Structural analysis of APOLLO, a candidate apomixis factor

A thesis submitted in fulfillment
of the requirements for the degree of
Master of Science in the Department of Plant Science
To the College of Graduate and Postdoctoral Studies
University of Saskatchewan
Saskatoon, Saskatchewan, Canada

Submitted by
Anne Nerbas
Bachelor of Science in Agriculture

April 2021

Permission of use

In the completion of this thesis in fulfillment of the requirements for a Master of Science in Plant Science, Postgraduate degree from the University of Saskatchewan, I permit that the libraries of this university may make the thesis available for inspection. I also permit the copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor who supervised my work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was performed. This thesis is protected by copyright, and any copying or publication or use of it or parts of it for financial gain shall not be allowed without my written permission. Recognition shall be given to me and the University of Saskatchewan in any scholarly use of the material of my thesis.

Requests for permission to copy or to make other uses of materials in this thesis/dissertation in whole or part should be addressed to:

Head of the Department of Plant Sciences
51 Campus Drive
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5A8 Canada

OR

Debby Burshtyn
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9 Canada

Acknowledgements

First, I would to express my heartfelt appreciation to Joanne Ernest, my co-supervisor, whose encouragement, support, advice, and counsel over the past four years has been relentless. Joanne's spirit for science and life has been a great source of inspiration. I am also grateful to Tim Sharbel, my supervisor, whose encouragement, counsel, expertise and enthusiasm has been invaluable to me. Tim's passion for knowledge and dedication to this research throughout the setbacks astounds me. They have left a marvelous mark on my life that I plan to carry with me into my career.

I would like to thank all the members of committee for their constructive advice and patience; Gordon Gray, Chris Ambrose and Steve Shirtliffe. I would like to acknowledge the Global Institute for Food Security for their support and access to laboratory space. I would also like to spread thanks to the entire Seed and Biology Development team. As well as the Global Institute for Food Security management staff for keeping the administrative foundations in check. Thank you to Michal Boniecki from the Protein Characterization and Crystallization Facility (College of Medicine, University of Saskatchewan) for providing the cells (BL21 and C41), vector and primers that got this project off the ground. Thank you to Jieyu Chen from the Root Soil Microbial Interaction group at the Global Institute for Food Security for providing me with GFP-4M8H that I used as a positive control to determine the efficacy of my protein production.

I am immensely grateful for the financial assistance I received to bring this project to completion. Thank you to the Martin Pedersen and Family for the generous scholarship that supported me through the early days of the project. Thank you to the S.N. Horner Scholarship in Agriculture for the considerate financial support through the final year. This work was supported by a grant given to Professor Tim Sharbel, Department of Plant Sciences, at the University of Saskatchewan.

I am incredibly thankful to my wonderful friends and my dearest sister for always being there to listen to my grievances and provide encouragement. To my loving husband, I am extremely grateful for you being there for every step forward and the many steps back. To my parents, for always being just a phone call away.

Finally, I would like to dedicate this thesis to my father, Tim Nerbas and my great-grandfather, Jack Bullock. Their hard work and sacrifices gave me the privilege and inspiration to pursue the knowledge that I seek.

Table of Contents

Abstract

Apomixis is a naturally occurring phenomenon in which plants reproduce without fertilization, resulting in the asexual production of seed. This phenomenon has been well-documented in *Boechera,* a genus in the Brassicacea family that includes members that reproduce through diplospory (gametophytic apomixis). The occurrence of apomixis in *Boechera* is associated with expression of the *APOLLO* gene. Apomictically-reproducing *Boechera* species carry one copy of an apomictic APOLLO allele and one copy of a sexual APOLLO allele. There are twelve apomictic-specific non-synonymous polymorphisms in the APOLLO coding sequence, compared to the sex-allele. The structure and function of the APOLLO exonuclease in *Boechera* may be affected by these polymorphisms. *APOLLO* has predicted regions of structural disorder and potential sites of post-translational modifications. An expression system was created to obtain the sex- and apo-alleles in *E. coli* cells. Whole cell lysates were examined by SDS-PAGE using stain-free technology. The proteins present on the PAGE were transferred to a PVDF membrane using a wet membrane transfer. The histidine tag present on the apo- and sex- proteins were visualized by a western blot. The results indicate that a functional expression vector system, pMCSG7 in BL21 (DE3) *E. coli* cells grown in LB media was identified for the apo-allele and pMCSG7 in C41 *E. coli* cells grown in auto-induction media was identified for the sex-allele. The predicted intrinsic disorder profile established that near the middle of the APOLLO protein there is a region of disorder, that appears to be smaller in the apo-allele than the sex-allele. This region is potentially a binding site for interactions with other cellular components. The difference in size of the binding site between the alleles is the first evidence of a proteomic difference between the sex- and apo alleles of APOLLO.

List of Figures

List of Tables

List of Abbreviations

2-ME	2-Mercaptoethanol
4CN	4-chloro-1-naphthol
α-His	anti N-terminal
α-R	anti rabbit
aa	amino acid
APOLLO	apomixis-linked locus
apo-allele	APOLLO apomictic allele
APS	ammonium persulfate
BBM	Baby Boom
BLAST	Basic Local Alignment Search Tool
bp	base pair
BSA	bovine serum albumin
cDNA	complementary deoxyribonucleic acid
CIAT	International Center for Tropical Agriculture
CNV	copy number variation
°C	degrees Celsius
dCTP	deoxycytidine triphosphate
dH_2O	distilled water
ddH_2O	double-distilled water
DEDDh	DnaQ-like exonuclease superfamily
dGTP	deoxyguanosine triphosphate
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DTT	dithiothreitol
EXOlll	exonuclease III
>	greater than
g	grams
GMO	genetically modified organism
His tag	histidine tag
HRP	horseradish peroxidase
hrs	hours
ID	intrinsic disorder
IDP	intrinsic disorder profile
IMAC	immobilized metal affinity chromatography
IPTG	isopropyl β-D-1-thiogalactopyranoside
kb	kilo base
kDa	kilodalton
kV	kilo volts
LB	Lysogeny broth, also known as Luria-bertani broth
LC	liquid chromatography
LF	low-fluorescence
LIC	ligation independent cloning
µl	micro litre

M	molar
MiMe	mitosis instead of meiosis
mins	minutes
mL	milliliter
mm dia	millimeter diameter
mRNA	messenger ribonucleic acid
#	number
N-terminal	amino-terminal
nmole	nano moles
NMR	nuclear magnetic resonance
NRC	National Research Center
OD	optical density
ORF	open reading frame
%	percent
PAGE	polyacrylamide gel by electrophoresis
PCR	polymerase chain reaction
PMD	plant mobile domains
pPICZ	*Pichia pastoris* expression vector
PSI-PRED (UCL)	Position Specific Iterated - Predict Secondary Structure (London's Global University)
PTM	post translational modification
PVDF	polyvinylidene difluoride
RNA	ribonucleic acid
rpm	revolutions per minute
SDS	sodium dodecyl sulphate
sex-allele	APOLLO sexual allele
secs	seconds
SND1	(Small RNA Degrading Nucleases)1 exonuclease
SOC	Super Optimal broth with Catabolite repression
ssMP	site-specific modification profile
SKP1	S-Phase kinase associated protein 1
SWITCH1	novel protein
TBFS	transcription factor-binding sites
TBST	Tris-Buffered Saline-Tween-20
TEMED	tetramethylethylenediamine
TEV	tobacco etch virus
TREX1	Gene encoding DNA exonuclease
Tris-HCL	trisaminomethane-hydrochloride
UPGRADE2	unreduced pollen grain development
UTR	untranslated region
V	volts
X	value

1. Introduction

Plants reproduce either sexually or asexually to produce offspring. Sexual reproduction is the more common method of reproduction among angiosperms, but asexual reproduction does occur in many families. Asexual reproduction is the generation of offspring from a single individual; the offspring is therefore a genetic clone. Vegetative reproduction, occurring in plants such as potatoes and apples, is a common type of asexual reproduction whereby a new plant grows from a tissue fragment of the parent plant (Spillane et al. 2004). Apomixis is another form of asexual reproduction which occurs through seeds whereby the offspring are genetically identical to the mother plant (Barcaccia and Albertini 2013). The plant science community first became aware of this reproductive mechanism during Gregor Mendel's inheritance experiments with *Hieracium* (Corral et al. 2013). The data he published from those genetic segregation experiments with *Hieracium*, although incorrect has become useful information in establishing and verifying apomixis as a method of reproduction (Bicknell and Koltunow 2004).

Apomixis is asexual seed formation by either gametophytic or sporophytic reproduction; in which an unreduced gametophyte egg forms and develops into an embryo without fertilization (Hand and Koltunow 2014). Apomixis is taxonomically widespread in angiosperms, and is associated with species that have genera-rich families (Hojsgaard et al. 2014b). This accounts for about 1.1% of all angiosperms, and hence the occurrence of apomixis is very rare (Mau et al. 2015). These numbers include all types of apomixis that occur, including adventitious embryony, apospory and diplospory (see Figure 1 in Conner and Ozias-Akins 2017). Apomixis is a very complex mechanism that is likely determined by different genetic factors in multiple species (Barcaccia and Albertini 2013). Apomixis has great potential to relieve the demands on crop breeding and food security, as some literature estimates that apomixis technology may have a greater impact on crop yields than the Green Revolution (Barcaccia and Albertini 2013), as genetically fixing F1 hybrid crops would significantly reduce the cost and time of breeding.

Apomixis is found to occur in nature repeatedly, as asexuality is an easier method to reproduce an offspring. The factors that influence the mechanism of apomixis include hybridization, polyploidy, genomic imprinting, cytogenetic alterations and geographical parthenogenesis (van Dijk 2003; Pupilli and Barcaccia 2012; Lovell et al. 2013; Mau et al. 2015). Despite the favorable option of asexuality, apomixis is not found in common crop species. Apomictic species are generally found

in the Asteraceae, Poaecea, Roseaceae and Orchidaceae families (Hojsgaard et al. 2014b). Apomixis is being investigated throughout many model plant species, and in particular within the genus *Boechera* (family Brassicaceae) which includes sexual and apomictic species that are closely related to both *Arabidopsis thaliana* (Rushworth et al. 2011) and canola. Canola is sold to producers as a F1 hybrid crop, and by using apomixis technology the cost and time of breeding could be reduced. This is an important course of action, as of 2016 the canola sector is worth C$26.7 billion of the Canadian economy (LMC International 2016).

APOLLO (APOmeiosis-Linked LOcus) is a candidate gene linked to the switch from meiosis to apomeiosis, one of the three main differences between sexual and apomictic seed formation (see below), in *Boechera* species (Corral et al. 2013). *APOLLO* exists as two different alleles; an apomictic-allele (apo-allele) that has 12 conserved amino acid changes in the protein primary sequence (Corral et al. 2013) relative to the sexual allele. Sexual *Boechera* are homozygous for the sex-allele, while all the apomictic *Boechera* that have been investigated to date are heterozygous for the apo-allele (Corral et al. 2013). The apo- and sex- APOLLO alleles contain essential functional differences that can be observed in their expression in a prokaryotic system. The *APOLLO* gene is approximately 2,538 nucleotides long which are divided into 5 exons, and contains an open reading frame that encodes for a 496 amino-acid protein. *APOLLO* is a putative DEDDh 3′ -> 5′ exonuclease (Corral et al. 2013) that belongs to a superfamily of exonucleases. The DEDDh superfamily have 3′ -> 5′ exonuclease activity that allows them to remove single nucleotides from the 3′ end of a RNA or DNA strand (Huang et al. 2016). Although little is known about the tertiary structure of *APOLLO*, the most similar DEDDh exonuclease with structure determined is TREX1 (an ortholog of APOLLO) through x-ray crystallography (Brucet et al. 2008).

Apomixis technology may be an essential resource to revolutionizing global food security. By utilizing the relationship between *Boechera* and Brassicaceae family, there is a chance for the development of apomixis technology in canola. However, there is still a great deal to determine about the functionality of *APOLLO* and the application of apomixis to our cropping systems. As a gene candidate associated with maternal apomeiosis, understanding the structure and function of *APOLLO* is an essential component of the potential deployment of apomixis.

From the preliminary structure that is determined, function can be predicted and differences identified.

The objectives of this project are to:

1) Clone both the sex- and apo-APOLLO alleles into a functional expression vector;
2) Identify an expression system for high level expression and purification of soluble APOLLO protein;
3) Determine potential areas of the expression system that indicate a functional difference between the alleles.

A series of hypothesis were used to determine if the data produced is a pre-translational or post-translational problem. The hypothesis tests are as follows:

Null hypothesis 1: the APOLLO alleles can be expressed and purified in a bacterial system. This hypothesis can be tested by cloning the alleles into a suite of bacterial systems and then measuring expression levels by SDS-PAGE and western blot. In this experiment, we were able to express one allele in the bacterial system but not the other, and could not generate soluble protein. Since it was impossible to test all bacterial systems under every condition combination, the null hypothesis cannot be rejected.

Null hypothesis 2: the apo- and sex- allele specific nucleotide variation in the coding regions affects the structure and/or function of APOLLO. This hypothesis can be tested by purifying the APOLLO alleles and crystallizing their protein for comparison of structure and function. In this experiment, we were unable to purify soluble protein of either allele, and could not conduct these experiments. Therefore, the null hypothesis cannot be rejected.

2. Literature Review

2.1 Apomixis

Apomixis is asexual reproduction via seed production. It bypasses many of the regulatory and fundamentals of reproduction such as meiosis and fertilization, and the embryo in the resulting seed is a maternal clone. Apomixis is seen as having many applications in plant breeding by being able to maintain or "fix" a desirable genotype, and associated phenotype, no matter how complex the nature of the genotype. The first step towards apomixis technology is to understand its evolution in nature.

2.1.1 Forms of apomixis

Apomixis is defined as either gametophytic or sporophytic (Hand and Koltunow 2014). While sporophytic apomixis (also known as adventitious embryony), follows the normal sexual pathway of angiosperms, until egg cell development where diploid somatic ovule cells on the outside of the embryo sac differentiate and begin forming multiple embryos (Hand and Koltunow 2014). In gametophytic apomixis (**Figure 2.1**), the maternal embryo develops without fertilization. The endosperm can develop with or without fertilization of the central cell; a typical sexual endosperm is triploid and comprised of two maternal genomes and one paternal genome, but variable maternal:paternal genome ratios are found in apomicts. Gametophytic apomixis can further be divided into diplospory or apospory, depending on the origin of the diploid zygote within the embryo sac. In diplospory, the megaspore mother cell (MMC) bypasses meiosis (or undergoes a modified meiosis) and the unreduced egg cell develops into the embryo. In apospory, a diploid somatic cell (instead of the mother cell) will remain in the embryo sac and begin mitosis. The processes of reproduction that are modified or "bypassed" in gametophytic apomixis are hypothesized to be controlled through genetic mechanisms acting on the sexual developmental pathway (Hand and Koltunow 2014).

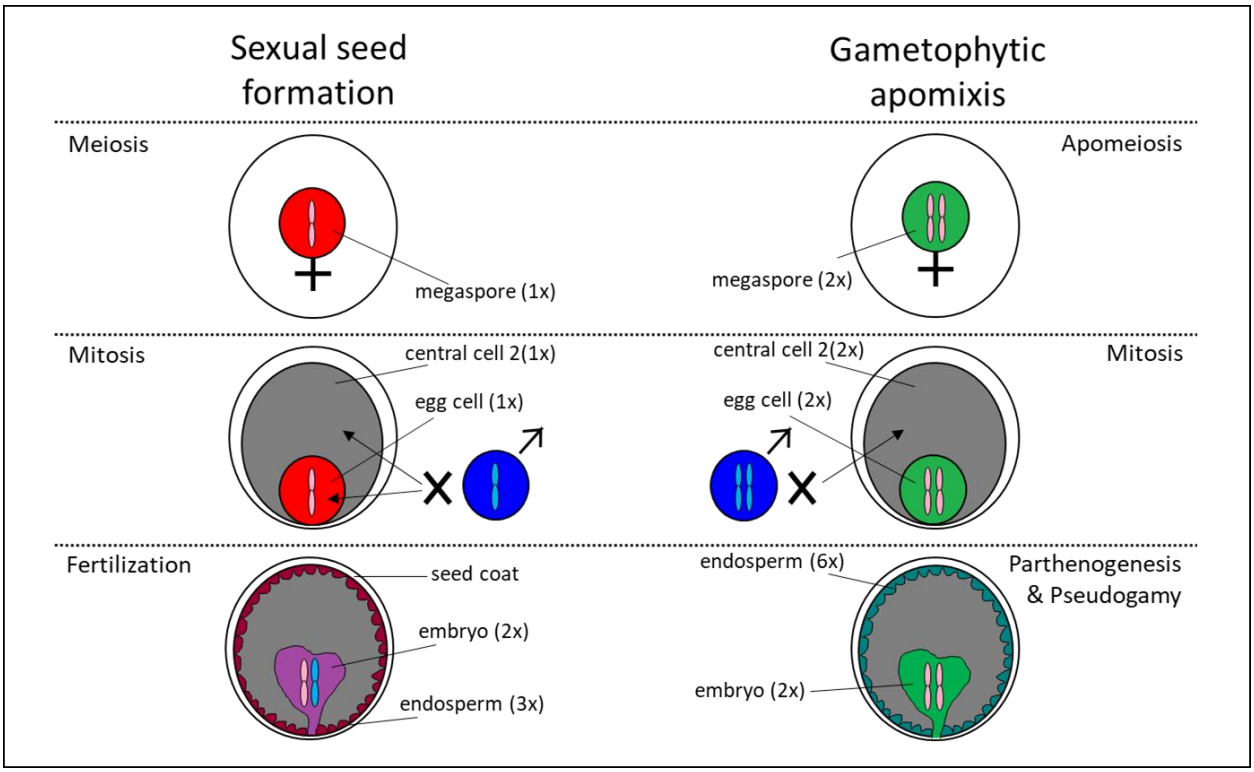


Figure 2.1. Gametophytic seed development pathways of apomixis as compared to sexual seed formation. The ovule is depicted after the embryo sac has matured; this is where sexual versus apomictic development diverge. The development of both the egg and embryo is shown. In the gametophytic pathways, the endosperm is depicted as having variable ploidy because it can develop through either the pseudogamous or the autonomous development pathways (Adapted from Hand and Koltunow 2014; Conner and Ozias-Akins 2017).

2.1.2 Phases of apomixis

The mode and complexity of apomixis inheritance is highly debated in the literature. The widely accepted genetic model proposed is that the inheritance of apomixis is controlled by a single dominant locus containing Mendelian linked factors (Barcaccia and Albertini 2013; Aguilera et al. 2015). However, with new molecular and cytogenetic technologies it has been revealed that apomixis is a very complex mechanism that is likely determined by different genetic factors in multiple species (Barcaccia and Albertini 2013). It is generally agreed that apomixis can be divided into three functional steps (Pupilli and Barcaccia 2012), whereby one or more loci control each component of apomixis expression. Gametophytic apomixis is characterized by apomeiosis, parthenogenesis and pseudogamy (Lovell et al. 2013). Apomeiosis is a derivation of meiosis without reduction to form unreduced male and female gametes (Lovell et al. 2013). This is followed by parthenogenesis, the development of an embryo from an unfertilized egg cell (Barcaccia and Albertini 2013). The formation of a viable endosperm that require fertilization by

a sperm cell is known as pseudogamous endosperm formation, although in some species autonomous (i.e. without fertilization) endosperm formation also occurs (Hand and Koltunow 2014). All three processes are essential for the development of clonal progeny that are identical to the maternal parent.

2.1.3 Distribution of apomixis

Apomixis is found in many taxa distributed among angiosperms. Approximately 2.2% of angiosperms are apomictic (Hojsgaard et al. 2014b) representing all types of apomixis, an increased relative to numbers cited in previous research (Carman 1997). The three most common families for gametophytic apomixis to occur in are Poaceae, Asteraceae and Rosaceae (Hojsgaard et al. 2014b). Of these families, some species of interest have been found and developed into model species to investigate apomixis mechanisms. This includes *Taraxacum* spp., *Panicum* spp., *Boechera* spp., *Ranunculus* spp., *Paspalum* spp., and *Hieracium* spp. (Barcaccia and Albertini 2013). However, currently other apomictic species such as citrus crops (Pupilli and Barcaccia 2012) and forages crops (Aguilera et al. 2015) are being exploited for their inherent apomictic abilities in agriculture.

Naturally-occurring apomixis in angiosperms is widely geographically distributed, from tropic to arctic environments. In nature, apomixis is an invasive trait that has been found on all continents (Hao et al. 2011) as it is strongly associated with biodiversity (Hojsgaard et al. 2014b). Within specific climates the different types of apomixis are more abundant than others, for example sporophytic apomictic species are more frequently found in tropical climates as compared to gametophytic apomictic species that are more abundant in all other climates (Hojsgaard et al. 2014b). This is not to state that each apomictic type is limited to each climate zone, but that the different types have prospered in the different climate zones due to their asexual nature. It is thought that apomixis limits a plant's evolutionary fitness by decreasing its potential to adapt to changing environments (Mau et al. 2015). Apomixis is distributed among large plant families that are genetically rich, but is not limited to these families as it has evolved repeatedly over time (Hojsgaard et al. 2014b). Hence, apomixis plays an important role in the evolutionary process of expanding distribution by occupying new ecological and geographical niches through geographical parthenogenesis (Horandl 2006).

2.1.4 Origins of apomixis

A number of factors have a combined effect on the evolution of apomixis, including geographical parthenogenesis, hybridity, polyploidy, mutational masking and epigenetic changes (van Dijk 2003; Lovell et al. 2013). Evolutionary evidence indicates that both hybridization and polyploidy may not merely be a mechanism to counteract the effect of mutation accumulation but rather act as an inducer of apomixis through advantageous mutations (Mau et al. 2015) and gene duplication (Pupilli and Barcaccia 2012). The association of geographical parthenogenesis, hybridization, polyploidy, genomic imprinting, mutation accumulation with apomixis may only be consequential, and not necessarily a prerequisite (Koltunow et al. 1995; Pupilli and Barcaccia 2012).

2.1.4.1 Geographical parthenogenesis

In some species there is a relationship between geographical diversity and occurrence of apomixis, a correlation referred to as geographical parthenogenesis. Geographical parthenogenesis is a phenomenon that occurs when asexual lineages exhibit geographical distributions that are different from that of their sexual relatives, including larger range, higher latitudes and higher elevation, which is indicative of harsher climates (Coughlan et al. 2014; Lovell et al. 2014; Hartmann et al. 2017). Although it has been proposed that geographical parthenogenesis may be the cause of a change in reproduction mode, other factors are believed to contribute to geographical parthenogenesis, such as biotic interactions, hybridization, polyploidy, resource allocation and method of dispersal. These concepts have been studied in many different apomictic genera which have both sexual and apomictic species, and demonstrate there are selection pressures that have a correlation with the presence of apomixis.

Apomictic species are commonly found in extreme environments. For example, a colder climate is often associated with apomictic individuals, likely due to the increased occurrence of apomictic individuals during glacial periods (Hojsgaard et al. 2014b; Kirchheimer et al. 2018). One theory suggests that a cold shock may trigger the formation of unreduced pollen by a disruption of microtubule formation during meiosis, which causes the gametes to defect into post-meiotic cytokinesis (Ramsey and Schemske 1998). This is the leading pathway towards sexual polyploidization (Schinkel et al. 2016). Increased levels of ploidy can also cause a breakdown in self-incompatibility, a common occurrence in sexual species which promotes outcrossing, leading to increased hybridization (Paule et al. 2018). As a result of altered gene expression (Paule et al. 2018), polyploidy allows for niche broadening which may have been a mechanism for widely-

distributed successful clonal lineages (Kirchheimer et al. 2018). Paule et al. (2018) determined that polyploids exhibited a greater expansion in distribution range than diploids during the latest glacial maximum. Another example of an extreme environment can be caused by soil conditions, which is frequently mentioned in the literature as associated with a higher presence of apomictic individuals. Acidic soils within the European Alps have a greater incidence of apomictic individuals and species (Schinkel et al. 2016; Stöcklin and Armbruster 2016). Saline soil conditions, having excessive salts in the soil (Gregor 2013) may cause cellular toxicity towards drought symptoms. Low soil moisture is often associated with higher incidence of apomixis, as drought conditions may force the plant to either abort flowers or quickly reproduce. These stressful environmental conditions provide examples for the hypothesis of geographical parthenogenesis being associated with apomixis. Soil conditions provide a reasonable hypothesis supporting selection for asexual reproduction and the requirement for fewer resources.

