEC1.1:HAP2-GFP/pDD65:HAP2-mCherry pistils.							
Female	Male	KanS	KanR	n	%KanR	Std Dev	# Crosses
Female-HAP2	<i>hap2-2/+</i>	166	0	166	0.00	0.00	5
Col0 (WT)	<i>hap2-2/+</i>	207	0	207	0.00	0.00	5
Female-HAP2	HAP2cds_25.2	123	81	204	39.71	8.15	4
Col0 (WT)	HAP2cds_25.2	117	102	219	46.58	6.29	4
Female-HAP2	HAP2cds_32.2	125	47	172	27.33	4.65	4
Col0 (WT)	HAP2cds_32.2	126	47	173	27.17	5.52	4
Female-HAP2	Δ FH_2.2	153	0	153	0.00	0.00	4
Col0 (WT)	Δ FH_2.2	159	0	159	0.00	0.00	4
Female-HAP2	F172A_6.3A	129	53	182	29.12	14.94	4
Col0 (WT)	F172A_6.3A	123	44	167	26.35	5.86	4
Female-HAP2	F172A_6.3B	62	23	85	27.06	1.50	2
Col0 (WT)	F172A_6.3B	49	12	61	19.67	3.24	2
Female-HAP2	IFI>AAA_1.1A	152	0	152	0.00	0.00	4
Col0 (WT)	IFI>AAA_1.1A	171	0	171	0.00	0.00	4
Female-HAP2	IFI>AAA_1.1B	165	0	165	0.00	0.00	4
Col0 (WT)	IFI>AAA_1.1B	176	0	176	0.00	0.00	4
Female-HAP2	IFI>AAA_2.4	102	0	102	0.00	0.00	4
Col0 (WT)	IFI>AAA_2.4	93	0	93	0.00	0.00	4
Female-HAP2	R163A_9.2	138	0	138	0.00	0.00	4
Col0 (WT)	R163A_9.2	100	0	100	0.00	0.00	3
Female-HAP2	K179A_2.3	111	46	157	29.30	5.49	4
Col0 (WT)	K179A_2.3	99	46	145	31.72	3.45	4
Female-HAP2 indicates pEC1.1:HAP2-YFP/pDD65:HAP2-mCherry pistils. Col0 (WT) indicates wildtype pistils. The pollen donors were HAP2-YFP variants in the <i>hap2-2/+</i> background (SALK_152706) and were in the second generation after transformation (T2). For the male donor, the HAP2 variant is listed first followed by the primary transformed line (T1). The A/B at the end of the male donor name indicates two different T2 plants were analyzed from one T1 line. The <i>hap2-2</i> allele has a kanamycin marker associated with it. KanS = # kanamycin sensitive seedlings. KanR = # kanamycin resistant seedlings. n = total number of sensitive and resistant seedlings for all crosses (indicated in # Crosses column) for that mating pair. %KanR = percent of seedlings that were kanamycin resistant and therefore received the <i>hap2-2</i> allele. Std Dev = standard deviation of %KanR for all crosses for each mating pair.							

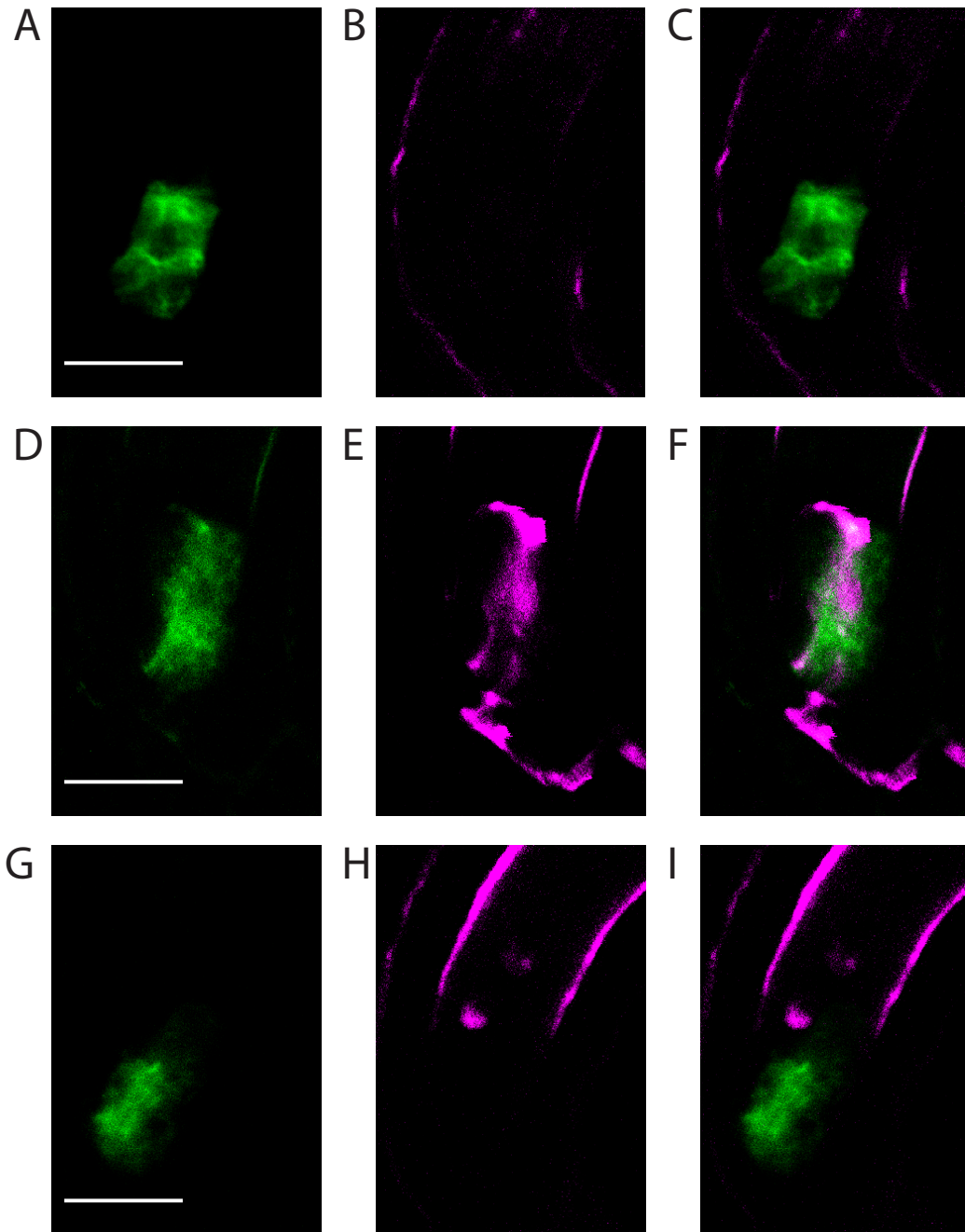


Figure 4. HAP2-GFP expression in synergid cells. All ovules shown here were from flowers that were emasculated and dissected the next day. Ovules in panels **D-I** were pollinated the morning after emasculatation and dissected ~7 hours after pollination. **A-C.** Confocal image of an ovule expressing HAP2-GFP in the synergid cells that was emasculated but not pollinated the next day. The GFP channel is shown in panel **A**, the RFP channel in panel **B**, and the merge of the two channels in panel **C**. **D-F.** Confocal image of an ovule expressing synergid-specific HAP2-GFP that was from a flower pollinated with LAT52:dsRED pollen (Francis et al., 2007), which labels the pollen tube

cytoplasm, and dissected approximately 7 hours after pollination. The confocal image is shown as GFP only (**D**), RFP only (Pollen tube cytoplasm, **E**), and RFP and GFP channels merged (**F**). **G-I**. Confocal image of an ovule from a pistil expressing synergid-specific HAP-GFP that was analyzed 7 hours after pollination with pollen expressing a sperm nuclei marker (HTR10:HTR10:mCherry). Synergid-GFP is shown in the GFP channel (**G**), while the sperm nuclei marker is shown in the RFP channel (**H**). Panel I shows the GFP and RFP channels from **G** and **H**, respectively, merged together. Scale bar represents 20um.

Discussion

In this study, we have ectopically expressed HAP2-GFP in the egg cell and the synergid cells of the female gametophyte from *A. thaliana*. We have shown that some HAP2-GFP in the egg cell may localize to the membrane, but its localization was not restricted to the plasma membrane. We were unable to detect HAP2 fusion proteins in the central cell, and we did not observe membrane localization of HAP2 in the synergid cells. Introduction of HAP2 fusion proteins in female gametes was not sufficient for fusion in our assay nor was it able to rescue low levels of fusion with sperm expressing partially functional HAP2-YFP variants. Combined, the results from this study demonstrate the difficulty in assessing the direct mechanism of HAP2-mediated gamete fusion and emphasize the importance in understanding the proteins and other elements involved in regulating HAP2 function.

Our goal was to use plant-based assays to address whether Arabidopsis HAP2 is sufficient to fuse plant cells when ectopically expressed and whether ectopic fusion events would occur unilaterally or bilaterally. Specifically, we also wanted to test a cooperative model for HAP2 fusion, where HAP2 expressed from both membranes would increase fusion levels over its expression on one membrane. We did not observe rescue of fusion-defective *hap2* sperm with HAP2 in the egg cell nor did we observe an increase in fusion levels when testing a cooperative model of HAP2 fusion. However, double fertilization in flowering plants requires two fusion events: sperm fusion with the egg cell and sperm

fusion with the central cell (Berger et al., 2008; Iwakawa et al., 2006). We introduced a single transgene construct directing expression of HAP2-GFP in the egg cell and HAP2-mCherry in the central cell, but we were unable to confirm HAP2 expression in the central cell.

We used seed set, resulting from successful double fertilization, as a method to assess HAP2 function in female gametes. However, we can assess whether a sperm-egg fusion event occurs without central-cell fusion using either confocal microscopy (Beale et al., 2012; Hamamura et al., 2011) or ovule clearing approaches (Wong et al., 2010). Alternatively, HAP2 expression in the egg cell may not have been present in high enough quantities to mediate fusion in our assay. However, HAP2 surface expression was low in the mammalian cell fusion assays (Valansi et al., 2017), and we do not currently know the amount of HAP2 required at the sperm surface for fusion to occur. It is also possible that enough HAP2 protein was present on the egg cell but HAP2 was not active for fusion. Therefore, we cannot rule out the possibility of a cooperative model of HAP2 fusion in vivo.

We observed the ability of synergid cells expressing HAP2-GFP to receive pollen tubes and to experience gamete fusion, and we observed no obvious aberrant effects or signs of ectopic cell fusion events. However, these observations were not quantified. It could prove useful to test whether more subtle effects of HAP2 expression in the synergid can be observed, such as effects on pollen tube attraction or delays in gamete fusion.

Altogether, our results demonstrate that while we can drive HAP2 expression in some cells within the female gametophyte to test HAP2 sufficiency in plant cells, there are likely components of the HAP2-mediate gamete fusion mechanism that are missing. For example, if a receptor for GEX2 was identified, we could try expressing the receptor on the sperm cell, while expressing HAP2 and GEX2 in the egg cell and central cell, thereby providing the specific adhesion interaction that may provide a more suitable environment

for gamete fusion to occur. Additionally, if the receptors for EC1 proteins were identified, perhaps we could force HAP2 to the plasma membrane of synergid cells. By understanding the requirements for HAP2 activation, we could perform fusion assays in a more conducive environment. Furthermore, identification of proteins necessary for proper HAP2 transport and expression at the cell membrane would allow for those proteins to be co-expressed with HAP2 in specific cells during fusion assays.

Methods

Plant growth conditions were as described previously ((Fedry*, Submitted); Chapter 2).

Generation of constructs for expression in female cells.

Constructs were generated using the Greengate cloning system (gift from Jan Lohmann: Addgene kit #1000000036; (Lampropoulos et al., 2013)). The HAP2 coding sequence was sub-cloned using primers HAP2cDNA_F1 and HAP2cDNA_F2 (see primer table, Table S3) from pgl290 (pHAP2:HAP2:YFP; (Fedry*, Submitted); chapter 2), and TA cloning was used to ligate the HAP2 PCR product into the Xcm-I digested C000 entry vector. EC1.1 promoter was cloned from Col-0 DNA using pEC1.1_F1 (Sprunck et al., 2012) and pEC1.1_R1 (Table S3). DD65 promoter was cloned from Col-0 DNA using pDD65_F1 and pDD65_R1 (Table S3; sequences adapted from (Denninger et al., 2014)). MYB98 promoter (Kasahara et al., 2005) was cloned from Col0 DNA using primers pMYB98_F2 (1559 basepair (bp) upstream of ATG) and pMYB98_R2 (88 bp upstream of ATG). All promoter sequences were TA cloned into XcmI-digested A000 entry vector (Lampropoulos et al., 2013). SUC9cDNA was sub-cloned from pgl691 (gift from Anke Reinders; (Leydon et al., 2017)) and TA cloned into C000. PCR reactions were performed using Phusion High-Fidelity DNA Polymerase (New England Biolabs (NEB) # M0530S) with 5X Hi-Fi buffer and dNTPs (NEB# N0447S). TA cloning was performed using GoTaq

Flexi DNA Polymerase (Promega# M8291) and dATP (NEB# N0440S), as well as T4 DNA ligase (NEB# M0202L). Constructs were generated ligating respective promoters, B-dummy (B003), HAP2cDNA in C000 (pgl986), either linker-GFP (D001) or linker-mCherry (D003), UBQ10 terminator (E009), and Hygromycin resistance (F005) into pGGZ003 destination vector, using the initial longer reaction conditions: 50 cycles of alternating between 37°C for 5 minutes and 16°C for 5 minutes then termination with 5 minutes at 50°C and 5 minutes at 80°C (Lampropoulos et al., 2013). The GreenGate reaction formula followed previously published recommendations (Lampropoulos et al., 2013), using CutSmart Buffer (NEB# B7204S), 10mM ATP (NEB# P0756S), T4 DNA Ligase (NEB# M0202L), and BsaI-HF (NEB# R3535S). For the combined EC1.1:HAP2:GFP and DD65:HAP2:mcherry construct, each fusion protein was generated in an intermediate vector (M000 and N000) then combined in the destination vector (Z003) according to the recommended protocol (Lampropoulos et al., 2013).

Generation of transgenic plants with constructs built using GreenGate cloning.

Plasmids generated by GreenGate cloning were transformed with the pSoup helper plasmid (Hellens et al., 2000) into *Agrobacterium* (strain GV3101; (Koncz and Schell, 1986)) cells by electroporation. Plants were transformed using the floral dip method (Clough and Bent, 1998). Primary transformed seeds (T1) were sterilized as described previously ((Fedry*, Submitted); chapter 2) and selected on Murashige and Skoog (MS) plates (Sigma Aldrich, St. Louis, MO) supplemented with the appropriate drug associated with their transgene: hygromycin B at 25mg/L (Corning, Tewksbury, MA). Female HAP2 constructs were transformed into wild-type Col0 plants. T1 seeds on plates were cold-treated for at least 2 days before growth for 7-14 days at 22°C under constant light prior to transfer to soil.

Generation of plants expressing HTR10:HTR10:mCherry construct..

HTR10 promoter was sub-cloned from pgl532 using HindIIIpHTR10F and SpelpHTR10R primers (Table S3) and TA cloned into the GreenGate pGGA000 promoter entry vector. HTR10/MGH3 coding sequence was sub-cloned from plasmid U89237 (Arabidopsis Biological Resource Center (ABRC)) using primers BaseMGH3F and BaseMGH3R and TA cloned into pGGC000. The new HTR10:HTR10:mChery transgene was built using the GreenGate reaction as described above, but with the following entry vectors: HTR10 promoter, B-dummy, HTR10cds, linker-mCherry, UBQ10 terminator, and Kanamycin resistance (F007) in the Z003 destination vector. The HTR10:HTR10:mCherry plasmid was transformed into Agrobacteria using pSoup helper plasmid as described above. HTR10:HTR10:mCherry was transformed into mDUO1 plants (mDUO1 seeds were a gift from Dr. David Twell) that contain a transgene encoding an estradiol-inducible mutated version of the DUO1 transcription factor that is resistant to microRNA159 (Brownfield et al., 2009). However, the HTR10:HTR10:mCherry;mDUO1 plants using in these experiments were not induced with estradiol and were instead used as a pollen donor expressing a sperm nuclei marker, similar to previously published experiments using an HTR10:HTR10:RFP sperm nuclei marker (Ingouff et al., 2007).

Imaging plants with fluorescent markers in the ovule.

Plants from the second generation after transformation (T_2) for both constructs were crossed with EC1.1:HAP2:GFP as the female and EC1.1:SUC9:mCherry as the male (pollen donor). Plants were imaged in the F_1 generation (first generation after crossing). For assessments of fertilization, plants were emasculated and pollinated the next day with either LAT52:dsRED to label the pollen tube (Francis et al., 2007) or HTR10:HTR10:RFP to label the sperm nuclei (Ingouff et al., 2007). Pollen was allowed to grow in vivo for ~7 hours (7hpp) before dissection and mounting in 80mM sorbitol. Images were taken using a Zeiss LSM 800 confocal with 488 laser for GFP and 561 laser

for mCherry and processed using FIJI version of ImageJ (Schindelin et al., 2012). Ovules expressing FGR7.0 or EC1.1:EC1.1:GFP were also imaged with the Zeiss 800 confocal.

Genetic transmission analysis of HAP2 in female gametes.

Plants in the first generation after transmission with either EC1.1:HAP2:GFP or EC1.1:HAP2:GFP/DD65:HAP2:mCherry were analyzed for transmission of the transgene. For EC1.1:HAP2:GFP/DD65:HAP2:mCherry, *male sterile-1* (*ms1*; (Wilson et al., 2001)) females were pollinated with pollen from T1 plants to assess transmission through the male and crosses were analyzed for hygromycin resistance associated with the transgene. Flowers from T1 plants were emasculated (all the anthers that produce the pollen were removed) then pollinated the next day with either wildtype pollen or pollen expressing HTR10-RFP in their nuclei (Ingouff et al., 2007) and analyzed for transmission through the female by testing hygromycin resistance. To test whether HAP2 expression in the female could rescue the *hap2* fusion defect, T1 plants were emasculated, pollinated with *hap2*-1/+ pollen (Johnson et al., 2004; Wong et al., 2010), and the resulting seeds were analyzed for Basta resistance (25mg/L; Glufosinate ammonium, Oakwood Chemicals, Estill, South Carolina).

Testing the effect of HAP2 in female gametes on transmission of HAP2:YFP variants.

HAP2:YFP variants ((Fedry*, Submitted); chapter 2) in the T2 generation were used as pollen donors. HAP2:YFP variant T2 plants were genotyped for the *hap2-2* allele (SALK_152706; (Alonso et al., 2003; von Besser et al., 2006)) as described previously ((Fedry*, Submitted); chapter 2). Pollen was also screened for the HAP2:YFP expression in pollen, but not for homozygous expression. Both Col0 flowers and EC1.1:HAP2:GFP/DD65:HAP2:mcherry T3 flowers were emasculated and pollinated the next day with pollen from HAP2:YFP variant T2 plants that were heterozygous for *hap2-2*, as indicated by genotyping for the *hap2-2* allele as described previously ((Fedry*,

Submitted);chapter 2; see primer tableTable S3). Crosses were allowed to dry before collection. EC1.1:HAP2:GFP/DD65:HAP2:mcherry plants suggested most plants were heterozygous for the transgene (data not shown). The F₁ seedlings were plated on MS plates with kanamycin sulfate and counted for sensitivity versus resistance to the antibiotic, with kanamycin resistance associated with transmission of the *hap2-2* allele (Alonso et al., 2003). Plants with the EC1.1:HAP2:GFP/DD65:HAP2:mcherry construct in the T3 generation (two generations after transformation) were selected for using hygromycin resistance.

Acknowledgements: We thank Dr. Anke Reinders and Dr. J.M. Ward for the SUC9 cDNA. We thank Brittany Johnson for help with preliminary female HAP2 by HAP2 variant crosses. We thank Dr. Rita Gross-Hardt for FGR7.0 seeds, Dr. Stefanie Sprunck for EC1.1:EC1.1:GFP over-expression seeds, Dr. David Twell for mDUO1 seeds, Dr. Greg Copenhagen for LAT52:dsRED seeds, and Dr. Fred Berger for HTR10:HTR10:RFP seeds.

Supplemental Data

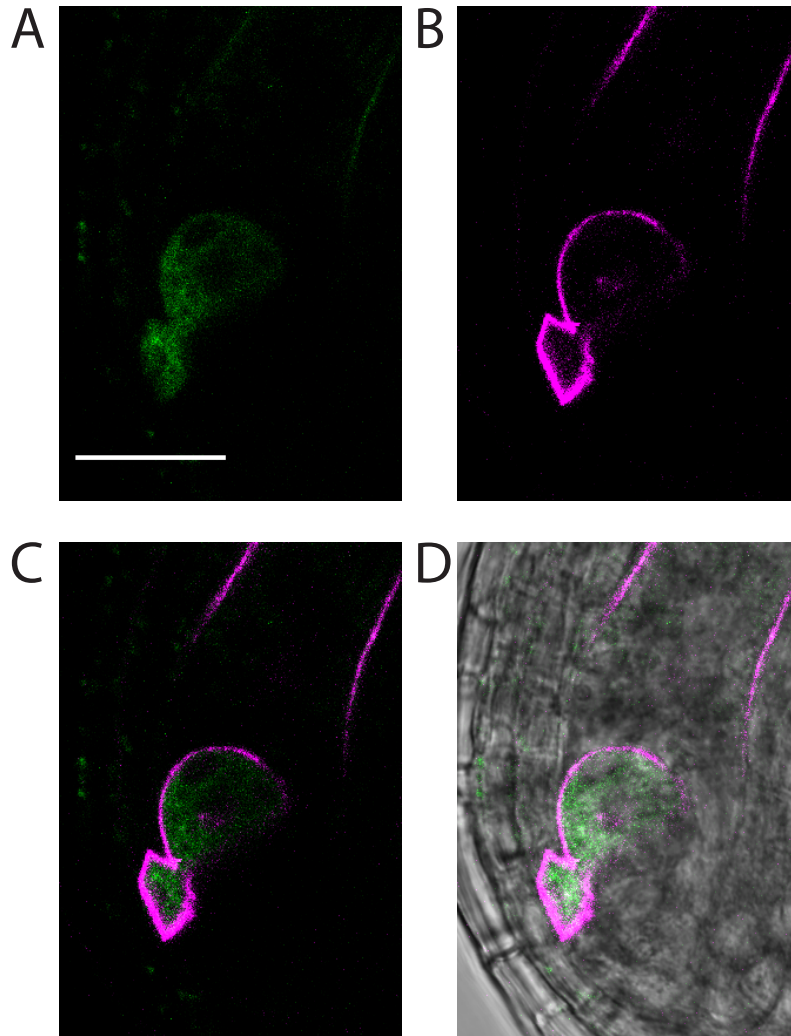


Figure S1. HAP2 expression in the egg cell after gamete fusion. A pistil expressing both HAP2-GFP and SUC9-mCherry in egg cells was pollinated with pollen expressing HTR10-RFP in sperm nuclei and analyzed ~8 hours after pollination. **A-D.** Confocal image of an ovule expressing both HAP2-GFP and SUC9-mCherry in the egg cell and showing de-condensed sperm nuclei RFP overlapping with the egg cell and the central cell. The GFP channel is shown in **A** with an absence of HAP2-GFP signal in the egg cell nucleus, while the RFP channel is shown in **B**. In panel **C**, RFP and GFP channels are merged, and an RFP-positive sperm nucleus is observed inside the egg cell, overlapping with the absence of HAP-GFP expression, and weaker RFP-positive sperm nucleus is shown in the central cell. **D.** DIC channel has been added to the merge in panel C. Scale bar represents 20um.

Tables S1. Transmission data for primary transformed plants expressing HAP2-GFP in the egg cell.

	T1 self-cross data					T1 x Col-0 data				
	#	#	%		#	#	#	%		#
T1	HygS	HygR	HygR	n	crosses	HygS	HygR	HygR	n	crosses
T1.1	28	80	74.07	108	2	92	93	50.27	185	4
T1.2	17	46	73.02	63	3	11	19	63.33	30	3
T1.3	n/a	n/a	n/a	n/a	n/a	14	17	54.84	31	3
T1.5	31	118	79.19	149	3	34	39	53.42	73	2
T1.6	34	105	75.54	139	3	56	61	52.14	117	3
T1.8	27	92	77.31	119	3	27	30	52.63	57	2
Genetic expectation			75.00					50.00		
T1 indicates the primary transformed line used for the analysis. Hygromycin sensivity (HygS) and resistance (HygR) are reported. n = total number of seedlings analyzed across the total number of crosses listed.										

Tables S2. Transmission data for primary transformed plants expressing HAP2-GFP in the egg cell pollinated with *hap2-1/+* pollen.

T1	# Basta S	# Basta R	% Basta R	n value	# crosses
T1.1	144	0	0	144	4
T1.2	36	0	0	36	3
T1.3	21	0	0	21	2
T1.5	65	0	0	65	2
T1.6	133	0	0	133	3
T1.8	56	0	0	56	2
T1 indicates the primary transformed line used in the analysis. Basta sensitivity (Basta S) and resistance (Basta R) is indicated. n = total number seedlings analyzed across the number of crosses listed.					

Table S3. Primers used in this study.

Primer Name	Primer	Reason
HAP2cDNA_F1	ATGGTGAACGCGATTTTAAT	cloning HAP2 cds
HAP2cDNA_R1	CTCTACTCTCACGTAGTCTTTGTT	cloning HAP2 cds
pEC1.1_F1	TGCCTTATGATTTCTTCGGTTTC	cloning EC1.1 promoter
pEC1.1_R1	TTCTCAACAGATTGATAAGGTCG	cloning EC1.1 promoter
pDD65_F1	AATCAAAATTTAACATTTAAATAAAGTC	cloning DD65 promoter
pDD65_R1	ATCCTTTTCTACTTTGTTTTGT	cloning DD65 promoter
pMYB98_F2	GATTGGATTTTTCAAAGATGGGTC	cloning MYB98 promoter
pMYB98_R2	GTGAAGTCTTTTTCCGATTTTGTG	cloning MYB98 promoter
HindIIIpHTR10F	ATA <u>AAGCTT</u> TACTTCTCCGACCAAAAACTTT	cloning HTR10 promoter
SpeI pHTR10R	ATA <u>CTAGT</u> TAACCGTCAATAACTTGCGAC	cloning HTR10 promoter
BaseMGH3F	ATGGCACGTACGAAGCAAAC	cloning HTR10cds
BaseMGH3R	AGCACGTTCCCCACGAAT	cloning HTR10cds
hap2SeqTR3	CAATCAAACGCGCAGAAGGAAGC	<i>hap2-2</i> genotyping
Hap2C2	GGCCTCACTCGTTCTCAATTGGAG	<i>hap2-2</i> genotyping
Lba1	TGGTTCACGTAGTGGGCCATCG	<i>hap2-2</i> genotyping

References

- Aguilar, P.S., Baylies, M.K., Fleissner, A., Helming, L., Inoue, N., Podbilewicz, B., Wang, H., and Wong, M. (2013). Genetic basis of cell-cell fusion mechanisms. *Trends Genet* 29:427-437.
- Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., et al. (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301:653-657.
- Avinoam, O., Fridman, K., Valansi, C., Abutbul, I., Zeev-Ben-Mordehai, T., Maurer, U.E., Sapir, A., Danino, D., Grunewald, K., White, J.M., et al. (2011). Conserved eukaryotic fusogens can fuse viral envelopes to cells. *Science* 332:589-592.
- Beale, K.M., Leydon, A.R., and Johnson, M.A. (2012). Gamete Fusion Is Required to Block Multiple Pollen Tubes from Entering an *Arabidopsis* Ovule. *Curr Biol*.
- Berger, F., Hamamura, Y., Ingouff, M., and Higashiyama, T. (2008). Double fertilization - caught in the act. *Trends Plant Sci* 13:437-443.
- Bressanelli, S., Stiasny, K., Allison, S.L., Stura, E.A., Duquerroy, S., Lescar, J., Heinz, F.X., and Rey, F.A. (2004). Structure of a flavivirus envelope glycoprotein in its low-pH-induced membrane fusion conformation. *The EMBO journal* 23:728-738.
- Brownfield, L., Hafidh, S., Borg, M., Sidorova, A., Mori, T., and Twell, D. (2009). A plant germline-specific integrator of sperm specification and cell cycle progression. *PLoS Genet* 5:e1000430.
- Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16:735-743.
- Cole, E.S., Cassidy-Hanley, D., Fricke Pinello, J., Zeng, H., Hsueh, M., Kolbin, D., Ozzello, C., Giddings, T., Jr., Winey, M., and Clark, T.G. (2014). Function of the male-gamete-specific fusion protein HAP2 in a seven-sexed ciliate. *Curr Biol* 24:2168-2173.
- Denninger, P., Bleckmann, A., Lausser, A., Vogler, F., Ott, T., Ehrhardt, D.W., Frommer, W.B., Sprunck, S., Dresselhaus, T., and Grossmann, G. (2014). Male-female communication triggers calcium signatures during fertilization in *Arabidopsis*. *Nat Commun* 5:4645.
- Dresselhaus, T., Sprunck, S., and Wessel, G.M. (2016). Fertilization Mechanisms in Flowering Plants. *Curr Biol* 26:R125-139.
- Drews, G.N., and Yadegari, R. (2002). Development and function of the angiosperm female gametophyte. *Annual Review of Genetics* 36:99-124.
- Faure, J.E., Rotman, N., Fortune, P., and Dumas, C. (2002). Fertilization in *Arabidopsis thaliana* wild type: developmental stages and time course. *Plant J* 30:481-488.
- Fedry, J., Liu, Y., Pehau-Arnaudet, G., Pei, J., Li, W., Tortorici, M.A., Traincard, F., Meola, A., Bricogne, G., Grishin, N.V., et al. (2017). The Ancient Gamete Fusogen HAP2 Is a Eukaryotic Class II Fusion Protein. *Cell* 168:904-915 e910.
- Fedry*, J., Forcina*, J., Legrand, P., Pehau-Arnaudet, G., Haouz, A., Johnson, M.A., Rey, F.A., and Krey, T. (Submitted). HAP2 drives gamete fusion via insertion of highly divergent fusion loops into female gametes.
- Francis, K.E., Lam, S.Y., Harrison, B.D., Bey, A.L., Berchowitz, L.E., and Copenhaver, G.P. (2007). Pollen tetrad-based visual assay for meiotic recombination in *Arabidopsis*. *Proc Natl Acad Sci U S A* 104:3913-3918.
- Gibbons, D.L., Erk, I., Reilly, B., Navaza, J., Kielian, M., Rey, F.A., and Lepault, J. (2003). Visualization of the target-membrane-inserted fusion protein of Semliki Forest virus by combined electron microscopy and crystallography. *Cell* 114:573-583.

- Gibbons, D.L., Vaney, M.C., Roussel, A., Vigouroux, A., Reilly, B., Lepault, J., Kielian, M., and Rey, F.A. (2004). Conformational change and protein-protein interactions of the fusion protein of Semliki Forest virus. *Nature* 427:320-325.
- Hamamura, Y., Saito, C., Awai, C., Kurihara, D., Miyawaki, A., Nakagawa, T., Kanaoka, M.M., Sasaki, N., Nakano, A., Berger, F., et al. (2011). Live-Cell Imaging Reveals the Dynamics of Two Sperm Cells during Double Fertilization in *Arabidopsis thaliana*. *Curr Biol* 21:497-502.
- Hellens, R.P., Edwards, E.A., Leyland, N.R., Bean, S., and Mullineaux, P.M. (2000). pGreen: a versatile and flexible binary Ti vector for *Agrobacterium*-mediated plant transformation. *Plant Mol Biol* 42:819-832.
- Huang, C.Y., Butrapet, S., Moss, K.J., Childers, T., Erb, S.M., Calvert, A.E., Silengo, S.J., Kinney, R.M., Blair, C.D., and Roehrig, J.T. (2010). The dengue virus type 2 envelope protein fusion peptide is essential for membrane fusion. *Virology* 396:305-315.
- Ingouff, M., Hamamura, Y., Gourgues, M., Higashiyama, T., and Berger, F. (2007). Distinct dynamics of HISTONE3 variants between the two fertilization products in plants. *Curr Biol* 17:1032-1037.
- Iwakawa, H., Shinmyo, A., and Sekine, M. (2006). *Arabidopsis* CDKA1;1, a cdc2 homologue, controls proliferation of generative cells in male gametogenesis. *Plant J* 45:819-831.
- Jahn, R., and Scheller, R.H. (2006). SNAREs - engines for membrane fusion. *Nat Rev Mol Cell Biol* 7:631-643.
- Johnson, M.A., von Besser, K., Zhou, Q., Smith, E., Aux, G., Patton, D., Levin, J.Z., and Preuss, D. (2004). *Arabidopsis hapless* mutations define essential gametophytic functions. *Genetics* 168:971-982.
- Kasahara, R.D., Portereiko, M.F., Sandaklie-Nikolova, L., Rabiger, D.S., and Drews, G.N. (2005). MYB98 is required for pollen tube guidance and synergid cell differentiation in *Arabidopsis*. *Plant Cell* 17:2981-2992.
- Kielian, M., and Rey, F.A. (2006). Virus membrane-fusion proteins: more than one way to make a hairpin. *Nat Rev Microbiol* 4:67-76.
- Koncz, C., and Schell, J. (1986). The Promoter of TI-DNA Gene 5 Controls the Tissue-Specific Expression of Chimeric Genes Carried by a Novel Type of *Agrobacterium* Binary Vector. *Molecular & General Genetics* 204:383-396.
- Lampropoulos, A., Sutikovic, Z., Wenzl, C., Maegele, I., Lohmann, J.U., and Forner, J. (2013). GreenGate---a novel, versatile, and efficient cloning system for plant transgenesis. *PloS one* 8:e83043.
- Lescar, J., Roussel, A., Wien, M.W., Navaza, J., Fuller, S.D., Wengler, G., Wengler, G., and Rey, F.A. (2001). The Fusion glycoprotein shell of Semliki Forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH. *Cell* 105:137-148.
- Leydon, A.R., Weinreb, C., Venable, E., Reinders, A., Ward, J.M., and Johnson, M.A. (2017). The Molecular Dialog between Flowering Plant Reproductive Partners Defined by SNP-Informed RNA-Sequencing. *Plant Cell* 29:984-1006.
- McCue, A.D., Cresti, M., Feijo, J.A., and Slotkin, R.K. (2011). Cytoplasmic connection of sperm cells to the pollen vegetative cell nucleus: potential roles of the male germ unit revisited. *J Exp Bot* 62:1621-1631.
- Modis, Y., Ogata, S., Clements, D., and Harrison, S.C. (2004). Structure of the dengue virus envelope protein after membrane fusion. *Nature* 427:313-319.
- Mohler, W.A., Shemer, G., del Campo, J.J., Valansi, C., Opoku-Serebuoh, E., Scranton, V., Assaf, N., White, J.G., and Podbilewicz, B. (2002). The type I membrane

- protein EFF-1 is essential for developmental cell fusion. *Developmental cell* 2:355-362.
- Mori, T., Hirai, M., Kuroiwa, T., and Miyagishima, S. (2010). The Functional Domain of GCS1-Based Gamete Fusion Resides in the Amino Terminus in Plant and Parasite Species. *PLoS one* 5.
- Mori, T., Kuroiwa, H., Higashiyama, T., and Kuroiwa, T. (2006). GENERATIVE CELL SPECIFIC 1 is essential for angiosperm fertilization. *Nat Cell Biol* 8:64-71.
- Oren-Suissa, M., and Podbilewicz, B. (2007). Cell fusion during development. *Trends Cell Biol* 17:537-546.
- Perez-Vargas, J., Krey, T., Valansi, C., Avinoam, O., Haouz, A., Jamin, M., Raveh-Barak, H., Podbilewicz, B., and Rey, F.A. (2014). Structural basis of eukaryotic cell-cell fusion. *Cell* 157:407-419.
- Pinello, J.F., Lai, A.L., Millet, J.K., Cassidy-Hanley, D., Freed, J.H., and Clark, T.G. (2017). Structure-Function Studies Link Class II Viral Fusogens with the Ancestral Gamete Fusion Protein HAP2. *Curr Biol* 27:651-660.
- Podbilewicz, B., Leikina, E., Sapir, A., Valansi, C., Suissa, M., Shemer, G., and Chernomordik, L.V. (2006). The *C. elegans* developmental fusogen EFF-1 mediates homotypic fusion in heterologous cells and in vivo. *Developmental cell* 11:471-481.
- Rey, F.A., Heinz, F.X., Mandl, C., Kunz, C., and Harrison, S.C. (1995). The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution. *Nature* 375:291-298.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nature methods* 9:676-682.
- Shemer, G., Suissa, M., Kolotuev, I., Nguyen, K.C., Hall, D.H., and Podbilewicz, B. (2004). EFF-1 is sufficient to initiate and execute tissue-specific cell fusion in *C. elegans*. *Curr Biol* 14:1587-1591.
- Sprunck, S., Rademacher, S., Vogler, F., Gheyselinck, J., Grossniklaus, U., and Dresselhaus, T. (2012). Egg cell-secreted EC1 triggers sperm cell activation during double fertilization. *Science* 338:1093-1097.
- Stiasny, K., Allison, S.L., Schalich, J., and Heinz, F.X. (2002). Membrane interactions of the tick-borne encephalitis virus fusion protein E at low pH. *Journal of virology* 76:3784-3790.
- Takeuchi, H., and Higashiyama, T. (2012). A species-specific cluster of defensin-like genes encodes diffusible pollen tube attractants in *Arabidopsis*. *PLoS Biol* 10:e1001449.
- Valansi, C., Moi, D., Leikina, E., Matveev, E., Grana, M., Chernomordik, L.V., Romero, H., Aguilar, P.S., and Podbilewicz, B. (2017). *Arabidopsis* HAP2/GCS1 is a gamete fusion protein homologous to somatic and viral fusogens. *J Cell Biol* 216:571-581.
- Volz, R., Heydlauff, J., Ripper, D., von Lyncker, L., and Gross-Hardt, R. (2013). Ethylene signaling is required for synergid degeneration and the establishment of a pollen tube block. *Developmental cell* 25:310-316.
- von Besser, K., Frank, A.C., Johnson, M.A., and Preuss, D. (2006). *Arabidopsis* HAP2(GCS1) is a sperm-specific gene required for pollen tube guidance and fertilization. *Development (Cambridge, England)* 133:4761-4769.
- White, J.M., Delos, S.E., Brecher, M., and Schornberg, K. (2008). Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme. *Crit Rev Biochem Mol Biol* 43:189-219.
- Wilson, Z.A., Morroll, S.M., Dawson, J., Swarup, R., and Tighe, P.J. (2001). The *Arabidopsis* MALE STERILITY1 (MS1) gene is a transcriptional regulator of male

- gametogenesis, with homology to the PHD-finger family of transcription factors. *Plant Journal* 28:27-39.
- Wong, J.L., and Johnson, M.A. (2010). Is HAP2-GCS1 an ancestral gamete fusogen? *Trends Cell Biol* 20:134-141.
- Wong, J.L., Leydon, A.R., and Johnson, M.A. (2010). HAP2(GCS1)-dependent gamete fusion requires a positively charged carboxy-terminal domain. *PLoS Genet* 6:e1000882.

Chapter 4: Assessment of pH as a trigger for HAP2-mediated gamete fusion

Jennifer Forcina, Samyukta Mallick, Jose Feijo, and Mark Johnson.

This chapter is currently unpublished. I performed all the imaging and calibration experiments, and I performed the genetic crosses. I generated the figures and drafted the manuscript with input from Dr. Mark Johnson. Samyukta Mallick helped with cloning and analysis of genetic transmission assays. Dr. Jose Feijo provided LAT52:pHluorin seeds, as well as pHluorin and pHuji plasmids.

Summary

In flowering plants, the sperm are delivered by a pollen tube to an ovule, which contains the female gametophyte. In *Arabidopsis*, the female gametophyte is composed of seven cells, including the two female gametes: the egg cell and the central cell. At the entrance to the female gametophyte are two synergid cells that secrete pollen tube attractant proteins. After pollen tube arrival at the ovule, one synergid cell degenerates and the pollen tube bursts, releasing the sperm to the site of fusion. The site of fusion is located adjacent to the egg cell and central cell, and the sperm experience a delay at the site of fusion before gamete fusion occurs. The released contents of the burst pollen tube and the remnants of the degenerating synergid cell contribute to the environment at the site of fusion. HAP2 is expressed in the sperm cell and directly mediates gamete membrane fusion with the egg cell and central cell. HAP2 is a class II fusion protein, with structural homology to class II viral fusion and somatic cell-cell fusion proteins. Class II viral fusion proteins require a decrease in pH as a trigger to initiate fusion with the target membrane. To test the hypothesis that the trigger for HAP2 is pH-dependent, we used a genetically encoded, ratio-metric pH sensor to test the pH of the pollen tube before and after burst. We observed a decrease in the pH of the pollen tube burst area in vivo compared to the pollen tube cytoplasm semi-in-vivo. A decrease at the site of fusion compared to the rest of the pollen tube burst area was also observed. We also generated plants that express pH sensors in their synergids cells with the goal of testing whether the pH of the intact synergid cell before pollen tube arrival differs from the degenerated synergid cell after pollen tube burst. Our results suggest that the sperm may be exposed to a decreased pH at the site of fusion.

Introduction

In flowering plants, sperm are delivered within a separate cell, the pollen tube, to the female gametophyte (ovule) that houses seven cells including the female gametes: the egg cell and central cell (Drews and Yadegari, 2002). Double fertilization is a specialized form of reproduction used by flowering plants, in which two sperm are delivered by the pollen tube and fuse with an egg cell and a central cell to produce an embryo and an endosperm, respectively (Berger et al., 2008). Within the pollen tube, the two sperm are connected to each other and to the vegetative nucleus of the pollen tube (McCue et al., 2011). A series of important signaling events and molecular interactions regulate the delivery of the sperm cells to the female gametes. Attraction proteins, called LUREs, are secreted from the synergid cells at the micropylar end of the ovule and attract a pollen tube in a process referred to as pollen tube reception (Okuda et al., 2009; Takeuchi and Higashiyama, 2012). The pollen tube enters the ovule and bursts, releasing the sperm and causing degeneration of the synergid cell (Faure et al., 2002; Leydon et al., 2015). Then, the sperm reside in the ovule at the site of fusion, in close proximity to the egg cell and central cell, for approximately 7-8 minutes before fusion occurs (Hamamura et al., 2011). During those 7-8 minutes, the sperm reside in an environment composed of the contents released from the burst pollen tube and potentially from the degenerated synergid cell (Berger et al., 2008; Dresselhaus et al., 2016).

After pollen tube burst, the egg cell secretes small cysteine rich EC1 proteins that have been shown to induce translocation of the sperm-specific fusion protein HAP2 to the sperm membrane in vitro (Sprunck et al., 2012). These EC1 proteins are important for efficient gamete fusion, yet both their receptor and mode of action remain unknown. HAP2 is a broadly conserved, gamete-specific transmembrane protein (Cole et al., 2014; Ebchuqin et al., 2014; Kawai-Toyooka et al., 2014; Liu et al., 2008; Mori et al., 2006; Okamoto et al., 2016; Steele and Dana, 2009; von Besser et al., 2006), and it is required

for fertilization in *Arabidopsis* (Johnson et al., 2004; Mori et al., 2006; von Besser et al., 2006) and a variety of other eukaryotic species (Blagborough and Sinden, 2009; Camacho-Nuez et al., 2017; Cole et al., 2014; Ebchuqin et al., 2014; Hirai et al., 2008; Liu et al., 2008; Okamoto et al., 2016). HAP2 has recently been shown to be a class II fusion protein (Fedry et al., 2017) and is structurally homologous to viral fusion proteins (VFPs; (Bressanelli et al., 2004; Dessau and Modis, 2013; Gibbons et al., 2004)) and the somatic cell fusion protein EFF-1 (Perez-Vargas et al., 2014). VFPs require a trigger that enables the insertion of the fusion loop into target membranes, and class II VFPs specifically require a decrease in pH, which occurs when they enter the endosome (White et al., 2008). A trigger has yet to be identified for EFF-1. Liposome insertion assays have demonstrated that HAP2 does not have the same pH requirements for membrane interaction in vitro as its viral counterparts (Fedry et al., 2017). Therefore, it will be important to determine whether there is a trigger for HAP2, whether that trigger is pH related, and whether the trigger is the same for all flowering plants.

The ectodomain of class II fusion proteins is composed of three main domains ((Gibbons et al., 2003; Lescar et al., 2001; Modis et al., 2003; Rey et al., 1995); Figure S1A). Domain II contains the membrane interacting region in HAP2 and the fusion loop in viruses (Gibbons et al., 2004; Kielian and Rey, 2006; Lescar et al., 2001; Modis et al., 2004; Rey et al., 1995). In viruses, the fusion protein inserts into the target membrane in an extended conformation but then the proteins fold back to form 'hairpins' and bring the two membranes closer together (Kielian and Rey, 2006). During this process, the location of domain III changes such that it is folded back closer to domain I (Kielian and Rey, 2006). If exogenous domain III is added during this process in vitro, it interferes with the ability of the fusion 'hairpin' to fold back properly and complete membrane fusion ((Liao and Kielian, 2005); Figure S1). While the interfering activity of domain III has been demonstrated for

viruses and EFF-1 (Liao and Kielian, 2005; Perez-Vargas et al., 2014), it has yet to be determined for HAP2.

The pH of the growing pollen tube in vitro has been investigated both by administering a chemical sensor to lily pollen (Feijo et al., 1999) and using a genetically encoded sensor, pHluorin, expressed in the tobacco pollen tube (Michard, 2008). Both studies reported localized acidity at the tip of the growing pollen tube in vitro (Feijo et al., 1999; Michard, 2008). While these studies have contributed to our understanding of both spatial and temporal pH changes during pollen tube growth in vitro, information regarding pH changes of the pollen tube growing within the ovule, particularly during pollen tube reception and burst, is lacking. In order to test whether a pH change occurs at the site of gamete fusion, we need to be able to measure pH in the environment surrounding the sperm both inside the pollen tube and after sperm deposition in the ovule.

In this study, we have investigated several important questions that were prompted by the recent structural analyses of HAP2 proteins. First, we tested whether the pH of the environment surrounding the sperm changes, which may serve as a trigger for HAP2 function. We used genetically encoded, ratio-metric pHluorin (Michard, 2008) to assess the pH in the pollen tube before and after burst and to further test whether there the pH differs at the site where the sperm are expected to fuse. We also introduced both pHluorin and a RFP-based pH sensor, pHuji (Shen, 2014), into the synergid cells to assess potential pH changes in the synergid cells before, during, and after pollen tube burst. Our goal was not only to characterize the fusion environment but to eventually manipulate it, so we tested whether we could deliver exogenous HAP2 domain III to the site of fusion from the degenerating synergid cell. Our results show that cytoplasmic proteins expressed in synergid cells appear to be delivered to the site of gamete fusion. However, some contents of the synergid cell remained closer to micropylar end of the embryo sac after synergid degeneration, which suggests they might not be present at the site of gamete

fusion. Our studies provide the first insights into the pH environment surrounding the sperm both before and after pollen tube burst, while demonstrating differences in delivering fusion proteins to the site of fusion.

Results

The pH of the growing pollen tube is greater than the pH of its released cytoplasm.

Expression of the ratio-metric pH sensor, pHluorin, in the cytoplasm of the pollen tube provides the ability to measure the pH of the pollen tube (Michard, 2008) in different biological conditions. First, we generated a calibration curve using semi-in-vivo (SIV) growing pollen tubes (Figure 2A; (Palanivelu and Preuss, 2006)) in buffers from pH 6 to pH 8 (Figure 1). The final calibration curve was the result of combining three independent calibration experiments (Figure S2). Fluorescence intensity was measured in the same detection range after excitation at two different wavelengths (488nm and 405nm). The calibration results showed a linear relationship between the ratio of fluorescence intensity (488nm/405nm) and the pH of the calibration buffers between pH 6 and pH 8 (Figure 1D, $R^2 = 0.998$). We found the average pH of the Arabidopsis SIV pollen tube (PT) to be 8.22 ± 0.19 ($\pm$ standard deviation; SIV PT, $n=15$ pollen tubes, Figure 2C-D). The pH of the pollen tube burst environment (Figure 2E) in ovules 7 hours after pollination (Figure 2B) with pHluorin-expressing pollen is 6.15 ± 0.45 ($n=22$ burst events; Figure 2C and 2G). Since the pollen tube burst environment (Figure 2E) exists within the tissue of the female gametophyte, we wanted to test whether the pH of the intact pollen tube differed within the ovule. Therefore, we tested the pH of a coiling pollen tube (Figure 2C, 2F, 2H), an event that occurs in mutants defective in pollen tube reception (Capron et al., 2008; Huck et al., 2003; Kessler et al., 2010). We found that the pH of a coiling pollen tube in the female mutant *lorelei* (*lre*; (Capron et al., 2008; Tsukamoto et al., 2010)) is 6.42 ± 0.57 ($n=3$ coiling tubes). The pH of both coiling pollen tubes and the burst environment differ from

the pH of the SIV growing pollen tube (Figure 2C). However, the pollen tube burst environment did not differ from that of the *lre* coiling pollen tube (Figure 2C). It is possible that a coiling pollen tube may have a different pH as a result of its reproductive defect. Additionally, there may be a difference in pollen tube pH between those grown in SIV (Figure 2A) and those grown in vivo (Figure 2B). Therefore, we grew pollen tubes in the SIV assay and compared the pH of pollen tube pieces that were outside the ovule versus inside the ovule. The average pH for the portions of the pollen tube inside the ovule (7.55 ± 0.32) did not appear to differ from that of the portions outside the ovule (7.63 ± 0.24 ; Figure S3). We expressed a sperm nuclei marker in pollen tubes already expressing pHluorin and observed that pHluorin signal does not overlap with sperm nuclei (Figure S4). Pollen tubes expressing both a sperm nuclei marker and pHluorin provide an opportunity to assess the pH environment immediately surrounding the sperm as they travel through the pollen tube and arrive at the site of fusion.

The pH at the site of gamete fusion is less than the average pH of the burst event.

The site of gamete fusion is expected to be smaller than the entire area consisting of the burst pollen tube contents (Figure 2E, site of fusion would be immediately surrounding the sperm cells shown in orange). We therefore tested whether the pH of the site of gamete fusion differed from that of the entire burst area (Figure 3A). The site of fusion can be visualized in the ovule as a thin strip of burst pollen tube cytoplasm that appears to wrap around the egg cell in a “hook”-like appearance (Figure 2G and Figure 3B-E). We found that the average pH of the hook (5.26 ± 0.89) to be less than the average pH of the whole burst area including the hook (6.29 ± 0.37 ; Figure 3A; $p < 0.001$). A decrease in pH in the hook of the pollen tube burst area suggests that a pH difference may occur at the site of gamete fusion.

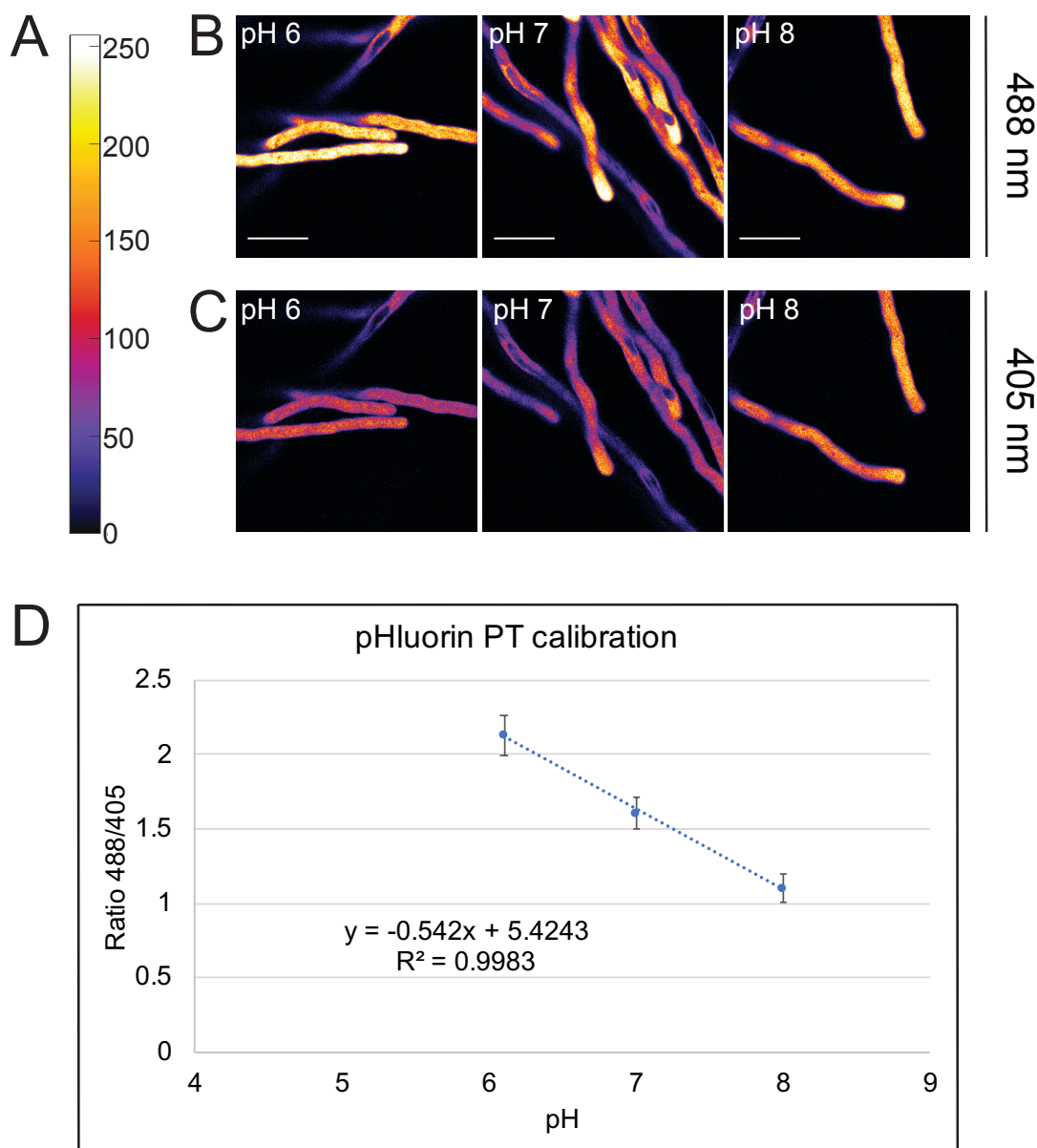


Figure 1. Calibration of pHluorin in Arabidopsis pollen tubes. pHluorin is under the pollen specific promoter LAT52 and pollen tubes grown in semi-in-vivo (SIV; see Figure 2A) were treated with three calibration buffers. **A.** Calibration bar generated in MatLab for FIRE look-up table (LUT) from ImageJ. **B-C.** Example confocal images of pollen tubes in pH calibration buffers for pH 6, 7, or 8. For panel **B**, images were taken at 488nm excitation, and the corresponding images taken at 405nm excitation are shown in panel **C**. Images are pseudo-colored with FIRE LUT from ImageJ (calibration bar shown in **B**). Scale bar represents 20um. **D.** Graph of ratio of background corrected mean fluorescence intensity at 488nm divided by corresponding values at 405nm for each calibration buffer: pH 6, 7, and 8. Calibration trials were performed in 3 different experiments with total pollen

tubes analyzed of 14, 21 and 19 for pH 6, 7, and 8, respectively. Individual trials included 3-9 pollen tubes for each pH buffer. Error bars represent standard deviation.

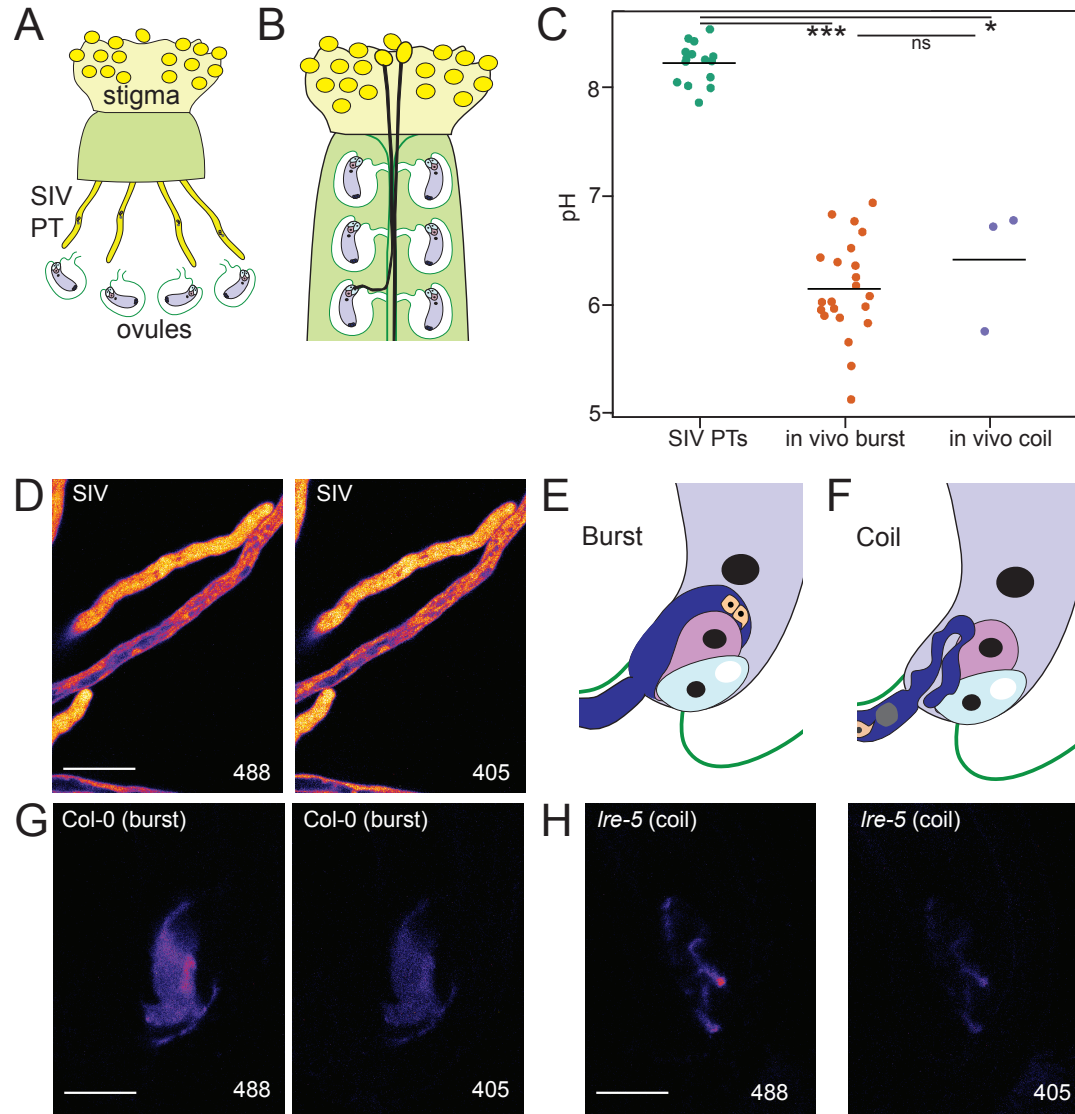


Figure 2. pH of pollen tubes grown both in SIV and in vivo for ~7 hours after pollination. **A.** Schematic of the semi-in-vivo (SIV) pollen tube growth assay. For panels **C-D**, the SIV assay was performed without ovules. **B.** Schematic of a pistil half after allowing pollen tubes to grow in vivo and then dissected for in vivo analysis (data in **C,G-H**). **C.** Graph representing calculated pH for different pHluorin samples using the equation from Figure 1A. Each dot represents an individual sample and the black line in each group represents the mean. SIV pollen tubes were analyzed in 3 different experiments with 4-6 pollen tubes each (n=15 total pollen tubes). In vivo burst events were analyzed in 4 different experiments with 2-11 events each with either Col0 females, females expressing an egg membrane marker, or females expressing a RFP-based pH sensor (see Figure 6;

n=22 total burst events). Data for coiling in *lre-5* ovules is from one experiment with 3 coiling events. **D.** Example SIV pollen tubes shown using the FIRE LUT (calibration bar shown in Figure 1B) at the indicated excitation (488nm on left, 405nm on right). **E.** Schematic showing pollen tube burst contents in dark blue and the persistent synergid in light blue. The egg is shown in pink and the central cell in light purple. The sperm are sitting at the 'site of fusion' and are shown in orange. **F.** Schematic of pollen tube coiling in an ovule. Pollen tube shown in dark blue. **G.** Example of a pHluorin pollen tube burst 7 hours after pollination (hpp). **H.** Example of a coiling pHluorin pollen tube inside a *lre-5* ovule 7hpp. **D,G-H.** Left image was taken with excitation at 488nm while the right image was taken at 405nm. All images shown are pseudo-colored with FIRE LUT as in Figure 1C-D; calibration bar shown in Figure 1B. Scale bar represents 20um.