2.1.4.2 Hybridization

Hybridization is the process of bringing together two different genomes into a hybrid organism. Interspecific hybridization in the genus *Boechera* may have altered the timing of reproductive organ development (*sensu* Carman 1997; Hojsgaard et al. 2014a). This is reflected by genome-wide gene deregulation in apomictic ovules (Polegri et al. 2010; Sharbel et al. 2010), which could alter the sexual pathway (Aliyu et al. 2013; Hojsgaard et al. 2014a), in addition to the effect of hybridization on the apomeiosis component of apomixis (Barke et al. 2018). Hybridization also can bring together genetic factors that are recessive in two sexual accessions to produce an apomictic offspring (Lovell et al. 2013). An example of this is the co-occurrence of *APOLLO* and *UPGRADE2* which are apomixis-specific genes that evolved independently and were brought together by hybridization (Mau et al. 2015). As hybrid organisms encounter many problems (e.g. genomic shock; Hojsgaard and Horandl 2019), for example maintaining balanced chromosome numbers, apomixis may have a mechanism through which such shock could be prevented (Beck et al. 2012). Nonetheless, hybridization and its association with apomixis has resulted in high levels of diversity within apomixis lineages (Beck et al. 2012).

2.1.4.3 Polyploidy

Polyploidy and hybridization commonly occur together in apomictic lineages (Lovell et al. 2013). Artificial polyploidization demonstrated that apomixis could be introduced into a sexual diploid accession through polyploidy in *Paspalum spp*. by crossing diploid individuals (Siena et al. 2008).

Polyploidy may not be required considering not all apomicts are polyploid (Bicknell and Koltunow 2004; Ozias-Akins 2006; Kantama et al. 2007; Lovell et al. 2013); the effects and correlation of polyploidy with apomixis is likely lineage specific. Polyploidy has also had important indirect effects on apomixis evolution. First, is that polyploidy provides a reproductive barrier against diploid individuals in the population. Second, polyploidy may cause the breakdown of genetic self-incompatibility, as a step towards the pseudogamy component. Lastly, polyploidy may help to establish an apomictic chromosome factor that would select the individual for a novel niche through a physiological or adaptive trait (Hojsgaard and Horandl 2019). Common phenotypic changes include variation in flowering time, flower size, seed size and seed number (Voigt-Zielinski et al. 2012). These phenotypic changes can be observed in the broad distribution pattern of apomicts (Hojsgaard et al. 2014b). Polyploid individuals are more competitive along the edges of their diploid population, which creates the potential for ecological differentiation (Karunarathne et al. 2018). This allows apomicts to be excellent invaders into new niches, with the ability to adapt to diverse ecologies such as harsh environments (Sochor et al. 2017) and other specialized niches (Adams 2007; Aliyu et al. 2013). This in turn leads to geographical parthenogenesis (van Dijk 2003) and the popup of new species (Hojsgaard et al. 2014b). Polyploidization is the driver of niche differentiation through mechanisms such as genetic composition, genome size and deleterious allele masking (Mau et al. 2015).

2.1.4.4 Epigenetics

The influence of epigenetic phenomenon has been hypothesized in some apomictic species. The three components of apomixis occurring together at the same time in the same individual is highly unlikely as each component alone is naturally selected against (van Dijk and Vijverberg 2005). This gap leaves room for the potential that an epigenetic influenced state provides the conditions that can be inherited for clonal reproduction. Specifically, the component parthenogenesis is considered to be spontaneous in most apomictic species (Hojsgaard and Horandl 2019). However, according to Karunarathne et al. (2018) apomictic *Boechera* individuals may have altered epigenetic profiles, change in the manner of genomic imprinting by methylation in the expression of parthenogenesis. For example, an early spring melt would initiate a delayed pollination period between pollinators and viable pollen (Hojsgaard and Horandl 2019). In this scenario, the stress of the delayed pollination would cause genomic imprinting that would be carried into the next generation to be coupled with another apomixis component, triggering apomixis.

2.1.4.5 Mutation accumulation

Another driver of apomixis is mutation accumulation, also known as *Muller's ratchet*. Muller's ratchet explains that asexual populations accumulate deleterious mutations in an irreversible manner that is similar to the movement of a ratchet because recombination (and reversal to a less mutational load) does not occur (Muller 1964). The accumulation of mutations can be either advantageous or deleterious. For example, the accumulation of mutations may lead to an evolutionary dead-end (according to Muller's ratchet) because they are limited by the lack of adaptive potential (Pupilli and Barcaccia 2012). Any new traits developed by mutation accumulation have the ability to escape apomictic populations by crossing with sexual populations, allowing the traits to escape extinction (Pupilli and Barcaccia 2012). For example, there are some mutations that can induce the development of parthenogenesis in embryos in a haploid environment (Pupilli and Barcaccia 2012; Barcaccia and Albertini 2013), which is a means towards apomixis. Mutations can also increase genetic diversity of the apomictic lineages, providing a fitness advantage in extreme conditions (Voigt-Zielinski et al. 2012) and increasing the heterozygosity of those lines (Paun et al. 2006). Heterozygosity will increase as apomictic lineages age because mutations accumulate relative to their sexual ancestors (Lovell et al. 2013). Mutation accumulation is therefore another driver of natural selection that can have a profound influence on the evolution of apomixis which could provide vital for the induction of apomixis technology into crop species.

2.1.5 Induction of apomixis

Apomixis does occur in many wild tropical grasses and a few minor crop species, but the majority of crop species are not natural apomicts (Ozias-Akins 2006; Hojsgaard et al. 2011). Therefore, there is a need to develop a method to insert apomixis technology into crop species. The literature indicates there are three possible methods to introduce apomixis: traditional breeding, mutagenesis and the synthetic approach (using a cassette).

Traditional breeding was the original method used to integrate apomixis into crops (Conner and Ozias-Akins 2017). Traditional breeding for apomixis has been attempted in pearl millet and maize (Conner and Ozias-Akins 2017). In pearl millet, Singh et al. (2010) used backcrossing against apomictic *Pennisetum squamulatum* (a wild relative of pearl millet) in order to eliminate linkage drag, resulting in a tetraploid apomictic pearl millet line which had a lower seed set and a reduction in fertility (Conner and Ozias-Akins 2017). This may have occurred due to epigenetic factors that

could limit introgression of apomixis by linkage drag of undesirable effects (Bicknell and Koltunow 2004; Conner and Ozias-Akins 2017). Traditional breeding can be considered a lab success but not a functional success for implementation towards food security.

Another approach to the integration of apomixis is artificial doubling. Artificial doubling is used by breeding programs to double the genome of sexual species in the hopes of incorporating apomixis traits. This is a common method used by institutes such as the International Center for Tropical Agriculture (CIAT) and Embrapa (Pinheiro et al. 2000; Simioni and do Valle 2009). Chromosome duplication in a sexual diploid *Brachiaria spp*. individual facilitated intraspecific crosses with an apomictic individual (Simioni and do Valle 2009). From these crosses, the authors obtained 14 apomictic seeds. However, the duplication of diploids to tetraploids produced many sterile embryo sacs. Additionally, after examining the resulting apomictic seeds, the authors determined that sterility was due to abnormalities in the female gamete, rather than the male gamete (Simioni and do Valle 2009). Using artificial doubling resulted in sterility and shrunken female gametes of the apomictic individuals due to chromosome abnormalities during the duplication event (Simioni and do Valle 2009). Despite these drawbacks, there is optimism in using this method to obtain apomictic crop species once the polyploid sexuals have been stabilized.

The final approach has gained much more interest in research. This method is to isolate genes from natural apomicts and mutants to synthesize an "apomixis cassette" that would contain all required genetic components for apomixis to be introduced into major crop species (Ozias-Akins 2006; Conner and Ozias-Akins 2017). This method includes identifying and combining the genes involved in apomeiosis, parthenogenesis, and autonomous endosperm development (Barcaccia and Albertini 2013). This "apomixis cassette" would function dominantly for all three developmental processes of apomixis. By using the candidate genes that have been identified, a transgenic plant could be developed to introduce the genes into other major crop plants (Conner and Ozias-Akins 2017). According to Conner and Ozias-Akins (2017), this is not yet a reality, as the candidate genes for the apomixis developmental processes in various plant species are still being identified. An example of this would be the development of the *MiMe (mitosis instead of meiosis)* genotype in *A. thaliana* that was developed from two previously identified genes: one eliminates the recombination event and the other modifies chromatid segregation (d'Erfurth et al. 2009). The *MiMe* genotype induces apomeiosis, which will serve as a stepping stone towards

apomixis engineering. This method is described as the most optimal method for inducing apomixis into major crop species such as maize or rice. Research continues to mine model species to understand the apomixis pathway in order to develop engineered crop species that would produce high yields of genetically pure crops. This technology has the ability to revolutionize hybrid crop species if apomixis can be understood and captured.

There have been many methods used to identify the most cited candidate genes for apomixis in the literature. The original *Baby Boom* (*BBM*) gene was identified in *Brassica napus* as a transcript (cDNA) from embryogenic cells in microspore-derived tissue culture (Jha and Kumar 2018). *BBM* induces embryogenesis (formation and development of an embryo) in differentiated cells within culture conditions in many plant species, including *A. thaliana* (Boutilier et al. 2002), *Glycine max*, *Capsicum annuum* and *Zea mays* (Jha and Kumar 2018). The *BBM* gene and its successors has been thoroughly reviewed (Conner et al. 2008; Conner and Ozias-Akins 2017). Another candidate gene, *APOLLO (Apomixis-Linked Locus*) from *Boechera* species, was identified using comparative transcriptome analysis of sexual and apomictic ovules in order to identify differences in gene expression (Corral et al. 2013). The *APOLLO* gene has a polymorphism that allows its upregulation in apomictic pre-meiotic ovules as compared to sexuals (Conner and Ozias-Akins 2017). The *APOLLO* gene and its relationship to *Boechera* species is discussed in the next section.

Another strategy to identify genes involved in apomixis is through targeted mutagenesis (Bicknell and Koltunow 2004). Research into mutagenesis on sexual plants has identified components that mimic the developmental steps of an apomictic pathway (Conner and Ozias-Akins 2017). For example, the dyad recessive mutation found in *A. thaliana* modifies the first meiotic event into a mitotic one, resulting in the formation of an unreduced, non-recombinant female gamete and creating a parthenogenesis-like event (Ravi et al. 2008). This mutation occurs in the *SWITCH1* gene (Conner and Ozias-Akins 2017). These are just a few examples of the many candidate genes that have been identified (Conner and Ozias-Akins 2017).

2.1.6 Value of apomixis

With the growing need to improve the performance of crop varieties, apomixis could be an important tool for our modern agriculture system by providing food security. Food security was defined by the World Food Summit in 1996 as follows: “when all people, at all times, have physical and economic access to sufficient, safe and nutritious food to meet their dietary needs and food

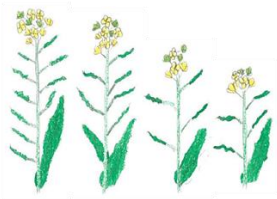
preferences for an active and healthy life". As the global population is predicted to reach almost 10 billion by 2050 (United Nations 2015), agricultural production will have to rise by 70% as of 2009 statistics (FAO 2009). As a technology, apomixis could serve as a major agricultural advantage by fixing heterosis in a single generation, which would be beneficial to plant breeders (Bicknell and Koltunow 2004). Apomixis introduced into F1 hybrid crops would generate a genetically fixed progeny for generations (**Figure 2.2**; Bicknell and Koltunow 2004; Spillane et al. 2004; Hojsgaard et al. 2011; Pupilli and Barcaccia 2012; Corral et al. 2013). The amount of time needed to breed novel traits into crops would be reduced which would also reduce the cost to produce new varieties (Bicknell and Koltunow 2004). Breeding programs would be both faster and able to rapidly react to changing needs and environments because apomixis would make maintaining a stable genome easier (Spillane et al. 2004). Apomixis provides a new perspective to developing superior varieties (Ozias-Akins 2006), as well as providing support to the food chain, not to mention new financial and technological opportunities.

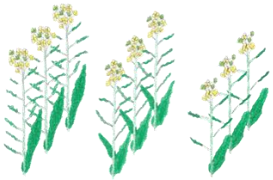
Commercial Hybrid Production

Apomictic Hybrid Production

Segregating F2 population

Is formed by allowing an F1 population to self fertilize or randomly mate



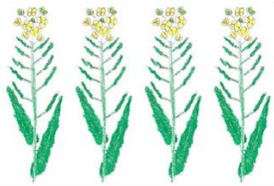


Breeding inbred lines

Genotypes that have desirable traits are selected and self fertilized

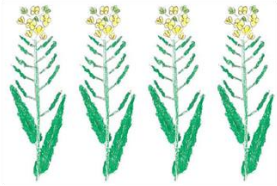
Production of sterile seed parents & fertile pollen donors

Inbred lines are tested for combining ability, the best varieties are selected and inbred to obtain pure lines, fertile pollinators (pollen donors) and seed parents that are emasculated



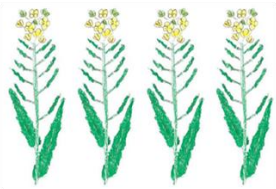
Commercial F1 hybrids

Crossing inbred lines will result in uniform, vigorous, and high yielding F1 hybrids that are sold for commercial production



Segregating F2 population

Is the result of the seed that farmers sell because it is highly variable from outcrossing



Fixed heterosis

Seed parents crossed with apomictic pollen donors will result in F1 hybrids with permanently fixed heterosis

Figure 2.2. Comparison between production of commercial F1 hybrids and apomictic F1 hybrids. The conventional breeding approach requires that pure lines of both parents are maintained, producing an offspring that will not maintain heterosis upon re-seeding, as seeds from the F1 population will be highly variable due to segregation and recombination. However, in apomictic production, the inbred lines are selected and then crossed with apomictic pollen donors. The F1 hybrid seeds are heterozygous, which can be maintained, creating permanently fixed heterosis. Adapted from Barcaccia and Albertini (2013).

Apomixis has particularly great potential for crops which are currently bred and sold as hybrids. F1 hybrid seeds represent a large portion of the $50 billion commercial seed market (Spillane et al. 2004). Hybrid seed production requires development and maintenance of highly inbred parental lines (Ramulu et al. 1999). Using apomixis to fix these parental genotypes would reduce this requirement, as well as minimize the number of generations required to generate these valuable inbred lines, thus accelerating the breeding process. As F1 hybrid seeds represent a large part of the market, this represents a substantial cost saving, estimated at almost $4 billion (inflation included) per year in the production of hybrid seed (Spillane et al. 2004). Despite the great potential of this technology, the data for these estimates are from dated sources and no new estimates have been made since. Apomixis technology could thus not only revolutionize crop production, but also have a large impact on social systems and economics around the globe.

Apomixis provides another option for plant breeders to improve food security for consumers everywhere. This technology may assist in meeting ever-increasing production targets by expanding the ecological niches available to grow productive crops, particularly in developing nations (Toenniessea 2001). Making this technology readily available for farmers worldwide puts the ownership of seed varieties for specific environments back into producers' hands, thereby improving local food security issues (Bicknell and Koltunow 2004) by providing higher quality seeds that fit local dietary needs and preferences. In the developed world, apomixis technology could enable breeders to respond more quickly to the changing demands of the consumer, decreasing the overall price of staple foods.

2.1.7 Risks of apomixis

The most pressing issues in apomixis technology are the biological issues of apomixis itself, which currently prevent this technology from being widely available. Consequences of apomixis technology include the increase in monoculture production that may result in a decrease in biodiversity (GRAIN 2001); this will be aggravated by reduced gene flow between individuals. However, this concern is countered by an increase in the level of heterozygosity and broader donor

genetic material from wild or distant relatives (Spillane et al. 2004) which could actually increase the genetic diversity of crops. There is still the potential for infectious apomixis to occur, whereby pollen of a dominant apomixis transgene spreads to closely related outcrossing species, resulting in the displacement of sexual relatives (van Dijk and Van Damme 2000; Spillane et al. 2004). This could result in an escape from sterility of the hybrids (Spillane et al. 2004). The theory of escapism is based on the assumption that apomixis will function as a non-conditional system, which is highly unlikely because of conditional apomixis. Conditional apomixis acts as a system that turns on male sterility to inhibit gene flow into the environment and then turns off for the incorporation of new genetic material through sexual crosses (Spillane et al. 2004). Research is well aware of the hurdles that apomixis faces before deployment.

However, the biggest challenges with apomixis may not be its genetics. First, there is the issue of maintaining public access to apomixis enabling technology (Spillane et al. 2004). Profit potential may be limited because a major target for development is developing world farmers (Bicknell and Koltunow 2004). It will be important to encourage research funding by governments into the public research sector, making apomixis available for all (Toenniessea 2001). Secondly, since apomixis is likely to be brought to market via genetic modifications, it will be opposed by anti-GMO groups that could place barriers to the accessibility of apomixis technology. Thirdly, it will be expensive to bring this technology to market given current regulatory requirements which will limit profitability. Despite the plethora of obstacles, the global need for apomixis may be greater than the challenges of its implementation.

2.2 *Boechera*

The genus *Boechera* (family Brassicaceae), commonly known as rockcress, are short-lived, perennial plants that commonly inbreed (Rushworth et al. 2011). *Boechera* (**Figure 2.3**) is an ideal model species for the development of apomixis technology due to its ease of propagation, small genome, short growth time, and large seed set (Ozias-Akins 2006). These traits are necessary for the development and implementation of apomixis into agriculture (Bicknell and Koltunow 2004). *Boechera* is also valuable as a model because of its close relationship to both *A. thaliana* (Rushworth et al. 2011) and canola (**Figure 2.4**). As a model for studying apomixis, the *Boechera* genus includes sexually and apomictically reproducing species. *Boechera* sexual plants are diploid and highly homozygous, while apomictic *Boechera* can be either diploid or polyploid, and likely occurred through hybridization (Lovell et al. 2013). *Boechera's* close relationship to *A. thaliana*

provides an extraordinary amount of information towards understanding sexual reproduction and the identification of a range of apomixis-like mutations (Koltunow and Grossniklaus 2003).

Figure 2.3. Photograph of *Boechera divaricarpa* (Boechera's rockcress) grown in laboratory conditions for research. Photo taken by Tim Sharbel.

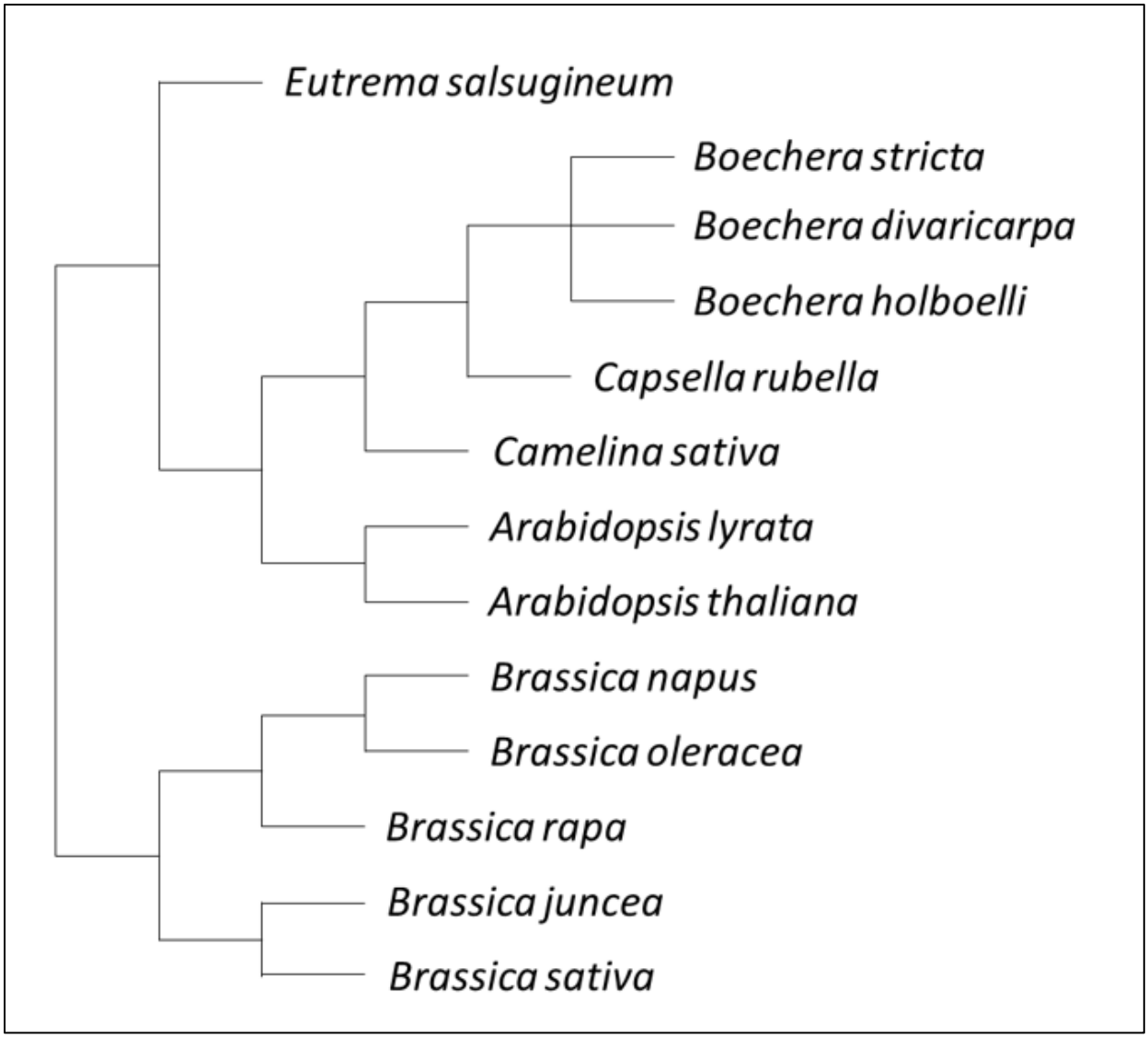


Figure 2.4. A phylogenetic tree of the relationship between *Boechera spp., Arabidopsis spp. and Brassica spp*. This tree indicates the close family relationship between these agriculturally important angiosperms. (Derived from Heng et al. 2014; Taskin et al. 2017; Yin et al. 2017; Kliver et al. 2018).

2.2.1 Distribution of *Boechera*

Boechera are commonly known for growing in undisturbed habitats throughout North America, ranging from desserts to alpine meadows (Rushworth et al. 2011), indicating that local adaption has occurred throughout many diverse environments. The occurrence of *Boechera* in many diverse habitats is the result of repeated cycles of glaciation that occurred in North America, causing geographic isolation and speciation to occur (Rushworth et al. 2011). The geographical patterns of dispersal indicate that the environmental conditions affected the distribution patterns of apomictic *Boechera* (Rushworth et al. 2011). From this, Lovell et al. (2014) determined that on average, there

is a greater occurrence of apomictic species than their sexual counterparts at higher elevations and lower latitudes.

2.2.2 Origins of apomictic *Boechera*

Research into *Boechera* species has demonstrated that hybridization is tightly linked to the origin and evolution of apomixis (Schranz et al. 2006). Apomictic *Boechera* accessions are most commonly derived from self-fertilizing, homozygous diploid sexual individuals (Beck et al. 2012). Further, the data collected in Lovell et al. (2013) shows that *Boechera* apomicts are most likely interspecific hybrids. This was suggested by the amount of heterozygosity detected in diploid sexuals, diploid apomicts and triploid apomicts that exhibited homozygosity, intermediate and high heterozygosity respectively (Lovell et al. 2013). High heterozygosity suggests that an apomictic lineage is derived from a recent hybridization event (Beck et al. 2012).

Hybridization is a large contributing factor towards apomictic evolution in *Boechera.* It is potentially involved with apomixis induction by producing regulatory changes to gene expression (Sharbel et al. 2010), since apomictic hybrids have different gene expression patterns than their sexual counterparts during ovule development (Sharbel et al. 2009, 2010). For example, there is evidence that interspecific hybridization occurred between diploid *Brassica stricta, Brassica holboellii* and *Brassica divaricarpa* (**Figure 2.4**), leading to apomictic offspring (Schranz et al. 2005; Kantama et al. 2007). Hybridization provides the necessary means to produce apomictic individuals through a lack of chromosomal pairing, changes in gene-expression and by potentially masking deleterious mutations (Kantama et al. 2007). The hybrid nature and widespread occurrence of diploid apomixis within *Boechera* species implies that there is a common origin of apomixis factors (Corral et al. 2013) that may have spread into sexual populations (Schranz et al. 2005).