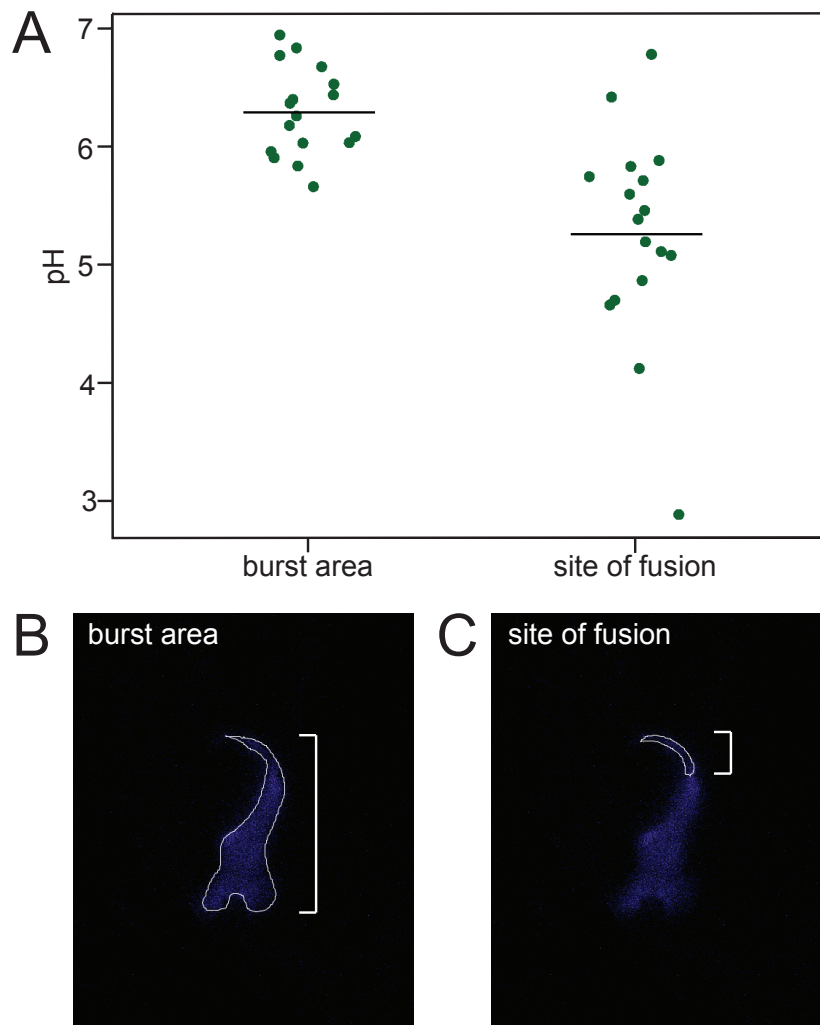


Figure 3. pH at the site of fusion. **A.** pH was calculated from ratios for the background-subtracted average intensity at 488 nm divided by 405 nm for the entire burst event (region of interest (ROI) example in panel **B**; white outline) and for the area at the site of gamete

fusion only (ROI example in panel **C**). Data is combined from 3 independent experiments (n=3-11). The ovules were from wild-type females, females expressing an egg membrane marker, and females expressing pHuji (see Figure 6). Each circle represents a single burst event. The black line represents the mean. $P < 0.001$, Student's paired t-test. **B-C**. Confocal images are from the same ovule and burst event. Panel **B** shows pHluorin signal using the FIRE LUT at 488 nm excitation with the burst ROI indicated, and panel **C** shows pHluorin signal at 488 nm excitation with the site of fusion ROI indicated.

pHluorin expressed in synergid cells is delivered to the site of gamete fusion.

The released contents of the burst pollen tube cell are expected to mix with the released contents of the degenerated synergid cell (DSYN). However, very little is known about whether the degenerated synergid cell, or even the intact one, contributes any molecules to the site of fusion. The pH of the synergid cell is unknown, both before and after pollen tube burst. Therefore, we tested whether the degenerated synergid cell contributed to the decreased pH we detected after pollen tube burst and particularly at the site of fusion. pHluorin was delivered to the site of fusion from the degenerating synergid cell after pollen tube burst (Figure 4C-H). We observed increased fluorescence in both channels in the degenerating synergid compared to persistent (PSYN) or untargeted synergid cells (Figure 4). We observed no significant difference between the ratio for the degenerating synergid cell compared to the persistent synergid cell (Student's t-test; Figure S5).

Calibration of pHluorin in the synergid cell may be more difficult than in the pollen tube. Calibration of a pollen tube expression pHluorin can be performed in vitro or in the SIV system. Both approaches to calibrating pollen tubes are likely more direct than trying to calibrate the synergid cells within the ovule due to the tissue surrounding the micropyle, or entrance to the ovule where the synergid cells rely. Also, the filiform apparatus which is an enrichment of cell wall material at the micropylar end of the ovule (Dresselhaus et al., 2016) may hinder the efficiency of the calibration buffer to reach the synergid cells. We used the SIV assay to test calibration of synergid cells expressing pHluorin. A linear

relationship between ratio and pH was observed after calibration with buffers at pH 6, 7, and 8 (Figure 5), but more analysis is needed to confirm that the synergid cells are being properly calibrated within the ovule. Furthermore, the difference in ratios across the three pH buffers was subtle. Therefore, rather than reporting the pH of the synergid, I have analyzed the change in ratio for pHluorin expression in the synergid cells (Figure S5).

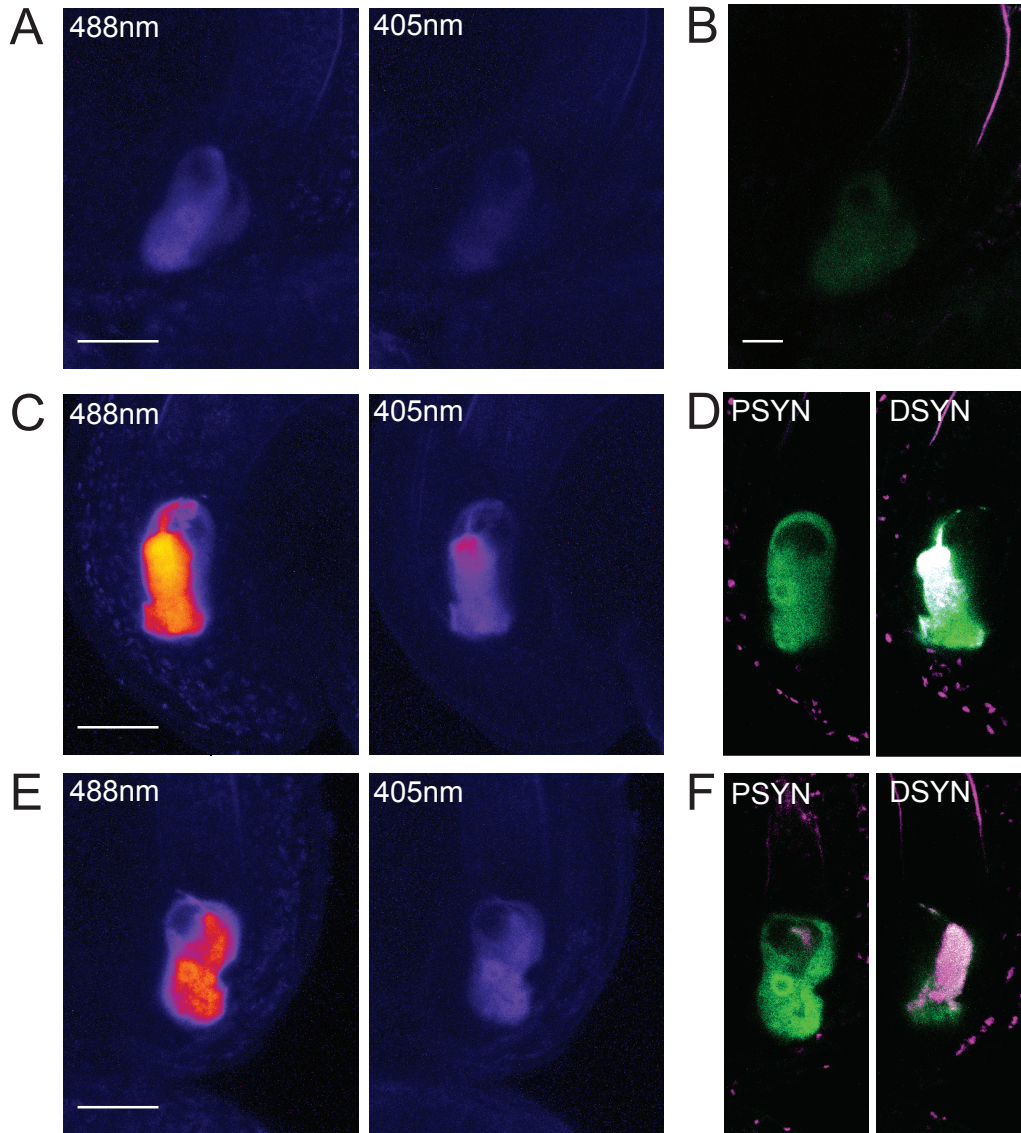


Figure 4. Expression of pHluorin in synergid cells. pHluorin is expressed in synergid cells under the MYB98 promoter. All analysis was performed with flowers from one primary transformed plant. Panels **A-B**, **C-D**, and **E-F** each represent one ovule expressing pHluorin in its synergid cells. Female flowers expressing pHluorin were pollinated with pollen expressing LAT52:dsRED and dissected ~7 hpp for analysis. Panels **A**, **C**, and **E**

are pseudo-colored with the FIRE LUT as in Figures 1A-C and 2D,G-H. The left panel is a mean projection of an image taken at 488nm while the right panel is a mean projection of the same sample imaged at 405nm excitation (**A,C,E**). Scale bar represents 20um. Panels **B,D**, and **F** show a merging of two channels: 488nm excitation in green and the RFP channel in magenta. Panels **A** and **B** shows both intact synergids in an ovule. Scale bar represents 10um. Panels **D** and **F** show the persistent synergid (PSYN) on the left and the degenerating synergid (DSYN) mixed pollen tube burst on the right.

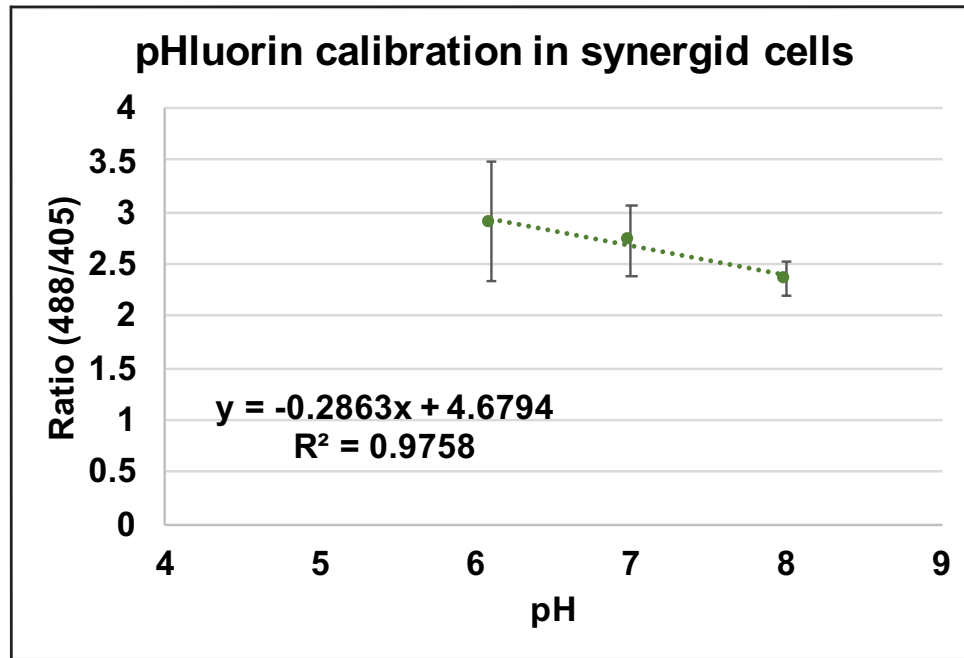


Figure 5. Calibration of pHluorin in synergid cells. Generation of a calibration curve for pHluorin expression in synergid cells. Calibration is the result of one experiment analyzing 3-5 ovules for pH 6, 7, and 8, including only intact synergid cells in ovules that did not show burst events. Ovules were from one primary transformed plant and were pollinated with LAT52:dsRED pollen tubes, allowed to grow in vivo ~6 hours then dissected and incubated in the calibration buffer for at least ~30-60 minutes before analysis. Error bars represent standard deviation.

pHuji expression in synergid cells provides an assay to assess the pH of the synergid cell and pollen tube with two different markers.

We observed pHuji expression in synergid cells (Figure 6) and pHuji signal was located at the site of gamete fusion (see 'hook' in Figure 6C). pHuji is a pH-sensitive, non-ratio-metric sensor, and the relative fluorescence intensity of pHuji changes with pH

(Shen, 2014). We observed high variability in fluorescence intensity across samples, which likely contributed to our difficulty in trying to calibrate pHuji at this stage (data not shown). Therefore, in order to assess whether pHuji will be suitable for pH measurements of the synergid cell more optimization is needed.

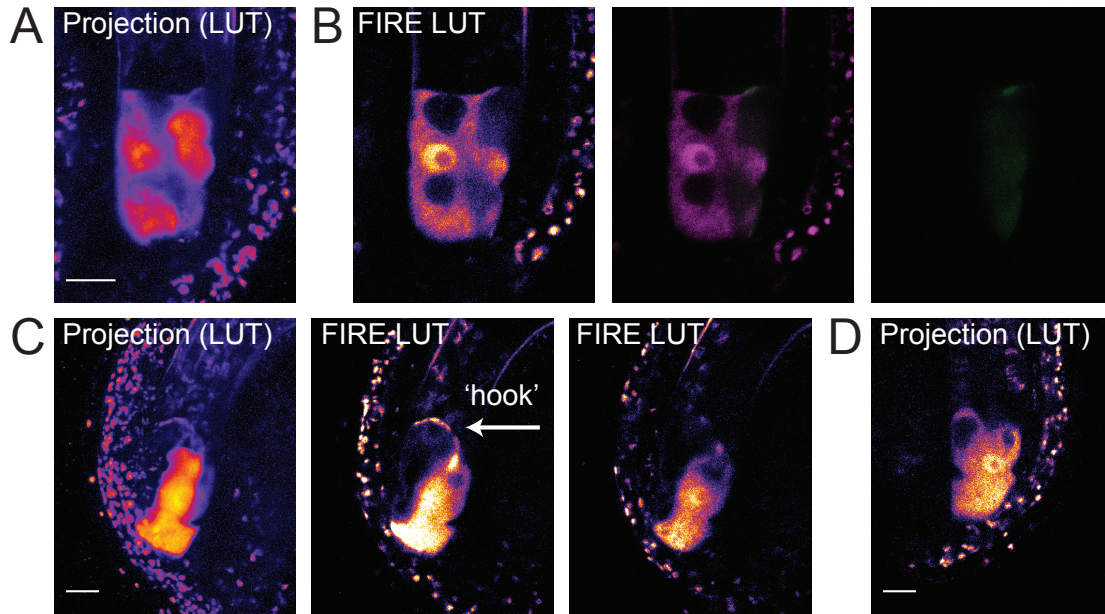


Figure 6. Expression of pHuji in synergid cells. Expression of pHuji under the MYB98 promoter was analyzed in ovules from one primary transformed plant. Pistils expressing pHuji in their ovules were pollinated with pHluorin-expressing pollen, allowed to proceed in vivo, and dissected and mounted ~7hpb. Panels **A-B** are from the same ovule. Panel **A** is a mean projection of a z-stack and the RFP channel is pseudo-colored with the FIRE LUT as in Figures 1B-C, 2C-D, 3A,3C,3E, and 3G. Scale bar represents 20um. Panel **B** represents a slice through the z-stack projected in **A**. **B**. Left panel is the RFP channel pseudo-colored using the FIRE LUT. Center panel is a merge of pHuji in magenta and the 488nm excitation of pHluorin in green. The right panel shows only the 488nm/GFP channel. **C**. The left panel is a mean projection of a z-stack through a pHuji-expressing ovule after pollen tube burst, while the center and right panels are individual slices in the z-stack. **D**. An example of an untargeted ovule expressing pHuji in both synergids. **C-D**. Images are pseudo-colored as in panel A. Scale bar represents 10um.

Delivery of HAP2 domain III to the site of fusion did not affect fertilization.

Since our experiments with pHluorin (Figure 4) and pHuji (Figure 6) in synergid cells suggested cytoplasmic contents of the synergid cell were released into the site of

gamete fusion, we expressed a HAP2 domain III fusion protein (HDIII; Figure S6) in synergid cells. HAP2 domain III was fused to GFP and had no signal sequence; therefore, it would not be expected to be secreted from the synergid cell and would likely accumulate in the cytoplasm. The HDIII protein was observed in the cytoplasm and nucleus but not vacuole of the synergid cells (Figure 7) and it was observed near the site of fusion (Figure 7B-C). We observed successful fertilization events by microscopy at approximately 7 hours post-pollination (7 hpp; Figure 7D) and 16 hpp (Figure S7) using a sperm nucle8i marker (HTR10:HTR10:RFP;). We also observed some sperm still sitting at the site of fusion (Figure 7E, center image) after 16hp, which may indicate a delay in fusion. However, sex-specific crosses revealed no negative effect on transmission through the female (Table S1-2), which would be expected if the ectopic HDIII was blocking HAP2 function and impacting fertilization rates.

We generated an extended domain III blocking construct (HDIII+) by 65 amino acids to include the stem region of HAP2 but not the transmembrane domain and expressed it in synergid cells (Figure 8). The extended domain III fusion product was no longer found in the nucleus of the synergid cell (Figure 8A) and was observed in potential vesicles (Figure 8). Additionally, when the protein appeared to be localized to these presumed vesicles, the vesicles did not make it to the site of fusion (Figure 8B-C and E). Some burst events did coincide with GFP expression near the site of fusion (Figure 8D). The specific localization of HDIII+ remains to be determined, and it is possible that expression levels of the protein may affect its localization. However, it is important to consider that proteins or molecules transported in vesicles in the synergid cells may not be released to the site of fusion. Since the extended domain III construct (HDIII+; Figure 8) does not appear at the site of fusion, it would not be expected to be able to block the fusion reaction (Figure S1). Indeed, no female-specific effect was observed in genetic transmission assays (Table S3).

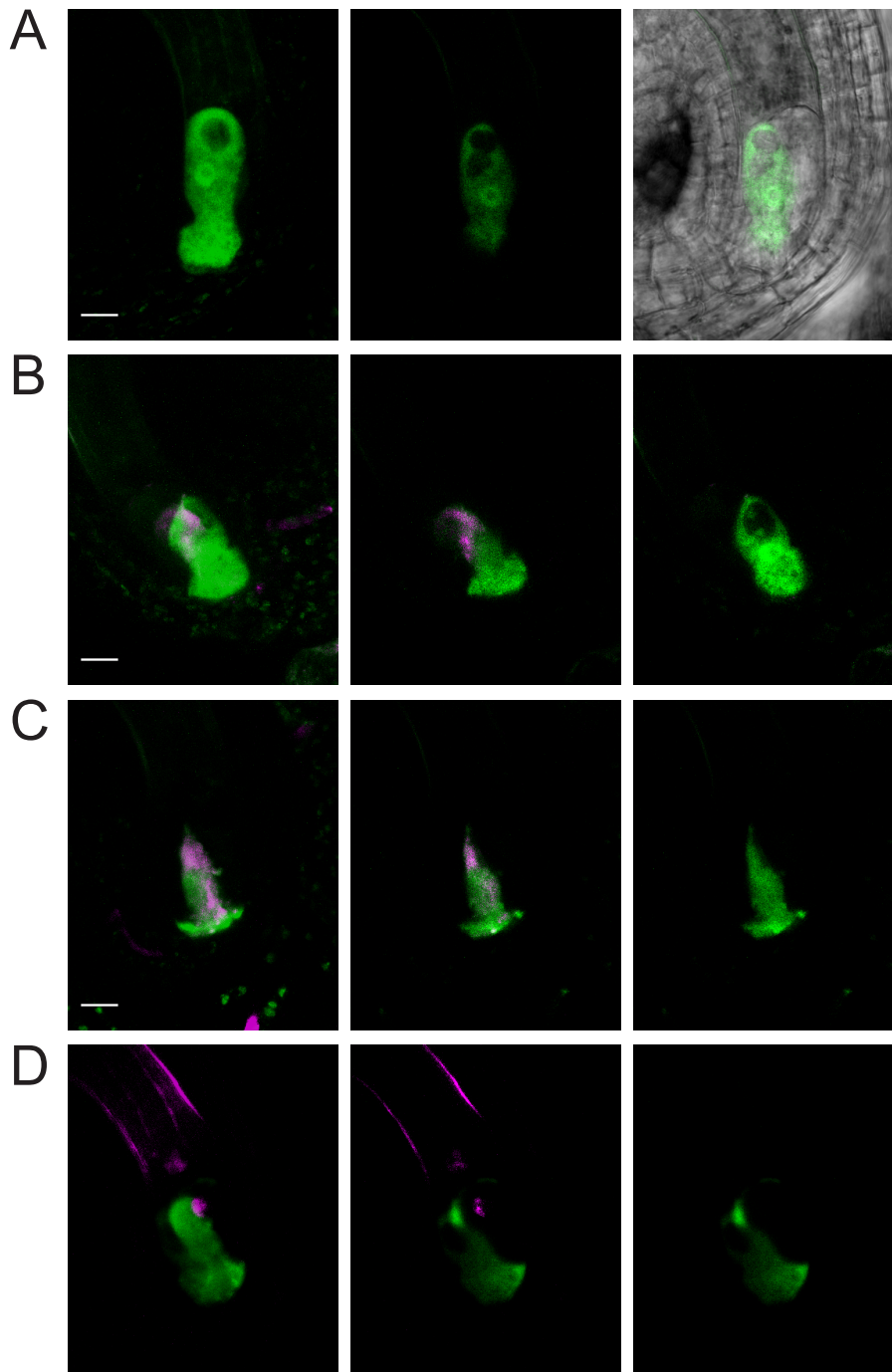


Figure 7. Expression of HAP2 Domain III in synergid cells. Domain III from Arabidopsis HAP2 was fused to GFP and expressed in ovules under the MYB98 promoter. All panels show the GFP channel as green and the RFP channel as magenta. **A-C.** Pistils expressing MYB98:HAP2-DIII:GFP were pollinated with LAT52:dsRED pollen and allowed to proceed in vivo until dissection after ~7-8hpp. Left images are projections of z-stacks consisting of merged GFP and RFP channels. Scale bar represents 10μm. **A.** Confocal image showing an untargeted ovule with two intact synergid cells expressing HAP2-DIII-

GFP. The center and right images are slices from the z-stack (left panel) showing either the GFP channel (center) or a merge of the GFP and DIC channels (right). **B.** The center image is a merge of RFP and GFP channels while the right image is GFP alone. **C.** The center and right images both show merged RFP and GFP channels, but they represent different slices in the z-stack from the left image. The center image shows the degenerating synergid cell while the right image shows the persistent synergid cell. **D.** Confocal image showing an ovule expressing HAP2-DIII-GFP ~7hpp with the sperm nuclei marker HTR10:HTR10:RFP. The left panel is a projection of a z-stack while the center and right panels are the same slice through the z-stack showing either the GFP and RFP channels merged (center) or GFP only channel (right). Scale bars represent 10um.

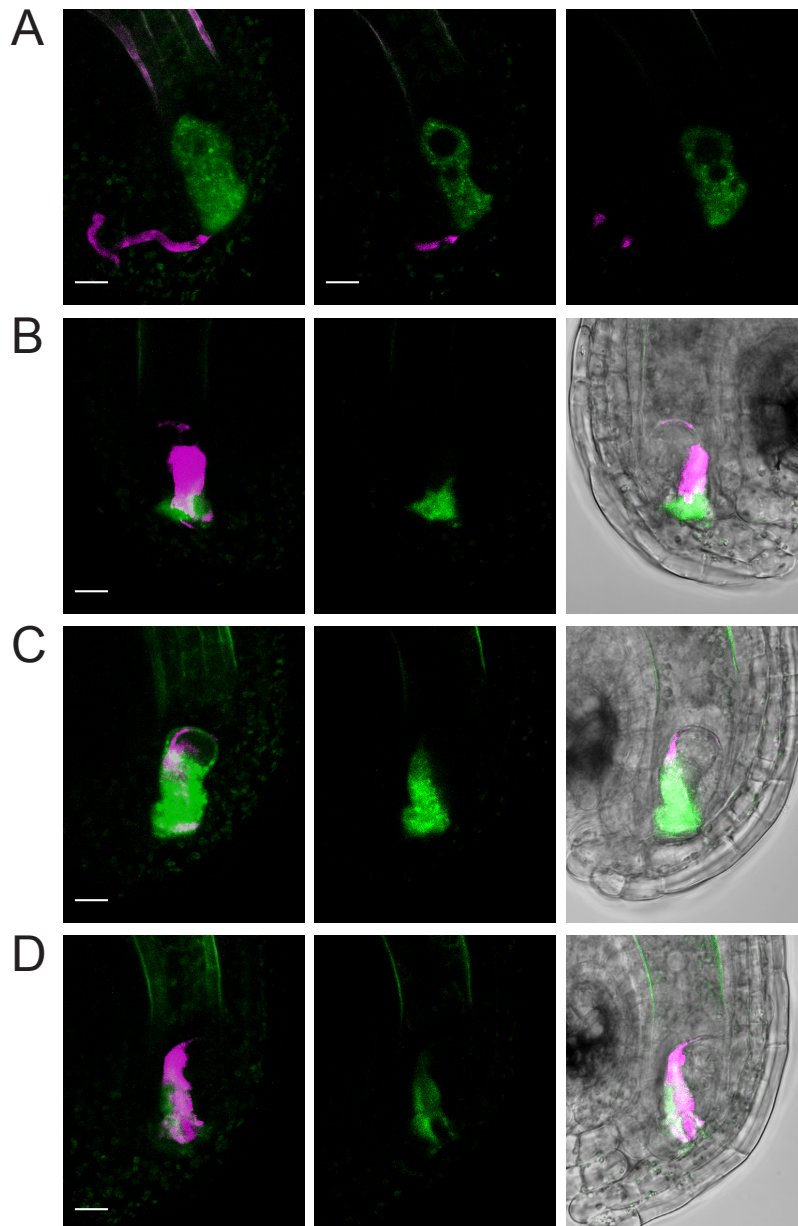


Figure 8. Expression of a HAP2 Domain III with an extended C-terminus in synergid cells. LAT52:dsRED was used to pollinate ovules expressing an extended DIII from HAP2 fused to GFP under the MYB98 promoter. Pollination was allowed to proceed in vivo and pistils were dissected ~7-8hpp for analysis. **A-D.** Left images shown projections of z-stacks while the other two images are slices of that z-stack. Left images show the merge of RFP and GFP channels. **A.** A pollen tube is at the micropyle of an ovule with two intact synergid cells. The center image shows one synergid cell while the right image shows the other synergid cell. **B-D.** The center image represents a slice through the z-stack but showing only the GFP channel, while the right image is a merge of the RFP, GFP, and DIC channels. Scale bar represents 10um.

Discussion

Our study has shown the first experiments assessing the pH of the gamete fusion environment for Arabidopsis. We have shown that the contents of the burst pollen tube have a lower average pH than the cytoplasm of the growing pollen tube, and we observed that the site of gamete fusion has a lower average pH than the overall burst event. We have also generated transgenic plants expressing the pH sensors pHluorin and pHuji in synergid cells and observed them before and after pollen tube burst. These genetic resources will enable more detailed investigation into changes in the pH of the environment immediately surrounding the sperm both before pollen tube burst and after sperm deposition. Since a trigger for HAP2 has yet to be identified, a change in pH, as we have observed here, remains a valid hypothesis that warrants additional testing.

It is important to consider that the optimal range for pHluorin is from pH 5.5 to pH 7.5 (Miesenbock et al., 1998; Moseyko and Feldman, 2001), yet two of our reported values, the average pH of the pollen tube and the average pH of the site of fusion, reside outside the optimal range and outside the buffers we used for calibration. However, other studies have performed calibrations with pH less than 5.5 (Miesenbock et al., 1998; Moseyko and Feldman, 2001), and experimental values have been reported below pH 5.5 (Miesenbock et al., 1998). Therefore, it will be useful to increase the number of calibration buffers and cover a wider range of pH values from 5.0 to 8.5, in order to include the

observed pH for both the site of gamete fusion and the pollen tube growing in the SIV assay. It will be important to confirm the pH change observed at the site of gamete fusion. Since the best indicator of the site of gamete fusion is the location of the sperm, these experiments should be repeated with pollen expressing both pHluorin and a sperm nuclei marker (as shown in Figure S4). To further test the synergid pH before pollen tube arrival and burst, it will be important to optimize pHluorin expression and calibration in synergid cells. With optimization, pHuji expression in synergid cells may be useful in combination with pHluorin expression in pollen tubes for simultaneous evaluation of the pH in the synergid cells and pollen tube during the process of fertilization in flowering plants.

While the average pH is an important first assessment of the environment of the site of fusion, a more detailed analysis including live-imaging will likely increase our knowledge of the pH during different steps in fertilization and gamete fusion. For example, the synergid cell may change its pH in response to the pollen tube communication or growth in the ovule. It would also be interesting to assess how the pH of the degenerating synergid changes over time, as well as whether the site of gamete fusion experiences fluctuations in pH. Since the sperm are waiting at the site of fusion for ~7-8 minutes (Hamamura et al., 2011) before fusion actually occurs, it will be essential to understand whether there are pH changes during that time. Additionally, we observed fairly high variability in the calculated pH at the site of fusion, which may be related to the timing of imaging relative to the burst event. Therefore, it will be necessary to follow pH changes at the site of fusion immediately following pollen tube burst over time to better understand the gamete fusion environment.

The plants generated in this study to investigate pHuji and pHluorin expression in synergid cells were analyzed in the first generation after transformation, and only one plant per construct was used for our assays. Therefore, it will be important to not only identify more individual plants expressing each construct, but to test plants in subsequent

generations for responses to pH during the fertilization process. It is of particular interest to observe pHuji in different plant lines, since pHuji analysis of pH is not ratio-metric and therefore could be subject to variation in expression levels and multiple insertions within the genome of a single plant. The calibration results (Figure 5) for pHluorin expressed in synergid cells come from only one experiment, but the small differences observed between each pH calibration point suggests the calibration buffer may not have been penetrating the ovule sufficiently. Calibration procedures for synergid cells expressing pH sensors will need to be optimized to provide more useful analysis in the future. However, our preliminary results suggest the pH of the intact synergid cell may be close to or even below pH 6 (Figures 5 and S5), which would be below the observed pH for the intact pollen tube.

While our domain III blocking experiments had no effect on genetic transmission, it would be interesting to assess whether they subtly impede the gamete fusion process. It is possible that not enough domain III was introduced to the site of fusion to noticeably affect fusion events. It is also possible that fusing GFP to the C-terminus of domain III impaired its ability to block fusion. For further studies, it will be interesting to test whether expressing domain III alone, without a GFP tag, can block fusion, either by delivering it from the synergid cell or from the pollen tube cytoplasm. Additionally, it would be interesting to add a signal peptide to HAP2 domain III and test whether it can be secreted from the egg cell when introduced under an EC1 promoter which may release domain III at a more effective time.

Currently, the environment at the site of gamete fusion in Arabidopsis remains poorly characterized. However, we now have tools to test whether pH plays an important role in the regulation of HAP2-mediated gamete fusion. We have also provided insights into whether proteins or molecules can be delivered to the site of fusion through the

degenerating synergid cell, which may be useful in manipulating the pH and other aspects of the environment at the site of fusion.

Methods

Plant growth conditions.

Plants were grown under a 16 hour light, 8 hour dark cycles as described previously ((Fedry*, Submitted); Chapter 2).

Constructs used for pH analysis.

Constructs were generated using the GreenGate cloning method (gift from Jan Lohmann; Addgene kit # 1000000036; (Lampropoulos et al., 2013)) as described previously (chapter 3). pHuji cds was sub-cloned into the C000 entry vector from plasmid p379 provided by Jose Feijo, using primers pHuiji_cds_F1 (ATGGTGAGCAAGGGCGAGGAGA) and pHuiji_cds_R2_with_stop (TTACTTGTACAGCTCGTCCATGCCG). The plasmid containing pHuji under the MYB98 promoter (chapter 3; (Kasahara et al., 2005)) was generated using B-dummy, D-dummy, and Hygromycin resistance entry vectors from Greengate in the pGGZ003 destination vector ((Lampropoulos et al., 2013)). pHluorin was sub-cloned from plasmid p534 provided by Jose Feijo using primers pHluorin_cds_F1 (ATGAGTAAAGGAGAAGAAGCTTTTCACTGG) and pHluorin_cds_R2_with_stop (TTATTTGTATAGTTCATCCATGCCATGTGTAA). An internal Bsal site was removed by site-directed mutagenesis (Q5 Site-Directed Mutagenesis Kit #E0554S, New England Biolabs (NEB)) to produce a synonymous mutation, using primers (pHluorin_Bsal.mut_F1: AGAGAGATCACATGGTCCTTCTTGAG and pHluorin_Bsal.mut_F2: AGAGAGATCACATGGTCCTTCTTGAG). pHluorin without Bsal site was placed under the MYB98 promoter using the same GreenGate constructs as the MYB98:pHuji. After plasmids were confirmed by Sanger sequencing (Genewiz, South Plainfield, NJ), they

were electroporated into *Agrobacterium* strain GV3101 (Koncz and Schell, 1986) using the pSoup helper plasmid (Hellens et al., 2000).

Generation of pH sensor expressing Arabidopsis plants.

Col0 plants were transformed using the floral drip method (Clough and Bent, 1998). Transformed (T1) seeds were sterilized as described previously ((Fedry*, Submitted); Chapter 2) and plated on Murashige and Skoog (MS) media supplemented with 25mg/L hygromycin B (Corning). Seeds on plates were cold-treated for at least 2 days then moved to 22°C with constant light for 7-14 days before transferring to soil. Only 1 T1 plant (out of 3-4) was found to be expressing pHuji, and two T1 plants (out of 3 screened, 4 total) expressing pHluorin were identified in the first T1 screening.

Imaging and analysis of pHluorin in pollen tubes.

All analysis of pollen tubes expressing pHluorin were performed using plants in the T4 generation (obtained in the T3 generation from the Feijo laboratory). Images were taken using a Zeiss LSM 800 with excitation at 488, 405, and 561 and detection wavelengths of approximately 475-526 for both 488nm and 405nm excitation and 577-700 for 561 excitation. DIC images were obtained with excitation by 488 laser. Images were captured at 1024x1024 and 16bit and processed in Image J (FIJI version). For calibration trials, calibration buffers were used at pH 5, 6.1, 7, and 8. However, due to higher variability at pH 5, and based on pH calibration results in previous papers (Michard, 2008; Miesenbock et al., 1998), we decided to perform the calibration analysis using pH 6.1, 7, and 8. Calibration buffers were generated using NaCl (30mM), KCl (100mM), and either MES (pH 6) or HEPES (pH 7 and 8). For calibration experiments, 2uM valinomycin (Sigma-Aldrich #V3639) and 5uM nigericin (Sigma-Aldrich #N7143) were added to the calibration buffers. SIV pollen tubes were incubated separately in each calibration buffer for at least 15 minutes prior to imaging.

For generation of calibration graphs, ROIs were drawn in the 488nm channel and used to measure the average fluorescence intensity at 488nm and 405nm excitation. A pollen tube expressing a red fluorescence marker in sperm nuclei but not pHluorin was used to obtain a background level of fluorescence in each channel, which was then subsequently used for background correction for the respective channel. The final calibration graph comprised of three trials using calibration buffers at pH 6.1, 7, and 8 and was generated by combining data from all three trials. The equation generated from this graph was used for calculating pH values of samples expressing pHluorin in the pollen tube.

SIV pollen tube samples (Figure 2A; (Palanivelu and Preuss, 2006)) were grown through an upright pistil on solid (1.5% low-melt agarose, Lonza) pollen growth media (PGM: 500mM KCl, 5mM CaCl₂, 1mM MgSO₄, 1% H₃BO₃, and 10% sucrose; pH adjusted to 7.5-7.8; (Boavida and McCormick, 2007)) for ~1.5-2 hours before lying them down and allowing pollen tubes to grow for another ~3-5 hours before adding liquid PGM and cover-slipping. For SIV pollen tube samples, an ROI was generated in the 488nm channel and used to measure average fluorescence intensity at 488nm and 405nm excitation. Background correction was performed using the negative control pollen tube and using the background fluorescence of synergid cells not expressing pHluorin for analysis of pollen tube burst and coil events within the ovule. For in vivo analysis, flowers were emasculated the day before and pollinated the next morning Pollen tubes were allowed to grow in vivo for 5-8 hours (depending on the experiment) before dissection and mounting in 80mM sorbitol. Calibration graphs were generated in Excel, while sample graphs were generated using R studio. Statistical analysis was also performed in R studio. Imaging analysis was performed in ImageJ (FIJI version; (Schindelin et al., 2012)), and the FIRE look up table (LUT) was used to display intensity differences.

Imaging and analysis of pHluorin in synergid cells.

All analysis of pHluorin in synergid cells was performed using one T1 plant, although a second T1 did express pHluorin in its synergid cells (data not shown). Image acquisition was performed as described above for pHluorin in pollen tubes, except that the laser lines needed to be increased for both 488nm and 405nm excitation. Therefore, new background fluorescence levels were recorded at these settings, using an ROI of the area in the embryo sac immediately above a burst event. Calibration was performed using the same buffers listed above, but calibration was performed on dissected pistils on agar pads used for the SIV assay. Samples were in calibration buffer for ~30-60 minutes before cover-slipping then imaging. For projections of z-stacks, the projections were done at mean fluorescence intensity.

Imaging and analysis of pHuji in synergid cells.

All analysis of pHuji in synergid cells was performed using the only T1 line (out of 3) expressing red signal in the synergid cells. Image acquisition was performed using the parameters described for pHluorin in pollen tubes. Some cross-over of pHuji signal or background fluorescence was detected in the synergids with the 405nm settings used for analyzing pHluorin in the synergid cells. Background fluorescence for pHuji was measured in the RFP channel in untargeted synergid cells expressing pHluorin but not pHuji and was used for background subtraction. For normalized fluorescence, mean fluorescence intensity of 60000 was set to 1.

Generation of HAP2 domain III constructs and plants expressing fusion proteins.

Constructs were generated using the Greengate Cloning system. Domain III was sub-cloned from pgl290 ((Fedry*, Submitted); Chapter 2) using forward primer (CCCGGTAAGATCATAAAC) and reverse primer (AGCCGTTGTACTAAATTGACA) by Phusion PCR (NEB) and TA cloned into C0000 entry vector as described previously (chapter 3). A start codon and three amino acids upstream of DIII in the linker region were

inserted using site-directed mutagenesis (HDIII). The HDIII+ construct that is 65 amino acids longer was generated using the same procedures, but the initial PCR product was generated using different forward (CCCGGTAAGATCATAAACATC) and reverse (GCTTAAGCAAACGTATTGAATG) primers. Plasmids were generated using the MYB98 promoter cassette (chapter 3), B-dummy, the respective coding sequence, linker-GFP, Hygromycin resistance, and the Z003 destination vector from GreenGate. Transformation into *Agrobacterium* and subsequently plants was performed as described above.

Analysis of HAP2 domain III fusion proteins in plants.

Genetic analysis was performed using plants from the T1 generation (selected on MS plates supplemented with Hygromycin). For crosses assessing effects of HAP2 domain III expression, flowers from the respective female plants were emasculated and then pollinated with either wildtype Col0 pollen or wildtype pollen expressing a sperm nuclei marker (HTR10:HTR10:RFP; (Ingouff et al., 2009)). Crosses were also performed onto *male-sterile 1* (*ms1*) plants to assess transmission of the transgene through the male, which should be unaffected. Dried crosses were collected, and then dried seeds were sterilized and plates on MS plates with Hygromycin. Seeds on plates were cold-treated for at least 2 days then incubated for 7-10 days at 22°C with constant light before counting for Hygromycin resistance versus sensitive seedlings.

For imaging analysis, plants from the T1 generation or second generation after transformation (T2) were used. Flowers expressing HAP2 domain III fusion constructs in their synergids were emasculated the day before and pollinated with either pollen expressing RFP in the cytoplasm (LAT52:dsRED) or RFP in the sperm nuclei (HTR10:HTR10:RFP). Pollen was allowed to grow in vivo before dissection ~7-8 hours after pollination (7hpp). For 16 hours after pollination (16hpp), flowers were pollinated in the evening and imaged the next morning. Dissected pistils were mounted in 80mM

sorbitol and cover-slipped before imaging on the Zeiss LSM 800. Images were processed in ImageJ (FIJI version) and projects are of maximum intensity.

Acknowledgements

We would like to thank Ethan Fitzgerald for his help making calibration buffers and preliminary experiments using pHluorin-expressing pollen tubes. We would also like to thank Dr. Eric Morrow and his laboratory for sharing their pH sensor protocols with us. We thank Dr. Fred Berger for HTR10:HTR10:RFP seeds and Dr. Greg Copenhaver for LAT52:dsRED seeds. We would also like to thank Geoff Williams for his help with image acquisition.

Supplemental Figures

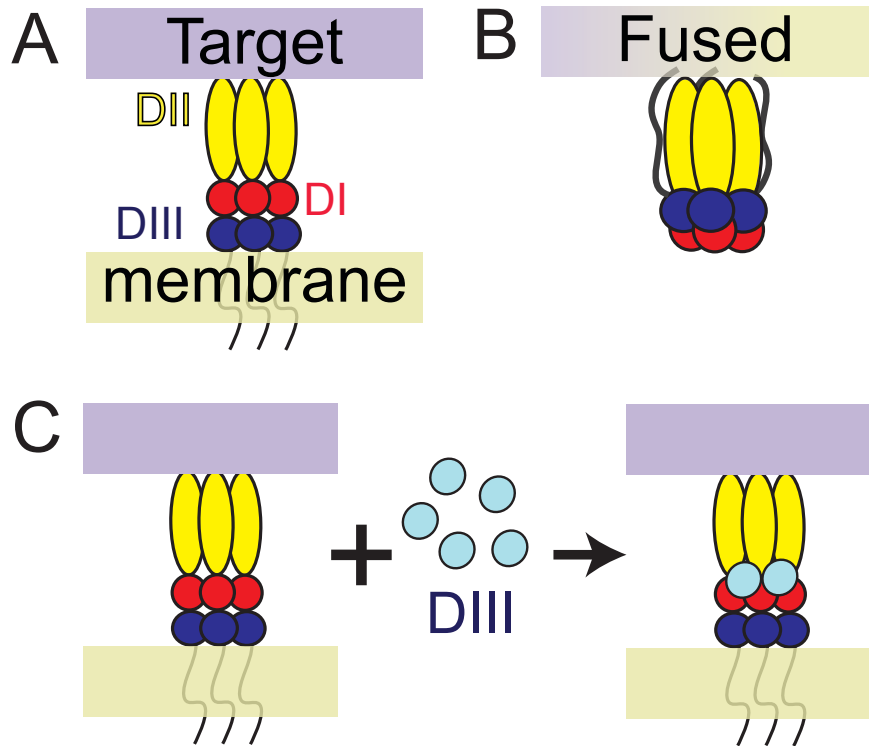
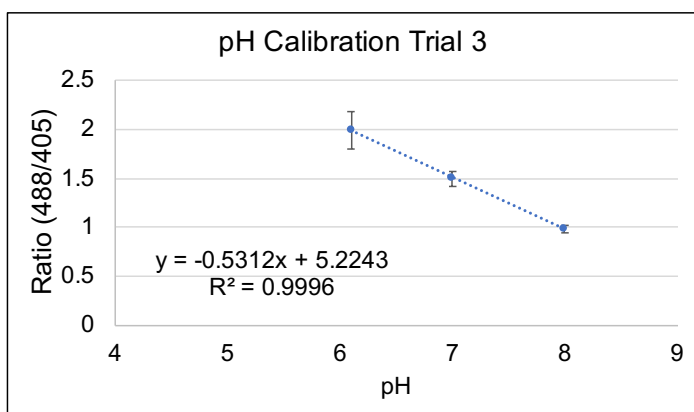
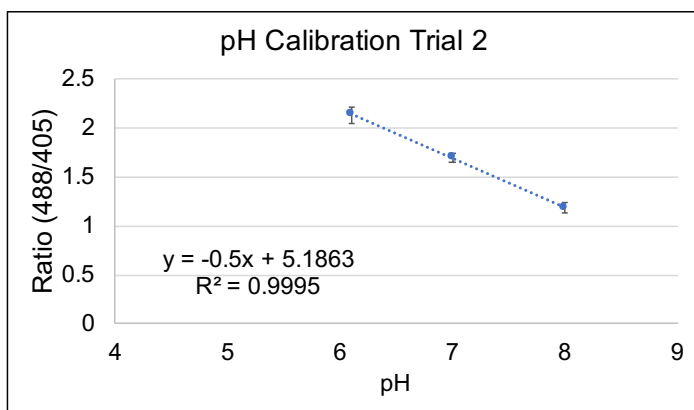
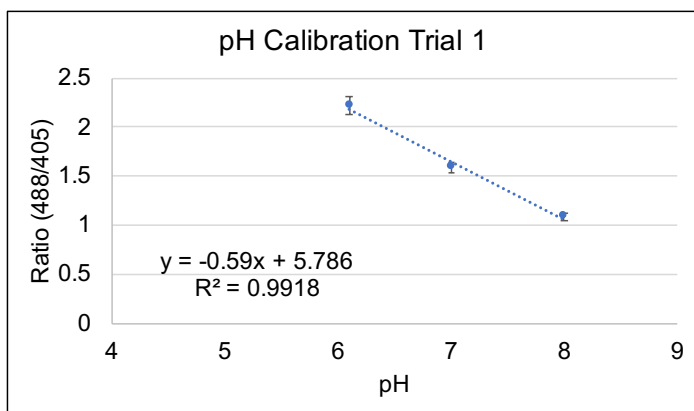
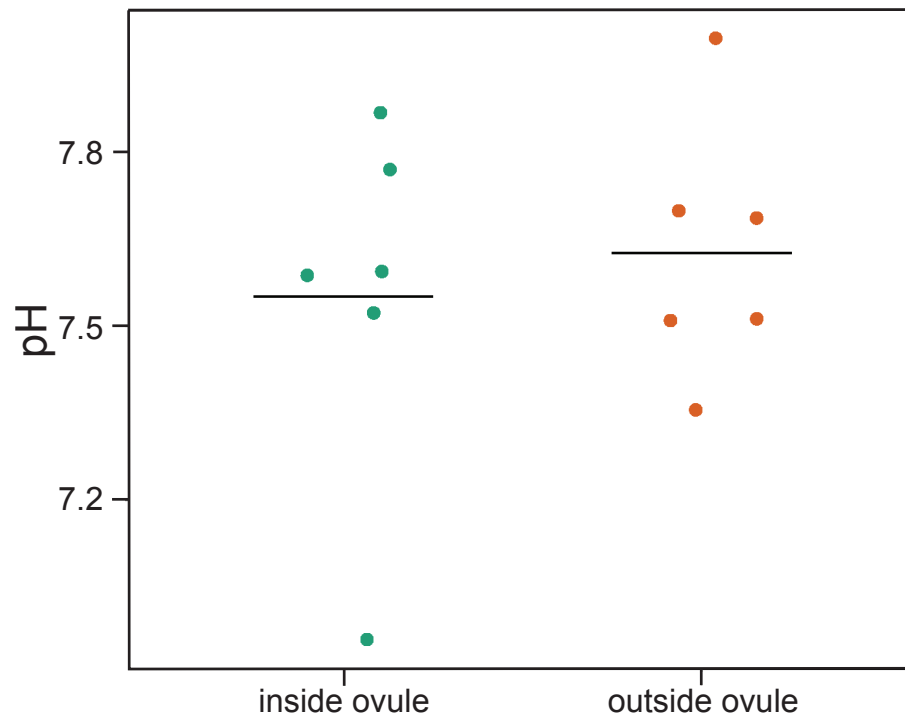


Figure S1. Schematic demonstrating blocking mechanism of exogenous domain III of class II viral fusion proteins. **A.** Three domains of class II fusion proteins are shown: domain I in red, domain II in yellow, domain III in blue (Gibbons et al., 2003; Lescar et al., 2001; Modis et al., 2003; Rey et al., 1995). The tip of domain II inserts into the target membrane (Kielian and Rey, 2006). **B.** After the ‘trimer-of-hairpins’ brings the membranes close together, membrane fusion occurs, resulting in the post-fusion trimer shown in B (Kielian and Rey, 2006). **C.** Addition of exogenous domain III blocks membrane fusion by preventing the ability of the protein to fold-back in the hairpin formation (Liao and Kielian, 2005).



S2. Calibration trials for three independent experiments. These three experiments were combined to generate Figure 1A. **A.** For trial 1: $n=5$, 5, and 6 for pH 6, 7, and 8 respectively. **B.** For trial 2: $n=6$, 9, and 8 for pH 6, 7, and 8, respectively. **C.** For trial 3: $n=3$, 7, and 5, for pH 6, 7, and 8, respectively. Error bars show standard deviation.



S3. Comparing pH of pollen tubes observed inside and outside ovules in the SIV assay. Each data point represents a portion of one tube. Tubes were used to calculate both 'inside' and 'outside' pH when possible. $n=6$ pollen tubes (partial) for both conditions. All data is from one experiment and includes both wildtype (Col0) and *lre* ovules; therefore, some inside events represent coiling. 'Inside' measurements are shown in green, while 'outside' measurements are shown in orange. The mean for each group is shown as a black bar.

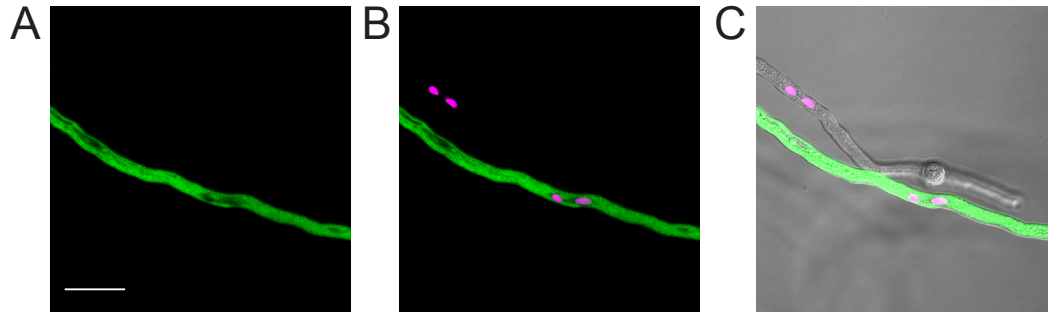
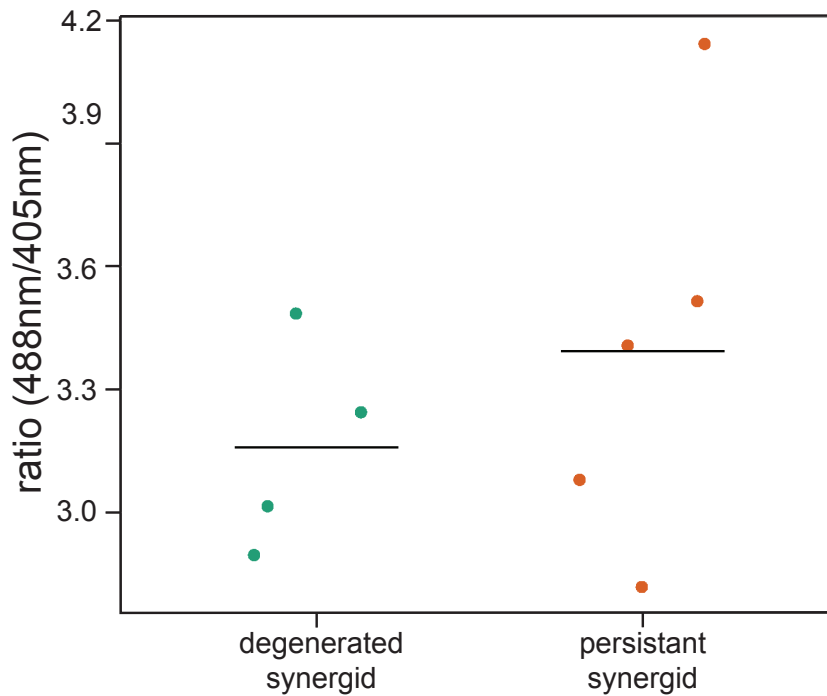


Figure S4. Expression of pHluorin in pollen tube with sperm nucleus marked with RFP. A plant expressing pHluorin in the pollen tube was crossed with one expressing a sperm nuclear RFP signal (HTR10 fused to RFP under the HTR10 promoter) and the pollen tubes shown here are from a plant resulting from that cross. The RFP sperm nuclei marker is shown as magenta. Panel **A** shows pHluorin signal when excited at 488nm. Panel **B** shows the signal from excitation at 488nm in the green channel merged with the sperm nuclei marker in the red channel. Panel **C** shows the merge of the GFP and RFP channels in **B** with DIC added. The upper pHluorin-negative pollen tube was used for background correction in figures 1-2. In panel **A**, the lower pHluorin-positive tube shows an absence of pHluorin signal where the sperm nuclei are observed in panel **A**. Scale bar represents 20um.



S5. Comparing pHluorin ratio for the degenerating synergid versus the persistent synergid. Data are combined from two experiments. We analyzed a total of 4 degenerating synergid cells (green) and 5 persistent synergid cells (orange). We also observed two ovules with untargeted synergid cells (ratio data not included here: average ratio = 3.2). Based on our calibrations of pHluorin in the synergid and in the pollen tube, the higher average ratio would suggest a potential lower pH in the persistent synergid cell, although there was no significant difference between the pH of the degenerated synergid cell and the pH of the persistent synergid cell (Student's t-test).

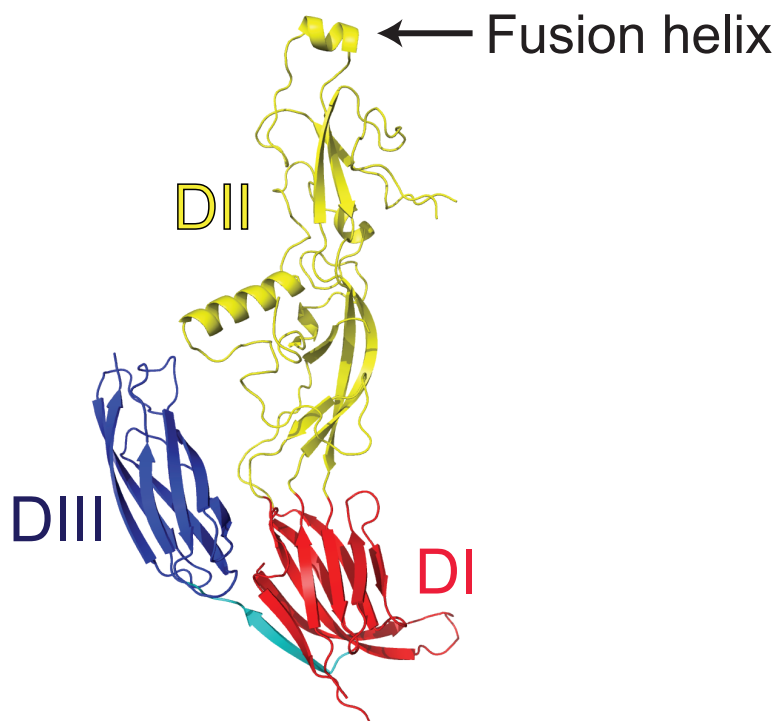


Figure S6. Arabidopsis HAP2 ectodomain monomer. Structure of one monomer of Arabidopsis HAP2 ectodomain with the fusion helix oriented towards the top of the page and contained within domain II (yellow). Domain III (DIII) is shown in blue (Figure S1A). The HAP2 DIII construct consisted of part of the linker region (shown in cyan) and all of domain III (Figure 7). The DIII longer construct also included 65 amino acids C-terminal to DIII within the stem region (not shown in the structure; Figure 8). PDB: 5OW3 ((Fedry*, Submitted), Chapter 2).

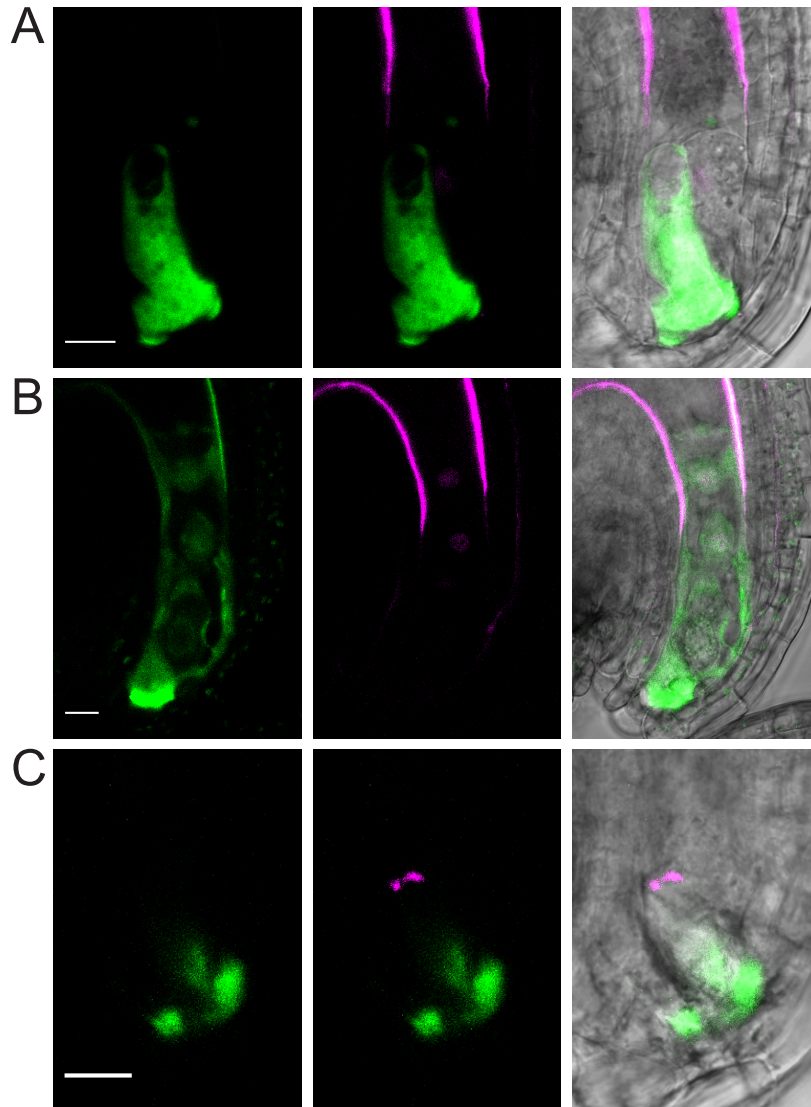


Figure S7. Ovules expressing HAP2-DIII-GFP 16 hours after pollination with a sperm nuclei marker. A-C. All images in each panel are the same slice through a confocal z-stack. The left image is the GFP only channel. The center image is a merge of GFP and RFP channels, and the right image adds DIC to the GFP and RFP channels. **A.** This ovule has undergone double fertilization and karyogamy, as indicated by the location and diffuse RFP signal from the sperm nuclei. **B.** This ovule shows endosperm proliferation after successful fertilization with sperm expressing the sperm nuclei marker (HTR10:HTR10:RFP), and HAP2-DIII-GFP signal is observed in the proliferating endosperm. **C.** This ovule has received a pollen tube and undergone synergid degeneration, but the sperm have not fused with the egg and central cell as indicated by their close proximity to each other and the compact RFP signal in the sperm nuclei. Scale bars represent 10 μ m.