2.3 Apomixis linked locus

APOLLO (APOmeiosis-Linked LOcus) is a candidate gene for the apomeiosis genetic factor in *Boechera* species (Corral et al. 2013). It was first identified in a microarray experiment profiling genome-wide expression in pre-meiotic ovules of apomictic and sexual *Boechera* (Corral et al. 2013). *APOLLO* exists as two different alleles; compared to the sex-allele, the apo-allele of APOLLO is commonly identified by the 20-nucleotide insertion polymorphism in the 5′ untranslated region (Corral et al. 2013). Sexual *Boechera* are homozygous for the sex-allele, while

all the apomictic *Boechera* that have been investigated to date are heterozygous for the sex-allele and apo-allele (**Figure 2.5**) (Corral et al. 2013). The APOLLO apo-allele is upregulated in apomictic pre-meiotic ovules, whereas the APOLLO sex-allele is ubiquitously expressed in somatic tissues, but is not expressed in either sexual or apomictic ovules.

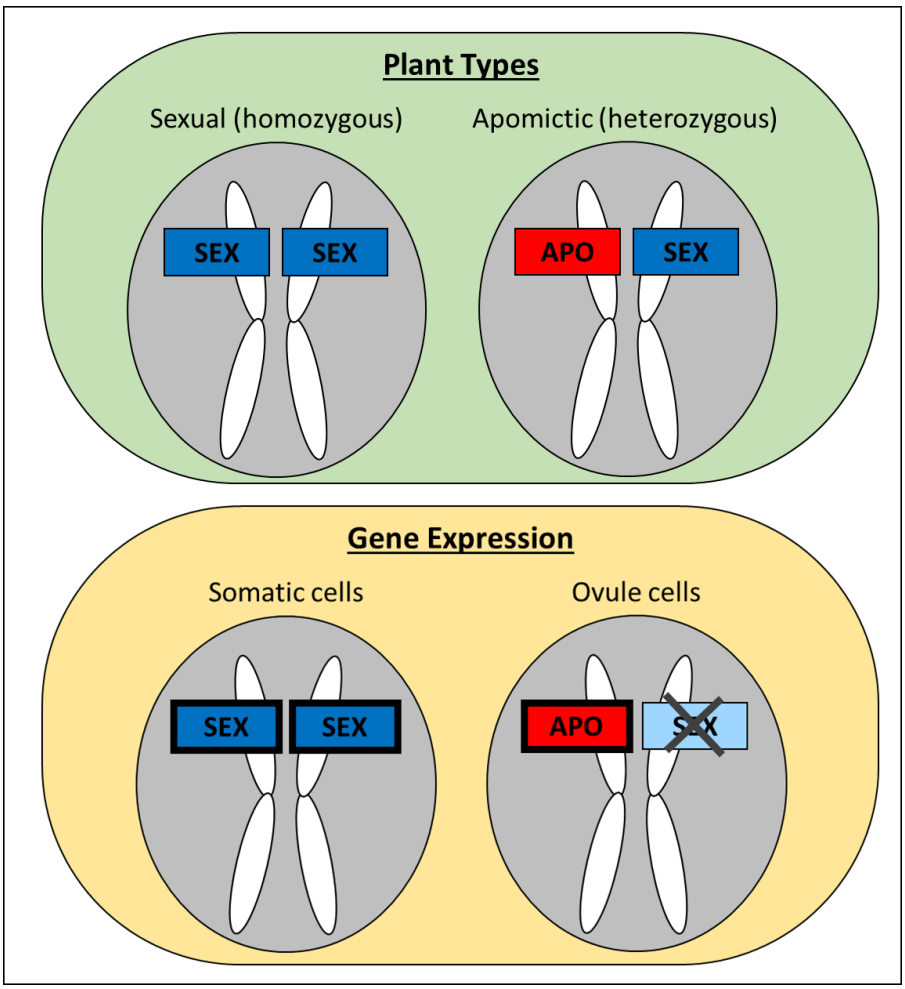


Figure 2.5. Allelic variation and expression of apomictic and sexual alleles of APOLLO. Sexual *Boechera spp*. are homozygous for the sex-allele, which is expressed in only somatic tissues of the plant. Apomictic *Boechera spp.* are heterozygous for the sex- and apo-allele, where only the apo-allele is upregulated in ovule tissues.

2.3.1 Polymorphisms of *APOLLO*

The conserved apomixis-specific polymorphisms in *APOLLO* can be categorized into three groups: a single 20-nucleotide insertion/deletion in the 5′ UTR; 16 nucleotide polymorphisms located in the intronic regions; and 12 amino acid polymorphisms in the protein sequence (Corral

et al. 2013). The single 20-nucleotide apomixis-specific insertion generated several putative plus-strand transcription factor-binding sites (TBFS) that were not present in the sex-allele (Corral et al. 2013). A 5′ -> 3′ transcription factor-binding site is where a transcription factor, a protein that controls the rate of transcription and patterns of expression, would bind to in the same direction. The 16 polymorphisms located in the intronic regions are synonymous and do not impact the coding sequence, as these regions of the DNA are removed during RNA transcription in the production of *APOLLO*. However, introns are important for alternative splicing, in that they influence and increase the proteins that are produced from a gene (Shaul 2017) and regulatory function that is independent of splicing (Gallegos and Rose, Alan 2019). Finally, there are 12 nonsynonymous polymorphisms in the coding sequence (**Table 2.1**), mutations which cause a change in the codon sequence to replace an amino acid within the sex-allele for a different amino acid found in the apo-allele. These amino acid polymorphisms found in the apo-allele may affect the degree of disorder, protein-protein interactions, and sites of post-translational modification, and may therefore change the structure and/or function of the APOLLO alleles.

Table 2.1. The location of the apo-allele specific amino acid changes compared to the sex-allele. The positions of amino acid changes between apo- and sex-allele were determined using 390S2_S16 (GenBank, KF705573) as a reference. The middle column indicates the relative nucleotide position of the polymorphism within the APOLLO sex-allele. The apo-allele amino acid composition was determined from the general consensus of APOLLO alleles in apomictic species. Derived from Corral et al. (2013).

Sex-allele amino acid state	Relative nucleotide position in allele	Apo-allele amino acid state
C	11	N
G	119	E
S	123	P
V	177	I
Y	179	H
K	195	T
F	200	L
S	201	F
T	243	A
R	285	T
G	331	R
E	412	K

2.3.2 Structure of *APOLLO*

The *APOLLO* gene is 2,538 nucleotides long and contains an open reading frame that encodes for a 496 amino-acid protein. The *Boechera APOLLO* sequences show similarities to the DEDDh

exonuclease family from *A. lyrata lyratae*, *A. thaliana* and *Brassica rapa* pekinensis (**Figure 2.4**). There are 12 conserved missense nonsynonymous polymorphisms within this coding sequence between the sex-allele and apo-allele (Corral et al. 2013). *APOLLO* has a conserved exon/intron structure, consisting of six exons and five introns (**Figure 2.46**). The 3′ UTR is about 160 nucleotides long. The 5′ UTR contains an apomixis-specific 20-nucleotide insertion that can be found in the gene orthologues of *A. thaliana* and *B. rapa pekinensis*. This insertion in the 5′ UTR provides APOLLO with a binding site for transcription factors that influence the differential gene expression seen in (**Figure 2.3**) (Corral et al. 2013)**.** The 5′ UTR and its 20-nucleotide insertion is being examined in a PhD study in the same lab (Honari, personal communication, unpublished results).

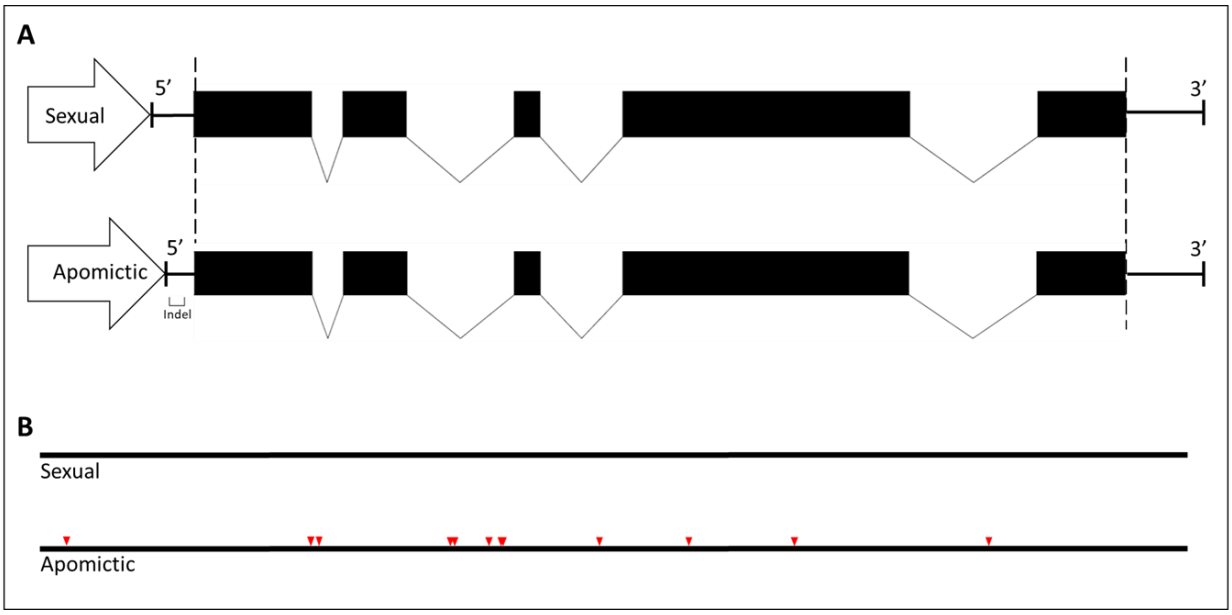


Figure 2.6. Map of the apomixis-specific polymorphisms of *APOLLO* gene (5′ -> 3′) compared to the sexual *APOLLO* gene. A, arrows indicate the promoter region, which occurs further downstream in the apomictic genome because of a truncated region in the 5′ UTR. In addition, the apomictic 5′ UTR has an apomixis-specific 20-nucleotide insertion. The black boxes mark the six coding exons and the "V" mark the five noncoding introns. Followed by a 3′ UTR that is approximately 160 nucleotides long. The vertical dotted lines indicate the region of *APOLLO* that is translated into protein which is expanded in B. B, red triangles indicate the 12 nonsynonymous polymorphisms within the coding exons in the apomictic APOLLO protein.

The APOLLO expression pattern in apomictic ovules indicates either activation of the apo-allele, repression of the sex-allele, or copy number variation for the apo-allele (Corral et al. 2013). Gene activation and repression are more commonly known as the process of turning genes on and off, which is referred to as gene regulation. Gene regulation occurs at the level of transcription which

is controlled via transcription factors. Copy number variation (CNV) occurs when a segment of DNA is present in a variable copy number compared to a reference genome of that individual (Redon et al. 2006). Interestingly, there is CNV between the apo- and sex-alleles of APOLLO (Corral et al. 2013) and this may be associated with *APOLLO* gene over-expression in ovules. CNV could also act on post-translational regulation where the duplicated copies are being transcribed but there is not an efficient mechanism to degrade the extra mRNAs (Corral et al. 2013). Another theory is that CNV is a method used by plants to prevent deleterious mutations from occurring by having more copies of the allele from mutation accumulation (Kantama et al. 2007).

APOLLO is a putative DEDDh 3′ -> 5′ exonuclease (Corral et al. 2013) that belongs to a superfamily of exonucleases known as DEDDh exonucleases. Members of this family have a conserved pattern of proteins, three Asp (D) and one Glu (E), distributed over three separate sequence segments (Moser et al. 1997; Corral et al. 2013). DEDDh exonucleases have preferences towards substrate binding sites (Huang et al. 2016). The DEDDh superfamily have 3′ -> 5′ exonuclease activity that allows them to remove single nucleotides from the 3′ end of a RNA or DNA strand (Huang et al. 2016). Enzymes with this type of activity are often involved in the maintenance of genome stability by digesting fragmented, modified and mismatches nucleotides; this is important in DNA synthesis during mitosis (Shevelev and Hübscher 2002). When first identified, the 3′ -> 5′ exonucleases were assumed to be proofreading enzymes that associate with DNA polymerases (Brucet et al. 2008). Within the 3′ -> 5′ exonuclease domain (**Figure 2.6**) there is the exonuclease (EXO) III motif. The EXOIII family is named after proteins predominantly found in *E. coli*; the EXOIII protein was first identified in mammals by Hadi et al. (2000). The EXOlll domain catalyzes the removal of mononucleotides from the 3′ termini of double-stranded DNA (Mol et al. 1995), which is where approximately 90% of the total cellular repair activity occurs (Hadi et al. 2000). The catalytic activity is dependent on the helical structure and the nucleotide sequence of the target DNA (Henikoff 1984; Moser et al. 1997). There is conservation of the primary sequence of *APOLLO* but little is known about its tertiary structure. However, APOLLO has an ortholog known as TREX1, found in mice (*Mus muscularis*) that has been studied.

Although little is known about the tertiary structure of *APOLLO,* the most similar DEDDh exonuclease with determined structure is TREX1. TREX1 has the highest exonuclease activity of the DEDDh exonuclease superfamily, and is required for maintenance of DNA involved in replication and repair (Brucet et al. 2008). An EXOIII protein and ortholog of *APOLLO,* TREX1 is a proofreading enzyme critical for removing nucleotides from the 3′ termini of DNA (Brucet et al. 2008). Many have suggested the function of TREX1 is an autonomous non-processive robust exonuclease; it may perform as a proofreading function (Brucet et al. 2007). This indicates that the TREX1 exonuclease is independent, will dissociate from the DNA after every catalytic reaction and is persistent under many conditions.

The structure of *TREX1* has been identified through x-ray crystallography (Brucet et al. 2008). Using the crystal structure, Brucet el al. (2007) identified that the hydrolytic reaction is initiated by metal ion-catalyzed phosphoryl transfer mechanisms with a histidine residue in the primary sequence. TREX1 enzymes and DNA intermediates suggest that TREX1 plays an important role in proofreading (de Silva et al. 2007), replication or repair (Brucet et al. 2008). *TREX1* is the only identified ortholog to *APOLLO* that has had its structure determined. Being an ortholog indicates that TREX1 and APOLLO may share a common ancestor but that they may not have the same function. There are several other orthologs of the APOLLO exonuclease that have been identified in crop and model plants, such as barley (*Hordeum vulgare*), maize (*Zea mays*) and soybeans (*Glycine max*), which can be used for further analysis and development (Corral et al. 2013).

2.3.3 Importance of *APOLLO*

There is conservation of polymorphisms of the APOLLO apo-allele which indicates importance. Apomictic *Boechera* characterized by these apomixis-specific polymorphisms are widespread throughout California and the American Midwest (Corral et al. 2013). Corral et al. (2013) hypothesized that the APOLLO apo-allele first evolved from a single ancestral sexual *Boechera* population that spread into sexual backgrounds by gene flow because the APOLLO apo-allele is invariant (conserved) among geographically and genetically diverse sexual *Boechera* species. Conserved functional polymorphisms are hypothesized to be important for apomictic reproduction despite having been subject to the accumulation of deleterious mutations according Muller's Ratchet (Lovell et al. 2017). However, according to the Hill-Robertson interference effect (Lovell et al. 2017) the conserved regions may have a higher frequency of linkage, meaning that it is not subject to the accumulation of deleterious mutations as Muller's Ratchet suggests. Hypothetically,

a conserved region that is so instrumental in the function of apomixis would show low genetic variation as any mutations in the conserved region are lethal. Hence, evolutionary importance of the conserved region of *APOLLO* is hypothesized because it violates the expectations of the Hill-Robertson interference and Muller's Ratchet effect.

It is unclear exactly how the *APOLLO* gene has integrated itself into the development of apomixis. Together, this data suggests functional differences between sex- and apo-alleles in apomictic ovules, and here I examine the structural differences between them. *APOLLO* shares the EXOlll domain, a functional component, with another exonuclease - *TREX1*, which is a proofreading enzyme. This similarity suggests that there might also be structural similarity between these two exonucleases. I used the methods outlined in Brucet et al. (2008), that determined the structure and functionality of *TREX1* (an ortholog of APOLLO), as a skeleton to build the methods for determining structure and function of the APOLLO exonuclease. They produced the *TREX1* protein in an *E. coli* expression system and purified it using affinity purification with a nickel-nitrilotriacetic acid-agarose resin. The purified protein was crystallized using the hanging drop vapor diffusion method with an ethylene glycol as cryoprotectant. Diffraction data was collected and processed to determine the structure. X-ray crystallography can be used as the bridge to understand the relationship between structure and function.

2.3.4 Development of methods to analyze *APOLLO*

2.3.4.1 Protein expression systems

Expression systems are used to produce protein expression via an artificial construct. Natural sources of novel proteins are often inefficient, and therefore an expression system is optimized to fit that protein (Mattanovich et al. 2012). The goal of an expression system is to produce a large amount of recombinant protein at a low cost within a system. Common expression systems used are higher eukaryotic cells (such as insect or mammalian), yeasts and prokaryotes such as *Escherichia coli*. Bacterial expression systems can produce large amounts of recombinant proteins at a low cost and are the rational choice to begin with when designing an expression system protocol (Terpe 2006). For this reason, *E. coli* is often used in initial investigation and protocol development for expression of a specific target protein, as it is a fast and inexpensive system. However, even with these modifications, prokaryotes lack the ability to process protein post translation modifications (PTMs) (Mattanovich et al. 2012). Alternatively, a yeast expression system is an ideal system to produce large amounts of recombinant protein that require PTMs,

while retaining a low production cost. Yeast is a unique system with advantages of both a unicellular and eukaryotic expression system (Mattanovich et al. 2012). This is an excellent alternative to an *E. coli* system because yeast can perform complex PTMs and provide protein folding pathways to produce correctly folded and fully functional proteins for structure determination (Verma et al. 1998). It is essential to consider the type of yeast cell that would be most suitable for the target protein. A cell-free protein synthesis system is derived from an *E. coli* expression system that both benefits from the large scale production process (Yokoyama 2003) and is used to quickly produce recombinant proteins (Carlson et al. 2012). A mammalian expression system utilizes an environment that mimics the native system, therefore producing a functional protein; however, production output is often low (Sarramegna et al. 2003).

As the demand for pharmaceutical products is increasing, consumers are also demanding safer production platforms. A mammalian expression system can often carry pathogens and viruses that can affect the quality of the pharmaceuticals (Abd et al. 2020). A plant-based expression system is the newest method used to produce proteins such as antibodies, vaccines and enzymes. A plant system can perform PTMs and produce complex functional proteins without the risk of contact with human pathogens (Abd et al. 2020). Despite the plant system being new, other platforms have been developed such as transgenic plants, hydroponic culture and *in vitro* culture (Abd et al. 2020). There are thus many options available to produce proteins in an artificial environment.

The ability to produce correctly folded and fully function proteins through PTM is an important aspect to consider when determining the expression system that is best suited for the gene. This is an important step because after the translation of mRNA to a primary polypeptide chain, many gene products undergo one or more PTMs to form a mature protein (Knorre et al. 2009). PTMs can occur immediately or shortly after translation, during or after folding or localization, or prior to degradation. Importantly, prokaryotic cells lack the machinery to carry out many eukaryotic PTMs; expression of these proteins in a bacterial system often lead to misfolding or insolubility (Verma et al. 1998). Assessing PTM status can be accomplished through experimental PTM analysis; however, this is a laborious, time consuming and expensive process (Hasan and Khatun 2018). An alternative to experimental PTM analysis is a computational analysis, as it is faster and less expensive. Computational analysis provides a prediction of PTM sites based on a database of known protein modifications using a statistical or machine learning algorithm (Hasan and Khatun

2018). In general, these analyses are based on the protein primary sequence and protein structures (for example, tertiary, quaternary, alpha helix and beta pleated sheet) that surround the protein functional site (Hasan and Khatun 2018). Unfortunately, the majority of these prediction tools are based on mammalian datasets; most commonly, human (Audagnotto and Peraro 2017). Some of these prediction tools do not specify the species that they use to train the model; while others are not a prediction tool, but rather a comparison tool of known PTMs in specific species. Importantly, the ability to predict potential areas of PTM may limit the possibility of a successful expression system.

2.3.4.2 Protein purification systems

To obtain structural and biochemical information, protein purification is required. Protein purification is a series of steps used to purify a specific protein from a complex biological sample. There are various methods that can be used to separate proteins depending on the properties of the sample and the target protein. Chromatography is a separation technique that is able to isolate, identify and purify molecules from a mixture (Coskun 2016). There have been many stationary phases developed including paper, volatile gases, immobilized silica on glass plates, and liquid; reviewed in Otter (2003) and Coskun (2016). In protein chemistry, liquid chromatography (LC) separates the target protein from other fragments using a column-based, stationary solid phase, which immobilizes the protein, and a mobile liquid phase (Otter 2003). Liquid chromatography can be adapted to the properties of the target protein. Ion-exchange chromatography separates proteins from other fragments by the affinity between the protein and stationary solid phase based on the charge (Coskun 2016). The stationary solid phase will contain an ionic charge on the surface that is opposite of the charged protein (Giri 2015). Affinity chromatography is another type of liquid chromatography that is used to purify specific proteins and other pharmaceutical needs (Wilchek and Chaiken 2000). This method is a more selective process than the other methods reviewed as it uses unique biological interactions and the chemical structure of the protein for purification. Immobilized metal affinity chromatography (IMAC) is a modified form of affinity chromatography. IMAC purifies proteins that have an affinity between metal ions (e.g. Zn^{2+}, Cu^{2+}, Cd^{2+}, Hg^{2+}, Co^{2+}, Ni^{2+} and Fe^{2+}) (Otter 2003) and the amino acids histidine, tryptophan, and cysteine (Hage et al. 2012). The amino acid is released from the column by reducing the pH, allowing the target protein to travel through the column in the mobile liquid phase. This method allows for purification of histidine-tagged proteins and it is the method to be used in this project.

2.3.4.3 Determination of structure

Protein structure can be determined by three different methods: x-ray crystallography, NMR spectroscopy and electron microscopy. X-ray crystallography is used to determine the structure of biomolecules by obtaining a three dimensional molecular structure from a formed crystal (Smyth and Martin 2000). Nuclear magnetic resonance (NMR) spectroscopy is an alternative approach, used to determine the structure of biomolecules when x-ray crystallography cannot be used. Unlike x-ray crystallography, NMR can generate high resolution structural information on intrinsically unstructured (flexible) proteins (Montelione et al. 2000). The remaining known protein structures have been determined by cryo-electron microscopy. Cryo-electron microscopy is an imaging technique used to determine the structure of molecules by freezing the samples at cryogenic temperatures, which are then imaged and processed into a three-dimensional image. When a sample is too flexible for x-ray crystallography and too large for NMR spectroscopy, cryo-electron microscopy is used to analyze the large, complex, flexible structure (Caston 2013; Earl et al. 2017; Vonck and Mills 2017).