Supplemental Table 1. Male-specific transmission results for three primary transformed plants expressing HAP2-DIII:GFP in synergid cells.

	<i>ms1</i> x T1				
T1	# Hyg S	# Hyg R	% Hyg R	n	# crosses
1.3	48	46	48.94	94	2
1.4	9	142	94.04	151	3
4.1	65	84	56.38	149	3

Number of hygromycin sensitive (Hyg S) and hygromycin resistant (Hyg R) seedlings counted are shown. The total number of seedlings analyzed (n) across all indicated crosses is provided. For a single-copy line, 50% hygromycin resistance (% Hyg R) is expected. T1 = primary transformed plant.

Supplemental Table 2. Female-specific transmission results for three primary transformed plants expressing HAP2-DIII:GFP in synergid cells.

	T1 x WT female				
T1	# Hyg S	# Hyg R	% Hyg R	n	# crosses
1.3	28	39	58.21	67	2
2.2*	49	57	53.77	106	2
4.1	69	64	48.12	133	3

*Crosses used HTR10:HTR10:RFP as the pollen donor, not Col-0.

Hygromycin sensitive (Hyg S) and resistant (Hyg R) seedling counts are included for each T1 plant listed. The total number of seedlings (n) analyzed is provided across all indicated crosses. If there is no effect on transmission, 50% hygromycin resistance (% Hyg R) is expected for a single-copy line. T1 = primary transformed plant.

Supplemental Table 3. Female-specific transmission results for three primary transformed plants expressing HAP2-DIII-extended:GFP in synergid cells.

	T1 x Col-0				
T1	# Hyg S	# Hyg R	% Hyg R	n	# crosses
2.1	46	52	53.06122	98	4
2.3	46	37	44.57831	83	2
2.4	44	53	54.63918	97	2
3.2	77	75	49.34211	152	3

Counts are provided for hygromycin sensitive (Hyg S) and hygromycin resistant (Hyg R) seedlings. The total number of seedlings analyzed (n) are provided across the number of indicated crosses. The genetic expectation for no effect on transmission would be 50% hygromycin resistant seedlings (% Hyg R). T1 = primary transformed plant.

References

- Berger, F., Hamamura, Y., Ingouff, M., and Higashiyama, T. (2008). Double fertilization - caught in the act. *Trends Plant Sci* 13:437-443.
- Blagborough, A.M., and Sinden, R.E. (2009). *Plasmodium berghei* HAP2 induces strong malaria transmission-blocking immunity in vivo and in vitro. *Vaccine*.
- Boavida, L.C., and McCormick, S. (2007). Temperature as a determinant factor for increased and reproducible in vitro pollen germination in *Arabidopsis thaliana*. *Plant J* 52:570-582.
- Bressanelli, S., Stiasny, K., Allison, S.L., Stura, E.A., Duquerroy, S., Lescar, J., Heinz, F.X., and Rey, F.A. (2004). Structure of a flavivirus envelope glycoprotein in its low-pH-induced membrane fusion conformation. *The EMBO journal* 23:728-738.
- Camacho-Nuez, M., Hernandez-Silva, D.J., Castaneda-Ortiz, E.J., Paredes-Martinez, M.E., Rocha-Martinez, M.K., Alvarez-Sanchez, M.E., Mercado-Curiel, R.F., Aguilar-Tipacamu, G., and Mosqueda, J. (2017). Hap2, a novel gene in *Babesia bigemina* is expressed in tick stages, and specific antibodies block zygote formation. *Parasit Vectors* 10:568.
- Capron, A., Gourgues, M., Neiva, L.S., Faure, J.E., Berger, F., Pagnussat, G., Krishnan, A., Alvarez-Mejia, C., Vielle-Calzada, J.P., Lee, Y.R., et al. (2008). Maternal Control of Male-Gamete Delivery in *Arabidopsis* Involves a Putative GPI-Anchored Protein Encoded by the LORELEI Gene. *Plant Cell* 20:3038-3049.
- Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16:735-743.
- Cole, E.S., Cassidy-Hanley, D., Fricke Pinello, J., Zeng, H., Hsueh, M., Kolbin, D., Ozzello, C., Giddings, T., Jr., Winey, M., and Clark, T.G. (2014). Function of the male-gamete-specific fusion protein HAP2 in a seven-sexed ciliate. *Curr Biol* 24:2168-2173.
- Dessau, M., and Modis, Y. (2013). Crystal structure of glycoprotein C from Rift Valley fever virus. *Proc Natl Acad Sci U S A* 110:1696-1701.
- Dresselhaus, T., Sprunck, S., and Wessel, G.M. (2016). Fertilization Mechanisms in Flowering Plants. *Curr Biol* 26:R125-139.
- Drews, G.N., and Yadegari, R. (2002). Development and function of the angiosperm female gametophyte. *Annual Review of Genetics* 36:99-124.
- Ebchuqin, E., Yokota, N., Yamada, L., Yasuoka, Y., Akasaka, M., Arakawa, M., Deguchi, R., Mori, T., and Sawada, H. (2014). Evidence for participation of GCS1 in fertilization of the starlet sea anemone *Nematostella vectensis*: implication of a common mechanism of sperm-egg fusion in plants and animals. *Biochem Bioph Res Co* 451:522-528.
- Faure, J.E., Rotman, N., Fortune, P., and Dumas, C. (2002). Fertilization in *Arabidopsis thaliana* wild type: developmental stages and time course. *Plant J* 30:481-488.
- Fedry, J., Liu, Y., Pehau-Arnaudet, G., Pei, J., Li, W., Tortorici, M.A., Traincard, F., Meola, A., Bricogne, G., Grishin, N.V., et al. (2017). The Ancient Gamete Fusogen HAP2 Is a Eukaryotic Class II Fusion Protein. *Cell* 168:904-915 e910.
- Fedry*, J., Forcina*, J., Legrand, P., Pehau-Arnaudet, G., Haouz, A., Johnson, M.A., Rey, F.A., and Krey, T. (Submitted). HAP2 drives gamete fusion via insertion of highly divergent fusion loops into female gametes.
- Feijo, J.A., Sainhas, J., Hackett, G.R., Kunkel, J.G., and Hepler, P.K. (1999). Growing pollen tubes possess a constitutive alkaline band in the clear zone and a growth-dependent acidic tip. *J Cell Biol* 144:483-496.

- Gibbons, D.L., Erk, I., Reilly, B., Navaza, J., Kielian, M., Rey, F.A., and Lepault, J. (2003). Visualization of the target-membrane-inserted fusion protein of Semliki Forest virus by combined electron microscopy and crystallography. *Cell* 114:573-583.
- Gibbons, D.L., Vaney, M.C., Roussel, A., Vigouroux, A., Reilly, B., Lepault, J., Kielian, M., and Rey, F.A. (2004). Conformational change and protein-protein interactions of the fusion protein of Semliki Forest virus. *Nature* 427:320-325.
- Hamamura, Y., Saito, C., Awai, C., Kurihara, D., Miyawaki, A., Nakagawa, T., Kanaoka, M.M., Sasaki, N., Nakano, A., Berger, F., et al. (2011). Live-Cell Imaging Reveals the Dynamics of Two Sperm Cells during Double Fertilization in *Arabidopsis thaliana*. *Curr Biol* 21:497-502.
- Hellens, R.P., Edwards, E.A., Leyland, N.R., Bean, S., and Mullineaux, P.M. (2000). pGreen: a versatile and flexible binary Ti vector for *Agrobacterium*-mediated plant transformation. *Plant Mol Biol* 42:819-832.
- Hirai, M., Arai, M., Mori, T., Miyagishima, S.Y., Kawai, S., Kita, K., Kuroiwa, T., Terenius, O., and Matsuoka, H. (2008). Male fertility of malaria parasites is determined by GCS1, a plant-type reproduction factor. *Curr Biol* 18:607-613.
- Huck, N., Moore, J.M., Federer, M., and Grossniklaus, U. (2003). The *Arabidopsis* mutant *feronia* disrupts the female gametophytic control of pollen tube reception. *Development (Cambridge, England)* 130:2149-2159.
- Ingouff, M., Sakata, T., Li, J., Sprunck, S., Dresselhaus, T., and Berger, F. (2009). The two male gametes share equal ability to fertilize the egg cell in *Arabidopsis thaliana*. *Curr Biol* 19:R19-20.
- Johnson, M.A., von Besser, K., Zhou, Q., Smith, E., Aux, G., Patton, D., Levin, J.Z., and Preuss, D. (2004). *Arabidopsis hapless* mutations define essential gametophytic functions. *Genetics* 168:971-982.
- Kasahara, R.D., Portereiko, M.F., Sandaklie-Nikolova, L., Rabiger, D.S., and Drews, G.N. (2005). MYB98 is required for pollen tube guidance and synergid cell differentiation in *Arabidopsis*. *Plant Cell* 17:2981-2992.
- Kawai-Toyooka, H., Mori, T., Hamaji, T., Suzuki, M., Olson, B.J., Uemura, T., Ueda, T., Nakano, A., Toyoda, A., Fujiyama, A., et al. (2014). Sex-specific posttranslational regulation of the gamete fusogen GCS1 in the isogamous volvocine alga *Gonium pectorale*. *Eukaryot Cell* 13:648-656.
- Kessler, S.A., Shimosato-Asano, H., Keinath, N.F., Wuest, S.E., Ingram, G., Panstruga, R., and Grossniklaus, U. (2010). Conserved molecular components for pollen tube reception and fungal invasion. *Science* 330:968-971.
- Kielian, M., and Rey, F.A. (2006). Virus membrane-fusion proteins: more than one way to make a hairpin. *Nat Rev Microbiol* 4:67-76.
- Koncz, C., and Schell, J. (1986). The Promoter of TI-DNA Gene 5 Controls the Tissue-Specific Expression of Chimeric Genes Carried by a Novel Type of *Agrobacterium* Binary Vector. *Molecular & General Genetics* 204:383-396.
- Lampropoulos, A., Sutikovic, Z., Wenzl, C., Maegele, I., Lohmann, J.U., and Forner, J. (2013). GreenGate---a novel, versatile, and efficient cloning system for plant transgenesis. *PloS one* 8:e83043.
- Lescar, J., Roussel, A., Wien, M.W., Navaza, J., Fuller, S.D., Wengler, G., Wengler, G., and Rey, F.A. (2001). The Fusion glycoprotein shell of Semliki Forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH. *Cell* 105:137-148.
- Leydon, A.R., Tsukamoto, T., Dunatunga, D., Qin, Y., Johnson, M.A., and Palanivelu, R. (2015). Pollen Tube Discharge Completes the Process of Synergid Degeneration That Is Initiated by Pollen Tube-Synergid Interaction in *Arabidopsis*. *Plant Physiol* 169:485-496.

- Liao, M., and Kielian, M. (2005). Domain III from class II fusion proteins functions as a dominant-negative inhibitor of virus membrane fusion. *J Cell Biol* 171:111-120.
- Liu, Y., Tewari, R., Ning, J., Blagborough, A.M., Garbom, S., Pei, J., Grishin, N.V., Steele, R.E., Sinden, R.E., Snell, W.J., et al. (2008). The conserved plant sterility gene HAP2 functions after attachment of fusogenic membranes in *Chlamydomonas* and *Plasmodium* gametes. *Genes Dev* 22:1051-1068.
- McCue, A.D., Cresti, M., Feijo, J.A., and Slotkin, R.K. (2011). Cytoplasmic connection of sperm cells to the pollen vegetative cell nucleus: potential roles of the male germ unit revisited. *J Exp Bot* 62:1621-1631.
- Michard, E., Dias, P., Feijo, J.A. (2008). Tobacco pollen tubes as cellular models for ion dynamics: improved spatial and temporal resolution of extracellular flux and free cytosolic concentration of calcium and protons using pHluorin and YC3.1 CaMeleon. *Sexual Plant Reproduction* 21:169-181.
- Miesenbock, G., De Angelis, D.A., and Rothman, J.E. (1998). Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins. *Nature* 394:192-195.
- Modis, Y., Ogata, S., Clements, D., and Harrison, S.C. (2003). A ligand-binding pocket in the dengue virus envelope glycoprotein. *Proc Natl Acad Sci U S A* 100:6986-6991.
- Modis, Y., Ogata, S., Clements, D., and Harrison, S.C. (2004). Structure of the dengue virus envelope protein after membrane fusion. *Nature* 427:313-319.
- Mori, T., Kuroiwa, H., Higashiyama, T., and Kuroiwa, T. (2006). GENERATIVE CELL SPECIFIC 1 is essential for angiosperm fertilization. *Nat Cell Biol* 8:64-71.
- Moseyko, N., and Feldman, L.J. (2001). Expression of pH-sensitive green fluorescent protein in *Arabidopsis thaliana*. *Plant Cell Environ* 24:557-563.
- Okamoto, M., Yamada, L., Fujisaki, Y., Bloomfield, G., Yoshida, K., Kuwayama, H., Sawada, H., Mori, T., and Urushihara, H. (2016). Two HAP2-GCS1 homologs responsible for gamete interactions in the cellular slime mold with multiple mating types: Implication for common mechanisms of sexual reproduction shared by plants and protozoa and for male-female differentiation. *Dev Biol* 415:6-13.
- Okuda, S., Tsutsui, H., Shiina, K., Sprunck, S., Takeuchi, H., Yui, R., Kasahara, R.D., Hamamura, Y., Mizukami, A., Susaki, D., et al. (2009). Defensin-like polypeptide LUREs are pollen tube attractants secreted from synergid cells. *Nature* 458:357-361.
- Palanivelu, R., and Preuss, D. (2006). Distinct short-range ovule signals attract or repel *Arabidopsis thaliana* pollen tubes in vitro. *BMC Plant Biol* 6:7.
- Perez-Vargas, J., Krey, T., Valansi, C., Avinoam, O., Haouz, A., Jamin, M., Raveh-Barak, H., Podbilewicz, B., and Rey, F.A. (2014). Structural basis of eukaryotic cell-cell fusion. *Cell* 157:407-419.
- Rey, F.A., Heinz, F.X., Mandl, C., Kunz, C., and Harrison, S.C. (1995). The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution. *Nature* 375:291-298.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nature methods* 9:676-682.
- Shen, Y., Rosendale, M., Campbell, R.E., Perrais, D. (2014). pHuji, a pH-sensitive red fluorescent protein for imaging of exo- and endocytosis. *Journal of Cell Biology* 207:419-432.
- Sprunck, S., Rademacher, S., Vogler, F., Gheyselinck, J., Grossniklaus, U., and Dresselhaus, T. (2012). Egg cell-secreted EC1 triggers sperm cell activation during double fertilization. *Science* 338:1093-1097.

- Steele, R.E., and Dana, C.E. (2009). Evolutionary history of the HAP2/GCS1 gene and sexual reproduction in metazoans. *PLoS one* 4:e7680.
- Takeuchi, H., and Higashiyama, T. (2012). A species-specific cluster of defensin-like genes encodes diffusible pollen tube attractants in *Arabidopsis*. *PLoS Biol* 10:e1001449.
- Tsukamoto, T., Qin, Y., Huang, Y., Dunatunga, D., and Palanivelu, R. (2010). A role for LORELEI, a putative glycosylphosphatidylinositol-anchored protein, in *Arabidopsis thaliana* double fertilization and early seed development. *Plant J* 62:571-588.
- von Besser, K., Frank, A.C., Johnson, M.A., and Preuss, D. (2006). *Arabidopsis HAP2(GCS1)* is a sperm-specific gene required for pollen tube guidance and fertilization. *Development (Cambridge, England)* 133:4761-4769.
- White, J.M., Delos, S.E., Brecher, M., and Schornberg, K. (2008). Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme. *Crit Rev Biochem Mol Biol* 43:189-219.

Chapter 5: HAP2 Fusion Helix: Conservation in Flowering Plants and Expression in Sperm

Jennifer Forcina, Samyukta Mallick, Kristin Beale, and Mark Johnson.

This chapter is currently unpublished. I performed all the transmission analysis, imaging, and image analysis. I drafted the manuscript. Samyukta Mallick helped with cloning and plant transformations. Kristin Beale generated the double cysteine mutant construct.

Summary

HAP2 is a gamete fusion protein, and our previous analysis of Arabidopsis HAP2 revealed an amphipathic fusion helix at the membrane interaction surface of the protein. The membrane-interacting region is located at the tip of the extracellular region of HAP2 and corresponds to the fusion loop of class II viral fusion proteins. Our previously generated structure-based alignment revealed high variability in the membrane interacting region of HAP2, but it also revealed potential membrane interacting motifs, including the amphipathic fusion helix in flowering plants. Previously, we demonstrated that the fusion helix from rice HAP2 is able to function in the context of Arabidopsis HAP2. We further tested the functional conservation of the fusion helix by replace the Arabidopsis fusion helix with the corresponding region in HAP2 from *Amborella trichopoda*, which is sister to the flowering plants. An extended region from *Trypanosoma cruzi* HAP2 loop 2, corresponding to the fusion helix in Arabidopsis HAP2, was unable to function in the context of Arabidopsis HAP2. We also identified a non-functional double cysteine mutant that disrupts two different disulfide bonds outside the fusion helix. To test whether HAP2-YFP fusion protein accumulated at the sperm membrane while in the pollen tube, we co-expressed different variants of HAP2-YFP with a sperm membrane marker. We observed decreased membrane localization of the double cysteine mutant YFP protein compared to wild-type HAP2-YFP. Our results demonstrate functional conservation of the flowering plant fusion helix motif and suggest that HAP2 protein can be localized to the sperm membrane within the growing pollen tube.

Introduction

Gamete fusion is an essential step in fertilization, but very little is known about how gamete membranes fuse. Currently, only one gamete fusion protein has been identified. HAP2(GCS1) is a class II gamete fusion protein (Fedry et al., 2017), and it is broadly expressed in species from four of the five major eukaryotic kingdoms, the exception being fungi (Cole et al., 2014; Ebchuqin et al., 2014; Hirai et al., 2008; Johnson et al., 2004; Kawai-Toyooka et al., 2014; Liu et al., 2008; Mori et al., 2006; Okamoto et al., 2016; Steele and Dana, 2009; von Besser et al., 2006). Biochemical studies using HAP2 from four different species demonstrates the ability of the membrane interaction surface of HAP2 to interact with artificial membranes in vitro ((Fedry et al., 2017; Fedry*, Submitted; Pinello et al., 2017); Chapter 2). The membrane interaction surface is predicted to be highly variable across the range of species that have HAP2 in their genome ((Fedry*, Submitted); Chapter 2). Variability in this region has been demonstrated for two HAP2 proteins: HAP2 from the flowering plant *Arabidopsis thaliana* and HAP2 from the kinetoplastid parasite *Trypanosoma cruzi* ((Fedry*, Submitted); Chapter 2).

Arabidopsis HAP2 uses an amphipathic helix at its membrane interaction surface ((Fedry*, Submitted); Chapter 2). The *Arabidopsis* fusion helix requires a patch of hydrophobic amino acids (I¹⁷¹, F¹⁷², I¹⁷⁶) to mediate its function in vivo and membrane interaction in vitro ((Fedry*, Submitted); Chapter 2). The characteristics of the fusion helix, including its amphipathicity, are conserved in flowering plants, and the fusion helix from rice (*Oryza sativa*) was able to functionally replace the *Arabidopsis* fusion helix ((Fedry*, Submitted); Chapter 2). However, chimeric HAP2 protein with the entire N-terminal ectodomain from rice in the context of the remaining *Arabidopsis* HAP2 protein was non-functional (Wong et al., 2010), suggesting some elements of the N-terminal domain, outside the fusion helix, regulated HAP2-mediated gamete fusion in vivo. Furthermore, a

lysine (K¹⁷⁹) found basal to the fusion helix is conserved in flowering plants and mutating K¹⁷⁹ to alanine affected HAP2 function in vivo but not in vitro ((Fedry*, Submitted); Chapter 2), suggesting that this particular charged amino acid may play a role in the fusion process after membrane insertion.

Double fertilization in flowering plants involves two sperm, contained within a pollen grain, that are delivered to the female gametes via polarized growth of the pollen tube (Berger et al., 2008; Dresselhaus et al., 2016). The sperm are attached to each other and the pollen tube nucleus as they travel within the pollen tube through the female plant tissue (Dresselhaus et al., 2016; McCue et al., 2011). When the pollen tube bursts, sperm are released to the site of fusion (Berger et al., 2008; Dresselhaus et al., 2016; Hamamura et al., 2011; Ingouff et al., 2007). After the sperm have remained at the site of fusion for ~7-8 minutes (Hamamura et al., 2011), one sperm fuses with the egg cell to form a zygote, and the other cell fuses with the central cell to form a triploid endosperm (Berger et al., 2008; Dresselhaus et al., 2016; Hamamura et al., 2011; Ingouff et al., 2007). The processes of regulation that prevent HAP2 from fusing the two sperm with other or with the same female gamete membrane, i.e. both with either the egg cell or the central cell, remain unclear. Proteins secreted by the egg cell, called EC1 proteins, have been shown to positively regulate movement of HAP2 protein to the sperm membrane in vitro and are important for gamete fusion in vivo (Sprunck et al., 2012). To function properly, HAP2 must be present at the membrane at the site of fusion, yet it should initiate fusion with any nearby membrane.

In order to properly assess HAP2 function, it is important to understand both how the protein interacts with the target membrane to initiate fusion and how those processes are regulated within the sperm cell prior to arrival at the site of fusion. Therefore, we were interested in analyzing the conservation of the membrane interaction surface of HAP2 from flowering plants and how variations in the HAP2 protein affect its expression in sperm

cells within the pollen tube. First, we assessed the role of several amino acids in and immediately surrounding the fusion helix. We found that the charge of lysine 179 is important for HAP2 function. We also replaced the fusion helix with the corresponding amino acids from *Amborella trichopoda* ((Fedry*, Submitted); Chapter 2), which is sister to all other flowering plants (Amborella Genome, 2013; Chaw et al., 2004), and we found that that the Amborella fusion helix can function in the context of Arabidopsis HAP2 ((Fedry*, Submitted); Chapter 2). Conversely, replacement of a the fusion helix with a sequence containing fusion loop 2 from *T. cruzi* ((Fedry*, Submitted); chapter 2) was not functional in the context of Arabidopsis HAP2. We expressed ER and plasma membrane marker proteins in sperm to facilitate localization assays for HAP2-YFP variants. We found a double cysteine mutant expected to disrupt two disulfide bonds appears to be trafficked to the membrane less efficiently than wild-type HAP2. Together, our data provide insights into the conservation of the membrane interaction surface of HAP2 in flowering plants and provide tools for characterizing factors that may affect HAP2 translocation to or accumulation at the sperm plasma membrane.

Results

A positively charged amino acid is required at the base of the fusion helix.

Previously, we demonstrated a requirement for a hydrophobic face at the tip of the Arabidopsis HAP2 fusion helix for function ((Fedry*, Submitted); Chapter 2; Figure S1). We found that the single amino acid changes resulting in the strongest effect on fusion were F172A, K179A, and D173P. Strikingly, K179A was the only mutant tested that only affect HAP2 function in vivo but not vitro ((Fedry*, Submitted); Chapter 2). A lysine at this position in the HAP2-membrane interacting region is conserved in the flowering plants, including Amborella ((Fedry*, Submitted); chapter 2; Figure 1A). Therefore, we mutated K179 to arginine and found minimal effect on HAP2 function, as indicated by an average

percent transmission of 29.9% compared to the expected 33% for a fully functional, single-copy HAP2 transgene (K179R; Figure 1B and Table S1). We compared the transmission data from K179R to both wild-type HAP2 (HAP2_{cds}) and HAP2 with the K179A mutation from our previous study ((Fedry*, Submitted); Chapter 2) and found that K179R differed from WT HAP2 at $p < 0.05$ but from K179A at $p < 0.001$ (Wilcoxon rank sum test with continuity correction). Therefore, it appears that a charged amino acid at the position of K179 is important for HAP2 function.

A

A. thaliana: ...QRRMPSSCGDIFDKMIKGGK
O. sativa:HRVQSKCGDIWSKLTGKGK
A. trichopoda:...KRRVPYPCHDFVKTWILGK
T. cruzi: .PVGSRGLQVCDIFR.....G
 *At+Tc.loop2: ...QRRMPSSCDIFRGAALAS.

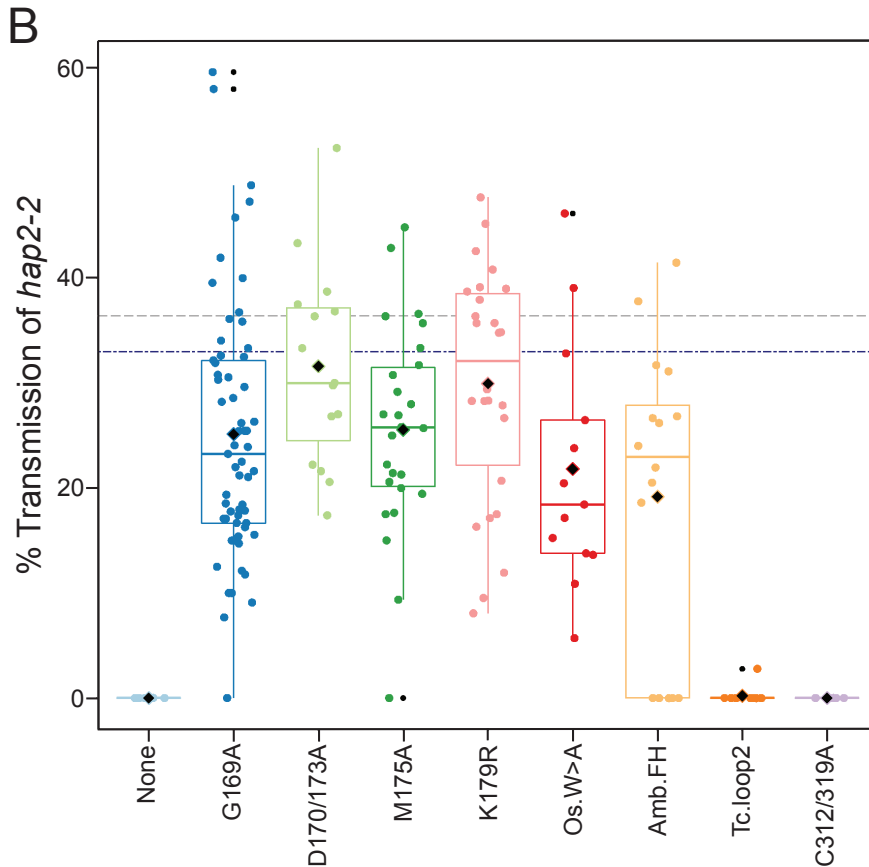


Figure 1. Genetic transmission assay results for HAP2-YFP mutants. A. This short alignment compares the *Arabidopsis thaliana* fusion helix (AtFH) region with the

corresponding regions in *Oryza sativa* (rice), *Amborella trichopoda*, and *Trypanosoma cruzi*. Asterisk indicates the sequence used in the Tc.loop2 replacement of the Arabidopsis fusion helix (“Tc.loop2”, panel **B**). **B**. This graph shows the rate of transmission of *hap2-2* as an assessment of HAP2-YFP variant function. All mutants used were in the *hap2-2/+* genotype, and *hap2-2/+* without a transgene was used as a pollen donor (“None”) as well. Each dot represents a cross, and this graph is a cumulative representation of all crosses for all lines for each construct (see Table S1 for total seed numbers and number of lines per construct). The black diamonds represent the mean for each group. The darker dotted line indicates the expected percent transmission for a single copy, fully functional transgene (33%). The lighter gray dotted line indicates the average percent transmission for HAP2cds previously reported ((Fedry*, Submitted); Chapter 2). The line within the boxplot indicates the median and the black dots indicate outlier data points.

Additional individual and double amino acid mutations have subtle effects on HAP2 function.

Since a glycine is not expected to participate in a helix (Pace and Scholtz, 1998) yet immediately precedes the Arabidopsis fusion helix, we tested whether G169 was important for HAP2 function and found a slight decrease in HAP2 function (G169A; Figure 1B and Table S1). We found a similar decrease in function when we mutated the methionine (M175) in the middle of the hydrophobic face at the tip of domain II to alanine (M175A; Figure 1B and Table S1). The result that the M175A mutation slightly affected HAP2 function, but did not eliminate it, is consistent with our previous finding that we needed to mutate the other three hydrophobic amino acids (I171, F172, I176) in combination (IFI>AAA) in that region to eliminate function ((Fedry*, Submitted); chapter 2; Figure S1). In addition, we found that generating a double aspartic acid mutant (D170.173A) had minimal effect on function (average transmission of 31.6%; Figure 1B and Table S1), which is consistent with our previous results that the single D173A mutation did not affect HAP2 function ((Fedry*, Submitted); Chapter 2). In combination with our previous HAP2-YFP mutant results ((Fedry*, Submitted); Chapter 2), our data demonstrate that the two aspartic acids on the charged side of the fusion helix are not

required for HAP2 function, but the hydrophobic face of the fusion helix is required. Additionally, no single amino acid mutation in the amphipathic helix tested so far has completely eliminated HAP2 function.

The fusion helix from Amborella is functional in Arabidopsis HAP2 but loop 2 from Trypanosoma is not.