3. Materials and Methods

In *Boechera*, *APOLLO* exists as two naturally occurring alleles: an apomictic-allele ("apo-allele") and a sexual allele ("sex-allele"). There are 12 nonsynonymous apomixis-specific polymorphisms in the protein sequence of the apo-allele (**Figure 2.6**) (Corral et al. 2013) (**Section 2.3.1** and here). An *E. coli* protein expression system (**Figure 3.1**) was used to determine if the apo- and sex-APOLLO alleles contain essential functional differences that can be observed in their expression in a prokaryotic system. The APOLLO amino acid sequence was analyzed using bioinformatics databases. The APOLLO alleles were then cloned into the pMCSG7 vector using ligation-independent cloning and transformed into BL21 (DE3) *E. coli* cells for protein expression and visualization via western blot. This information will be used to advance the creation of apomixis technology into cropping systems such as canola, a relative of *Boechera.*

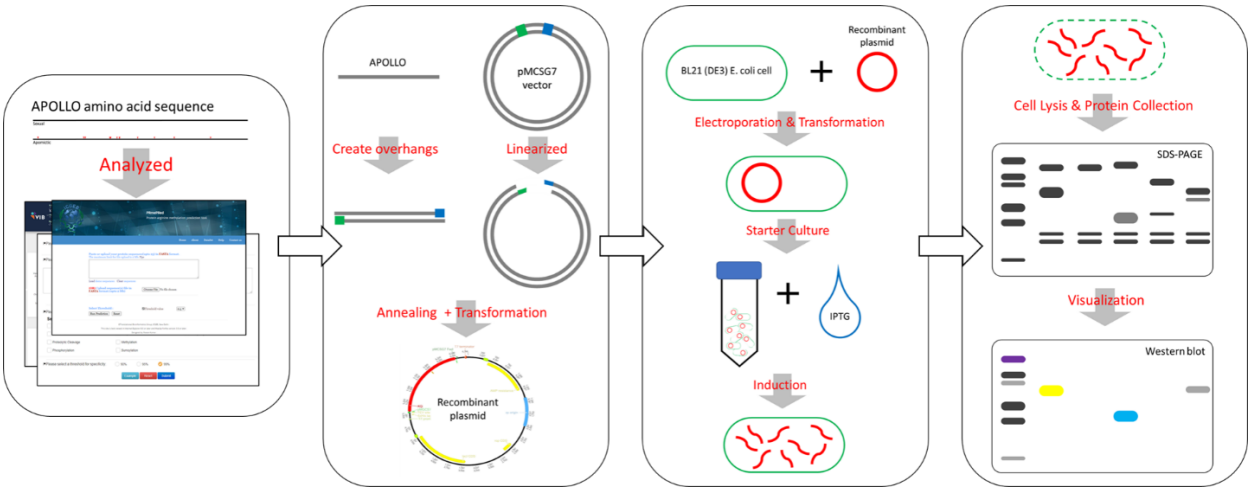


Figure 3.1. Flow chart of the analysis, cloning, transformation and protein expression techniques used in the expression system for the apo- and sex allele. The analysis protocol can be seen in the first box from the left (**Section 3.1** *Analysis of predicted bioinformatics for experimental design*). The cloning protocols can be seen in the second box from the left (**Section 3.2** *Cloning into destination vectors*). The bacterial transformation protocols can be seen in the third box from the left (**Section 3.3** *Bacterial transformation via heat shock for plasmid amplification and sequencing*). The protein expression protocols used can be seen in the final box on the right (**Section 3.4** *Bacterial protein expression of the APOLLO alleles*).

3.1 Analysis of predicted bioinformatics for experimental design

Using the APOLLO amino acid sequence, a series of bioinformatics analyses were used to determine potential characteristics of *APOLLO* that would influence the design of the following experiments. The PSI-PRED Workbench was used to determine predicted areas of intrinsic

disorder (http://bioinf.cs.ucl.ac.uk/psipred/), which led to the partitioning of *APOLLO* in order to remove these areas. Additionally, the APOLLO alleles were scanned using a series of databases which predict potential sites of post-translational modifications; these databases and search parameters can be found in **Table 3.1**. Finally, the locations of the 12 polymorphisms were compared using structural prediction tools that may improve the ability to express and purify *APOLLO*. These data guided the project and troubleshooting methods used.

3.2 Cloning into destination vectors

Phusion High Fidelity DNA Polymerase (ThermoFisher, Cat. #F530S) mix (**Table 3.2**) was used for sequence specific amplification by polymerase chain reaction (PCR); all PCR primers can be found in **Table 3.3** and **Table 3.4**. Template and fragment design are described in **Section 4.3** Ligation Independent Cloning of APOLLO alleles**.** A 1% agarose gel and a 1kb full scale DNA ladder (Bioline, Cat. #33026) was used to separate and visualize the PCR product. DNA bands were excised and purified from the agarose gel using a Wizard SV Gel and PCR cleanup kit as per the manufacturer's instructions. The DNA obtained was stored at 4°C.

3.2.1 *APOLLO* fragment cloning via ligation-independent cloning

APOLLO (**Table 3.3**; **Figure 2.6**) full length open reading frames (ORF) were cloned into pMCSG7 using ligation-independent cloning (LIC) (Aslanidis and de Jong 1990). The development of the ORF sequence is described in **Section 4.2** Fragmenting *APOLLO* for x-ray crystallization **and 4.3** Ligation Independent Cloning of APOLLO alleles. The primers (**Table 3.4**) designed for this project are about 40 base pairs long, including a 20bp annealing region that is complementary to the overhang region at the pMCSG7 vector insertion site. The *APOLLO* inserts were amplified from the supplied pUC57 vectors (Genscript; **Section 4.3** Ligation Independent Cloning of APOLLO alleles), and pMCSG7 vector was linearized, by proofreading PCR (see **Section 3.2** Cloning into destination vectors). The Phusion HiFi PCR resulted in products with blunt ends. LIC uses 3′ -> 5′ exonuclease activity to create complementary overhangs known as “sticky ends” (Aslanidis and de Jong 1990). Sticky ends were generated using a T4 polymerase (NEB, M0203S) digest reaction (**Table 3.5**). The T4 polymerase digest reaction was carried out at 22°C for 60 mins, stopped by heating to 75°C for 20 mins and stored at 4°C.

The vector and insert were annealed using a ligase free method. Vector and insert were combined at approximately 1:1 molar ratio, calculated based on their relative sizes (bp) (**Table 3.6**). Annealing was carried out in a 1.5 mL tube by incubating at 37°C for 10 mins, then on ice for 60 mins. The vector and insert were then transformed into DH5α *E. coli* where ligation was completed by the host (**Section 3.3** Bacterial transformation via heat shock for plasmid amplification and sequencing).

3.2.2 SKP1 cloning via Gateway cloning

The positive control, SKP1, was prepared for *E. coli* expression by Gateway cloning and was kindly generated by Dr. Joanne Ernest (GIFS). SKP1 was previously prepared in a Gateway entry vector by TOPO cloning and transformed into One Shot *E. coli* cells following the Gateway Cloning Protocols (ThermoFisher Scientific 2020a). SKP1 was then subcloned into the pDEST17 destination vector by Gateway LR cloning. A 1.5 mL reaction was prepared by mixing 1 µL of SKP1 entry clone, 2 µL of pDEST17, 3 µL of 1X TE buffer and 2 µL of LR Clonase II enzyme mix. The reaction was mixed by vortexing briefly and incubated at 25°C for 60 mins. Once complete, the reaction was stopped by the addition of 1 µL of Proteinase K solution. The sample was briefly vortexed and incubated at 37°C for 10 mins. For *E. coli* transformation, 4 µL of the LR reaction was transformed with 50 µL of One Shot OmniMAX 2 T1 Phage-Resistant Cells (Invitrogen, Cat. #C8540-03) (**Section 3.3** Bacterial transformation via heat shock for plasmid amplification and sequencing).

3.3 Bacterial transformation via heat shock for plasmid amplification and sequencing

The freshly annealed plasmids were transformed into 25 µL competent MAX Efficiency DH5α competent *E. coli* cells (Thermofisher scientific, Cat. # 12-297-016) or 50 µL One Shot OmniMax 2 T1 Phage-Resistant Cells (Invitrogen, Cat. #C8540-03) in pre-cooled Eppendorf tubes. 5µL of DNA was added and the cell-DNA mix was incubated on ice for 15 mins, then heat shocked at 42°C for 30 secs and placed immediately back on ice for two minutes. 200 µL of SOC media (Thermofisher scientific, Cat. #15544034) was added to the cells and incubated for 60 mins at 37°C with shaking. 50 µL of the SOC broth was spread on a LB plate containing ampicillin selection (100 µg/mL) and incubated overnight at 37°C. A PCR was performed directly from colonies to confirm the presence of the insert. Positive cultures were inoculated into LB broth with

ampicillin and grown overnight at 37°C. Plasmid DNA was isolated from broth cultures by miniprep (Wizard Miniprep DNA Purification System, Promega, Cat. # A1330) using the manufacturer's instructions. The plasmids were sequenced using T7 sequencing primers (**Table 3.4**) at the NRC in-house Sanger sequencing facility.

3.4 Bacterial protein expression of the APOLLO alleles

BL21 (DE3) *E. coli* cells were kindly provided by Dr. Michal Boniecki (Protein Characterization and Crystallization Facility, USask). Cells were freshly transformed with the sequenced plasmids by electroporation. Briefly, 2 µL of plasmid preparation was added to 25 µL of cells and incubated on ice for 15 mins. Cells were transferred to pre-cooled electroporation cuvettes (Bio-Rad, Cat. #1652089), then electroporated (Bio-Rad, MicroPulser^{TM}) at 0.22kV for 30 secs, and placed immediately back on ice for 2 mins. 200 µL of SOC media (Thermofisher scientific, Cat. #15544034) was added to the cells and they were incubated for 60 mins at 37°C with shaking at a speed of 220 rpm. The total volume of broth was spread on a LB plates containing ampicillin and incubated overnight at 37°C.

Single, ampicillin-resistant colonies were inoculated into 5 mL of LB containing 50 µg/mL ampicillin to make a starter culture. After incubation overnight at 37°C, 0.2 mL of the starter culture was inoculated into 20 mL of LB containing 50 µg/mL ampicillin. The culture was incubated at 37°C with agitation for 4 hrs, or until reaching an OD_{600} of 0.6 – 0.8 (Thermofisher scientific, NanoDrop One^{C}). A 10 mL aliquot of the culture was removed and stored at 4°C (uninduced control). Isopropyl β-D-1-thiogalactopyranoside (IPTG, 1M stock) was added to a final concentration of 1 mM and the cells were incubated with shaking at 22°C overnight (12 hrs) in 50 mL tubes. The cells were collected by centrifugation for 15 mins at 4000g, the supernatant was removed, and the pellet was re-suspended in 1 mL of lysis buffer (**Table 3.7**). The cell suspension was stored at -80°C.

The cell suspension was thawed on ice by hand shaking until fully defrosted. A dash of 0.1mm diameter Zirconia/Silica beads (BioSpec, Cat. #11079101z) was added to each sample. Each sample was sonicated in ice for 50 secs at an amplitude of 50% capacity for five pulses of 10 secs on, 20 secs off. A 100 µL aliquot was obtained from the lysed cells (total protein). The remaining sample was centrifuged at 4000g for 10 mins at 4°C and the remaining supernatant was taken (soluble protein). Total and soluble aliquots were stored at -80°C.

3.4.1 Separation of *APOLLO* fragments by 1D SDS-PAGE

Cell lysates were separated by SDS-PAGE to determine protein expression. Gels were either hand poured or pre-cast. Hand-poured SDS-PAGE gel were created by mixing 4.4 mL of ddH_2O, 3 mL of acrylamide-bis (40%, 19:1), 2.5 mL of Tris-HCL (pH 8.8), 100 µL of SDS (10%), 10 µL of TEMED and 50 µL of APS (10%) in a 50 mL falcon tube. The separation gel was poured into gel apparatus using a pipette and was layered with water until set (~1 hour). The water was removed off the top after the separation gel had solidified. The stacking gel was created by mixing 3.5 mL ddH_2O, 500 µL acrylamide (40%), 1.25 mL Tris-HCL (pH 8.8), 50 µL SDS (10%), 5 µL TEMED and 25 µL APS (10%) in a tube and cooled to 4°C. The stacking gel was poured over top of the solidified separation gel and the comb was placed in the gel apparatus. The entire gel was solidified and ready for use after 30 mins. For pre-cast gels, Mini-PROTEAN TGX ANY-kD Stain Free™ Precast Gels (Bio-Rad, Cat #. 456-8123) were used.

Cell extract samples (**Section 3.4** Bacterial protein expression of the APOLLO alleles) were denatured by adding 5 µL of 6X Laemmli buffer (Alfa Aesar, Lot #S06E03) containing 2-Mercaptoethanol (2-ME) (**Table 3.8**) to 25 µL of each in a 0.2 mL tube. They were then boiled at 95°C for 10 mins and stored overnight at 4°C if necessary.

The gel cassette was placed into the Mini-PROTEAN Tetra Cell (Bio-Rad, 1658004EDU) and 1X running buffer (Bio-Rad, Cat. #1610732) was added to the inner and outer chambers as described by the manufacturer. Each well was rinsed with 1X running buffer prior to loading samples. The denatured samples (25 µL) were loaded into the 30 µL wells and the gel was run at 200V at room temperature until the dye reached the reference line (about 40 mins). Gels were then removed from the plastic casing and either visualized directly using a ChemiDoc™ MP Imaging System (Bio-Rad, 731BR03414) and stain-free gel protocol that activated the gels for 1 min (stain-free pre-cast gels) or stained with Coomassie staining (hand-poured gels) (**Table 3.9**). For Coomassie staining, fixing solution was poured over the gel and rocked for 10 mins, then discarded. The Coomassie stain was poured over the gel and rocked until the bands were adequately stained (~2 hours). The Coomassie de-stain was poured over the gel and rocked for 2 hrs. Then the gels were rinsed with water and imaged using a ChemiDoc™ MP Imaging System.

3.4.2 Transfer of proteins to a PDVF membrane

For Western blot, proteins separated by SDS-PAGE were transferred onto a PVDF membrane using the Mini Trans-Blot Electrophoretic Transfer Cell (Bio-Rad) according to the manufacturer's instructions. Proteins were transferred onto a TransBlot® Turbo™ Midi-size LF PVDF membrane (Bio-Rad, L002048). First, the membrane was cut to size and activated in 100% methanol for 10 mins, then transferred to cool 1X transfer buffer (**Table 3.10**). The other components of the transfer sandwich were also pre-soaked in 1X transfer buffer. The membrane sandwich was prepared, starting from the negative electrode: bottom sponge, one filter paper (Bio-Rad, Cat. #1703965), gel, PVDF membrane, one filter paper, and top sponge. The sandwich was finished by closing the positive electrode. The membrane sandwich and ice pack were placed in the transfer tank and filled with 1X transfer buffer. The transfer was run for 1 hr at 100V at 4°C. The PVDF membrane was retrieved by dismantling the transfer sandwich and placed in blocking buffer (**Table 3.2**) for 1hr rocking at room temperature.

3.4.3 Visual confirmation of APOLLO alleles via Western Blot

The primary antibody solution was prepared: 5 µL of rabbit α-His antibody (1/2000) was added to 10 mL of blocking buffer. The blocking buffer was removed from the membrane. The primary antibody solution was poured over the membrane and rocked overnight at 4°C. The membrane was washed in TBST and rocked for 5 mins three times, with fresh TBST each time. The secondary antibody solution was prepared: 5 µL of goat α-Rabbit antibody (1/2000) was added to 10 mL of blocking buffer (**Table 3.3**). The secondary antibody solution was poured over the membrane and rocked for 1 hr at room temperature. The membrane was washed in TBST and rocked for 5 mins five times, with fresh TBST each time.

The histidine-tag was detected with Opti-4CN, which is a colorimetric horseradish peroxidase (HRP) substrate, that uses 4-chloro-1-naphthol (4CN) as the detection reagent. The detection substrate was prepared as described by the manufacturer: 0.2 mL of Opti-4CN™ Substrate (Bio-Rad, L008338), 1 mL of Opti-4CN™ Diluent (Bio-Rad, L008339) and 9 mL of dH_2O per 10 mL stain solution (enough for 1 membrane). The detection substrate was poured over the membrane and rocked until the desired level of sensitivity was obtained, up to 30 mins. The membrane was then placed on the ChemiDoc™ MP Imaging System and the protein bands were visualized with the colorimetric protocols.

4. Results

4.1 Analytical genomics of APOLLO alleles

Bioinformatics data were generated from online databases to guide the project. Bioinformatics is a tool that can be used to interpret biological data. Multiple online databases were used to generate bioinformatics data that influenced the direction of the project. The PSI-PRED Workbench (http://bioinf.cs.ucl.ac.uk/psipred/) was used to predict areas of intrinsic disorder, which guided the design of the various *APOLLO* fragments (**Section 4.2** Fragmenting *APOLLO* for x-ray crystallization). Multiple databases were used to generate the post translation modifications (PTM) profiles of both APOLLO alleles; these databases and parameter descriptions can be found in **Table 3.1**. Although these tools are not exhaustive, they are helpful when conducting research into an entirely new unknown protein such as *APOLLO,* where very little experimental information exists. The data generated is based on the 12 conserved nonsynonymous polymorphisms that are present between the apo- and sex-allele (**Figure 4.1.A**) which are distributed across the entire length of the APOLLO gene.

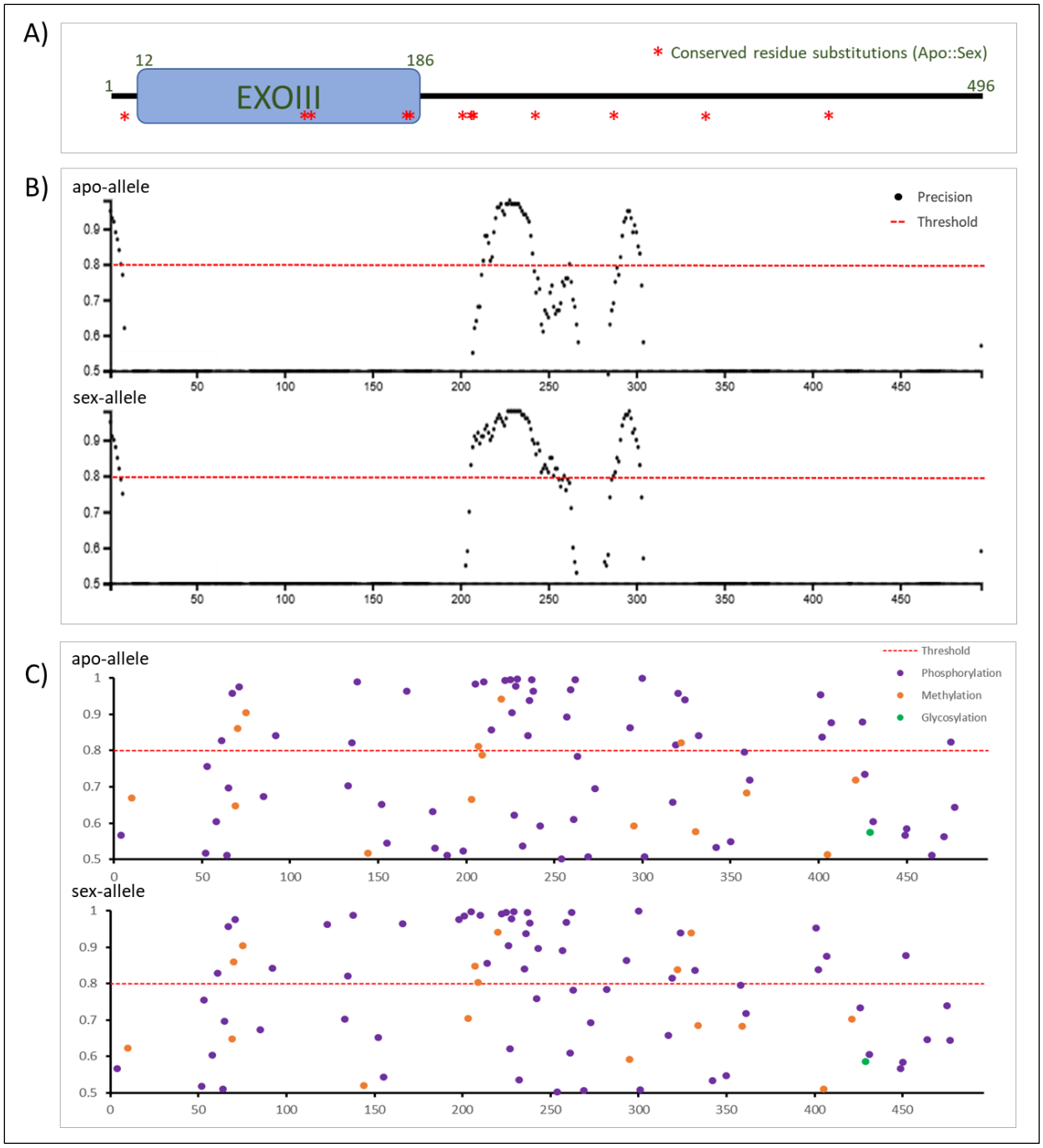


Figure 4.1. The locations of the conserved apo-allele amino acid changes (A) compared to the predicted sites for intrinsic disorder (B) and post-translational modifications (C). A) The 496 aa APOLLO protein sequence, with the region corresponding to the EXOIII domain indicated. Red asterisks indicate amino acid changes conserved in the apo-allele as specified in **Table 2.1**. B) Top; APOLLO apo-allele intrinsic disorder profile relative to the aa position, bottom; APOLLO sex-allele intrinsic disorder profile relative to the aa position. The confidence score indicates the likeliness of disorder; predicted regions of disorder show up as peaks. The threshold (red) sits at a probability of 80%, indicating there is a highly predicted site of disorder. Intrinsic disorder and domains were predicted using PSI-PRED, UCL (http://bioinf.cs.ucl.ac.uk/psipred/). C) Top; APOLLO apo-allele profile with sites of three predicted PTMs

(phosphorylation, methylation and glycosylation) relative to the aa position, bottom; APOLLO sex-allele profile with sites of predicted PTM relative to the aa position. The confidence score indicates the likeliness of a PTM occurring at that sites. Data points falling above the 80% probability threshold (red) indicates that there is a strongly predicted site of a PTM. The databases used to generate the PTM profile can be found in **Table 3.1**. The profiles were measured on a scale of 0% (0.0) to 100% (1.0).

APOLLO is predicted to have exonuclease activity and exonuclease enzymes typically have processing capabilities that could enclose DNA, this prediction suggests that there might be areas of disorder for the purpose of binding nucleic acids (Breyer and Matthews 2000). Regions of disorder are often associated with protein:protein interactions. These regions of disorder may interfere with the ability to crystallize the protein. Therefore, the intrinsic disorder profile (**Figure 4.1.B)** was first generated to determine if there were disordered regions. Two main areas of disorder were identified; one region at the very beginning of the protein and another near the middle of the protein that has multiple peaks. The areas in which disorder is predicted are broadly similar between the two alleles, although there are differences in the extent of the region and level of predicted disorder. There is also a notably wider region of predicted disorder in the sex-allele between the aa positions 200-275. However, without question there is a high probability of disorder around position 235 that is common to both alleles. Using the intrinsic disorder profile that identified these disordered regions, the fragments were designed to eliminate those regions.

The predicted intrinsic disorder profiles (**Figure 4.1.B)** of the apo- and sex-allele were compared to the 12 polymorphisms (**Figure 4.1.A)** that exist between the two alleles. The majority of the polymorphisms are located within regions of the protein that are predicted to be highly ordered, regions of which in an exonuclease are used as machinery for cleaving. Although most polymorphisms fall within highly ordered regions, there are three (C11N, T243A and R285T) that are located in regions with a high probably for disorder. A wider region of disorder is observed in the sex-allele. This wider area corresponds to the location of three polymorphisms (K195T, F200L and S201F) that may affect the degree of (dis)order in that region. As well, there are four polymorphisms (G119E, S1223P, V177I and Y179H) that occur within the predicted (highly structured) EXOIII domain. The corresponding nature of these polymorphisms at these critical areas of order and disorder suggest a potential influence on post translational modifications (PTMs).

Protein PTMs are a suite of covalent enzymatic modifications to proteins after synthesis at the ribosome. These modifications may alter protein folding and structure, regulate function, and/or

determine interaction with other proteins (Xie et al. 2015). The most common PTMs include phosphorylation, glycosylation, and ubiquitination, which add a side chain to a specific amino acid residue, as well as proteolytic cleavage and deamination (Liu et al. 2018).

As an exonuclease, APOLLO may have sites of post-translation modifications (PTMs) that influence the structure and function of the protein, including its potential for disorder. Therefore, multiple databases were searched for predicted sites of PTM (**Table 3.1**) and combined to generate a predicted PTM profile (**Figure 4.1.C**). Originally, the alleles were only scanned through PTM-ssMP, which is a database that predicts nine types of PTM in human proteins, including phosphorylation, sulfation, ubiquitination, methylation, and sumoylation (Liu et al. 2018). This database did not predict any PTMs in either allele of APOLLO. However, subsequent to this, a series of other databases used to predict sites of PTM (**Table 3.1**) were found and the alleles were scanned and plotted (**Figure 4.1.C**). The majority of the hits are sites of phosphorylation, in addition to several methylation sites and one glycosylation site. Above the 80% probability threshold, there are 8 points of predicted S-phosphorylation that differ between the sex-allele and apo-allele, as well as 2 sites of T-phosphorylation, 1 site of Y-phosphorylation, and 4 sites of R-methylation. However, at the 300 aa position, two different databases (PTM Viewer and NetPhos 3.1 Server; **Table 3.1**) predicted a site of phosphorylation with very high probability. As well, between aa 200 and aa 250 there is a cluster of phosphorylation sites that correspond with the disordered region. Within this region, the sex-allele appears to have a wide region of disorder. This correlates with the altered disorder profile which was associated with the three polymorphisms (K195T, F200L and S201F). **Figure 4.1** displays the correlation between known polymorphisms and predicted sites of disorder and phosphorylation. The identification and analysis of prediction sites of PTMs in the APOLLO alleles may provide insight into the protein function and structure.

4.2 Fragmenting *APOLLO* for x-ray crystallization

The APOLLO protein was broken down into 16 fragments based on the intrinsic disorder profile (IDP) (**Figure 4.2**). Protein fragments that exhibit intrinsic disorder indicate that there is a lack of order within the secondary and tertiary structure (Stender et al. 2015). The presence of intrinsic disorder (ID) can interfere with creating crystal structures from proteins, and can result in mobile domains or proteolysis (Oldfield et al. 2013). The fragments of *APOLLO* were designed to eliminate regions of disorder that would interfere with the process of crystallization based on the predicted IDP. By removing areas of predicted disorder, the remaining sequence could be

crystallized to examine its structure and predict function of the individual fragments; structural fragments can be bioinformatically reconstituted to determine a complete structure. Furthermore, this fragmenting method provided a potential opportunity to identify specific regions of the apo-allele-derived protein that function differently than those derived from the sexual allele.

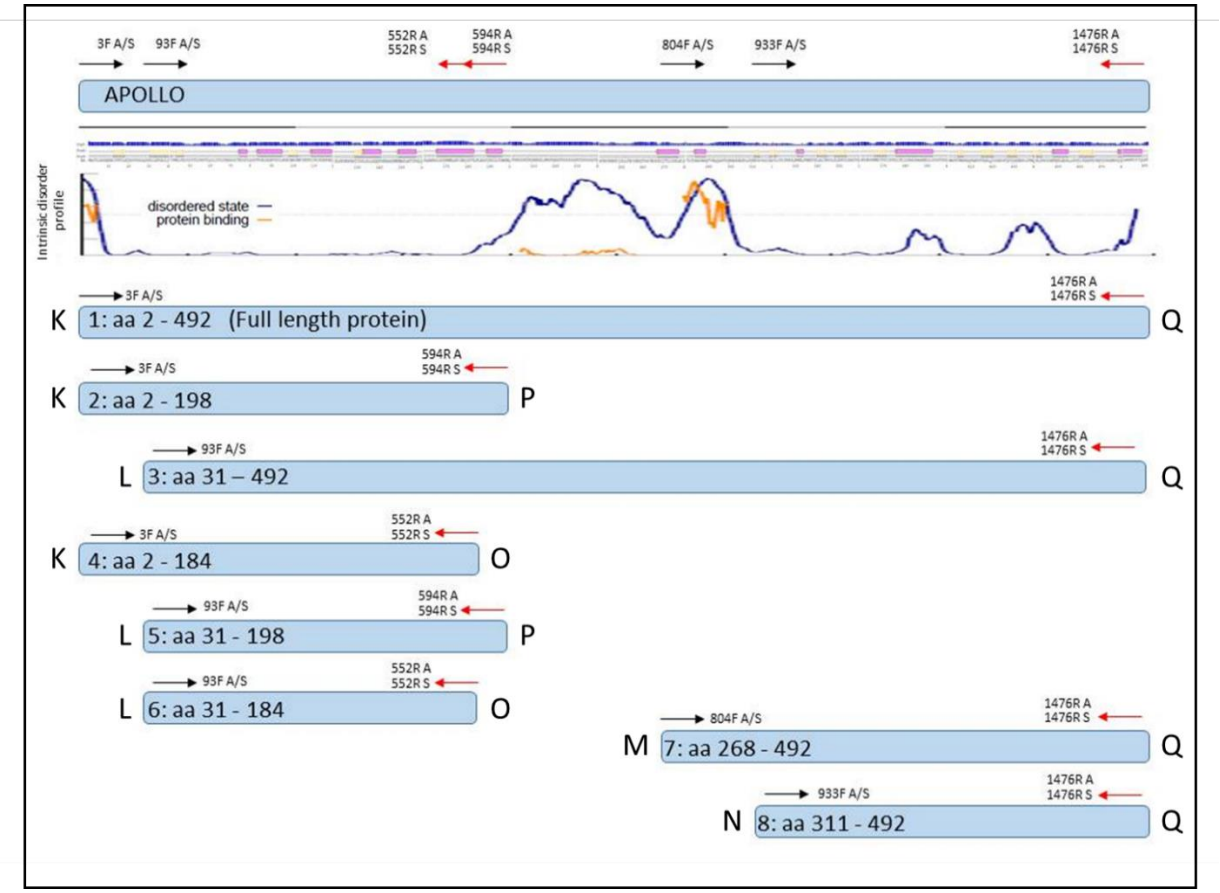


Figure 4.2. The predicted intrinsic disorder profile of *APOLLO* was used to design fragments for cloning and expression. *APOLLO* contains three regions of substantial disorder, which may interfere with crystallization. Primers were designed to amplify the full-length APOLLO protein, as well as seven truncated proteins with the various regions of disorder eliminated. The primers used for cloning are shown in black (forward) or red (reverse). The nicknames of each primer (**Table 3.4**) is located at the end of each fragment. The fragments are often referred to by their primer nickname.

Each fragment (**Figure 4.2**) was designed with specific and nonspecific primers (**Table 3.4**; **Figure 4.2**) to eliminate a different region and the following questions were asked about the functionality of *APOLLO*: (1) Do any of the fragments express in a manner that make it an essential part of the gene, (2) do any of the fragments express the same as the total length, and (3) does having the gene fragmented change the functionality of the gene all together. All of the forward primers are nonspecific for the apo- and sex-allele but all the reverse primers are specific to either

the apo-allele or sex-allele (**Figure 4.2**). The first fragment is not technically a fragment but the full-length sequence of *APOLLO* but is regarded as one in this context. However, this full-length sequence was named based on the primers that were used to create it. By having specific primers for each of the fragments created, there are now a total of 16 fragments, each of which having an apo-allele and sex-allele variant.

Each fragment removes a specific region that exhibits either a disordered state or protein binding. KQ (**Figure 4.2**) does not eliminate any regions and is used to determine if the total sequence is necessary for successful expression, which would indicate that the regions require each other for total function. The other 7 fragments remove different parts of the disorder to test if a specific region of disorder is indeed disordered or if that disordered region is required for future crystallization studies. The fragments KP, KO, LP, LO, MQ and NQ were designed to eliminate the disorder peaks in the center of the gene at different intervals. The fragments were created to generate halves of the gene independently and use bioinformatics to recreate the structure of the entire protein. The fragments LQ, LO and LP were designed to eliminate the disorder peak that occurs at the start of the profile. By this method, the different fragments that are crystallized are puzzles pieces of the entire APOLLO protein that would be available for structure and function determination.

4.3 Ligation Independent Cloning of APOLLO alleles

Consensus open reading frame (ORF) sequences of sex- and apo-alleles of APOLLO were generated from all known *APOLLO* mRNA sequences described in Corral et al. (2013) by Dr. Joanne Ernest (GIFS). These consensus sequences were codon optimized for *E. coli* expression and synthesized by Genscript (USA; www.genscript.com), with the apo-*APOLLO* and sex-*APOLLO* fragments in a pUC57 plasmid backbone (**Table 4.4**). A colony PCR was performed to confirm that the correct size of fragment was present in the pUC57 vector (**Figure 4.3**). Ligation independent cloning was then used to clone the 16 fragments into a destination vector for protein expression in *E. coli*.

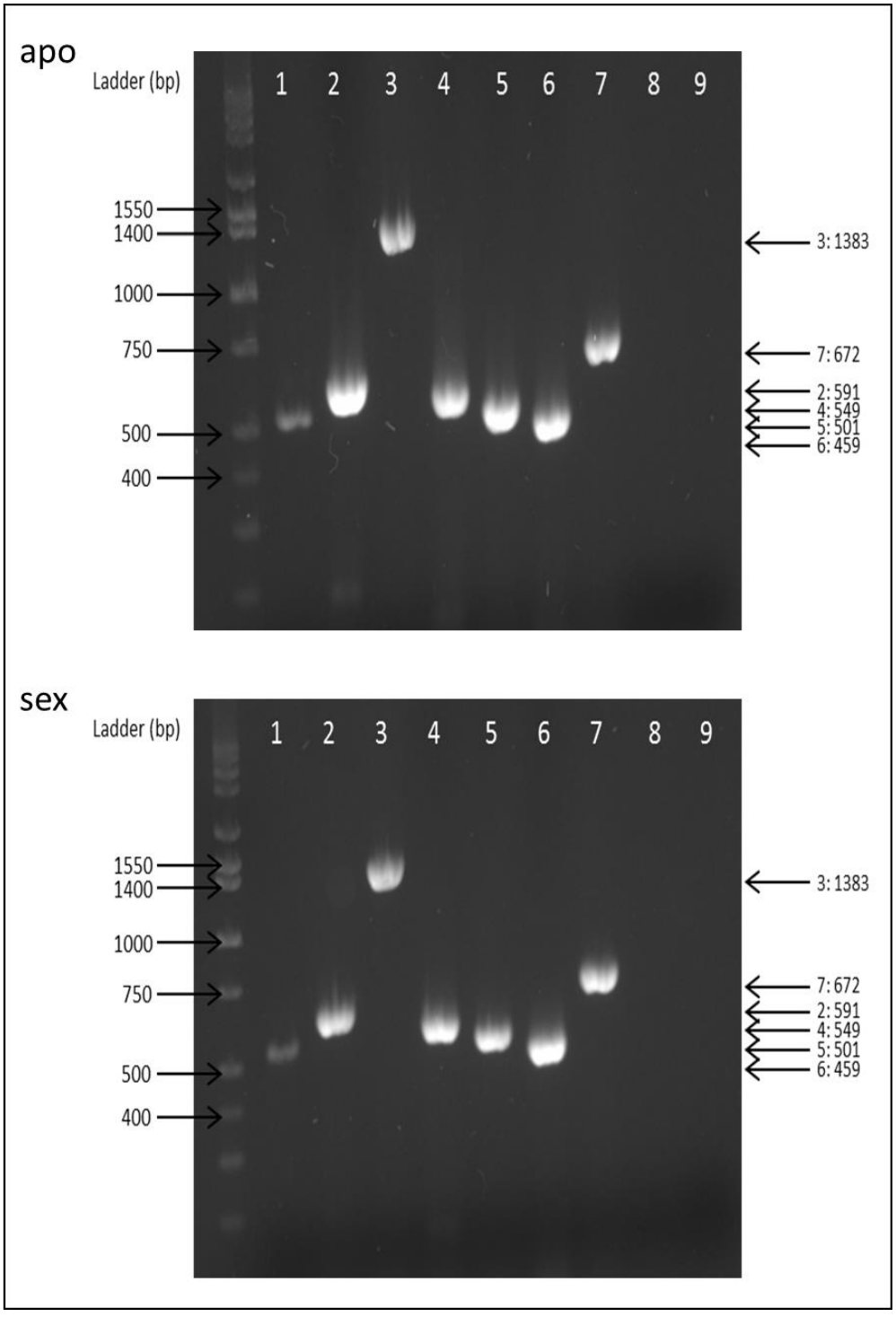


Figure 4.3. Colony PCR by gel electrophoresis for confirmation of plasmid size in vector. The top gel shows the apo-allele fragments and the bottom gels shows the sex-allele fragments that were amplified using primers in **Table 3.4**. Columns by number: 2:KP; 3:LQ; 4:KO; 5:LP; 6:LO; 7:MQ (**Table 3.3**); 9: no template control. The KQ colony PCR was performed on a different gel and confirmed the plasmid size was correct. The length of the DNA is measured in basepairs (bp) shown on the left. The ladder is a DNA

HyperLadder™ 50bp (Bioline, Cat. # BIO-33054). Fragment sizes are indicated on the right. Bands from lanes 2-7 were excised and purified (**Section 3.2** *Cloning into destination vectors*).

The *APOLLO* fragments were cloned into a pMCSG7 vector (**Figure 4.4**) using ligation-independent cloning. The pMCSG7 vector is a bacterial expression vector that contains an N-terminal His-tag with a tobacco etch virus (TEV) protease site, driven by a T7 promoter. pMCSG7 carries an ampicillin selectable marker. This vector is often used for high-throughput expression of recombinant proteins for the purpose of protein crystallization (Stols et al. 2002). The inclusion of a TEV protease cleavage site immediately upstream of the inserted enables the cleavage of the His-tag from the expressed protein during or after protein purification (Stols et al. 2002).

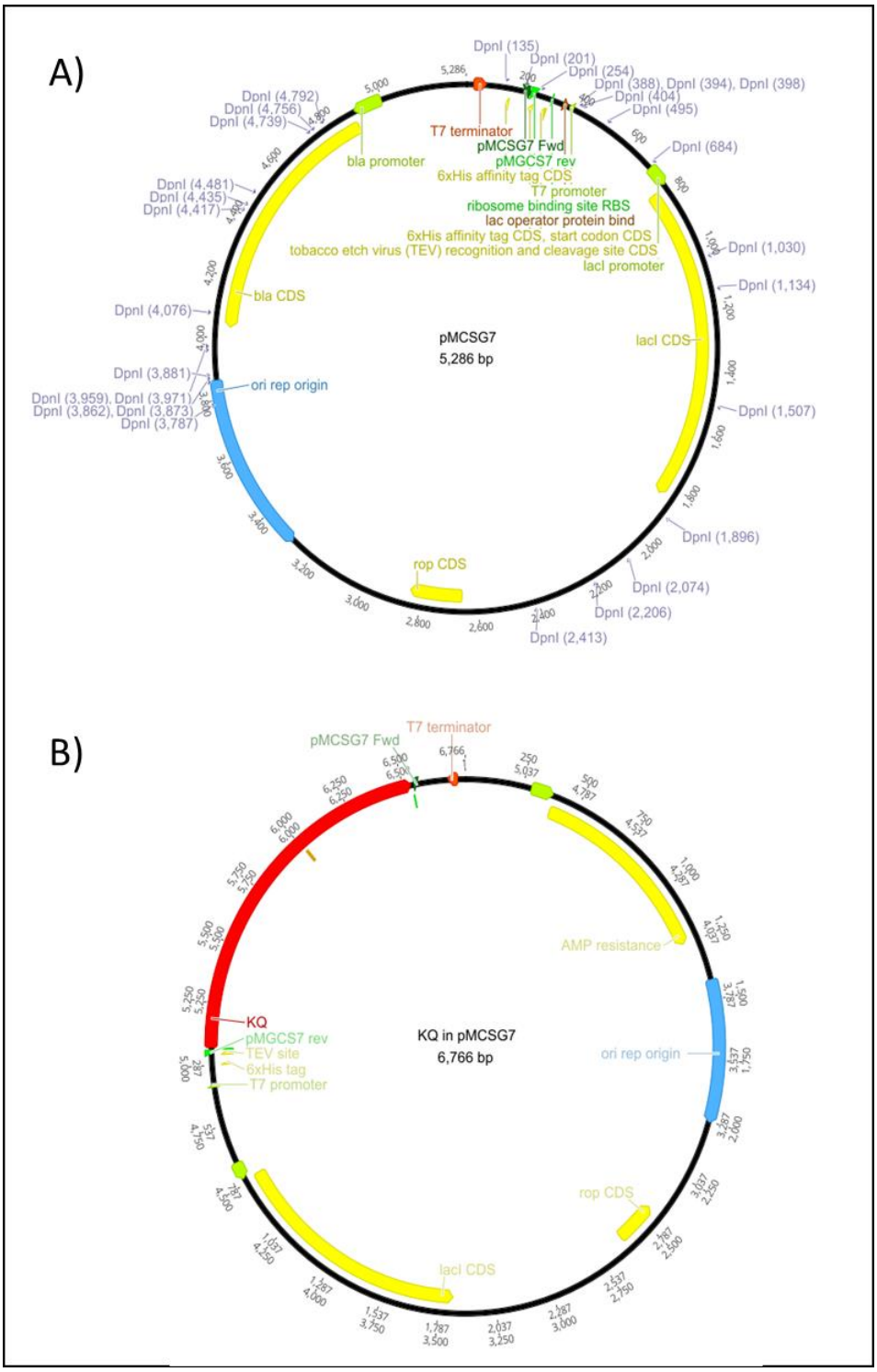


Figure 4.4. pMCSG7 vector used for ligation independent cloning. A) The empty bacterial expression vector, pMCSG7, which was used specifically with ligation independent cloning. Each of the 16 fragments of *APOLLO* was ligated into the vector via LIC. B) The fragment KQ (full length APOLLO protein in red) after LIC into the pMCSG7 vector. Images were produced using "Geneious R 11.1.5 (https://www.geneious.com)", a bioinformatics software platform for sequence data analysis in molecular biology.

4.4 Protein Expression

4.4.1 Troubleshooting

The proteins of the APOLLO alleles were expressed in BL21 (DE3) *E. coli* cells (OneShot). Individual colonies of each fragment were grown in LB media (with selection) until OD_{600} and induced for protein expression with IPTG (**Section 3.4** Bacterial protein expression of the APOLLO alleles). A sample of the total protein and soluble protein was taken. The solubility of proteins is dependent on the solvent it is to be dissolved in and the conditions (i.e. pH and temperature) that the protein is conformed into its tertiary structure (Zayas 1997). In order to be soluble, the protein has hydrophilic polar groups that encompass the hydrophilic groups via the creation of alpha helices or beta sheets. A total of 32 samples were separated using PAGE (**Figure 4.5**) to determine if protein expression occurred, how strongly, and whether the predicted size was obtained.

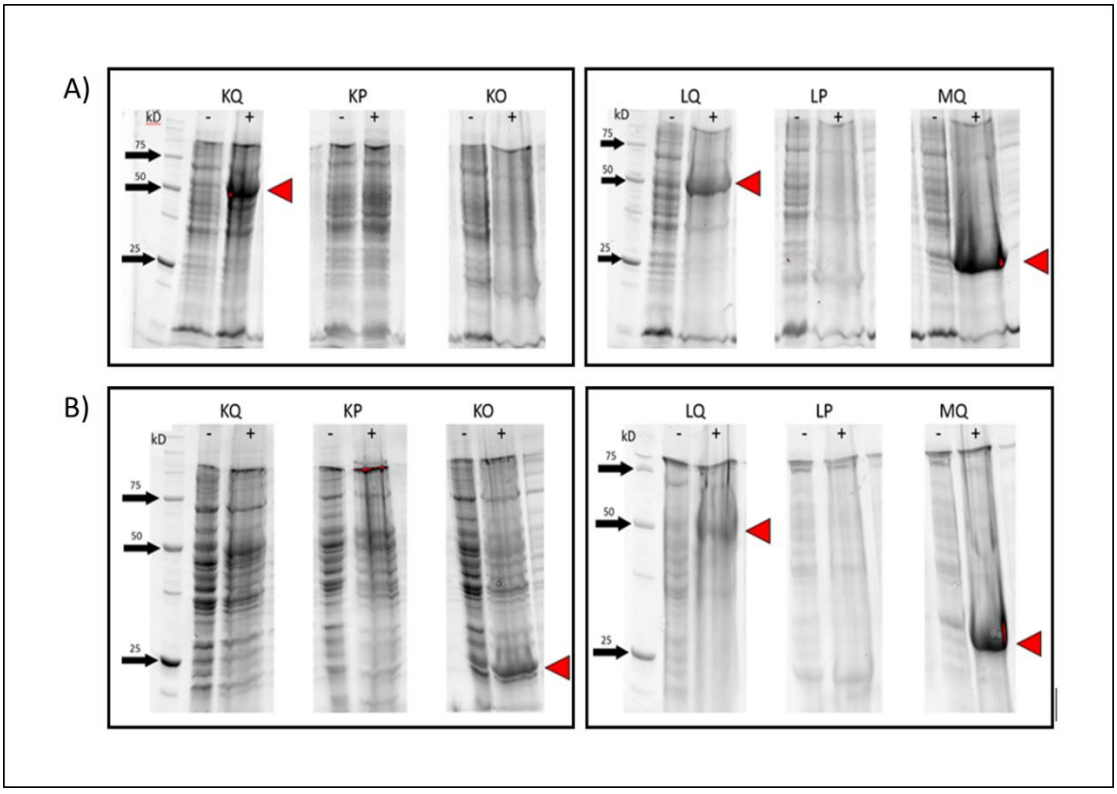


Figure 4.5. SDS-PAGE of 12 of the 16 APOLLO total protein fragments expressed in an *E. coli* system. The cloned fragments of the apo- (A) and sex- (B) APOLLO alleles were grown in BL21 (DE3) *E. coli* cells and LB media that was induced with IPTG. The proteins were separated on a Mini-PROTEAN®TGX Stain Free™ Precast Gels by SDS-PAGE. 10 µL of each sample was loaded into a well; 5 µL of Precision Plus (BioRad, Cat. #161-0363) ladder was also loaded into the first well; the gel ran for 40 mins at 200V. Each fragment has two lanes; total protein from induced cells (+) and un-induced control

(-). Fragment sizes are indicated (black arrows; kDa). KQA, KOS, LQA, LQS, MQA, MQS fragments were successfully induced, as indicated by the red arrows. KQS, KPA, KPS, KOA, LPA, LPS fragments could not be successfully induced in the *E. coli* system.

The gels with total protein samples had higher level expression than the soluble protein samples; the soluble samples that did not show any expression and were therefore excluded from **Figure 4.5**. The red arrows in **Figure 4.5** indicate which fragments (KQ, LQ and MQ) had strong enough expression to be seen on a SDS-PAGE at the intended size of the fragment (54.4, 51.2 and 24.9 kDa; **Table 3.3**) of the apo-allele. Another interesting observation of the fragments that can be observed from the apo-allele is that they all share the Q primer which is present at the 3′ termini. There are also three red arrows that indicate which fragments (KO, LQ and MQ) had strong enough expression on the SDS-PAGE at the intended size of the fragment (54.4, 20.2 and 24.9 kDa) of the sex-allele. What is striking about these results is that the LQ and MQ were both detectable in both of the alleles. As well, the full-length apo-allele that can be detected and the full-length sex-allele that cannot be detected is a consistent observation throughout the experiment and troubleshooting procedures. Due to the unsuccessful induction of many of the fragments, further research into x-ray crystallization determined that it would not be an effective method to use on proteins with ID. The fragments were then removed from this project and focus shifted to exclusively the full-length apo- and sex-alleles.

BL21 (DE3) cells are chemically competent *E. coli* cells that are suitable for transformation and protein expression with a T7 expression system *E. coli* strain. C41 cells are derived from BL21 cells, but the C41 cells have a mutation to prevent cell death with toxic proteins present, which may impact the performance of the cell's expression of the APOLLO alleles. In order to optimize the auto-induction media, this alternate cell line (C41) was used to determine the effect on protein expression. Other factors which can slow protein expression and limit inclusion body formation include a lower expression temperature and the presence of ethanol in the media; these were also tested. The full-length APOLLO alleles (KQ) were expressed in auto-induction media having been grown in both BL21 (DE3) *E. coli* cells and C41 *E. coli* cells (**Figure 4.6**). The APOLLO alleles were expressed at 20°C and 37°C, with and without the presence of ethanol (**Figure 4.6**).