The finding that the rice fusion helix could function in the context of Arabidopsis HAP2 raised some interesting questions. First, would the presence of a tryptophan in the rice fusion helix eliminate the need for a hydrophobic face? Second, how interchangeable is the fusion helix in flowering plants? Finally, how flexible are the restraints on which helices will function within Arabidopsis HAP2? To address these questions, we tested several variations in the fusion helix. We found that mutating the tryptophan in the rice fusion helix resulted in a partially functional HAP2 (Os.WtoA: average transmission of 21.8%; Figure 1B and Table S1). Replacing the fusion helix with the corresponding region in Amborella HAP2 (Figure 1A) also decreased Arabidopsis HAP2 function (Amb.FH: average transmission of 19.2%; Figure 1B and Table S1). We note however that the median transmission for Amb.FH is above the mean, and the crosses resulting in 0% transmission were all produced from the same primary transformed plant (Figure 1B). Since one primary transformed plant (out of four tested, Table S1) resulted in 0% transmission of the *hap2-2* allele, it is likely that the effect is the result of where the transgene landed in the genome rather than due to the replacement of the fusion helix with that from Amborella HAP2. However, this could be investigated further by following this particular plant into the next generation and by analyzing more primary transformed plants. Lastly, we generated an extended version of the *T. cruzi* HAP2 loop 2 (Figure 1A) that added five amino acids C-terminal to loop 2 that contribute to the *d* strand, which is a conserved component of *b*, *c*, *d* core element of the HAP2 structure ((Fedry*, Submitted);

Chapter 2). However, this *T. cruzi* loop 2 derivative (Tc.loop2) is predicted to be capable of forming a helix but with a low hydrophobic moment (0.251; Figure S2; (Gautier et al., 2008)) compared to the hydrophobic moments for Arabidopsis HAP2 (0.75) and rice HAP2 (0.73) reported previously ((Fedry*, Submitted); Chapter 2). Interestingly, the predicted hydrophobic moment for Amborella HAP2 is also fairly low (0.35; (Fedry*, Submitted); Chapter 2). However, the *T. cruzi* loop2 variant was non-functional in our in vivo genetics assay (Figure 1B and Table S1).

Several HAP2 fusion helix variants have similar expression patterns to wild-type HAP2 and correlate well with a sperm plasma membrane marker.

In order to test whether mutants in HAP2 affect its localization at the sperm plasma membrane, we generated a sperm membrane marker with SUC9cds (Leydon et al., 2017) fused to mCherry under the sperm-specific HTR10 promoter (Ingouff et al., 2007) and introduced this marker into plants expressing different HAP2-YFP fusion helix variants (Figure 2; variants from: (Fedry*, Submitted); Chapter 2). We observed a similar expression pattern with our sperm membrane marker (Figure 2) as reported previously for other membrane proteins expressed in sperm, specifically tetraspanin proteins TET9 (Sprunck et al., 2012) and TET12 (Boavida et al., 2013). Both our sperm membrane marker and the HAP2 fusion helix variants tested (HAP2cds, HAP2 Δ FH, IFI>AAA, and F172A; Figure 2) were observed in similar patterns in sperm cells within the pollen tube; both mCherry and YFP accumulated in the region between the two sperm and along the connection to the vegetative nucleus (Figure 2). It is important to note that both fluorescent proteins exhibited variable expression levels. However, we observed similar correlations between HAP2-YFP expression and sperm membrane expression for all HAP2 fusion helix variants tested in this study (Figure S3).

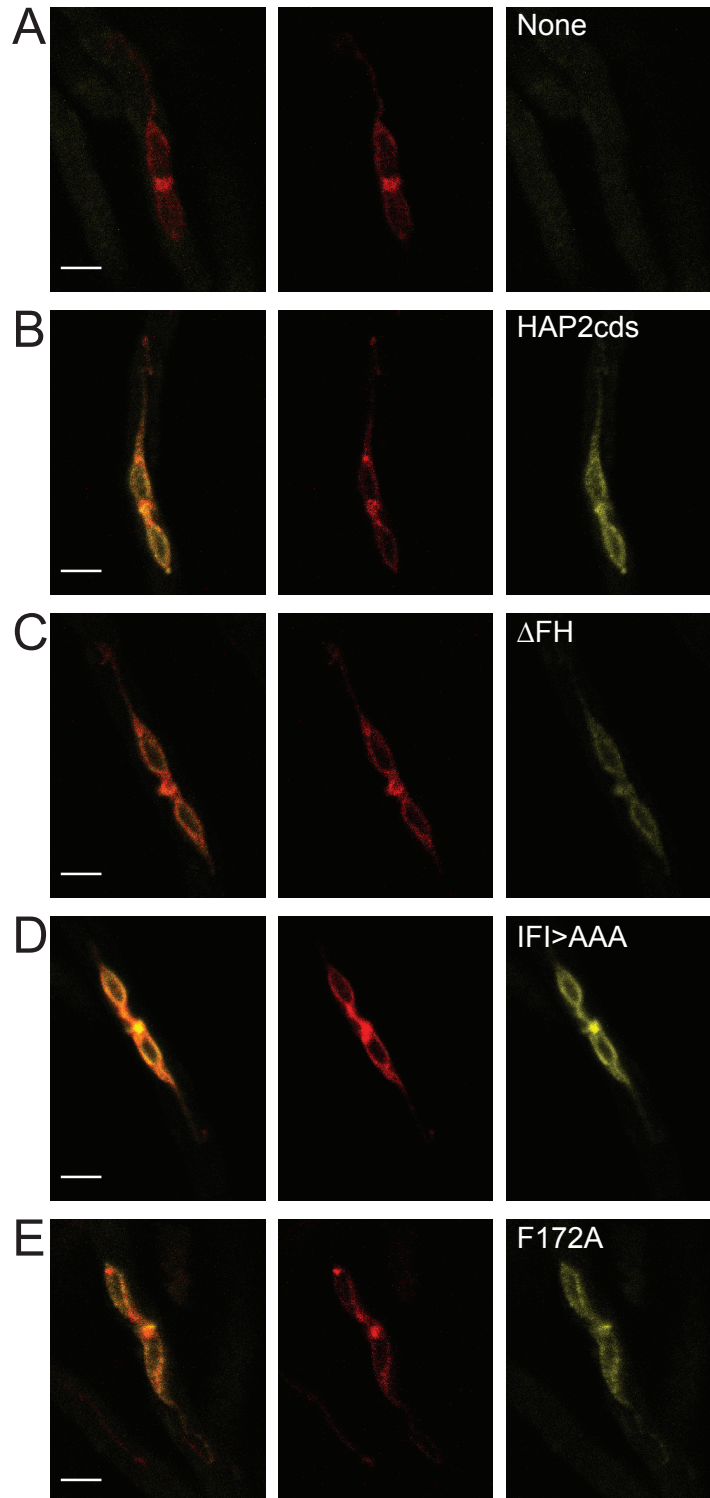


Figure 2. Expression of a sperm membrane marker with different HAP2 fusion helix variants. Sperm for this figure were imaged within growing pollen tubes in the semi-in-vivo (SIV) assay without ovules (see chapter 4, figure 2A). All panels are confocal images with merged RFP and YFP channels on the left, the RFP channel only in the middle, and

the YFP channel only on the right. The HAP2-YFP variant expressed in each pair of sperm is indicated in the upper left corner of the right panel. The sperm membrane marker is SUC9 fused to mCherry under the sperm-specific HTR10 promoter. **A.** A representative pair of sperm from a plant expressing the sperm membrane marker without any HAP2-YFP fusion protein is shown (“None”). **B.** A confocal image of a pair of sperm expressing the sperm membrane marker and the HAP2cds fused to YFP. **C-E.** Confocal images of example sperm expressing both the sperm membrane marker and HAP2-YFP variants ((Fedry*, Submitted); Chapter 2). **C.** This panel shows a sperm pair with the HAP2 Δ FH mutation and the sperm membrane marker. **D.** This confocal image shows an example pair of sperm expressing the HAP2-YFP triple hydrophobic variant (I¹⁷¹F¹⁷²I¹⁷⁶>AAA) and the sperm membrane marker. **E.** An example pair of sperm is shown expressing both the sperm membrane marker and HAP2-YFP with the F172A mutation. All panels are from primary plants after transformation with the sperm membrane marker, except panel B (HAP2cds) which is from a plant in the second generation after transformation. For all panels, scale bar represents 5 μ m.

Mutating two cysteines that contribute to different disulfide bonds eliminates HAP2 function and affects HAP2 localization.

We generated a HAP2 variant containing a double mutation, in which both C312 and C319 are mutated to alanine. The double cysteine mutant (C312A/C319A) is expected to disrupt two different disulfide bonds (Figure S4). The double cysteine mutant (C312/319A) was non-functional in our genetic transmission assay (C312/C319A; Figure 1B and Table S1). We also observed more variation in the pattern of YFP expression for the double cysteine mutant in sperm, both alone (data not shown) and in combination with our sperm membrane marker (Figure 3). Interestingly, the variation in HAP2-YFP expression ranged from overlapping in the connecting region between the sperm and around the edges of the sperm (Figure 3A and 3C) to an absence of HAP2-YFP expression in the connection between the two sperm (Figure 3E). Our co-localization analysis suggests a lower overall correlation between the HAP2 double cysteine mutant (C312/319A) and the sperm membrane marker compared to wild-type HAP2 and the sperm membrane maker (Figure S3).

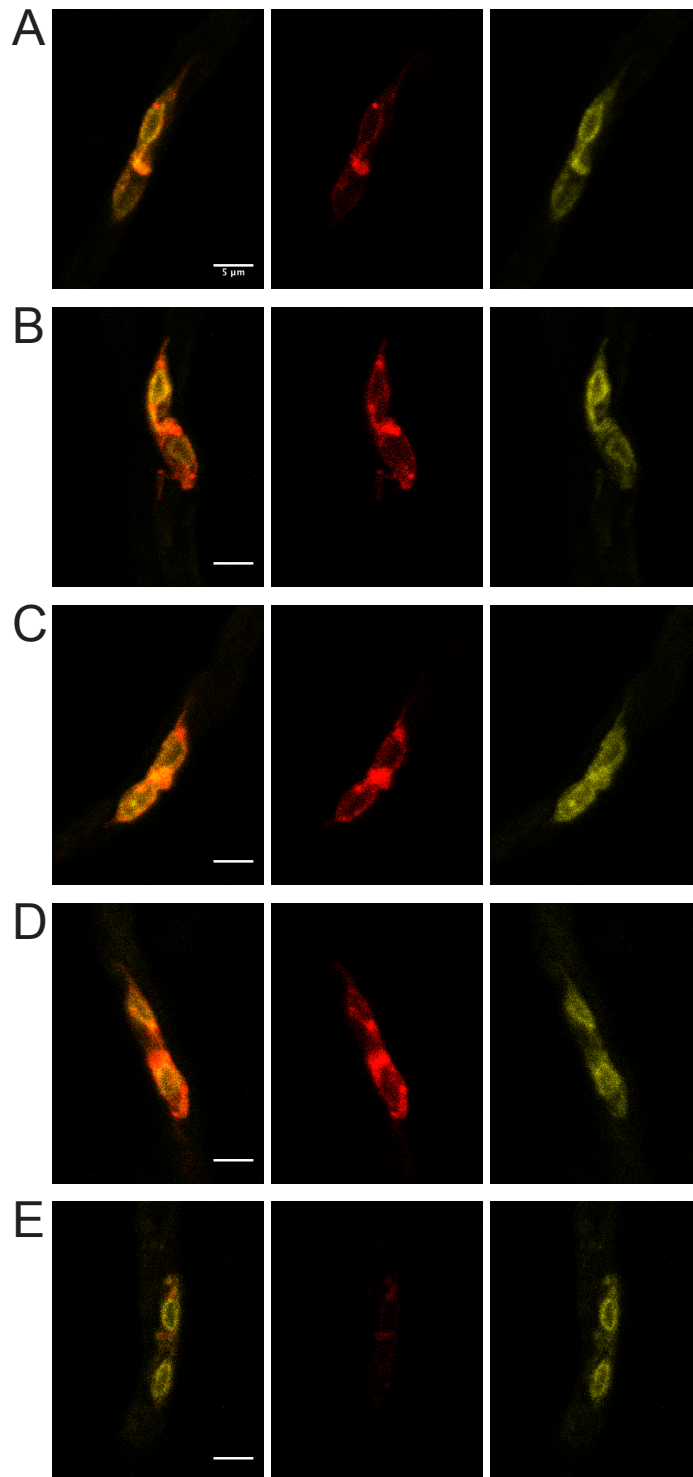


Figure 3. Expression of a sperm membrane marker with a HAP2-YFP double cysteine mutant. Sperm were imaged growing in pollen tubes in the SIV assay, as

described in Figure 2. All sperm shown are expressing both the sperm membrane mCherry marker and a HAP2-YFP variant with a double cysteine mutation (C312A/C319A) which is expected to affect cysteines in two different disulfide bonds (see Figure S1). All panels show the RFP and YFP channels merged in the left image. The RFP only channel is shown in the center image, and the YFP only channel is shown in the right image. The sperm membrane marker was transformed into plants expressing HAP2^{C312A/C319A}-YFP. Panels **A-B** are from one primary transformed plant. Panels **C-D** are from another primary transformed plant, and panel **E** is from a third plant. Panels **A-E** are examples of the different expression patterns observed in the HAP2^{C312A/C319A}-YFP sperm. Scale bar represents 5µm.

Expression of an ER retention signal in sperm reveals a different pattern than observed for either the sperm membrane marker or HAP2-YFP variants.

We have observed a pattern of highly expressed HAP2genomic-YFP that can appear as a single more perinuclear ‘ring’ inside of the sperm cell or as a ‘double-ring’ that suggests HAP2-YFP expression may be found both perinuclear and in the sperm membrane (Figure S5). This ‘ring’ pattern has also been described previously for isolated sperm expressing HAP2genomic-YFP in vitro (Sprunck et al., 2012). We have also observed this pattern for higher expressing lines of other HAP2-YFP variants (data not shown). Therefore, we are interested in whether any HAP2-YFP signal overlaps with the endoplasmic reticulum (ER) in sperm. We therefore expressed mCherry fused to an ER retention marker under the sperm-specific HTR10 promoter and observed a more punctate-like pattern distributed throughout the sperm (Figure 4). These puncta could also be found accumulating near the midpoint between the two sperm and at the edges of the sperm cell as well (Figure 4). Without co-expression with a plasma membrane marker or HAP2-YFP, we cannot make any strong conclusions; however, we can now use this marker to address the question of whether any HAP2-YFP co-localizes with an ER marker in Arabidopsis sperm cells. We can also use co-localization analysis of different cell compartments to interrogate where HAP2 is expressed in the sperm cell at different times

during fertilization, and these tools could provide opportunities to address how HAP2 is regulated.

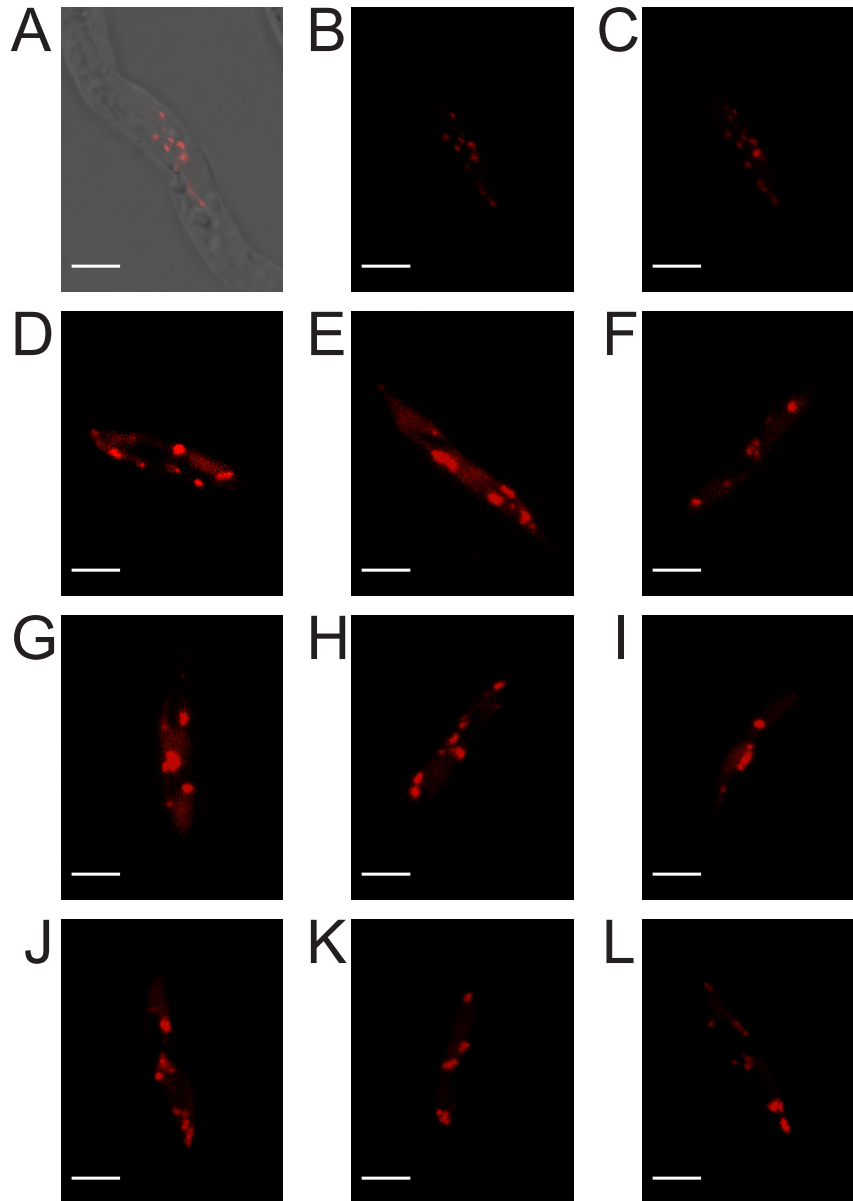


Figure 4. Expression of a sperm ER marker. All sperm shown are from plants expressing mCherry fused to a KDEL sequence under the sperm-specific HTR10 promoter. All images are of sperm growing in SIV pollen tubes. Panel **A** is a merge of the RFP and DIC channels. Panels **A-B** are the same slice from the z-stack projected in **C**. Panels **D-G** show examples of sperm pairs from the same primary transformed plant. Panel **D** is a single confocal image of a pair of sperm, while panels **E-L** are maximum

projections of confocal z-stacks of sperm pairs. The sperm pairs shown in panels **A-C** and **H-I** are from a different primary transformed plant than the one used for panels **D-G**. Panels **J-L** are from a third primary transformed plant. Scale bar represents 5µm.

Discussion

The membrane interaction region of HAP2 is at the tip of the ectodomain of the protein and has been demonstrated to interact directly with artificial membranes in vitro, for several different HAP2 proteins ((Fedry et al., 2017; Fedry*, Submitted; Pinello et al., 2017); Chapter 2). We have previously shown that there are amino acid motifs within the membrane interaction region of HAP2 that appear to be lineage-specific ((Fedry*, Submitted); Chapter 2). Specifically, Arabidopsis HAP2 uses an amphipathic fusion helix as its membrane interaction surface, and the overall character of this fusion helix is conserved in the flowering plants ((Fedry*, Submitted); Chapter 2). Here we show that the HAP2 fusion helix from Amborella is able to function within the context of the Arabidopsis HAP2 protein. We also demonstrated that the tryptophan in the rice (*O. sativa*) fusion helix is not the sole contributor to its function in the context of Arabidopsis HAP2. Furthermore, we have demonstrated that an extended version of the *T. cruzi* loop 2, which corresponds to the region of the fusion helix in flowering plants, is not a functional replacement for the Arabidopsis fusion helix.

Our *T. cruzi* loop 2 data suggests that the Arabidopsis fusion helix region cannot be replaced by any random sequence of similar length (10-11 amino acids). Our *T. cruzi* loop 2 replacement helix consists of the loop 2 sequence extended into the *d* strand, and this 10 amino acid segment has the potential to form a helix with a low hydrophobic moment (Figure S2; Heliquest server: (Gautier et al., 2008)). Based on these results, it will be interesting to test whether we can mimic the membrane interaction surface of non-flowering plant HAP2 proteins in the context of Arabidopsis HAP2. HAP2 proteins from non-flowering plants different in length and amino acid composition from the fusion helix

motif observed in flowering plants, and some have the potential for two fusion loops at their membrane interaction surface ((Fedry*, Submitted); Chapter 2). It is important to note that the entire membrane interaction surface of *T. cruzi* was not used in this experiment, so it will be interesting to test whether replacing the Arabidopsis membrane interaction surface with the full regions corresponding to fusion loops 1 and 2 from *T. cruzi* ((Fedry*, Submitted); Chapter 2), *Chlamydomonas reinhardtii* ((Fedry et al., 2017; Fedry*, Submitted); Chapter 2), and non-flowering plants can result in functional HAP2 proteins ((Fedry*, Submitted); Chapter 2).

Our co-expression analysis suggests that even the non-functional fusion helix variants from Arabidopsis HAP2 are localized similarly to our sperm membrane marker within the growing pollen tube. Expression of a non-functional double cysteine mutant that is predicted to disrupt two disulfide bonds and therefore could result in misfolded protein did not correlated well with the expression of the plasma membrane marker. Interesting, an ER marker in sperm appears to have a different pattern from any we have observed with HAP2-YFP variants. We acknowledge that expression levels may affect localization of fusion proteins in sperm, and therefore, it will be important to continue this analyses with more lines and differing levels of expression. While resolution within the sperm cells is limited in this study, it is important to assess HAP2 expression and localization in sperm and specifically within growing pollen tubes since expression of HAP2 in other cells may not be localized properly (Chapter 3) due to the absence of potential regulatory proteins found in sperm.

Our analysis of HAP2-YFP fusion helix variants and the sperm membrane marker suggest that HAP2-YFP protein is accumulating at the sperm plasma membrane. HAP2-YFP even appears to be enriched in the area between the two sperm cells and was observed in the connection to the vegetative nucleus of the pollen tube. Together, these observations highlight important areas of HAP2 regulation that warrant further study. First,

the proposed mechanism of EC1 proteins is to activate sperm by driving translocation of HAP2 to the sperm plasma membrane (Dresselhaus et al., 2016; Sprunck et al., 2012). However, in this study, we observed HAP2-YFP expression at the sperm plasma membrane, often without accompanying perinuclear, or 'ring'-like, expression, which suggests that HAP2 in the pollen tube may already be localized to the plasma membrane. One possible explanation is that EC1 proteins activate sperm by ensuring that HAP2 accumulates in greater quantities at the sperm plasma membrane prior to gamete fusion. Alternatively, EC1 could activate sperm by initiating translocation of other proteins to the sperm membrane that may participate in the HAP2-mediated gamete fusion mechanism, and any HAP2 that is not membrane-localized may be driven to the plasma membrane as a result of this more general sperm activation process. Another interesting question that remains is why the two sperm do not fuse with each other if they are both expressing HAP2, and HAP2 appears to accumulate in the connecting area between the two sperm. Additionally, why do the sperm wait at the site of fusion for an extended period of time before fusing? In order to address both of these questions, we need to determine how HAP2 is regulated. For example, regulatory proteins may bind to HAP2 and block its fusion function by concealing the fusion helix or maintaining HAP2 in an inactive conformation. Identification of proteins that regulate or interact with HAP2 will contribute to our understanding of the HAP2-mediated gamete membrane fusion process, which may also require other proteins to coordinate the membrane environment at the site of gamete fusion.

Methods

Growth of transgenic plants.

Plants used in this study were germinated on plates and maintained in a 16 hour/8 hour as described previously ((Fedry*, Submitted); chapter 2).

Generation and Analysis of HAP2-YFP variants.

Site-directed mutagenesis of the HAP2 coding sequence fused to YFP and under its native promoter was performed as described previously (NEB Q5 Site-Directed Mutagenesis Kit; (Fedry*, Submitted); Chapter 2). Mutagenesis was performed using primers listed in Table S2, and sequencing was confirmed using Sanger sequencing (Genewiz). Transformation of HAP2-YFP variants into *hap2-2/+* (SALK_152706; (Alonso et al., 2003; von Besser et al., 2006)) plants was performed as described previously ((Fedry*, Submitted); Chapter 2).

Genetic transmission assay was performed as described previously ((Fedry*, Submitted); chapter 2). In brief, after transformation, primary transformed seeds (T1) were selected on Murashige and Skoog (MS) solid media (Sigma Aldrich) supplemented with Basta (glufosinate ammonium at 25mg/L; Oakwood Chemical) to select for plants transformed with the HAP2-YFP variant transgene. Selected T1 seeds were screened for the *hap2-2* allele (see primers in Table S2) by PCR and for the presence of YFP in pollen grains by fluorescence microscopy (Zeiss Axiovert 200M). Pistils from *male-sterile 1 (ms1)* plants (Wilson et al., 2001) were pollinated with HAP2-YFP variants, and seedlings resulting from these crosses were analyzed for segregation of kanamycin (50mg/L; ThermoFisher/ThermoScientific) resistance associated with the *hap2-2* allele.

Sperm specific expression constructs.

Constructs were generated using the GreenGate cloning method (gift from Jan Lohmann; Addgene kit # 1000000036; (Lampropoulos et al., 2013)) as described previously (chapter 3). The sperm membrane marker was generated with the HTR10 promoter entry vector (chapter 3), B-dummy (GreenGate: B003), SUC9cDNA entry vector (Chapter 3), linker-mCherry (D003), UBQ terminator (E009), and Hygromycin resistance (F005) in the pGGZ003 destination vector. The sperm ER marker was generated using the HTR10 promoter entry vector, B-dummy, 3xmCherry (C026), HDEL (D008), UBQ

terminator, and Hygromycin resistance in pGGZ003. Plasmids were confirmed by Sanger sequencing and transformed into plants as described previously (Chapter 3). The sperm membrane marker was transformed into *hap2-2/+* plants and T2 plants from one T1 line of each HAP2-YFP variant, which were segregating both the HAP2-YFP variant transgene and *hap2-2/+*. The sperm ER marker was transformed into Col0 plants. Primary transformed seeds (T1) were selected on MS media supplemented with Hygromycin at 25mg/L (Corning). Analysis was performed on pollen from the T1 generation.

Image analysis of pollen expressing both the sperm membrane marker and HAP2-YFP variants.

Pollen grains were imaged in the semi-in-vitro assay without ovules (Palanivelu and Preuss, 2006), as described previously (Chapter 4). Pollen tubes were allowed to grow outside the cut pistils on solid pollen growth media (PGM; chapter 4; (Boavida and McCormick, 2007)) for ~3-5 hours at 22C before liquid PGM was added for mounting with coverslips. Images of sperm markers and HAP2-YFP variants were taken with a Zeiss LSM 800 confocal microscope and processed in ImageJ (FIJI version). HAP2-YFP images were imaged using the 488 nm laser line, and sperm membrane mCherry markers were imaged using the 561 nm laser line (587 nm excitation). Images were captured at 1024x1024, 16 bit, 40x water magnification.

For generating sperm membrane marker and HAP2-YFP images (Figures 2-3), background subtraction was performed with a rolling ball radius of 30 pixels. For analysis of co-localization, the Coloc2 ImageJ plugin was used with an ROI drawn around the sperm pair and the threshold regression of Costes; PSF and Costes randomizations were set to default (3.0 and 10, respectively). Col-localization data were filtered to include only those with a Pearson's R value below threshold that equaled zero and a Costes P -value > 0.95. The Pearson's R value (with no threshold) was used to generate the Pearson's R correlation graph. Sperm images from two T1 lines (including T2 data from one line) were

used for the HAP2cds co-localization analysis (n=9 sperm pairs). Sperm images from 1 T1 line (including T2 data) was used for HAP2ΔFH (n=8). Sperm images were from 3-6 T1 lines for F172A (n=8, IFI>AAA (n=13), and C312/319A (n=14), respectively.

Sperm Isolation and imaging.

Sperm isolated protocol was adapted from a previously published method (Borges et al., 2012). Flowers were collected into 15mL tubes, and filter-sterilized sperm extraction buffer (SEB: 7mM MOPS pH6, 438 mM sucrose, 3.6mM CaCl₂, 5.83mM MgSO₄, 1.3mM H₃BO₃, 0.74mM KH₂PO₄) was added at a ratio of 1:2 with the volume of flowers. Flowers in SEB are vortexed at medium speed for 3 minutes at room temperature. The plant debris is then filtered over a 40um mesh into a 50mL conical tube, and 1-2mL additional SEB is used to rinse the tube and mesh. Pollen collection was then spun at ~2000 rpm for 1-2 minutes at 4C. Almost all supernatant was removed (except ~250ul), and pollen was re-suspended with 50-100ul acid washed beads (425-600um, Sigma). Another 500ul SEB is used to rinse the tube and is added to the pollen/bead mixture. Each tube was then wrapped in parafilm and vortexed at maximum speed for 4 minutes at room temperature. A 20um μMACS column was pre-wet with SEB and then used to filter the pollen debris. Sperm were then rested on ice for at least 30 minutes to recover. Samples can be concentrated by gently spinning at 200xg for 2 minutes at room temperature, but this may damage sperm. Sperm were live-imaged in the SEB buffer on a Zeiss LSM 800 confocal. Debris from the pollen tube and glass beads is observed.

Acknowledgements: We thank Dr. Anke Reinders and Dr. J.M. Ward for the SUC9 cDNA. We thank Samantha Paul and Alexander Leydon for cloning SUC9 into the C000 GreenGate entry vector.

Supplemental Figures

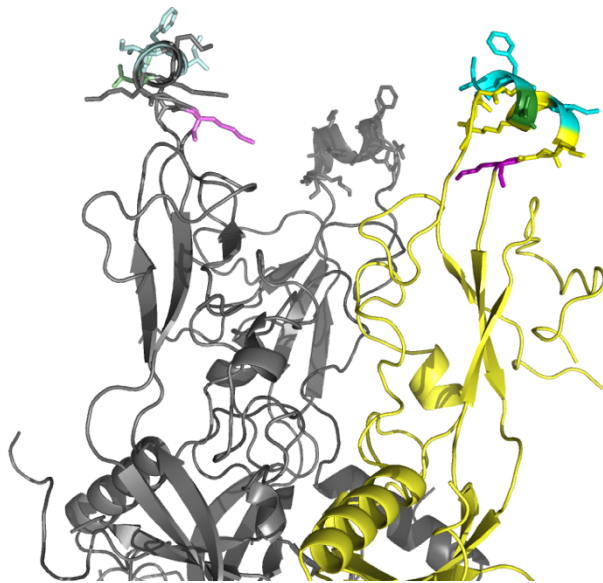


Figure S1. Arabidopsis HAP2 fusion helix at the tip of domain II. This is a limited view of the Arabidopsis HAP2 trimer structure ((Fedry*, Submitted); Chapter 2) with only an upper portion of the ectodomain shown to highlight the fusion helix. One monomer within the trimer has domain II colored in yellow, while the other two monomers are shown in gray. The sidechains are shown for the region that is replaced in the Arabidopsis fusion helix swap experiments. The hydrophobic amino acids identified previously in the non-functional triple mutant (IFI>AAA) are shown in cyan (I¹⁷¹/F¹⁷²/I¹⁷⁶), and the lysine (K¹⁷⁹) shown to be important for function in vivo is shown in magenta ((Fedry*, Submitted); Chapter 2).