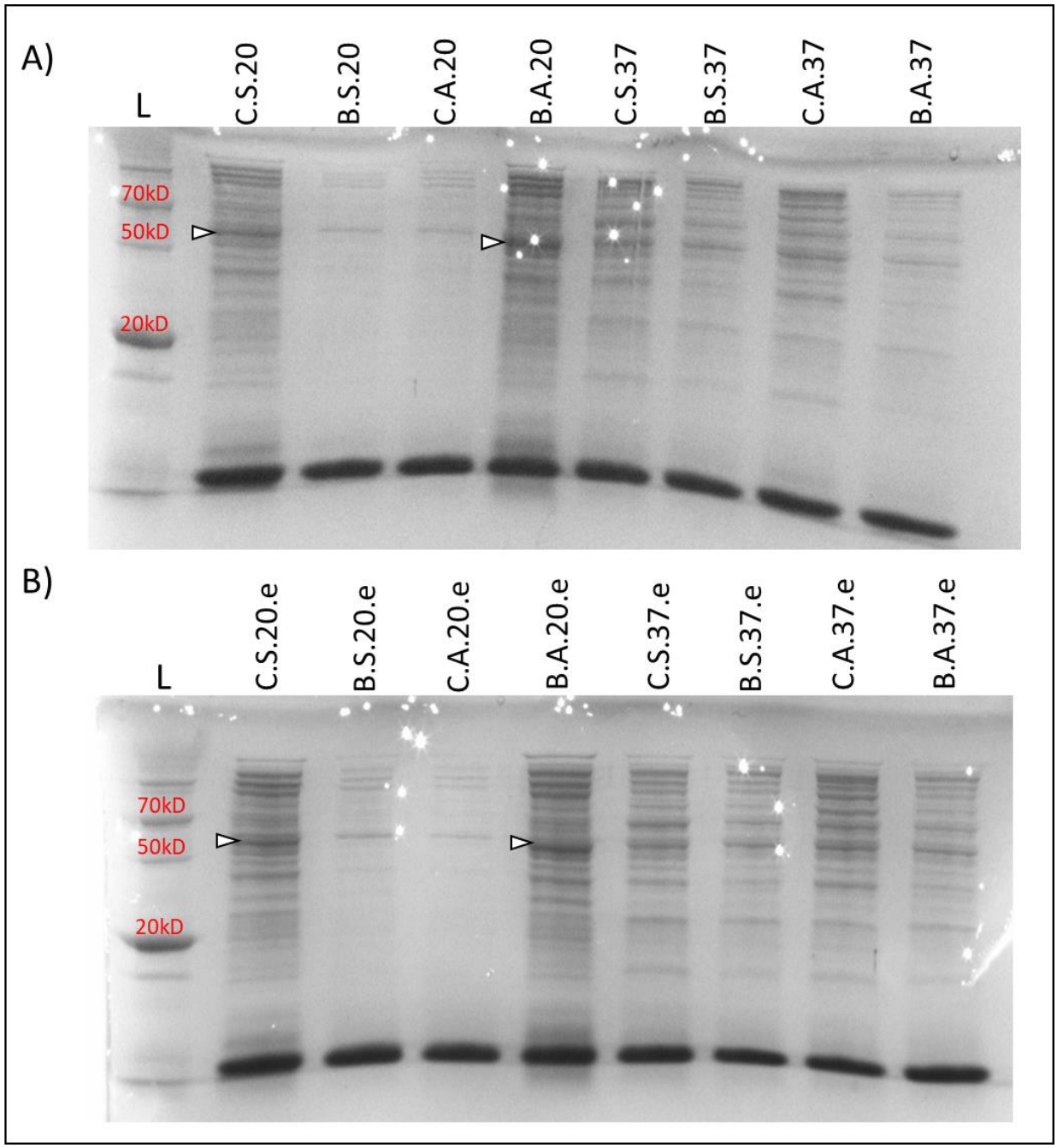


Figure 4.6. The influence of auto-induction media, *E. coli* cell type and expression conditions on APOLLO expression (SDS-PAGE) of the full length protein. Total protein from *E. coli* in auto-induction media separated on a 12% polyacrylamide gel. 10 µL of each sample was loaded into the gel. 5µL of Precision-Plus ladder (BioRad, L001650A) was loaded; the gel ran for 40mins at 200V. Specific molecular weight of proteins in kilodaltons (kD) shown in red. L: ladder; C: C41 cells; B: BL21(DE3) cells; S: sex-allele; A: apo-allele; 20: room temperature; 37: 37°C; e: 3% ethanol (V/V) present. The A) gel did not have ethanol present in the samples, but the B) gel did have ethanol present in the samples. The white arrows show the strongest bands from the gels. This was used to determine the optimal conditions to express each allele.

Protein that is expressed in *E coli* is often insoluble because the expression conditions yield high amounts of protein. Under these high expression conditions, proteins tend to aggregate into inclusion bodies to protect the bacteria from protein toxicity and to protect the proteins from proteolytic degradation (Gutierrez-Gonzalez et al. 2019). Once the proteins have aggregated into inclusion bodies, they are difficult to access and are not part of the soluble fraction anymore. Although harsh solubilization methods can recover proteins from inclusion bodies, these processes often denature the protein of interest, rendering it unsuitable for structural or functional analysis. To minimize formation of inclusion bodies, optimization of the expression conditions is often required. Expression is normally induced with the addition of IPTG when grown in LB media, which can cause a rapid production of protein (Studier 2005). In order to decrease the rate of protein expression, another media can be used, such as the lactose driven auto-induction media. Auto-induction media contains three carbon sources (glycerol, glucose, and lactose) and can be used with any vector strain that possess the *lac* operon (Fox and Blommel 2009). In this media, the glucose and glycerol sugars feed the cells initially, but as they run out, the cells switch to lactose as a carbon source. This drives expression of the T7 promoter, initiating the production of the APOLLO protein in this case. A comparison of the induction rate between auto-induction media and LB media with IPTG was made to determine the suitability of this media (**Figure 4.7**).

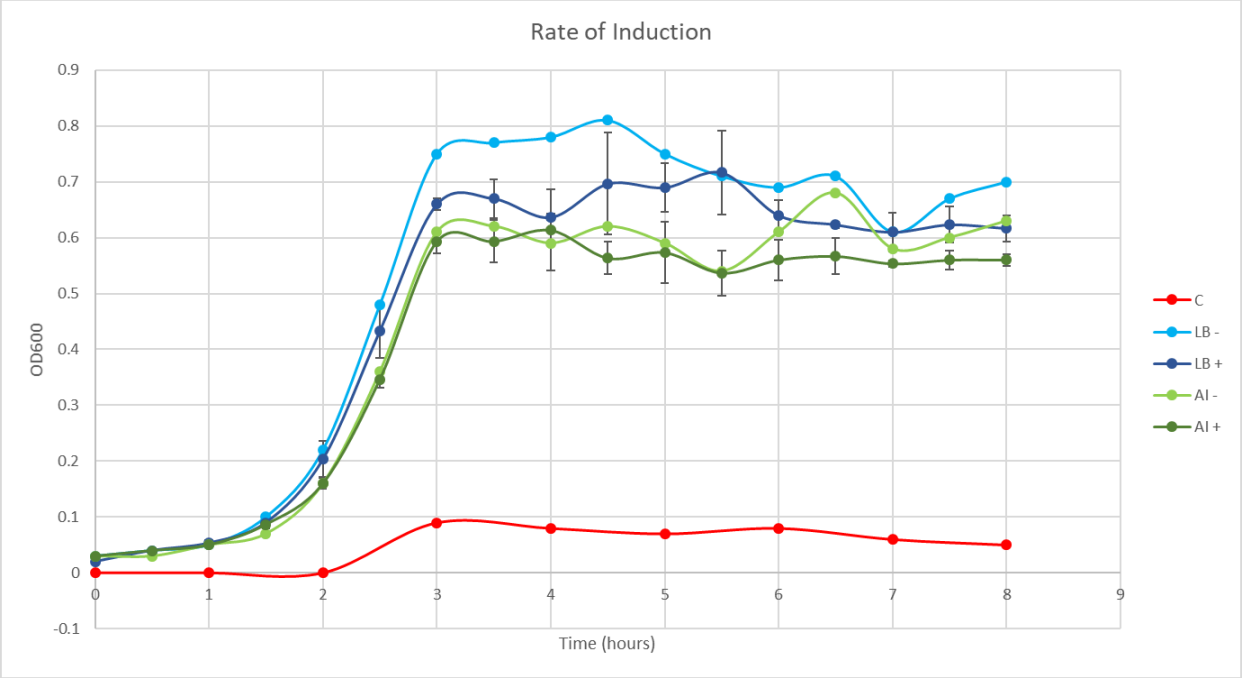


Figure 4.7. Comparison of cell growth (OD_{600}) over time between auto-induction media and Lysogeny broth. The positive control, SKP1 was grown in auto-induction media and LB media with their OD_{600} measured every half hour for 8hrs. C: negative control was distilled water; LB-: cell production without a template in LB media; LB+: cell production of the positive control, SKP1 in LB media; AI-: cell production without a template in auto-induction media; AI+: cell production of the positive control SKP1 in auto-induction media. There were three repetitions of both LB+ and AI+ samples, the means of all three were plotted with the error bars that represent their standard deviation. Single LB- and AI- samples were plotted and do not have an error bar. OD_{600} readings were measured using a NanoDrop One.

Unsurprisingly, the highest rate of cell growth was observed in the absence of a plasmid, in LB media. However, the plasmid-containing cells in LB media reached the same OD_{600} as the no-plasmid control by 4.5hrs after inoculation. This drop in OD_{600} observed at 5.5hrs suggests that there was cell lysis or that the cells may have settled to the bottom of the sample tube and did not represent a correct measurement. Interestingly, this effect was not seen in the auto-induction media; here, a similar growth rate was observed in cells with or without with the presence of a template. It was observed that cells grew more slowly in auto-induction media than in LB media and did not reach as high an OD_{600} after 8 hours, regardless of the presence or absence of an extrachromosomal plasmid.

At about 56kDa (expected size 55.5kDa) the apo- and sex- KQ can be seen being expressed at the correct size (**Figure 4.6**; white arrows). The same level of expression can be seen in both alleles, with and without the presence of ethanol. Therefore, ethanol did not affect the level of expression

of the APOLLO alleles. A higher expression of the APOLLO alleles was observed at room temperature (20°C) compared to 37°C (**Figure 4.6.A** vs **.B**) on the SDS-PAGE.

Interestingly, there was cell preference for each of the alleles. The APOLLO sex-allele has a darker band, indicating a higher level of expression, when grown in C41 cells as compared to BL21 (DE3) cells (**Figure 4.6.B**; column C.S.20 vs column B.S.20). The opposite observation is made for the APOLLO apo-allele, which exhibited a higher level of expression when grown in BL21 (DE3) cells as compared to C41 cells (**Figure 4.6.B**, column C.A.20 to column B.A.20

4.4.2 Calibrated

The APOLLO expression vectors were freshly transformed into BL21 (DE3) and C41 cells for protein expression. Single colony PCR was used on plasmid miniprep extractions to determine if the intended plasmid was present in the *E. coli* expression cells. The sex-allele did not produce any colonies when transformed into the BL21 (DE3) cells and the apo-allele did not produce colonies when transformed into the C41 cells. The SKP1 plasmid (positive control) which was in a pDEST 17 backbone could not be confirmed by colony PCR; therefore, another positive control (GFP-4M8H) was used. GFP-4M8H was kindly provided by Dr. Jieyu Chen (GIFS) and consisted of a GFP expression vector in a pET309 backbone, grown in Rosetta™(DE3) competent cells. This vector was confirmed by colony PCR.

SDS-PAGE was used to observe the difference in protein expression levels between the total protein samples and the soluble protein fractions. The positive control, GFP-4M8H, shows a high level of total protein expression where a "ghost" band appears at about 46kDa (**Figure 4.8**, white arrow). Ghosting is often observed when the amount of protein in a single band is very high. The band is also observed in the GFP positive control soluble fraction, although this band is not as strong as for the total protein fraction. For both GFP samples, the size of this band is larger than the predicted size of 34kDa. The apparent size of GFP on this PAGE may be skewed due to the large amount of protein present in the sample.

For the APOLLO expression samples, a dark band can be seen at approximately 54 kDa in the total protein fraction of the apo-allele (**Figure 4.8**, lane 3, white arrow). This indicates that the total protein sample of the apo-allele is present at the correct size (55.5kDa). However, this band was not observed in the soluble fraction, suggesting that the expressed protein may be aggregated into inclusion bodies.

Conversely, no specific band was observed at the predicted size for APOLLO of 55.5kDa in either fraction of the sex-allele. This may be due to degradation or cell toxicity of the sex-allele in BL21 (DE3) cells.

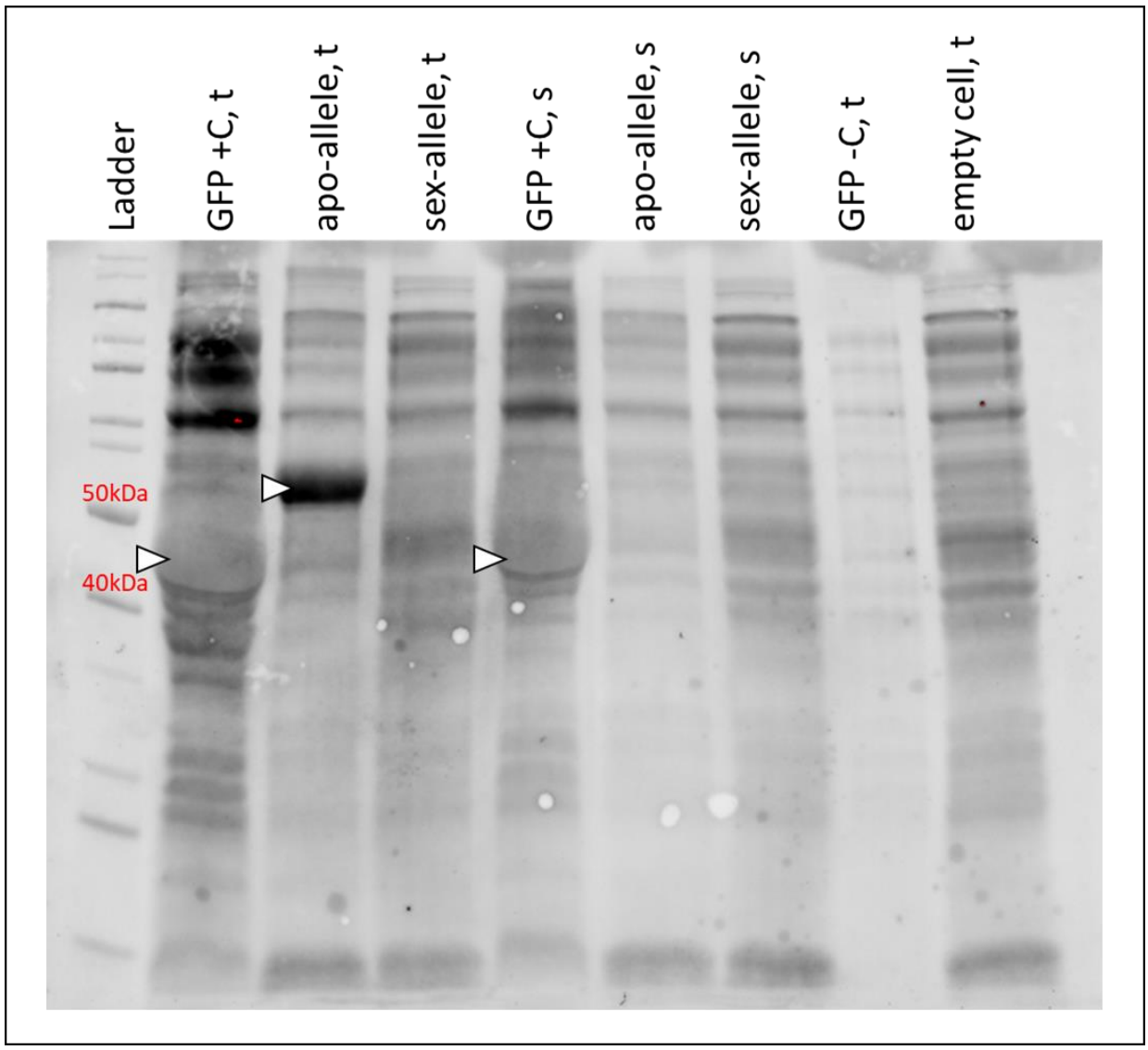


Figure 4.8. The total protein compared to the soluble fraction of the APOLLO apo and sex-allele ran on SDS-PAGE. The apo-allele was grown in BL21 (DE3) cells and the sex-allele was grown in C41 cells; the cells were placed in LB media until they were induced with IPTG. The proteins were separated on a Mini-PROTEAN®TGX Stain Free™ Precast Gels by SDS-PAGE. 20µL of each sample was loaded into a well; 5µL of PageRuler™ Unstained Protein Ladder (Thermo Scientific™, Cat. #26614) was loaded into the first well; the gel ran for 40mins at 200V. Specific molecular weight of proteins in kilodaltons (kDa) shown in red. L: ladder; GFP: GFP-4M8H; +C; positive control; -C: negative control; t: total protein; s: soluble fraction. The white arrows show the strongest bands from the gels; the first arrow in the second lane indicates that the correct size of the total protein sample of the GFP positive control is working properly in

the SDS-PAGE; the second arrow in the third lane illustrates that the total protein sample of the apo-allele is present on the PAGE; the third arrow in the fifth lane displays that the soluble fraction of the GFP positive control is also present on the PAGE. This SDS-PAGE reveals that the apo- and sex-allele is not present in the soluble fraction of proteins in the sample.

Once the protein extracts had been visualized by SDS-PAGE (**Figure 4.8**), a western blot was conducted to confirm the identity of the band. A histidine tag had been attached to both the apo- and the sex-allele and was also present on the GFP-4M8H positive control. The GFP gave a strong ECL signal in lane two from the total protein content and in lane five from the soluble fraction at about 46kDa (**Figure 4.9**, white arrow); this size is larger than the predicted size of 34kDa, but correlates to the signal observed in the SDS-PAGE. The GFP positive control shows a clear large band due to high concentration of protein present at the size. A faint band is also observable in third lane which correlates with the expression of the APOLLO apo-allele (total protein) (**Figure 4.9**, white arrow) near the correct size of 55.5kDa. This band, though faint, does not appear in the total protein content of the sex-allele or in either lanes that contained the alleles in a soluble fraction. Unfortunately, the inherently low specificity of the anti-histidine antibody gives a poor signal-to-noise ratio. The results of the western blot (**Figure 4.9**) correspond with the results of the SDS-PAGE (**Figure 4.8**) that indicate with the ECL signal of the histidine tag to confirm the identity of the apo-allele protein. Although the apo-allele is not present in the soluble fraction, this western blot indicates that the apo-allele is being expressed and therefore, could be extracted through protein purification means as described in **Section 2.3.4.2** *Protein purification systems*.

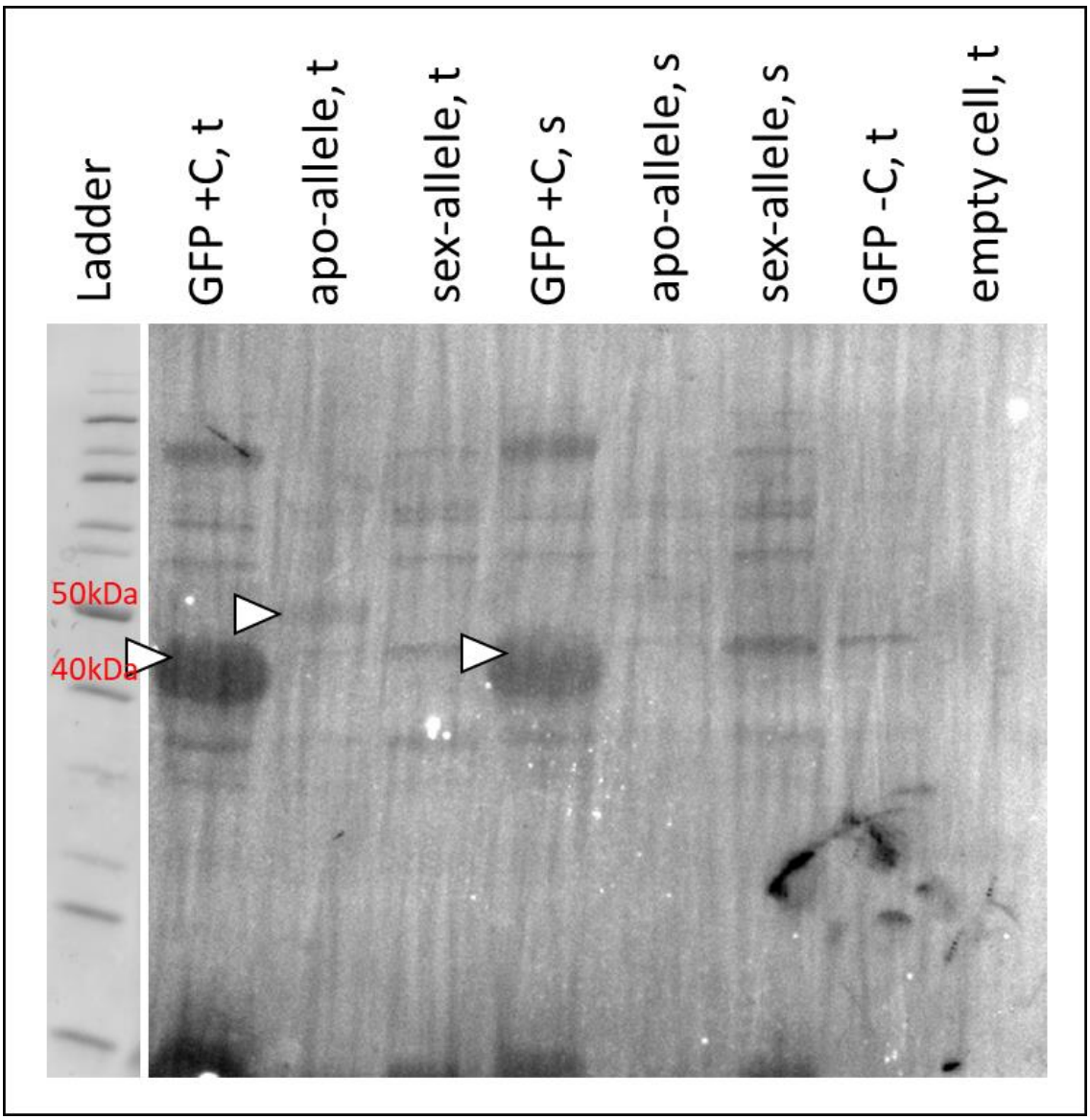


Figure 4.9. Western blot comparing the presence of the APOLLO apo- and sex-allele protein from the total protein content and the soluble fraction. The apo-allele was grown in BL21 (DE3) cells and the sex-allele was grown in C41 cells. The proteins were separated on a Mini-PROTEAN®TGX Stain Free™ Precast Gels by SDS-PAGE. 20µL of each sample was loaded into a well; 5µL of PageRuler™ Unstained Protein Ladder (Thermo Scientific™, Cat. #26614) was loaded into the first well; the gel ran for 40mins at 200V. Transfer at 100V for 1 hr, room temp; detection via Opti-CN colorimetric development, 30 minutes. Specific molecular weight of proteins in kilodaltons (kDa) shown in red. L: ladder; GFP: GFP-4M8H; +C; positive control; -C: negative control; t: total protein; s: soluble fraction. The white arrows display where the histidine-tag fluoresced on the western blot; the first arrow in the second lane indicates that the correct size of GFP was present in the total protein sample; the second arrow in the third lane illustrates that the apo-allele is present in the total protein sample; the third arrow in the fifth lane displays that GFP was also present in the soluble fraction. This Western blot reveals that the apo- and sex-allele is not present in the soluble fraction of proteins in the sample because there was no fluorescence in those lanes which indicate that a histidine-tag is present.

5. Discussion

5.1 Regions of interest

5.1.1 Exonuclease Activity

APOLLO is predicted to be a DEDDh 3′ -> 5′ exonuclease (Corral et al. 2013), and as such may be an enzyme that has the ability to remove nucleotides from a polynucleotide molecule. DEDDh exonucleases have preferences for substrate binding and are essential in the metabolism of DNA and RNA, for example by acting in a 3′ -> 5′ direction to remove nucleotides from the 3′ end of a RNA or DNA strand (Huang et al. 2016). Further, there is the presence of an EXOIII motif within *APOLLO* (Corral et al. 2013), which has been implicated in the removal of mononucleotides from the 3′ termini of double-stranded DNA (Mol et al. 1995). Taken together, these known functions can be used to make inferences with respect to the observations made here.

Thus, it is hypothesized that APOLLO removes nucleotides from the 3′ end of double-stranded DNA. This directionality of APOLLO that is predicted to be 3′ -> 5′ is also consistent with the directionality of its ortholog TREX1 (Brucet et al. 2008). The 3′ end is referred to as the tail end of a DNA strand because nucleic acids are synthesized (via ligation) in a 5′ -> 3′ direction by a polymerase. APOLLO may catalyze in a 3′ -> 5′ direction, the same direction in which transcription takes place to produce mRNA. Therefore, if APOLLO acts on double stranded DNA, which is most commonly found in the nuclei of a cell, this may indicate that APOLLO catalyzes its reaction in the nucleus, before the mRNA is transported out. Further evidence of being located in the nucleus can be found in Brucet et al. (2007) with work on the ortholog TREX1. They determined that the TREX1 protein is found in the cytoplasm. However, when the C-terminus is removed it activates a translocation event that transport TREX1 into the nucleus (Brucet et al. 2007). This ancestral relationship between APOLLO and TREX1 provokes how they may share other similarities as well, such as what APOLLO interacts (i.e. binds) with to catalyze the reaction. In order to perform as an exonuclease APOLLO must have interactions with other molecules, more specifically, it may even contain areas of disorder for the purpose of cleaving nucleotides.

5.1.2 Intrinsic Disorder Activity

The areas of intrinsic disorder that have been identified are intriguing but they are not a guarantee of intrinsic disorder in *APOLLO*. Intrinsic disorder predictors have been used to assist in the prediction of functional properties of proteins such as post-translational modification sites. For example, Iakoucheva et al. (2004) used the intrinsic disorder profile and amino acid frequency

positions to predict sites of phosphorylation by differentiating between phosphorylation and non-phosphorylation sites in eukaryotic cells. They determined that the phosphorylation of a protein generally occurs in areas within intrinsic disorder (Iakoucheva et al. 2004) which is supported by **Figure 4.1**. The determination of a region as intrinsically disordered or ordered is based on the amino acid sequence bias, which has high accuracy (Radivojac et al. 2007). The length of an intrinsically disordered region is dependent on its method of characterization; the longer the region (> 30 residues) the more accurate the prediction is (Peng et al. 2006). Although there are holes in this method to predict intrinsic disordered regions of proteins, the determination of intrinsic disorder (ID) is an important model for research to use when establishing areas of interest in a faster manner (Radivojac et al. 2007).

It is a characteristic of an intrinsically disordered region that preclude structure determination, as a disordered region does not have secondary structure. Protein regions of intrinsic disorder are natively disordered as compared to the neighboring regions of the protein. Unlike ordered protein regions, they do not have α-helix and β-sheet at the secondary structure level (Peng et al. 2006). Therefore, these regions also do not have a fixed tertiary structure (Stender et al. 2015). These regions are highly flexible and may form a local structure when binding to DNA (Stender et al. 2015). Disordered regions are often associated with protein:protein binding, as these functions require flexibility in structure. This is why many intrinsically disordered proteins are over-represented in signaling and transcription functions within cellular regulatory networks (Stender et al. 2015). Radivojac et al. (2007) also indicates that these regions are involved in regulation, signaling and control pathways through catalytic reactions. The function of the protein will determine the protein form; for example, the molten globule (a secondary protein structure) will have a collapsed disordered form and the pre-molten globule (intermediate state of secondary structure) will have an extended disordered form (Peng et al. 2006). The local protein structure that forms does not affect the ability to predict and characterize the disordered protein. Therefore, a profile of the predicted intrinsic disorder was generated.

The *APOLLO* IDP (**Figure 4.1**) that was predicted using PSI-PRED indicated regions of disorder that differ between the apo- and sex-allele. The first region of disorder that can be observed is at the very beginning of the protein (**Figure 4.1.B**). The disorder predicted appears to be almost identical between the alleles. Interestingly, this region occurs before the EXOIII domain. The next

region of disorder is located near the middle of the protein. This center region of disorder is located at about aa position 200 to 300 in the sex-allele (**Figure 4.2**). In the apo-allele this region is smaller, starting at about aa position 210. This observation could be explained by the presence of a smaller putative binding region in the apo-allele. This center region bears two peaks in both alleles, where the potential for disorder is high, but the alleles vary on the shape of the peaks they take. This observation may indicate that the alleles interact with different substrate partners, which is creating the differences in the potential for intrinsic disorder in the binding region. However, the IDP is not the only indicator of a binding region, the PTM profile can also provide insight into substrate differences between the alleles.

5.1.3 Post-translational Modification Activity

A PTM can characterize the change in substrate by the different types of PTMS that can occur. There are three types of PTMs predicted; glycosylation (asparagine), phosphorylation (threonine, serine, tyrosine) and methylation (arginine). A methylation site is often found in a region of flexibility which can allow the site to be methylated once or twice (McBride and Silver 2001). The presence of an methylation site would allow for a number of interactions with DNA, RNA, other proteins and even chromatin structure that can affect the rate of transcription (McBride and Silver 2001). The methylation sites are scattered throughout the protein in both alleles, but it is intriguing to find these PTMs within an exonuclease that needs to be flexible when binding to DNA. However, the most common type of PTM recorded was the phosphorylation sites in both of the alleles (**Figure 4.1**). The presence of a phosphorylation site would enable a change in metabolic reaction of the intended protein (Blom et al. 1999). More recent research demonstrates another proposed mechanism; the presence of a phosphorylation site will downregulate or inhibit exonuclease activity as a repair tool for damage caused to DNA (Kijas et al. 2015; Morafraile et al. 2020). This research was performed in the EXOI sequence segment, which is one of three sequence segments (including EXOIII present in APOLLO) that are distributed throughout the DEDDh exonuclease superfamily (Corral et al. 2013). The high number of phosphorylation sites predicted could indicate that they have changed the function of APOLLO as an exonuclease.

A prokaryotic system of *E. coli* was chosen as the protein expression system based on the methods that were used with TREX1 (Brucet et al. 2008). Some variants of *E. coli* have the potential to carry out basic PTMs and folding but often lack the ability to perform more complex modifications such as phosphorylation and glycosylation (Fernandez and Hoeffler 1999). The PTM-ssMP

prediction tool that was used referred to a training set based on well-known PTM sites in humans. This leaves a large gap for misinterpretation in the prediction model output, as the database uses human PTM and may not be optimized for predicting plant PTMs. Although no PTM sites were predicted, there are also no PTMs sites identified in the homolog TREX1, with which *APOLLO* shares an EXOIII domain. However, recent data showed experimentally that TREXI does undergo phosphorylation and ubiquitination *in vivo* (Kucej et al. 2017). This indicates the potential inaccuracy of the PTM-ssMP prediction tool, and may suggest that the results generated from the other PTM prediction tools may not be accurate.

There are differences and similarities between PTM sites of the sex- and apo-alleles that have a potential to reveal potential sites of interest in APOLLO. Of the predicted sites of PTM (**Figure 4.1.C**) the alleles share the same position of a predicted site and the same value (to the tenths) of 71 out of the 89 sites of predicted PTMs. Of these similarities, the PTM at position 300 has a predicted site of a serine-phosphorylation that was predicted by two of the PTM prediction tools (Plant PTM Viewer and NetPhos 3.1 Server) used. The value predicted by NetPhos 3.1 Server was nearly 1 but the Plant PTM Viewer predicted an exact value of 1. The Plant PTM Viewer additionally confirmed that there was a phosphorylation site at that position based on data from *Arabidopsis*. Interestingly, a serine is present at this 300 aa position, which also occurs at the exact same position in *Arabidopsis*. However, following the serine, where there is a confirmed phosphorylation site, and when comparing *APOLLO* to the *Arabidopsis* line DNA length differences. This similarity of this PTM in these two related species indicates common ancestry, and supports the hybridization hypothesis of Corral et al. (2013).

Despite the alleles having many similar sites of predicted PTMs, there is also a few sites that differ between these alleles that can provide insights into how they are differently expressed. From the remaining 8 other PTMs sites that were predicted to be different between the alleles, there are 6 sites where the alleles share the same aa position but have a different predicted value (to the tenths). Of the 6 sites, only position 262 which occurs in both alleles, where the sex-allele has a predicted serine-phosphorylation but the apo-allele has a predicted threonine-phosphorylation. Finally, there are 11 sites of a predicted PTMs that differ between the alleles in their aa position and type of PTM: sex- S123, S201, T243, T282, R334, N452; and apo- S181, T182, S189, S425, S471. Many of the sites are either scattered around the center of the protein near the intrinsic disorder or near

the end of protein where the C-termini region has different allelic versions due to the 12 polymorphisms. All of these predicted sites of intrinsic disorder or PTMs are not concrete evidence of actual differences between the alleles, but these observations can be backed by the 12 apo-allele specific polymorphisms.

The initial results indicated that the full-length sequence was not expressed in the sex-allele but was expressed in the apo-allele (**Figure 4.5**). There are only 12 amino acid changes between sex- and apo-allele, indicating that they are very similar. Therefore, these amino acids changes may underly critical changes in the functionality, structure or stability of the different APOLLO alleles. For example, a change of amino acid may affect protein synthesis or protein structure**.** An additional cysteine in a protein sequence will oxidize and form a disulfide bond which increases the stability of a protein. Less lysine in a gene sequence indicates that the sequence will be less prone to degradation because lysine promotes the presence of degrading enzymes that could target the protein (Ramachandran et al. 2007). Both of these aa changes occur between the APOLLO alleles; at aa position 11, a cysteine becomes an asparagine in the apo-allele, and at aa position 200, a phenylalanine becomes a lysine in the apo-allele. While the positions of these changes do not correspond with predicted PTM sites (**Figure 4.1.C**), the C11N (**Table 2.1**) change corresponds with the beginning of the EXOIII domain and the F200L change corresponds with the beginning of the first ID region in the sex-allele. It is by observing all of the data as a whole that inferences can be made.

5.1.4 Corresponding Regions

Regions that correspond between the polymorphisms, IDP and the PTM profile can be examined. The first region of interest occurs at the very beginning of the protein. This region has conserved PTMs, there is a methylation site which can indicate flexibility (McBride and Silver 2001), that also occurs within the first peak of ID in both alleles. This indicates that there is conservation of this PTM within the protein.

The next region of interest is that there are three polymorphisms that correspond directly to a particular site of a predicted phosphorylation PTM in the sex-allele, they are aa position 123, 201 and 243. All of these have the potential to cause a change in the regulation of the metabolic pathway (Blom et al. 1999) or inhibit exonuclease activity (Kijas et al. 2015; Morafraile et al.

2020). Incidentally, the polymorphism at aa position 123 occurs within the EXOIII domain, the same EXOIII domain that looks identical in the IDP of both alleles.

The next region of interest is the central region of the protein where there are two peaks of intrinsic disorder and a cluster of PTMs. In this central region there is only one polymorphism at aa position 243, but there are three polymorphisms that correspond with the wider ID region of the sex-allele (**Figure 4.1.A**) at aa positions 195, 200 and 201. This may indicate that the alleles are interacting with different substrates as they are characterized by a different size in binding region. The IDP indicates that the central region of smaller in the apo-allele than the sex-allele (**Figure 4.1.B**). The span of this region corresponds with a cluster of PTMs, which is larger in the sex-allele (**Figure 4.1.C**). The correlation of this region between all the data indicates significance to this central region. The most obvious difference between the two alleles is the length of the central region, where there might be binding sites (**Figure 4.2**) that are associated with phosphorylation. This central region contains a cluster of PTMs which are mainly phosphorylation, which has a different pattern for each allele but does occur in the same region.

The final region of interest in the second peak in the central region of ID. In this region there is a phosphorylation site that was predicted by two PTM prediction tools at aa position 300. This may indicate there is a downstream effect of the ID that can be seen in the predicted PTM sites or that that the polymorphisms may affect the regions of disorder, and thereby change the characteristics of the alleles which can be observed in their protein expression. This correlation between species at aa 300 may also indicate a characteristic that allows for APOLLO to function differently in both species. These correlations between data deserve more attention in the future research of this gene.

Despite the interest of this project being focused on the 12 polymorphisms effects on APOLLO there is also further research being done into the indel polymorphism located in the 5′ UTR. As it is this indel that is also present in both *Arabidopsis* and *Boechera spp.* which is characterized by a plus-strand transcription binding site present only in the apo-allele. It is by the collaboration of this project and the 5′ UTR indel project that APOLLO can be fully understood.

5.2 Fragmenting *APOLLO*

The initial project began with 16 fragments of the *APOLLO* gene that had removed regions of predicted intrinsic disorder that may interfere with crystallization because the best method to determine regions of intrinsic disorder in *APOLLO* is by structure determination of purified

protein. This series of fragments would be crystallized and have structure determined, then reconstituted to determine the overall structure of *APOLLO*. The most commonly used method, x-ray diffraction, would likely be unsuccessful in this case because the inability to diffract x-rays through a crystal with disordered regions can result from the heterogeneous or dynamic structure of the protein, which can cause missing density within the crystal (Oldfield et al. 2013). For these reasons x-ray crystallization is not commonly successful on proteins with ID. However, there has been evidence that by removing these areas of ID, crystallization and diffraction can be successful (Bonecki et al. 2003). Therefore, *APOLLO* was fragmented to remove regions of predicted disorder (**Figure 4.2**). However, there is a lack of information on intrinsic disordered proteins because it is difficult to accurately characterize these regions well through x-ray crystallization.

This method proved to be ineffective when trying to determine protein structure and function of the *APOLLO* gene. All fragments were successfully cloned into the pMCGS7 expression vector, but only six out of the 16 fragments were expressed successfully in an *E. coli* system under the original protocol conditions (**Figure 4.5**), and these results were inconsistent. The two fragments MQ and LQ were able to be expressed in the *E. coli* system for both of the alleles. These fragments both exhibit the Q primer sequence which is found at the C-terminus, the same that when removed activates a translocation event that transports TREX1 into the nucleus (Brucet et al. 2007). This region performed the most consistently out of all of the fragments being that in addition to those two fragments, KQ was also successful but only for the apo-allele. This suggests that the C-terminus region may be critical for protein stability or mRNA stability, thereby making any of the fragments without the end reverse primer unstable. This may also indicate that the fragments containing the EXOIII domain did not work in this protein expression system or that the domain inhibits protein expression.

However, these results may also indicate that there is nothing wrong with the C-terminus in the Q primer but that the problem lies in the N-terminus where the EXOIII domain is located. For example, both allelic versions of KP and LP, which are the only fragments with the P primer region, were unable to be expressed. This may indicate that the presence of the region that the P primer resides in may have a toxic effect on the cells in which the plasmid is produced. These results suggest that the protein expression did not work due to the presence of the central ID region. However, lack of protein expression may have nothing to do with ID if the EXOIII sequence

segment present in in the N-terminus is inhibiting protein expression or exonuclease activity (Kijas et al. 2015; Morafraile et al. 2020).

Despite these interesting results that were not examined until after the results were completed, the regions of disorder might inhibit the fragmenting approach to structural determination by x-ray crystallization. Regions with ID indicate, these areas within the protein are more flexible which would interfere with x-ray crystallization. Therefore, the 16 fragments of *APOLLO* may not have represented that same structure as the full-length protein. Unfortunately, this approach would not be able to produce an accurate model of the APOLLO protein structure, which could have affected the results here. Hence, subsequent experiments eliminated many of the fragment samples and streamlined the protocols using only the full length (KQ) apo- and sex-allele, and provide the first steps to the eventual full characterization of this gene.

5.3 Troubleshooting protein expression

5.3.1 Insolubility

One of the first issues with the protein expression was with solubility. In that, the expressed APOLLO alleles would be present in the total protein samples but would not appear in the soluble protein samples, indicating that the protein alleles were not in a soluble form. This may be due to two different reasons; (1) that APOLLO may have a translocation event occur through the nucleus membrane in which it would need to be hydrophobic and not water soluble (Rawlings 2016), or (2) that a heterologous protein (which can be a common problem) require PTMs as afunctional or structural requirements(Wingfield et al. 2014).

PTMs are essential for correct protein function and without them the protein may mis-fold and aggregate into what are known as “inclusion bodies”. Inclusion bodies may form when recombinant proteins are overexpressed and accumulate within the cytoplasm (Terpe 2006). These aggregates are protected from solubilization or protease degradation (Sarramegna et al. 2003). The formation of inclusion bodies often results in a loss of enzymatic activity and structure (Mattanovich et al. 2012). Thus, purification is not efficient with insoluble proteins. In order to extract these proteins by purification in their soluble form would require the use of strong denaturants such as guanidine·HCl, acetic acid and urea to resolubilize them (Wingfield et al. 2014). Although these substances are good at making protein soluble the problems begin when the proteins need to re-fold. The re-folding process can result in the formation of disulfide-bonds,

protein deamination and reduced protein folding (Wingfield et al. 2014), that will not give accurate results of the protein structure and function. Therefore, methods to increase the solubility of the allele proteins were implemented.

The most recommended technique to improve soluble protein content was to slow down the reaction of protein induction. This included methods such as using an autoinduction media instead of inducing expression with IPTG, reducing the expression temperature (Womack 2011) and adding a low concentration of ethanol, as tested in **Figure 4.6**. Although these methods did reduce the time to induce protein production and illustrate the ideal conditions for each of the alleles, it did not result in more available soluble protein and may not be an effective method to troubleshoot this experiment.

5.3.2 Optimization

The apo-allele was more reliable in its development than the sex-allele throughout this project. This trend was observed throughout the cloning and expression protocols. This may indicate that there are some aspects that characteristically change between the two alleles. For example, the apo-allele expressed high levels of protein content in the BL21 (DE3) despite those cells being not ideal for protein expression. Conversely, the sex-allele did not produce any protein in that cell expression system, it produced higher levels of protein in the C41 cells that can tolerate toxic protein better than BL21. This relationship was observed throughout the results because there might have been a cell toxicity or specificity that prevented consistent expression.

Heterologous proteins expressed in bacterial systems is limited by the toxicity of the overexpressed protein to the bacterial host cell. Toxicity may limit the production of protein yields by reducing the cell density or by limiting the amount of protein that the cell can produce (Donnelly et al. 2001). The first step in optimizing protein expression was to codon optimize the alleles for the host cell that they would be produced in to insure optimal translation. The *APOLLO* apo and sex nucleotide sequences were codon optimized using the Codon Optimization Tool (https://www.idtdna.com/CodonOpt) that optimized the *Boechera* DNA sequences for *E. coli*. The resulting DNA nucleotide sequences were synthesized by Genscript. Another measure in optimization is to use a protease inhibitor cocktail during processing of the proteins. Proteases are found in most cells and are crucial for the degradation function they serve, but will catalyze proteolysis (Razzaq et al. 2019). Proteolysis is the process of breaking down proteins into small

peptides and is catalyzed by enzymes known as proteases. Under these conditions, proteases will degrade the target proteins. As such, two protease inhibitors were used throughout the project to prevent proteolysis. The final means in which protein expression was optimized was to use a different strain of *E. coli* cells. BL21 (DE3) cells are commonly used for expression but are not recommended for cloning because there can be protein degradation by endonuclease activity. However, BL21 (DE3) cells have a high level of basal T7 expression that may result in "leaky" expression, unstable clones and an intolerance to toxic proteins (Womack 2011). Therefore, the use of the BL21 (DE3) cells needed to be examined and alternatives implemented.

The APOLLO alleles may be toxic to the BL21 (DE3) cells, making these cells not ideal for their protein expression. The first indication that cell toxicity might be affecting the expression results was with the reverse primers 552R A and 552R S (**Table 3.4**), that are nicknamed O. Both allelic version of the O primer was used to express two of the fragments that did not express any protein. Indicating that this section of the *APOLLO* sequence was toxic or not compatible with the *E. coli* cells being used or with *E. coli* in general. Another indication of cell toxicity was that when KQ S was expressed in BL21 it produced almost no protein, and was hence moved into C41 cells. The KQ S plasmid was transformed into both BL21 and C41 cells, but produced more colonies when in the C41 cells. The C41 cells are derived from BL21 cells, but with a mutation to prevent cell death with toxic proteins present. Successful transformation into C41 but not BL21 cells suggested that the APOLLO KQ S protein may be toxic to the BL21 cells. Interestingly, previous work at the John Innes Center that had attempted *APOLLO* transformation in *Arabidopsis* had encountered cell toxicity issues with the apo-allele (Sharbel, personal communication, unpublished results). The presence of toxic protein production is a post-translation explanation for the low protein production of KQ S. However, this may as easily be a pre-translation issue, as the point in the translational process where expression failed cannot be determined from this data alone. Therefore, it may be necessary to look at other protein expression systems such as yeast or a cell-free expression system. A yeast expression system can perform most complex PTMs and provide protein folding pathways to produce correctly folded and fully functional proteins for structure determination (Verma et al. 1998). A cell-free expression system is often used to express a protein that is highly toxic in other cells system and does not contain nucleases and proteases that can lower protein content (Womack 2011). Cell-free systems also avoid production of inclusion

bodies. Other methods such as these could be employed based on the similarities that *APOLLO* shares with genes such as *SDN1* in *Arabidopsis* and *TREX1* in humans (Brucet et al. 2008).

The cells that were used to express these proteins with similarities to APOLLO may be valuable. TREX1 is a proofreading enzyme that is found in humans that shares the EXOIII domain with APOLLO (Brucet et al. 2008). *TREX1* was cloned into the expression vector pETM-10 and its protein was expressed in the *E. coli* Rosetta cells (DE3) (Novagen) (Brucet et al. 2007). Another exonuclease SDN1 is also a part of the family of DEDDh 3′→5′ exonucleases that trims microRNAs from the 3′ end that is found in *Arabidopsis*. *SDN1* was cloned into the expression vector pET28-SMT3 and its protein was expressed in the *E. coli* BL21 (DE3) cells (Chen et al. 2018). Despite BL21 (DE3) cells being used for *SDN1* expression from *Arabidopsis* which is a relative of *Boechera*, our data shows that the BL21 (DE3) cells are not ideal for APOLLO expression. Therefore, the Rosetta (DE3) cells that were used for TREX1 expression from humans, which shares the EXOIII domain with APOLLO may be better suited.

5.3.3 Protein Detection

During the first few trials of Western blotting, a semi-dry transfer system was used but ultimately did not work. Once proteins of the alleles were separated and present on a PAGE, the proteins bands are then transferred to a carrier membrane. A semi dry transfer system was used because they can provide a faster transfer efficiency onto a membrane and the apparatus was available. The semi-dry transfer system, though fast, is also prone to failure because it uses less buffer than a wet transfer system, causing heat to build up to ensure faster protein transfer than the conventional wet transfer (MacPhee 2010; Gibbons 2014). After multiple attempts of using the semi-dry transfer, this method was discarded. Successful blotting depends on the carrier membrane being used and the gel acrylamide percentage (MacPhee 2010). Therefore, the type of membrane or transfer conditions that were being used with the semi dry transfer system may have been incompatible for the APOLLO protein.

The membrane being used to hold the proteins is a large determining factor in the ability to detect said proteins. A polyvinylidene difluoride (PVDF) membrane with a pore size of 0.45µm was used. PVDF has greater protein binding capacity than a nitrocellulose membrane, that offers a higher sensitivity; however this higher sensitivity can cause a higher background noise in the

antibody detection (Kurien et al. 2011; Xu et al. 2019). Both membranes come in a variety of pore sizes that range from 0.1 to 0.45µm, reflecting particle diameter based on the protein molecular weight (Arkhangelsky et al. 2012). A nitrocellulose membrane commonly has a pore size of 0.2µm with a higher affinity and retention of proteins (Mahmood and Yang 2012). *APOLLO* is about 56 kDa in size and can thus fit on a 0.45µm pore size membrane. Another large problem with the Western blot test may not have been the transfer process, but rather the antibodies being used in the immunoblot detection of the His-tags.

Once the proteins have successfully transferred to the immunoblot, then detection can occur using antibodies that bind to the proteins. The two-antibody detection system is usually used in a Western blot assay. The first antibody to be exposed to the proteins present on the membrane is known as the primary antibody. The primary antibody is produced by a different species than the host species that will bind to the antigen (Kramer and Pellenz 2013). There are a number of different types of primary antibodies that can be used for western blotting, including polyclonal, monoclonal and recombinant. In this experiment a polyclonal rabbit anti-6X His-tag® antibody (Abcam, ab9108) was used as the primary antibody. A polyclonal primary antibody is a collection of antibodies that can recognize multiple epitopes (antibody binding site of an antigen) on an antigen, and this detection system is highly sensitive but may have inconsistent detection (ThermoFisher Scientific 2020b). This may be a contributing factor for the earlier results that did not have any protein detection. The second antibody to be used in the detection system is known as the secondary antibody; the secondary antibody must be produced from a different species than that of the protein and the primary antibody in order not to produce false positive results. This assay utilized a polyclonal goat anti-rabbit IgG H&L (HRP) secondary antibody (Abcam, ab205718). HRP allows for proteins to be detection at very low concentrations by having the secondary antibody conjugate to HRP (Mahmood and Yang 2012). Proteins are then detected by the naked eye or light camera on the membrane after the substrate is cleaved by HRP due to the colorimetric signal of the target protein.