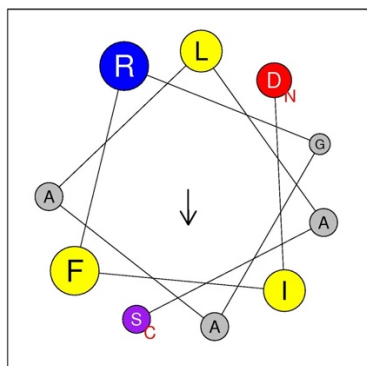


Figure S2. Predicted helix for the Tc.loop2 sequence used as a replacement for the Arabidopsis HAP2 fusion helix. This predicted helix was generated using the Heliquest server (Gautier et al., 2008) for the amino acids used in the Tc.loop2 construct that replaced the Arabidopsis fusion helix with the corresponding amino acids in Trypanosoma HAP2. The predicted hydrophobic moment for Tc.loop2 is low (0.251; (Gautier et al., 2008)).

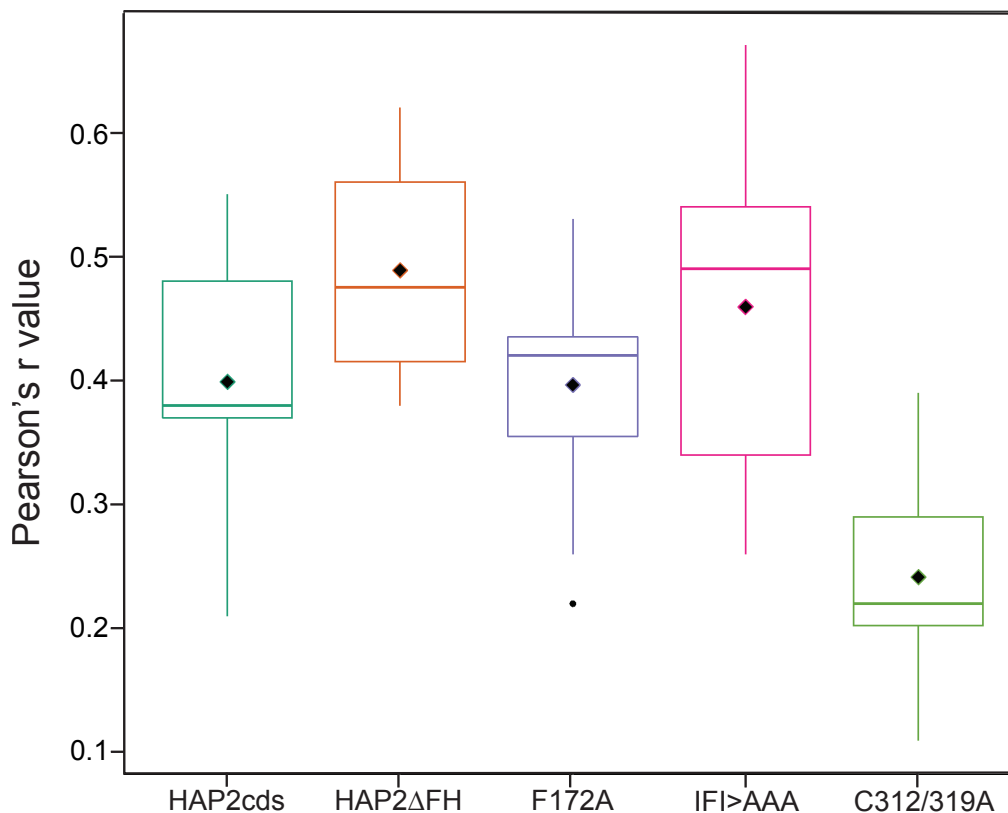


Figure S3. Pearson's correlation coefficient for different HAP2 variants with the sperm membrane marker. The Pearson's r value was determined by the coloc2 plugin in ImageJ (FIJI version; (Schindelin et al., 2012)). All data are combined for each sample. The mean is shown by the black diamond. The line inside the box indicates the median. The black dot represents an outlier data point. For HAP2cds, 9 total sperm pairs were analyzed. Sperm were from two different primary lines after transformation with the sperm membrane marker, and some sperm were from the second generation of one of those lines. For HAP2ΔFH, which is the deletion of the fusion helix, 8 sperm were analyzed, and all were from the same primary line and some were in the second generation. For F172A, 11 sperm were analyzed from 6 different primary transformed lines. For the triple hydrophobic mutant IFI>AAA, 13 sperm were analyzed from 3 different primary lines. For the double cysteine mutant C312A/C319A, 14 sperm were analyzed from 6 different primary lines.

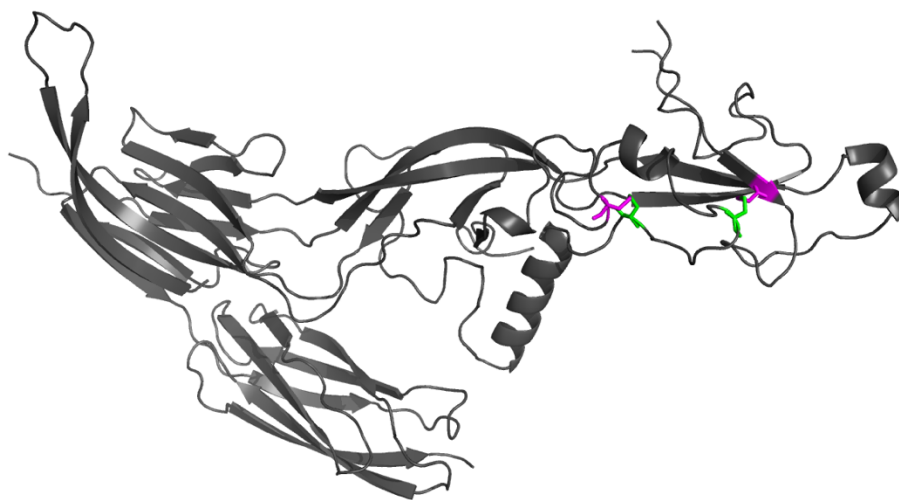


Figure S4. HAP2 double cysteine mutant C312A/C319A is expected to disrupt two disulfide bonds. A monomer of AtHAP2 is shown in gray. C312 and C319 are shown in green. The partner cysteines are shown in magenta. PDB: 5OW3 ((Fedry*, Submitted), Chapter 2).

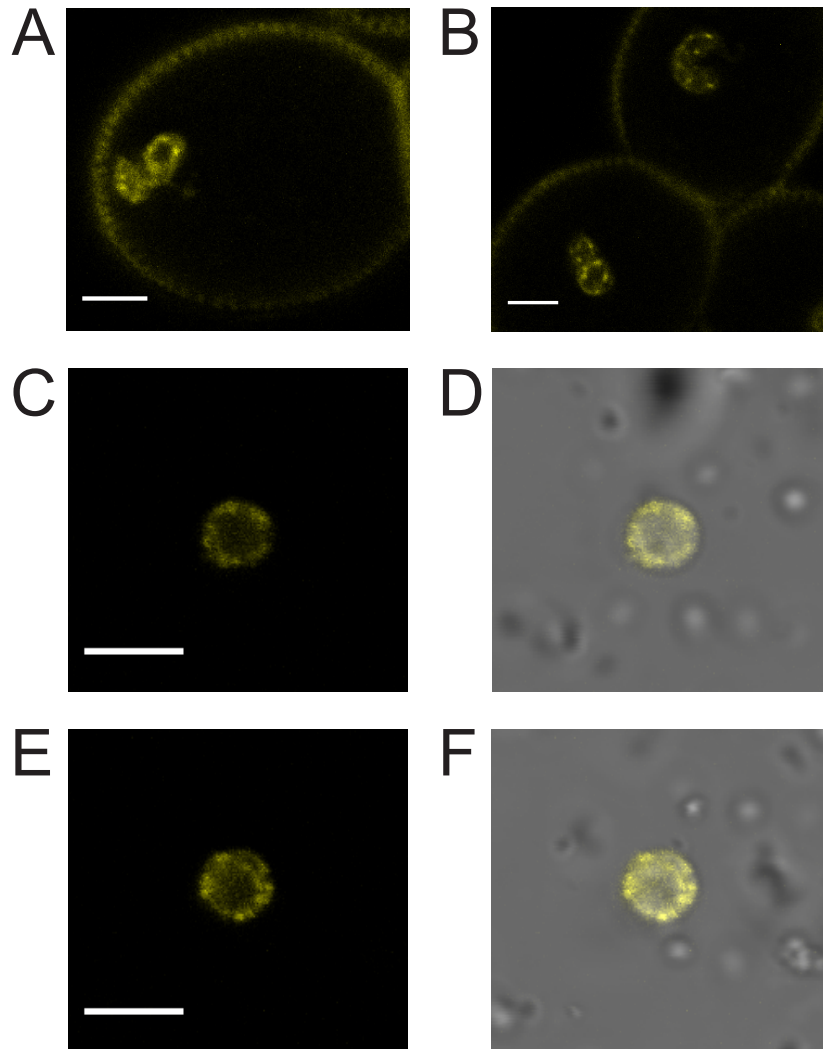


Figure S5. Examples of highly expressing HAP2 genomic fusion protein in sperm.

We have observed a line with high levels of HAP2genomic-YFP fusion protein (von Besser et al., 2006) expression in sperm. **A-B.** Examples of YFP expression pattern in sperm within pollen grains observed with highly expressing HAP2genomic-YFP constructs. Specifically, a ring-like pattern can be observed, which may have another 'ring' of signal outside it. The center 'ring' is similar to that reported for HAP2genomic-YFP in sperm in vitro (Sprunck et al., 2012). **C-F.** We manually isolated sperm cells expressing HAP2genomic-YFP from pollen and observed a double-ring pattern by in most cells (panels E-F, same sperm), although we have also included one that appears not to (panel **C-D**, same sperm). For **C-F**, the left image is from the GFP channel and the right image is a merge of GFP and DIC. For all images, scale bar indicates 5 μ m.

Table S1. HAP2-YFP variant plants used for the genetic transmission analysis.

Plasmid ID	Sample	# crosses	# seedlings	average seedlings/ cross	% KanR (avg)	# T1 lines
none	<i>hap2-2/+</i>	12	481	40.1	0.00	n/a
1060	G169A	61	2493	40.9	25.12	12
1432	D170.173A	15	510	34.0	31.60	4
1430	M175A	26	797	30.7	25.55	5
1679	K179R	26	1501	57.7	29.94	6
1723	Os.WtoA	13	721	55.5	21.81	3
1642	Amb.FH	16	664	41.5	19.18	4
1482	Tc.loop2	13	477	36.7	0.21	2
1062	C312.319A	7	122	17.4	0.00	2

All HAP2-YFP plants used for the analysis were in the *hap2-2/+* background. The total number of seedlings analyzed all from the number of crosses indicated across the number of primary transformed plants (# T1 lines) analyzed. The average percent transmission of kanamycin resistance (% KanR), and therefore the *hap2-2* allele, is provided for each HAP2-YFP variant.

Table S2. Primers used for site-directed mutagenesis of HAP2-YFP.

Primer Name	Primer	Reason
HAP2_FL_G169A_F1	GCCTTCTTCTTGTGCAGATATTTTGT	G169A
HAP2_FL_G169A_R1	ATTCTTCGTTGCGGTCCACAT	G169A
HAP2_D170A_F1	CTTCTTGTGGAGCTATTTTGTATAAGATG ATTAAAGG	D170A
HAP2_D170A_R1	AAGGCATTCTTCGTTGCGGTC	D170A
HAP2_D173A_in_D170A_F1	GCTATTTTGTCTAAGATGATTAAAGGGAA GGC	D170A/D173A
HAP2_D173A_in_D170A_R1	TCCACAAGAAGAAGGCATTCT	D170A/D173A
HAP2_M175A_F1	TTTGATAAGGCGATTAAAGGGAAGGC	M175A
HAP2_M175A_R1	AATATCTCCACAAGAAGAAGGCATTC	M175A
HAP2_K179R_F1	ATGATTAAAGGGAGGGCAAATACCGC	K179R
HAP2_K179R_R1	CTTATCAAAAATATCTCCACAAGAAGAAG GCAT	K179R
OsFH_WtoA_F2	GACAAAAGGGAAGGCAAATACCGCG	OsFH_WtoA
OsFH_WtoA_R2	AACTTACTTGCAATATCTCCACAAGAAGA A	OsFH_WtoA
HAP2_pFL.re.Amb_F2	ATTCTCGGGAAGGCAAATACCGCGCATT	Amborella FH
HAP2_pFL.re.Amb_R2	CCACGTCTTCACAAAATGGTCACAAGAAG AAGGC	Amborella FH
HAP2_re.Tc.pFL_F2	GCACTTGCAATCAGCAAATACCGCGCATT	Tc.loop2
HAP2_re.Tc.pFL_R2	AGCACCTCTAAAAATATCACAAGAAGAAG GCATTCTTC	Tc.loop2
HAP2Ex7R	GATTAACCTTTGAGAAAGTTATCATTTG	sequencing variants
ECORI_HAP2_CVFP_F1	gaattcCTTACTTACATTAGGGATGTACCT	sequencing variants
hap2SeqTR3	CAATCAAACCTGCGCAGAAGGAAGC	<i>hap2-2</i> genotyping
Hap2C2	GGCCTCACTCGTTCTCAATTGGAG	<i>hap2-2</i> genotyping
Lba1	TGGTTCACGTAGTGGGCCATCG	<i>hap2-2</i> genotyping

References

- Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., et al. (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301:653-657.
- Amborella Genome, P. (2013). The Amborella genome and the evolution of flowering plants. *Science* 342:1241089.
- Berger, F., Hamamura, Y., Ingouff, M., and Higashiyama, T. (2008). Double fertilization - caught in the act. *Trends Plant Sci* 13:437-443.
- Boavida, L.C., and McCormick, S. (2007). Temperature as a determinant factor for increased and reproducible in vitro pollen germination in *Arabidopsis thaliana*. *Plant J* 52:570-582.
- Boavida, L.C., Qin, P., Broz, M., Becker, J.D., and McCormick, S. (2013). *Arabidopsis* tetraspanins are confined to discrete expression domains and cell types in reproductive tissues and form homo- and heterodimers when expressed in yeast. *Plant Physiol* 163:696-712.
- Borges, F., Gardner, R., Lopes, T., Calarco, J.P., Boavida, L.C., Slotkin, R.K., Martienssen, R.A., and Becker, J.D. (2012). FACS-based purification of *Arabidopsis* microspores, sperm cells and vegetative nuclei. *Plant methods* 8:44.
- Chaw, S.M., Chang, C.C., Chen, H.L., and Li, W.H. (2004). Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. *J Mol Evol* 58:424-441.
- Cole, E.S., Cassidy-Hanley, D., Fricke Pinello, J., Zeng, H., Hsueh, M., Kolbin, D., Ozzello, C., Giddings, T., Jr., Winey, M., and Clark, T.G. (2014). Function of the male-gamete-specific fusion protein HAP2 in a seven-sexed ciliate. *Curr Biol* 24:2168-2173.
- Dresselhaus, T., Sprunck, S., and Wessel, G.M. (2016). Fertilization Mechanisms in Flowering Plants. *Curr Biol* 26:R125-139.
- Ebchuqin, E., Yokota, N., Yamada, L., Yasuoka, Y., Akasaka, M., Arakawa, M., Deguchi, R., Mori, T., and Sawada, H. (2014). Evidence for participation of GCS1 in fertilization of the starlet sea anemone *Nematostella vectensis*: implication of a common mechanism of sperm-egg fusion in plants and animals. *Biochem Bioph Res Co* 451:522-528.
- Fedry, J., Liu, Y., Pehau-Arnaudet, G., Pei, J., Li, W., Tortorici, M.A., Traincard, F., Meola, A., Bricogne, G., Grishin, N.V., et al. (2017). The Ancient Gamete Fusogen HAP2 Is a Eukaryotic Class II Fusion Protein. *Cell* 168:904-915 e910.
- Fedry*, J., Forcina*, J., Legrand, P., Péhau-Arnaudet, G., Haouz, A., Johnson, M.A., Rey, F.A., and Krey, T. (Submitted). HAP2 drives gamete fusion via insertion of highly divergent fusion loops into female gametes.
- Gautier, R., Douguet, D., Antonny, B., and Drin, G. (2008). HELIQUEST: a web server to screen sequences with specific alpha-helical properties. *Bioinformatics* 24:2101-2102.
- Hamamura, Y., Saito, C., Awai, C., Kurihara, D., Miyawaki, A., Nakagawa, T., Kanaoka, M.M., Sasaki, N., Nakano, A., Berger, F., et al. (2011). Live-Cell Imaging Reveals the Dynamics of Two Sperm Cells during Double Fertilization in *Arabidopsis thaliana*. *Curr Biol* 21:497-502.
- Hirai, M., Arai, M., Mori, T., Miyagishima, S.Y., Kawai, S., Kita, K., Kuroiwa, T., Terenius, O., and Matsuoka, H. (2008). Male fertility of malaria parasites is determined by GCS1, a plant-type reproduction factor. *Curr Biol* 18:607-613.

- Ingouff, M., Hamamura, Y., Gourgues, M., Higashiyama, T., and Berger, F. (2007). Distinct dynamics of HISTONE3 variants between the two fertilization products in plants. *Curr Biol* 17:1032-1037.
- Johnson, M.A., von Besser, K., Zhou, Q., Smith, E., Aux, G., Patton, D., Levin, J.Z., and Preuss, D. (2004). *Arabidopsis hapless* mutations define essential gametophytic functions. *Genetics* 168:971-982.
- Kawai-Toyooka, H., Mori, T., Hamaji, T., Suzuki, M., Olson, B.J., Uemura, T., Ueda, T., Nakano, A., Toyoda, A., Fujiyama, A., et al. (2014). Sex-specific posttranslational regulation of the gamete fusogen GCS1 in the isogamous volvocine alga *Gonium pectorale*. *Eukaryot Cell* 13:648-656.
- Lampropoulos, A., Sutikovic, Z., Wenzl, C., Maegele, I., Lohmann, J.U., and Forner, J. (2013). GreenGate---a novel, versatile, and efficient cloning system for plant transgenesis. *PloS one* 8:e83043.
- Leydon, A.R., Weinreb, C., Venable, E., Reinders, A., Ward, J.M., and Johnson, M.A. (2017). The Molecular Dialog between Flowering Plant Reproductive Partners Defined by SNP-Informed RNA-Sequencing. *Plant Cell* 29:984-1006.
- Liu, Y., Tewari, R., Ning, J., Blagborough, A.M., Garbom, S., Pei, J., Grishin, N.V., Steele, R.E., Sinden, R.E., Snell, W.J., et al. (2008). The conserved plant sterility gene HAP2 functions after attachment of fusogenic membranes in *Chlamydomonas* and *Plasmodium* gametes. *Genes Dev* 22:1051-1068.
- McCue, A.D., Cresti, M., Feijo, J.A., and Slotkin, R.K. (2011). Cytoplasmic connection of sperm cells to the pollen vegetative cell nucleus: potential roles of the male germ unit revisited. *J Exp Bot* 62:1621-1631.
- Mori, T., Kuroiwa, H., Higashiyama, T., and Kuroiwa, T. (2006). GENERATIVE CELL SPECIFIC 1 is essential for angiosperm fertilization. *Nat Cell Biol* 8:64-71.
- Okamoto, M., Yamada, L., Fujisaki, Y., Bloomfield, G., Yoshida, K., Kuwayama, H., Sawada, H., Mori, T., and Urushihara, H. (2016). Two HAP2-GCS1 homologs responsible for gamete interactions in the cellular slime mold with multiple mating types: Implication for common mechanisms of sexual reproduction shared by plants and protozoa and for male-female differentiation. *Dev Biol* 415:6-13.
- Pace, C.N., and Scholtz, J.M. (1998). A helix propensity scale based on experimental studies of peptides and proteins. *Biophys J* 75:422-427.
- Palanivelu, R., and Preuss, D. (2006). Distinct short-range ovule signals attract or repel *Arabidopsis thaliana* pollen tubes in vitro. *BMC Plant Biol* 6:7.
- Pinello, J.F., Lai, A.L., Millet, J.K., Cassidy-Hanley, D., Freed, J.H., and Clark, T.G. (2017). Structure-Function Studies Link Class II Viral Fusogens with the Ancestral Gamete Fusion Protein HAP2. *Curr Biol* 27:651-660.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nature methods* 9:676-682.
- Sprunck, S., Rademacher, S., Vogler, F., Gheyselinck, J., Grossniklaus, U., and Dresselhaus, T. (2012). Egg cell-secreted EC1 triggers sperm cell activation during double fertilization. *Science* 338:1093-1097.
- Steele, R.E., and Dana, C.E. (2009). Evolutionary history of the HAP2/GCS1 gene and sexual reproduction in metazoans. *PloS one* 4:e7680.
- von Besser, K., Frank, A.C., Johnson, M.A., and Preuss, D. (2006). *Arabidopsis HAP2(GCS1)* is a sperm-specific gene required for pollen tube guidance and fertilization. *Development (Cambridge, England)* 133:4761-4769.
- Wilson, Z.A., Morroll, S.M., Dawson, J., Swarup, R., and Tighe, P.J. (2001). The *Arabidopsis MALE STERILITY1 (MS1)* gene is a transcriptional regulator of male

- gametogenesis, with homology to the PHD-finger family of transcription factors. *Plant Journal* 28:27-39.
- Wong, J.L., Leydon, A.R., and Johnson, M.A. (2010). HAP2(GCS1)-dependent gamete fusion requires a positively charged carboxy-terminal domain. *PLoS Genet* 6:e1000882.

Chapter 6: Discussion: Arabidopsis HAP2 fuses gametes via an amphipathic helix and a mechanism reminiscent of viral fusion proteins

Arabidopsis HAP2 presents an amphipathic helix at its membrane interaction surface.

Successful reproduction requires two gametes to fuse, and HAP2 was the only candidate gamete fusion protein in a variety of species, including *Arabidopsis thaliana* (Wong and Johnson, 2010), when I started my thesis research. Classification as a fusion protein requires clear evidence that the protein is directly responsible for fusing membranes (Oren-Suissa and Podbilewicz, 2007). Our goal was to determine whether HAP2 was a fusion protein. To this end, we used the structural prediction server Phyre2 (Kelley et al., 2015) and found that Arabidopsis HAP2 was predicted to have structural similarity to class II viral fusion proteins ((Beale, 2014); Figure 1A-B). Recent structural studies have demonstrated that HAP2 is a class II fusion protein ((Fedry et al., 2017; Fedry*, Submitted); Chapter 2). The first reported structure was from the ectodomain of *Chlamydomonas reinhardtii* HAP2 and revealed structural homology to class II fusion proteins (Fedry et al., 2017). However, the *Chlamydomonas* HAP2 structure did have resolution at the membrane interaction region, where the two predicted fusion loops are located (Fedry et al., 2017).

Our structure-function analysis revealed an amphipathic helix at the membrane interaction surface of Arabidopsis HAP2, and a distinctly different, more compact fusion loop region for *Trypanosoma cruzi* HAP2 ((Fedry*, Submitted); Chapter 2). Our bioinformatics analysis revealed a highly variable membrane interaction region across HAP2 species, but there also appeared to be lineage-specific motifs, such as an

amphipathic helix in flowering plants ((Fedry*, Submitted); Chapter 2). We demonstrated that the helical character of the membrane interaction region is important but not essential for Arabidopsis HAP2 function by replacing a central amino acid (aspartic acid 173) with a proline (D173P; (Fedry*, Submitted); Chapter 2). We also demonstrated the requirement for a hydrophobic patch of amino acids presented at the tip of the Arabidopsis HAP2 fusion helix (Figure 1C) that interacts with the target membrane ((Fedry*, Submitted); Chapter 2). Results from our study, combined with structure-function analysis for *Chlamydomonas* HAP2 (Fedry et al., 2017) and *Tetrahymena* HAP2 (Pinello et al., 2017), suggest HAP2 may function in a unilateral fusion mechanism, similar to viral fusion mechanisms, by interacting directly with target membranes.

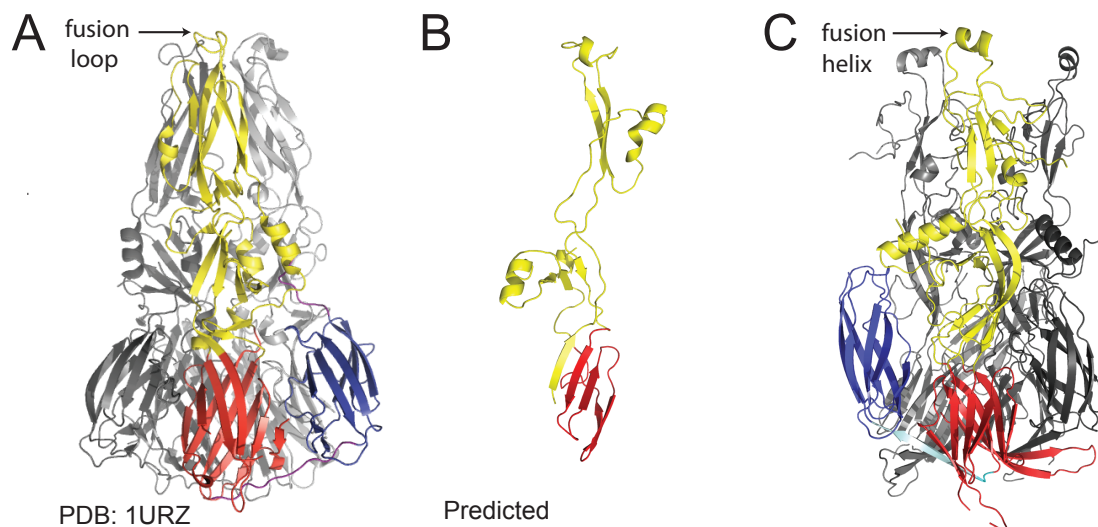


Figure 1. Structures of TBEV E and Arabidopsis HAP2 (predicted and solved). **A.** The TBEV E fusion protein in its post-fusion trimer (PDB: 1URZ; (Bressanelli et al., 2004)) was the top hit from the Phyre² (Kelley et al., 2015) analysis. The fusion loop is indicated at the tip of domain II (yellow). **B.** The Phyre² predicted structure for the portion of Arabidopsis HAP2 that was aligned to the viral fusion protein in the Phyre² analysis. **C.** The trimer structure for Arabidopsis HAP2 is shown ((Fedry*, Submitted); Chapter 2). For panels **A** and **C**, one monomer is shown with the standard domain coloring for class II viral fusion proteins: with domain I in red, domain II in yellow, and domain III in blue (Kielian and Rey, 2006; Rey et al., 1995). The other two monomers are shaded in grey. In panel

B, the portions of the predicted HAP2 structure that align to domain II or I are indicated by their respective colors (yellow or red).

The HAP2 fusion helix is conserved in flowering plants.

My *in vivo* analysis demonstrated that the fusion helix from rice can functionally replace the fusion helix in *Arabidopsis* HAP2 ((Fedry*, Submitted); Chapter 2). This result was particularly interesting considering previous work in our lab that demonstrated the entire N-terminal region (or ectodomain) from rice cannot replace the corresponding portion in *Arabidopsis* HAP2 (Wong et al., 2010). Interestingly, the rice fusion helix contains a tryptophan, and I was interested in whether the tryptophan would be the only amino acid in the fusion helix required for function. For the class II fusion protein from dengue virus, a tryptophan (W101) in its fusion loop is required for function (Huang et al., 2010). A mutant version of the rice fusion helix containing an alanine instead of the tryptophan was partially functional in the context of *Arabidopsis* HAP2 (Os.FH_W>A; Chapter 5), suggesting that the rice fusion helix may also require a hydrophobic patch for interaction with the target membrane, as opposed to relying on a single tryptophan residue. However, this result does not rule out the possibility that the tryptophan in the rice fusion helix is critical in the context of HAP2-mediated gamete fusion during fertilization in rice.

Arabidopsis and rice are members of two distinct groups of flowering plants: dicots and monocots, respectively (Chaw et al., 2004). *Amborella trichopoda* represents a sister group to all flowering plants (Amborella Genome, 2013), and *Amborella* HAP2 has a potential fusion helix with the conserved number of amino acids observed in the membrane interaction motif for flowering plants ((Fedry*, Submitted); Chapter 2). Introducing the *Amborella* fusion helix into *Arabidopsis* HAP2 resulted in decreased fusion levels but demonstrated functional conservation of the fusion helix in flowering plants

(Amb.FH; Chapter 5). The membrane interaction surface of *Chlamydomonas* HAP2 is predicted to have two fusion loops (Fedry et al., 2017), and the flowering plant fusion helix corresponds to loop 2 while the loop 1 region in flowering plants, and *Amborella*, is essentially non-existent ((Fedry*, Submitted); Chapter 2). Interestingly, HAP2 has yet to be found in the genome of any gymnosperm, yet it is present in more basal plants ((Fedry*, Submitted); Chapter 2). Both *Marchantia polymorpha* (liverwort) and *Physcomitrella patens* (moss) have a loop 1 sequence similar in length to *Chlamydomonas* as well as variability in their loop 2 region compared to flowering plants ((Fedry*, Submitted); Chapter 2). It will be interesting to determine whether the loop 2 / fusion helix region from *Marchantia* or *Physcomitrella* is sufficient for fusion in the context of *Arabidopsis* HAP2 or whether they require both loop regions at their membrane interaction surface. It would also be interesting to test the requirement of the two fusion loop regions in the context of their native HAP2 protein and fusion environment.

An interesting question highlighted by the functional conservation of the flowering plant fusion helix is whether there are aspects of the female gamete membrane, such as receptors, attachment proteins, or lipid domains, that are species-specific. We still do not know of any female proteins in flowering plants that are involved in gamete adhesion or may function in a potential gamete fusion complex. It is possible that HAP2 participates in an initial protein complex that spans the two gamete membranes, as proposed for some viral fusion proteins (White et al., 2008). It is also possible that female membrane proteins could contribute to the organization of a domain of the membrane specialized for fusion, as observed for CD9 on mammalian eggs (Jegou et al., 2011).

Sub-cellular localization of HAP2 may be highly regulated.

We were interested in expressing *Arabidopsis* HAP2 in female gametes to investigate several questions. Since the Phyre² results suggested structural similarity to

viral fusion proteins, and HAP2 is only required on the sperm (Johnson et al., 2004; Mori et al., 2006), we wanted to test the hypothesis that HAP2 could fuse cells unilaterally (Figure 2). We wanted to address whether HAP2 is sufficient to fuse cells unilaterally by testing whether HAP2 expression in female gametes is able to rescue the *hap2* gamete fusion defect (Johnson et al., 2004; Mori et al., 2006; von Besser et al., 2006) from the female side, since one requirement of fusion proteins is that they can fuse cells ectopically (Oren-Suissa and Podbilewicz, 2007; Wong and Johnson, 2010). Sufficiency of HAP2 to fuse cells has only recently been demonstrated in a mammalian ectopic expression system, and that study showed a requirement for HAP2 on both fusing membranes (Valansi et al., 2017). To us, this study raised the interesting question of whether HAP2 can act both unilaterally and cooperatively (Figure 2)? Another possible explanation for the bilateral ectopic fusion results is that HAP2 expression at the cell membrane was limited, and HAP2 expression in both cells provided high enough combined HAP2 levels to enable fusion to occur in the assay. In *Tetrahymena thermophila*, which is a ciliate with seven mating types, HAP2 is expressed on all gametes but fusion can still occur when it is only expressed on one of the two interacting gametes (Cole et al., 2014). Therefore, we wanted to also test whether HAP2 expression in the egg cell could increase the level of fusion in HAP2 variants that are partially functional ((Fedry*, Submitted); Chapter 2).

Previous attempts to express HAP2 in different cells within Arabidopsis were unsuccessful, with the exception of HAP2 expression in a compartment within the pollen tube (Frank, 2010). Here, we were able to express HAP2 in the egg cell and the synergid cells (Chapter 3). While some HAP2 may be at the membrane in the egg cell, most of the protein is not (Chapter 3). More analysis with organelles markers would be necessary to define ectopic HAP2 localization in the egg cell and the synergid cells. HAP2 was not expressed at the membrane but appeared to be in a sub-cellular compartment in synergid cells (Chapter 3). Additionally, expression of HAP2 in the egg cell was not able to rescue

fusion defects in HAP2-deficient sperm (Chapter 5). We also expressing HAP2 in the central cell, but we were unable to observe HAP2 fusion proteins accumulating in the central cell (Chapter 5). Plants expressing HAP2 in the egg, and containing a transgene designed to also express HAP2 in the central cell, were not able to rescue the *hap2* defect or increase levels of fusion for HAP2-YFP variants expressed in sperm (Chapter 5). It is important to consider the caveat that HAP2 protein expression in the egg was not only localized to the plasma membrane, and there may not have been sufficient amounts of HAP2 at the membrane. It is also possible that enough HAP2 was present but that it was not activated for fusion. Therefore, it would be interesting to try to force more HAP2 to the membrane in the egg and central cell by using a cell-specific signal sequence or membrane-anchoring sequence.

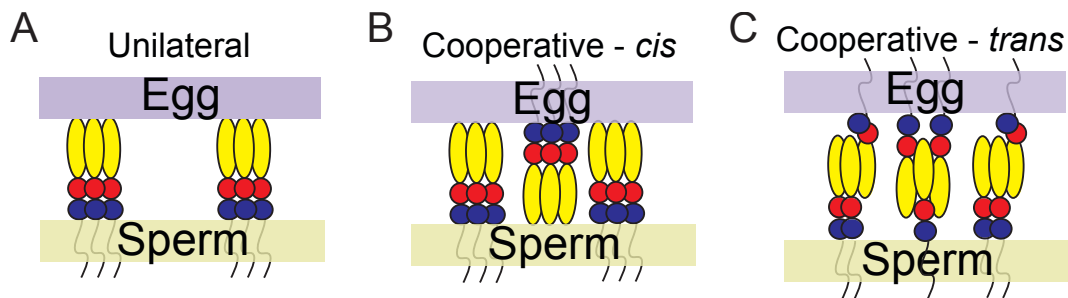


Figure 2. Models for potential HAP2 fusion mechanisms. A-C. The sperm membrane indicates the endogenous HAP2-expressing membrane. HAP2 expression in the egg membrane in panels **B** and **C** represents ectopic expression of HAP2, similar to our experiments in Chapter 3. **A.** Our current proposed hypothesis that HAP2 is acting as a unilateral fusion protein and interacts directly with target membranes. This hypothesis was also suggested previously based on the recent *Chlamydomonas* HAP2 structure (Fedry et al., 2017). **B.** A schematic of a proposed mechanism for HAP2 mediating fusion from both membranes and acting in *cis* to form trimers that insert into the membrane. **C.** A Schematic for HAP2 interacting with itself in *trans* to mediate membrane fusion. The way HAP2 interacts with itself in *trans* is organized in different ways in panel **C**, because we do not have a good understanding of how a class II fusion protein may interact with itself in *trans* to mediate fusion. However, trimerization of HAP2 in *trans* has been proposed for EFF-1 (Perez-Vargas et al., 2014).

Mutations in the fusion helix do not affect HAP2-YFP localization in sperm within pollen tubes.

The only known regulator of HAP2 function are EC1 proteins, which are secreted by the egg cell after pollen tube burst ((Sprunck et al., 2012); see Chapter 1: Figure 2). Synthetic EC1 proteins have been shown to induce movement of HAP2 to the sperm membrane in vitro, and this regulation of HAP2 movement has been proposed as a role of EC1 proteins in activating sperm cells ((Sprunck et al., 2012); Figure 3B). However, our confocal microscopy of sperm co-expressing HAP2-YFP variants and a sperm membrane marker in pollen tubes, suggests much of the protein is localized to the plasma membrane, unless HAP2 is expressed at high levels or has a mutation expected to disrupt two different disulfide bonds (C312A/C319A; Chapter 5). It may be that HAP2 localization is affected by expression level in transgenic plants or by release of the sperm from the pollen tube. We do not currently know how much HAP2 is needed at the membrane to mediate fusion, and therefore, it will be important to test whether enough HAP2 is present in the sperm membrane prior to pollen tube burst. However, we will first need to determine the amount of HAP2 necessary for fusion to occur, and this experiment may require an in vitro fusion assay in which the amount of HAP2 expression and membrane localization could be controlled or adjusted. If enough HAP2 is present at the sperm membrane prior to burst, it would suggest an additional regulatory mechanism to control HAP2-mediated membrane fusion, such as a trigger for HAP2 conformation change (Figure 3B). If enough HAP2 is present, it would be interesting to investigate other effects EC1 activation may have on the sperm and/or HAP2. An ER-localized fluorescent protein in sperm does not produce the same 'ring'-like pattern observed for HAP2-YFP not at the membrane (Chapter 5), which suggests further analysis of HAP2 localization is needed.

The environment at the site of gamete fusion contains factors that regulate HAP2 function.

The sperm are connected to each other as they travel through the pollen tube cytoplasm (McCue et al., 2011), raising the question of why they do not fuse with each other, especially with the observed enrichment of HAP2 at the membranous region between the two sperm (Chapter 5). One explanation is that the sperm, or potentially HAP2, is not activated by EC1 yet (Rademacher and Sprunck, 2013; Sprunck et al., 2012). Another possibility, which does not preclude sperm activation, is that HAP2 responds to a trigger after it is released to the site of fusion (Figure 3B). Class II viral fusion proteins require a pH trigger that induces a conformation change prior to insertion into the target membrane (Acuna et al., 2015; DuBois et al., 2013; Fritz et al., 2008; Gibbons et al., 2003; Stiasny et al., 2002; White et al., 2008). However, other viral fusion proteins have other triggering mechanisms (White et al., 2008). The environment surrounding the sperm changes as the sperm are released from the burst pollen tube, where they are not only surrounded by the contents of the burst pollen tube but also of the degenerated synergid cell ((Higashiyama et al., 1997); Figure 3A). The sperm also remain at the site of fusion for ~7-8 minutes before fusion occurs (Hamamura et al., 2011), during which time the sperm could become activated by EC1 and HAP2 could experience a trigger (Figure 3B).

We investigated whether the pH of the environment surrounding the sperm cells changes after pollen tube burst using the ratio-metric pH sensor pHluorin (Chapter 4). pHluorin has been used previously to study the pH of growing tobacco pollen tubes, and the pollen tube was characterized as having a different pH (more acidic) at its tip and as it grows (Michard, 2008). We were more interested in characterizing the different environments surrounding the sperm cells, so we analyzed the average pH of both the Arabidopsis pollen tube and the area of the pollen tube burst. We found that the pH of the pollen tube burst area was less than that of the pollen tube growing semi-in-vivo (Chapter

4). However, the pH of the pollen tube burst appeared to be similar to that of pollen tubes that continue to grow and coil inside mutant ovules without bursting, although our analysis of coiling pollen tubes used a low sample size and a female mutant to induce pollen tube coiling (Chapter 4). It is possible that the pollen tube pH is influenced by both the transmitting tract and the ovule. However, it is also possible that pollen tubes growing within a mutant ovule differ from those growing in wild-type ovules, since we observed no noticeable difference for the pH of portions of pollen tubes either just outside the ovule or just inside the ovule in semi-in-vivo conditions (Chapter 4).

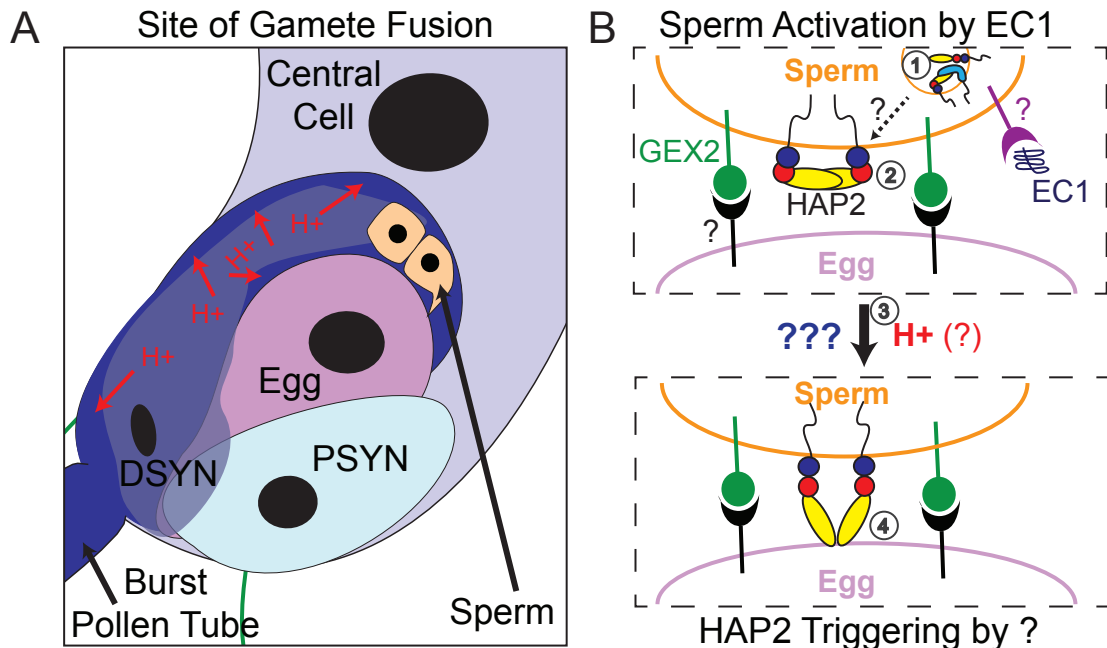


Figure 3. Models for potential regulation of the HAP2-mediated gamete fusion mechanism in Arabidopsis. **A.** This schematic highlights the location of the sperm (light orange) at the site of gamete fusion in an Arabidopsis ovule after pollen tube burst and sperm release. The egg cell is shown in pink and the central cell is shown in light purple. The remaining intact synergid cell (PSYN) is shown in light blue while the degenerated synergid cell (DSYN) is shown in a translucent purple-gray and overlaps with the burst pollen tube (dark blue). The red arrows with hydrogen ions represent the potential for the degenerating synergid cell to contribute to the lower pH at the site of gamete fusion compared to pH of the intact pollen tube, which is a hypothesis generated from preliminary data with a pH sensor expressed in synergid cells (Chapter 4). **B.** This panel highlights the proteins known to be involved in regulating sperm-egg attachment and fusion in Arabidopsis. The top panel depicts the egg and sperm cell after both pollen tube burst and the subsequent secretion of EC1 proteins (shown in dark purple bound to an unknown receptor in purple) by the egg cell (Sprunck et al., 2012). Interaction of the sperm-specific

gamete attachment protein GEX2 (green; (Mori et al., 2014)) with an unknown receptor (black) on the egg cell is also shown. HAP2 is shown both in a vesicle within the sperm cell and at the surface of the sperm cell. HAP2 is shown both unbound and potentially bound to an unknown regulatory protein within vesicles, because the regulation of HAP2 within vesicles is unknown, including whether it is bound to other proteins or has its membrane interaction surface hidden. The gray numbers within gray circles highlight important areas of HAP2 regulation that future studies should focus on. HAP2 is shown as a homo-dimer on the surface of the sperm cell, but the oligomerization state of HAP2 on the sperm surface is unknown (see Figure 4). The bottom panel is a model for HAP2 insertion of its fusion helix into the egg membrane after a potential conformational change in HAP2 while at the sperm membrane, which may result from a potential HAP2 trigger.

Live-imaging analysis will be important to determine the timing of pH changes as the pollen tube enters the ovule and bursts. The difference between the pollen tube pH as it travels through the pistil would also be interesting to assess, although difficult since the transmitting tract tissue appears auto-fluorescent (observations, data not shown). Since the site of gamete fusion is expected to be less than the entire area of pollen tube burst (Dresselhaus et al., 2016), we did a focused analysis of the 'hook' region of pollen tube burst, which is observed between the egg and central cell. We found a significant decrease in pH at the site of gamete fusion compared to the entire burst area (Chapter 4). Our data suggests the sperm may experience a decrease in pH at the site of fusion. We have also generated transgenic plants to evaluate the pH of the synergid cell, using two different pH sensors. Preliminary experiments using synergid cells expressing pHluorin suggest that the synergid cell pH may be lower than that of the pollen tube, and therefore, the degenerated synergid cell may contribute to the lower pH detected after pollen tube burst in the ovule (Chapter 4; Figure 3A). Our transgenic plants expressing different pH sensors in the synergid cells offer the possibility of imaging pollen tube pH and synergid pH simultaneously (Chapter 4).

Perspectives: What is the pre-fusion form of HAP2?

Both our recent study on *Arabidopsis* and *Trypanosoma* HAP2 (Fedry*, Submitted) and the previous study on *Chlamydomonas* HAP2 (Fedry et al., 2017) have identified a trimer structure of the HAP2 ectodomain that is structurally homologous to class II fusion proteins. However, the pre-fusion form of HAP2 remains unknown (schematics of potential pre-fusion forms in Figure 4). In class II viral fusion proteins, the pre-fusion form is often homodimeric (Dessau and Modis, 2013; Modis et al., 2003; Rey et al., 1995) or in a dimeric complex (Lescar et al., 2001) with a regulatory protein. A monomeric pre-fusion form has been suggested for EFF-1 (Zeev-Ben-Mordehai et al., 2014). Therefore, future experiments should include attempts to identify the pre-fusion form of HAP2, both through structural analysis but also through biochemical analysis of HAP2 interactions at the sperm membrane. It is possible that *Arabidopsis* HAP2 will not be in its pre-fusion form at the membrane until after sperm activation, which would require biochemical analysis to occur after sperm activation.

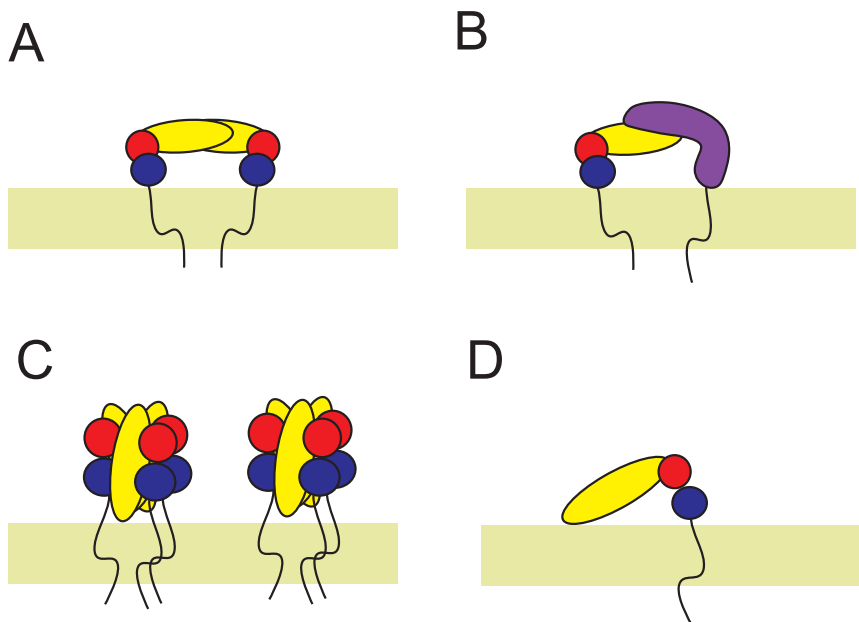


Figure 4. Schematics representing potential pre-fusion models for HAP2. A. Model for a potential pre-fusion homodimer based off class II viral fusion proteins (Dessau and Modis, 2013; Harrison, 2015; Kielian and Rey, 2006; Modis et al., 2003; Rey et al., 1995).

B. A model for a potential heterodimeric interaction. **C.** A potential trimeric pre-fusion structure, based on pre-fusion structures of class I viral fusion proteins (Apellaniz et al., 2014; Harrison, 2015; Lamb and Jardetzky, 2007; White et al., 2008; Wilson et al., 1981). **D.** A potential monomeric arrangement at the membrane based on the proposed monomeric pre-fusion form of EFF-1 (Zeev-Ben-Mordehai et al., 2014).

How is HAP2 regulated?

An exciting area of future research will be investigating how HAP2 is regulated: both which proteins and molecules regulate HAP2 function and when they regulate it (Figure 3B). An important regulatory step in the production of mature viruses that use class II viral fusion proteins is the prevention of exposure of class II viral fusion loops as the viruses are transported through the host cell before release outside the cell to continue the infection (Sanchez-San Martin et al., 2009). It is likely that HAP2 will require regulation, for example a partner protein, to prevent exposure of its fusion helix (or loops) as it is transported within the cell to the plasma membrane (regulatory step of interest number 1, Figure 3B). As discussed above, the organization of HAP2 on the sperm membrane pre-fusion is unknown (see Figure 4), and it may represent an important step in regulating HAP2 function, potentially through a partner protein or through interacting with itself to keep the fusion helix buried until its exposure is required for gamete fusion (regulatory step of interest number 2, Figure 3B). Viral fusion proteins experience a trigger that induces a conformational change that enables the fusion protein to interact directly with the host target membrane (White et al., 2008); however, a trigger for HAP2 has yet to be identified in any species (potential regulatory step 3, Figure 3B). Other proteins on either the sperm membrane or the egg and central cell membranes may contribute to the gamete fusion process by forming a membrane fusion complex or organizing the membrane environment to enable fusion to occur, and it would be essential to identify these proteins to fully elucidate the HAP2-mediated gamete fusion mechanism (potential regulatory step 4, Figure 3B).

Concluding Remarks

Our structure prediction analysis (Figure 1B) and the identification of HAP2 as a class II fusion protein (Fedry et al., 2017) provided the opportunity to investigate structure-based hypotheses. Our structure-function analyses ((Fedry*, Submitted); Chapter 2) revealed variability in the membrane interacting region of HAP2, which differs from the more conserved fusion loop of class II viral fusion proteins. Arabidopsis HAP2 uses a novel amphipathic helix to mediate membrane fusion, and the helix is functionally conserved in flowering plants. We still do not know the pre-fusion form of HAP2 or how it is regulated within the Arabidopsis sperm cell. Future studies should address questions of regulation of HAP2 before membrane localization, the arrangement of HAP2 at the membrane, potential HAP2 triggers, including pH, and the target membrane environment, such as lipid composition.

References

- Acuna, R., Bignon, E.A., Mancini, R., Lozach, P.Y., and Tischler, N.D. (2015). Acidification triggers Andes hantavirus membrane fusion and rearrangement of Gc into a stable post-fusion homotrimer. *J Gen Virol* 96:3192-3197.
- Amborella Genome, P. (2013). The Amborella genome and the evolution of flowering plants. *Science* 342:1241089.
- Apellaniz, B., Huarte, N., Largo, E., and Nieva, J.L. (2014). The three lives of viral fusion peptides. *Chem Phys Lipids* 181:40-55.
- Beale, K.M. (2014). Coupling Gamete Fusion with Sperm Cell Delivery Maximizes Reproductive Success in Flowering Plants. In: *Molecular Biology, Cell Biology, and Biochemistry Theses and Dissertations*. Brown Digital Repository: Brown University.
- Bressanelli, S., Stiasny, K., Allison, S.L., Stura, E.A., Duquerroy, S., Lescar, J., Heinz, F.X., and Rey, F.A. (2004). Structure of a flavivirus envelope glycoprotein in its low-pH-induced membrane fusion conformation. *The EMBO journal* 23:728-738.
- Chaw, S.M., Chang, C.C., Chen, H.L., and Li, W.H. (2004). Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. *J Mol Evol* 58:424-441.
- Cole, E.S., Cassidy-Hanley, D., Fricke Pinello, J., Zeng, H., Hsueh, M., Kolbin, D., Ozzello, C., Giddings, T., Jr., Winey, M., and Clark, T.G. (2014). Function of the male-gamete-specific fusion protein HAP2 in a seven-sexed ciliate. *Curr Biol* 24:2168-2173.
- Dessau, M., and Modis, Y. (2013). Crystal structure of glycoprotein C from Rift Valley fever virus. *Proc Natl Acad Sci U S A* 110:1696-1701.
- Dresselhaus, T., Sprunck, S., and Wessel, G.M. (2016). Fertilization Mechanisms in Flowering Plants. *Curr Biol* 26:R125-139.
- DuBois, R.M., Vaney, M.C., Tortorici, M.A., Kurdi, R.A., Barba-Spaeth, G., Krey, T., and Rey, F.A. (2013). Functional and evolutionary insight from the crystal structure of rubella virus protein E1. *Nature* 493:552-556.
- Fedry, J., Liu, Y., Pehau-Arnaudet, G., Pei, J., Li, W., Tortorici, M.A., Traincard, F., Meola, A., Bricogne, G., Grishin, N.V., et al. (2017). The Ancient Gamete Fusogen HAP2 Is a Eukaryotic Class II Fusion Protein. *Cell* 168:904-915 e910.
- Fedry*, J., Forcina*, J., Legrand, P., Pehau-Arnaudet, G., Haouz, A., Johnson, M.A., Rey, F.A., and Krey, T. (Submitted). HAP2 drives gamete fusion via insertion of highly divergent fusion loops into female gametes.
- Frank, A.C. (2010). HAP2(GCS1) is a conserved sperm-expressed gene required for fertilization. In: *Molecular Biology, Cell Biology, and Biochemistry Theses and Dissertations*. Brown Digital Repository Brown University.
- Fritz, R., Stiasny, K., and Heinz, F.X. (2008). Identification of specific histidines as pH sensors in flavivirus membrane fusion. *J Cell Biol* 183:353-361.
- Gibbons, D.L., Erk, I., Reilly, B., Navaza, J., Kielian, M., Rey, F.A., and Lepault, J. (2003). Visualization of the target-membrane-inserted fusion protein of Semliki Forest virus by combined electron microscopy and crystallography. *Cell* 114:573-583.
- Hamamura, Y., Saito, C., Awai, C., Kurihara, D., Miyawaki, A., Nakagawa, T., Kanaoka, M.M., Sasaki, N., Nakano, A., Berger, F., et al. (2011). Live-Cell Imaging Reveals the Dynamics of Two Sperm Cells during Double Fertilization in *Arabidopsis thaliana*. *Curr Biol* 21:497-502.
- Harrison, S.C. (2015). Viral membrane fusion. *Virology* 479-480:498-507.

- Higashiyama, T., Kuroiwa, H., Kawano, S., and Kuroiwa, T. (1997). Kinetics of double fertilization in *Tourenia fournieri* based on direct observations of the naked embryo sac. *Planta* 203:101-110.
- Huang, C.Y., Butrapet, S., Moss, K.J., Childers, T., Erb, S.M., Calvert, A.E., Silengo, S.J., Kinney, R.M., Blair, C.D., and Roehrig, J.T. (2010). The dengue virus type 2 envelope protein fusion peptide is essential for membrane fusion. *Virology* 396:305-315.
- Jegou, A., Ziyat, A., Barraud-Lange, V., Perez, E., Wolf, J.P., Pincet, F., and Gourier, C. (2011). CD9 tetraspanin generates fusion competent sites on the egg membrane for mammalian fertilization. *Proc Natl Acad Sci U S A* 108:10946-10951.
- Johnson, M.A., von Besser, K., Zhou, Q., Smith, E., Aux, G., Patton, D., Levin, J.Z., and Preuss, D. (2004). *Arabidopsis* *hapless* mutations define essential gametophytic functions. *Genetics* 168:971-982.
- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10:845-858.
- Kielian, M., and Rey, F.A. (2006). Virus membrane-fusion proteins: more than one way to make a hairpin. *Nat Rev Microbiol* 4:67-76.
- Lamb, R.A., and Jardetzky, T.S. (2007). Structural basis of viral invasion: lessons from paramyxovirus F. *Curr Opin Struct Biol* 17:427-436.
- Lescar, J., Roussel, A., Wien, M.W., Navaza, J., Fuller, S.D., Wengler, G., Wengler, G., and Rey, F.A. (2001). The Fusion glycoprotein shell of Semliki Forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH. *Cell* 105:137-148.
- McCue, A.D., Cresti, M., Feijo, J.A., and Slotkin, R.K. (2011). Cytoplasmic connection of sperm cells to the pollen vegetative cell nucleus: potential roles of the male germ unit revisited. *J Exp Bot* 62:1621-1631.
- Michard, E., Dias, P., Feijo, J.A. (2008). Tobacco pollen tubes as cellular models for ion dynamics: improved spatial and temporal resolution of extracellular flux and free cytosolic concentration of calcium and protons using pHluorin and YC3.1 CaMeleon. *Sexual Plant Reproduction* 21:169-181.
- Modis, Y., Ogata, S., Clements, D., and Harrison, S.C. (2003). A ligand-binding pocket in the dengue virus envelope glycoprotein. *Proc Natl Acad Sci U S A* 100:6986-6991.
- Mori, T., Igawa, T., Tamiya, G., Miyagishima, S.Y., and Berger, F. (2014). Gamete attachment requires GEX2 for successful fertilization in *Arabidopsis*. *Curr Biol* 24:170-175.
- Mori, T., Kuroiwa, H., Higashiyama, T., and Kuroiwa, T. (2006). GENERATIVE CELL SPECIFIC 1 is essential for angiosperm fertilization. *Nat Cell Biol* 8:64-71.
- Oren-Suissa, M., and Podbilewicz, B. (2007). Cell fusion during development. *Trends Cell Biol* 17:537-546.
- Perez-Vargas, J., Krey, T., Valansi, C., Avinoam, O., Haouz, A., Jamin, M., Raveh-Barak, H., Podbilewicz, B., and Rey, F.A. (2014). Structural basis of eukaryotic cell-cell fusion. *Cell* 157:407-419.
- Pinello, J.F., Lai, A.L., Millet, J.K., Cassidy-Hanley, D., Freed, J.H., and Clark, T.G. (2017). Structure-Function Studies Link Class II Viral Fusogens with the Ancestral Gamete Fusion Protein HAP2. *Curr Biol* 27:651-660.
- Rademacher, S., and Sprunck, S. (2013). Downregulation of egg cell-secreted EC1 is accompanied with delayed gamete fusion and polytubey. *Plant signaling & behavior* 8:e27377.

- Rey, F.A., Heinz, F.X., Mandl, C., Kunz, C., and Harrison, S.C. (1995). The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution. *Nature* 375:291-298.
- Sanchez-San Martin, C., Liu, C.Y., and Kielian, M. (2009). Dealing with low pH: entry and exit of alphaviruses and flaviviruses. *Trends Microbiol* 17:514-521.
- Sprunck, S., Rademacher, S., Vogler, F., Gheyselinck, J., Grossniklaus, U., and Dresselhaus, T. (2012). Egg cell-secreted EC1 triggers sperm cell activation during double fertilization. *Science* 338:1093-1097.
- Stiasny, K., Allison, S.L., Schalich, J., and Heinz, F.X. (2002). Membrane interactions of the tick-borne encephalitis virus fusion protein E at low pH. *Journal of virology* 76:3784-3790.
- Valansi, C., Moi, D., Leikina, E., Matveev, E., Grana, M., Chernomordik, L.V., Romero, H., Aguilar, P.S., and Podbilewicz, B. (2017). Arabidopsis HAP2/GCS1 is a gamete fusion protein homologous to somatic and viral fusogens. *J Cell Biol* 216:571-581.
- von Besser, K., Frank, A.C., Johnson, M.A., and Preuss, D. (2006). Arabidopsis *HAP2(GCS1)* is a sperm-specific gene required for pollen tube guidance and fertilization. *Development (Cambridge, England)* 133:4761-4769.
- White, J.M., Delos, S.E., Brecher, M., and Schornberg, K. (2008). Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme. *Crit Rev Biochem Mol Biol* 43:189-219.
- Wilson, I.A., Skehel, J.J., and Wiley, D.C. (1981). Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. *Nature* 289:366-373.
- Wong, J.L., and Johnson, M.A. (2010). Is HAP2-GCS1 an ancestral gamete fusogen? *Trends Cell Biol* 20:134-141.
- Wong, J.L., Leydon, A.R., and Johnson, M.A. (2010). HAP2(GCS1)-dependent gamete fusion requires a positively charged carboxy-terminal domain. *PLoS Genet* 6:e1000882.
- Zeev-Ben-Mordehai, T., Vasishtan, D., Siebert, C.A., and Grunewald, K. (2014). The full-length cell-cell fusogen EFF-1 is monomeric and upright on the membrane. *Nat Commun* 5:3912.