5.3.4 Alternate approaches

Since *APOLLO* is a newly-discovered gene there is little research and development into its expression, therefore the starting point of this project was to use an *E. coli* expression system. There has been no previously published research into the expression and purification of the *APOLLO* gene from *Boechera* species. Therefore, the project began with an *E. coli* expression

system because it the most commonly used organism for protein production (Terpe 2006). However, the *E. coli* expression system lacks the ability to process most PTM of proteins and does not have the capacity to withstand expression of a toxic protein. An alternative expression system may need to be considered that would be better suited to deal with APOLLO. For example, a cell-free expression system is often used to express toxic proteins as it is free of membranes which contribute to inclusion body formation, and contaminates such as nucleases and proteases (Womack 2011) that interfere with protein production. Another system is the yeast expression system that is often employed because it can perform most complex PTM but is still fast, affordable and easy to work with. These alternate expression systems are more difficult to work with than an *E. coli* system but may be able to provide the necessary requirements for successful protein expression of APOLLO.

5.3.4.1 Yeast Expression System

There was some work into using yeast as an expression system for *APOLLO*, but ultimately ran out of time to complete with the arrival of the COVID pandemic. Although this section of work was executed, it is incomplete in the approach presented here.

A yeast expression system was chosen for APOLLO expression. *Pichia pastoris* was chosen as the eukaryotic expression system, as it is easier to manipulate, faster, less expensive and has a higher heterologous protein expression level than the more commonly known *Saccharomyces cerevisiae*) (Invitrogen, Cat. #K1740-01). The total length (KQ) APOLLO alleles were cloned into pPICZ α C using ligation cloning. A restriction digest PCR was then used with FastDigest (Thermo Scientific, Lot #00443221), which resulted in the attachment of restriction enzymes NotI (Thermo Scientific, Lot #00438572) and ClaI (BSU151) (Thermo Scientific, Lot #00639480) to the ends of the vector (pPICZ α C) and inserts (APOLLO apo-allele and sex-allele). A T4 DNA polymerase and 10X T4 DNA ligase was used in the ligation between vector and insert for both the apo- and sex-allele. The reaction was transformed into DH5α cells by incubation on ice and heat shock. The product was then plated onto Low Salt LB media with Zeocin (Invitrogen, Lot #1771446). Many of the plates did not produce colonies, and those that did appear were negative for plasmid by colony PCR. Despite extensive efforts into troubleshooting this ligation method, it was ultimately unsuccessful. In-Fusion Cloning, an alternative cloning method was attempted, but was also unsuccessful. At this stage, it was determined that the work required to develop APOLLO protein expression in a yeast system was beyond the scope of this project.

5.3.5 Data summary

The data produced from these experiments is not enough to answer the null hypothesis that were initially proposed. The data is inconclusive in determining if the complications with the expression system is a pre-translational or post-translational problem.

The first null hypothesis proposed that the APOLLO alleles can be expressed and purified in a bacterial system. In this experiment, we were able to express both of the alleles in the bacterial system but only the apo-allele was confirmed by a Western blot, and the system could not generate soluble protein. Since it was impossible to determine at what stage of the transcription process is failing, the null hypothesis cannot be rejected. However, inferences can be made for possible solutions. If the problem is pre-translational this could indicate that there is either no transcription occurring because the plasmid if faulty or the cells are not compatible, and therefore no translation is occurring. If the problems is post-translational this may indicate that the protein is degrading quickly or that the protein is inaccessible because it aggregates into an insoluble fraction. The fact that the sex-allele could not be expressed and that the apo-allele could not be expressed in a soluble form indicates that each allele is suffering from a different complication in the expression process at pre- or post-translation.

The second null hypothesis proposed that the 12 polymorphisms present in the coding sequence of the apo-allele affects the structure and/or function of *APOLLO* in apomictic individuals. In this experiment, structure was unable to be determined because a purified protein of the alleles was not obtained. Therefore, this null hypothesis cannot be rejected in that the answer has yet to be determined. However, from the analytical genomics that were completed a set of more specific parameters and hypothesis can be presented and tested.

5.4 Future implications for apomixis

Although apomixis is a hot topic in biotechnology, there has not been much development of *APOLLO* that has been published since 2015. Corral et al. (2013) has been cited in most published works into apomixis development but not the development of *APOLLO*. The most recent paper to describe *APOLLO* has been Kliver et al. (2018) that explored a proposed model of its evolution through the *Boechera* genome and its relatives. Three more copies of *APOLLO* were identified in other Brassicaceae species; these orthologs of *APOLLO* were then searched through its phylogenetic tree (Figure 3 from Kliver et al. 2018) to conclude that diploid apomictic *Boechera*

arose from a duplication event, rather than a polyploidy event. The latest work into the development of *APOLLO* explored the association that heterochromatic chromosomes have with apomixis traits in *Boechera stricta* (Koch 2015; Mandakova et al. 2015). This reduction in published articles in the development of *APOLLO* may suggest that there is work being done and are about to see a surge of information on *APOLLO*.

The most obvious next steps in the development of *APOLLO* from this project is complete the work that was originally laid out for this project. This includes expression in a cell-free or eukaryotic system, followed by protein purification by immobilized metal affinity chromatography (IMAC). The protein purification would isolate the apo- and sex-alleles in their protein form which is an essential role in characterizing the function, interactions that the alleles have with other proteins and confirm the location and types of PTMs present in the protein. The next step would be to determine the structure of the apo- and sex-alleles by nuclear magnetic resonance (NMR) spectroscopy. The determination of protein structure would provide a three-dimensional visualization of the arrangement of atoms that make of the apo- and sex-alleles (Wang et al. 2009). The structure of a protein influences its function due to the conformational changes that exist in molecular biology (Wang et al. 2009). Additionally, knowing the structure would help determine if the alleles bind to other cellular components, and if there could be signaling or catalysis. All of these data would contribute to an understanding of the molecular evolution of these two alleles and potentially determine its origins and influence on both apomixis and *Boechera* genus. The most important information to be obtained from purification and structure determination would be to compare the structure of the alleles from this artificial environment compared to its actual cellular environment, which again would influence structure and thereby function.

Other than the work that was intended for this project, there is also other developmental work that would help in understanding *APOLLO*. This includes using CRISPR-Cas9 gene editing technology, comparative proteomics, and investigating the 5′ UTR polymorphism of *APOLLO*. CRISPR-Cas9 gene editing could be used to knockout one of the sex-alleles at the *APOLLO* locus or to transform one of the alleles to an apo-allele. Both methods have potential to create an apomictic plant from a sexual seed. This method could also determine a causative link between *APOLLO* and apomixis that has not been established before. Another method to establish a correlative link would be to use mass spectrometry to compare the proteins that are present in

ovules at the time meiosis, to determine if either APOLLO allele is present in these tissues and find any correlation present. This method is also known as comparative complex mixture proteomics and is often used for identifying proteins in various conditions by quantitative methods (Booth et al. 2011). Finally, the work being done with *APOLLO*'s 5′ UTR polymorphism will help in learning the expression patterns and effects on alternative splicing. It is hypothesized that the 5′ UTR insertion that is found in the apomictic allele of *APOLLO* has an important role in gene regulation and is therefore critical in the development of the apomeiotic egg cell. Although the 5′ UTR and 3′ UTR do not affect the protein coding sequence, they do influence regulatory aspects in a way that is not fully understood. These characteristic developments are all a part of fully understanding the functional and evolutionary aspects of *APOLLO* and apomixis in plant biology in order to create the "apomixis cassette" that would be used to turn on apomixis in canola.

The ability to create clonal seeds, as is accomplished by apomixis, has been long sought after in the agriculture community. The development of an apomixis trait would have an enormous impact on plant breeding. It would shorten the amount of breeding it takes to maintain a hybrid line and would decrease costs to breeding programs. Although this movement may seem unlikely for large breeding companies to make, as they would be losing out on profits from the F1 hybrid seed they sell every year to producers, the savings benefit and increase in their ability to combat new issues such as a pest or consumer demands would make it impossible for them not to adopt this technology. However, the outcomes of introducing an apomixis trait into a natural or cultivated environment are largely unknown.

Despite the obvious downsides of an apomixis trait in a cultivated crop system, the genetic potential is intriguing. Cultivated crop systems consist of genetically uniform individuals and highly heterozygous plants (Hojsgaard and Horandl 2019). The most likely systems for apomixis to be first introduced into are cereal crops such as wheat and rice of the Poaceae family (Fei et al. 2019). In a cultivated crop system, plant reproductive times are crucial to the impact on crop yield and performance, which are influenced by climatic factors that we are unable to control. Thereby, having a genetically pure crop, as apomictic individuals, would eliminate genetic variability influenced by environmental factors, reducing the climate-impact on crop production (Kukal and Irmak 2018) or potentially exacerbate them. Reducing climate effects on crop production would

increase production, making food more available and affordable. This also gives breeders an edge going forward into a state of climate change where the environmental conditions are unpredictable.

6. Conclusion

The purpose of this project was to identify differences between the apo- and sex-alleles in relation to determining the function of *APOLLO*. The results indicate that a functional expression vector system, pMCSG7 in BL21 (DE3) *E. coli* cells grown in LB media was identified for the apo-allele and pMCSG7 in C41 *E. coli* cells grown in auto-induction media was identified for the sex-allele. Despite the challenges of this project working with the *Boechera* species, that is not a model organism, which made for slow development, differences between the alleles were identified. The predicted intrinsic disorder profile determined that near the middle of the APOLLO protein there is a region of disorder, and that this region appears to be shorter in the apo-allele than the sex-allele. This region is potentially a binding site for interactions with other cellular components. This difference in the size of the center binding region indicates that the APOLLO alleles could be interacting with different cellular components during meiosis of ovule development in apomictic individuals. This is the first indication that *APOLLO* may be interacting with other cellular components and that the alleles may have different cellular functions. This is the first step in determining the function/s of the *APOLLO* alleles in apomictic and sexually reproducing *Boechera* individuals.

Appendix

Reference Tables

Table 3.1. The databases used to predict sites of PTM in the APOLLO apo- and sex-allele. The first column contains the names of each of the websites used in the production of **Figure 4.1**. The second column bears the specific PTMs that each website was designed to find in the sequence when uploaded. The third column consists of the specific parameters that were entered into the program that generated the results seen in **Figure 4.1**. The fourth column contains the exact web link for each database. Finally, the fifth column lists examples of papers that used these databases in their research. Using these databases, the apo- and sex-allele sequence was uploaded into the program in the FASTA format.

Website	PTMs	Parameters	Link	Reference
PTM-ssMP	Ser- Thr- Tyr-phosphorylation Tyr-sulfation Lys-acetylation, ubiquitination, methylation, sumoylation	Threshold value of 0.9 Human	http://bioinformatics.ustc.edu.cn/PTM-ssMP/server/	(Liu et al. 2018)
NetPhos 3.1 Server	Ser- Thr- Tyr-phosphorylation	Threshold value of 0.5 Accuracy of 69% Humans	http://www.cbs.dtu.dk/services/NetPhos/	(Blom et al. 1999; Audagnotto and Peraro 2017)
Plant PTM Viewer	many		https://www.psb.ugent.be/webtools/ptm-viewer/index.php	(Willems et al. 2019)
NetNGlyc 1.0 Server	Asn-glycosylation	Threshold value of 0.5 Accuracy of 76% Humans	http://www.cbs.dtu.dk/services/NetNGlyc/	(Gupta et al. 2004; Audagnotto and Peraro 2017)
PRmePRed	Arg-methylation	Threshold value of 0.5 Accuracy of 84.10% Eukaryotes, *in vivo*	http://bioinfo.icgeb.res.in/PRmePRed/	(Audagnotto and Peraro 2017; Kumar et al. 2017)

Table 3.2. Reaction mixture for Phusion High Fidelity DNA Polymerase PCR. A master mix was prepared where appropriate. This reaction mixture was amplified to create a large enough mixture for every set of PCRs that were tested. The second column indicate the amounts used for the inserts and the third column indicates the amounts used for the vector.

Components	1x (µl) for inserts	1x (µl) for vector
5x buffer (HF/GC)	10	4
Forward primer: variable	2.5	1
Reverse primer: variable	2.5	1
dNTPS	1	0.4
Phusion polymerase ($^{1}/_{10}$)	5	0.2
DMSO	1.5	0.6
H_2O	25	11.8
Template: insert/vector	2.5	1

	(50µl per insert)	(20µl per sample)

Table 3.3. Primers used for the inserts and vector that were targeted by the primers described in Table 3.4. The first column are the nicknames of each of the targets that are described fully in **Table 3.4. Primers used for fragmenting and sequencing.** and **Figure 4.2**. The second and third column are the individual nicknames of each of the primers used to create the inserts and vector.

Target/Insert	Forward primer	Reverse primer	Length (bp)	Length (aa)	Length (kDa)
KQ_A	K	Q_A	1470	490	54.39
KQ_S	K	Q_S	1470	490	54.39
KP_A	K	P_A	588	196	21.756
KP_S	K	P_S	588	196	21.756
KO_A	K	O_A	546	182	20.202
KO_S	K	O_S	546	182	20.202
LQ_A	L	Q_A	1383	461	51.171
LQ_S	L	Q_S	1383	461	51.171
LP_A	L	P_A	501	167	18.537
LP_S	L	P_S	501	167	18.537
LO_A	L	O_A	459	153	16.983
LO_S	L	O_S	459	153	16.983
MQ_A	M	Q_A	672	224	24.864
MQ_S	M	Q_S	672	224	24.864
NQ_A	N	Q_A	543	181	20.091
NQ_S	N	Q_S	543	181	20.091
pMCSG7	T7	T7	5286	1762	195.582

Table 3.4. Primers used for fragmenting and sequencing. The APOLLO primers that were sequenced for the six inserts that were a part of this project. All of the primers were measured in a scale of 100nmole. The purification of all the inserts was desalted. The lower-case letters indicate where the sticky end is created in the primers during ligation independent incorporation between the inserts and the vector, pMCSG7.

Name of the oligo	Sequence, from 5′ to 3′	NB	Nickname
Apollo 3F A/S	TACTTCCAATCCAATgccGCGAGCACCCTGGGCGGCG AC	39	K
Apollo 93F A/S	TACTTCCAATCCAATgccGCGATCCTGGAGTTTGGTGC GATTC	43	L
Apollo 552R A	TTATCCACTTCCAATgTTACAGCACGGTGCTGCAGTG TTTGATC	44	O (A)
Apollo 552R S	TTATCCACTTCCAATgTTACAGCACGGTCGCGCAATA TTTCACC	44	O (S)
Apollo 594R A	TTATCCACTTCCAATgTTAGCTCATATCGGTCAGAAT GTCCGGA	44	P (A)
Apollo 594R S	TTATCCACTTCCAATgTTAGCTCATATCTTTCAGAAT GTCCGGAAC	46	P (S)

Apollo 804F A/S	TACTTCCAATCCAATgccAGCGACATCACCACCCTGA TTAGCA	43	M
Apollo 933F A/S	TACTTCCAATCCAATgccAACGAAGTGAGCGTTAGCA GCATCC	43	N
Apollo 1476R A	TTATCCACTTCCAATgTTACTTCTTCAGTTGGATAACC AGTTTACG	46	Q (A)
Apollo 1476R S	TTATCCACTTCCAATgTTACTTCTTCAGATGGATAAC CAGTTTACG	46	Q (S)
pMCSG7 F	ATT GGA AGT GGA TAA CGG ATC CGA ATT CGA GC	32	
pMCSG7 R	ATT GGA TTG GAA GTA CAG GTT CTC GGT ACC CA	32	

Table 3.5. Reaction mixture that was used for T4 polymerase digest reactions of the inserts and vector. Change between inserts using dCTP and the vector using dGTP. These reaction mixtures were amplified to create a large enough mixture for every set of reactions. *The template amount is variable depending on the concentration of the insert. **The water amount is dependent on the amount of template added to the sample and added to equal 20µl in each sample. NEBuffer 2 (New England Biolabs, Cat. #B7002S), DTT (Fisher Scientific, Cat. #BP172-5), and BSA (New England Biolabs, Cat. #B9000S).

Inserts	Vector	1x (µl)
NEBuffer 2 (5x)	NEBuffer 2 (5x)	4
DTT (1M)	DTT (1M)	0.5
BSA (2mg/mL)	BSA (2mg/mL)	0.5
dCTP (10mM)	dGTP (10mM)	1
ddH_2O	ddH_2O	**variable
T4 polymerase 3U/µL	T4 polymerase 3U/µL	1
Template: variable inserts	Template: pMCSG7 vector	*variable
		(20µl per sample)

Table 3.6. Concentration of vector and inserts used for ligation-independent cloning during the T4 polymerase digest reactions. The amount added to each reaction is variable depending on the concentration obtained from the Phusion Polymerase reactions with purification.

	Vector (µl)	Insert (µl)
KQ (A/S)	3	0.5
KP (A/S)	5	0.4
LQ (A)	3	0.9
LQ (S)	3	0.5
KO (A/S)	5	0.4
LP (A/S)	5	0.4
MQ (A)	5	0.7
MQ (S)	5	0.4

Table 3.7. The lysis buffer recipe used for re-suspending the *E. coli* cells. This specific buffer is used to break open the cells in order to obtain the induced protein inside. *These ingredients are added to the buffer mixture immediately before use. Pefabloc (Sigma, Cat. #11429868001), lysozyme (Fisher Scientific, Cat. #P189833), protease inhibitor cocktail (Roche, Cat. #4693159001), and DTT (Fisher Scientific, Cat. #BP172-5).

Components	Concentration
Tris-HCl (pH 8)	50mM
NaCl	150mM
TritonX-100	1% (v/v)
EDTA	5mM
*Pefabloc	0.1mM
*Lysozyme	0.5mg/ml
*Protease inhibitor cocktail (Roche)	2 tablet/10ml
*DNase	1µ/ml

Table 3.8. 2-ME Laemmli buffer. The 2-ME Laemmli buffer is used for visualizing and separating proteins in the SDS-PAGE. The Laemmli sample buffer was from Bio-Rad. The buffer at a 5% concentration is stored at room temperature. When the buffer is added the SDS-PAGE samples the final concentration will be 1%.

Components	Concentration	Amount
Tris-HCl, pH 6.8	31.5mM	950µl
Glycerol	10%	
SDS	1%	
Bromophenol blue	0.005%	
2-Mercaptoethanol		50µl

Table 3.9. Coomassie blue staining solutions. The three solutions were used on the SDS-PAGE that were hand-poured to visualize the bands present. The Coomassie blue staining was from ThermoFisher scientific (Cat. # LC6065)

Solution	Components	Concentration
Coomassie Blue Stain	Coomassie Blue R-250	0.1%
	Methanol	40%
	Acetic acid	10%
Fixing Solution	Methanol	50%
	Acetic acid	10%
Destaining Solution	Methanol	40%
	Acetic acid	10%

Table 3.10. 1X Transfer Buffer. The Transfer Buffer is used as the buffer in western blotting for the transfer of proteins from the SDS-PAGE to a membrane. The 10X Tris/Glycine buffer, pH 8.3 was from Bio-Rad. The buffer is stored at a 1X concentration in the fridge at 4°C.

Components	Concentration	Amount
Tris	25mM	100mL
Glycine	192mM	
Methanol	20%	200mL
dH_2O		700mL

Table 3.2. Blocking buffer. The blocking buffer is used to improve the sensitivity of the test by blocking out background noise that would interfere with detection. The 10X TBST recipe is found in Table K. The skim milk powder used was can be found at any grocery store.

Components	Amount
10X TBST	50mL
Skim milk powder	1.25g

Table 3.3. 10X TBST (Tris-Buffered Saline-Tween-20). This solution is used as a wash buffer in western blotting. The solution was adjusted to a pH 7.6 and stored at room temperature.

Components	Concentration	Amount
Tris-HCl	10mM	12.115g
NaCl	150mM	40.03g
Tween-20	0.05%	5mL
dH_2O		400mL

Table 4.4. Codon optimized consensus sequences *E. coli* expression. The sequences were synthesized by Genscript (USA; www.genscript.com) and were kindly generated by Dr. Joanne Ernest (GIFS). The sequences were optimized for the apo-APOLLO and sex-APOLLO fragments in a pUC57 plasmid backbone

	apo-allele	sex-allele
Protein sequence	MASTLGGDERNEIVFFDLETAVPTKSG QPFAILEFGAILVCPMKLVELYSYSTLV RPTDLSLISTLTKRRSGITRDGVLSAPT FSEIADEVYDILHGRIWAGHNIKRFDCV RIRDAFAEIGLPPPEPKATIDSLSLLSQK FGKRAGDMKMASLATYFGLGDQAHRS LDDVRMNLEVIKHCSTVLFLESSVPDIL TDMSWLFPRKSPRTRSNEKSLPNGVR ESPTSSSSSPKTDPSSSSVDATAVKNH PIISLLTECSESDTSSCEDPSDITTLISKL HIGTLTDAADEAKTVRQQGSTDPNAKD ESFLGVNEVSVSSIRASLIPLYRRSLRM ELFHNDTPLHLCWYSLKIRFGISRKYVD HVGRPKMNIVVDIPPDLCKILDASDAAA HNLLIDSSTSSDWRPTVMRKKGFANYP TARLQISSESGTQVYQKEEPLGTNQKL DFSSDNFEKLESALLPGTLVDAFFSVEP YDYK KMVGIRLAARKLVIQLKK	MASTLGGDGRCEIVFFDLETAVPTKSG XQPFAILEFGAILVCPMKLVELYSYSTL VRPTDLSLISTLTKRRSGITRDGVLSAP TFSEIADEVYDILHGRIWAGHNIKRFDC VRIRDAFAGIGLSPPEPKATIDSLSLLSQ KFGKRAGDMKMASLATYFGLGDQAHR SLDDVRMNLEVVKYCATVLFLESSVPD ILKDMSWFSPRKSPRTRSNEKSLPNGV RESPTSSSSSPKTDPSSSSVDATTVKN HPIISLLTECSESDTSSCEDPSDITTLISK LHIGTLRDAADEAKTVRQQGSTDPNAK DESFLGVNEVSVSSIRASLIPLYRGGLR MELFHNDTPLHLCWYSLKIRFGISRKYV DHVGRPKMNIVVDIPPDLCKILDASDAA AHNLLIDSSTSSDWRPTVMRKEGFANY PTARLQISSESGTQVHQKEEPLGTNQK LDFSSDNFEKLESALLPGTLVDAFFSLE PYDYKKMVGIRLAARKLVIHLKK
Genscript optimized sequence	ATGGCGAGCACCCTGGGCGGCGAC GAGCGTAATGAGATTGTGTTCTTC GACCTGGAGACCGCGGTTCCGACC AAAAGCGGCCAACCGTTCGCGATC CTGGAGTTTGGTGCGATTCTGGTG TGCCCGATGAAGCTGGTTGAACTG TACAGCTATAGCACCCTGGTGCGT CCGACCGACCTGAGCCTGATCAGC ACCCTGACCAAACGTCGTAGCGGT ATTACCCGTGATGGTGTGCTGAGC GCGCCGACCTTCAGCGAGATTGCG	ATGGCGAGCACCCTGGGCGGCGAC GGCCGTTGCGAGATTGTGTTCTTC GACCTGGAGACCGCGGTTCCGACC AAAAGCGGCCAACCGTTCGCGATC CTGGAGTTTGGTGCGATTCTGGTG TGCCCGATGAAGCTGGTTGAACTG TACAGCTATAGCACCCTGGTGCGT CCGACCGACCTGAGCCTGATCAGC ACCCTGACCAAACGTCGTAGCGGT ATTACCCGTGATGGTGTGCTGAGC GCGCCGACCTTCAGCGAGATTGCG

	GACGAAGTTTACGACATCCTGCAC	GACGAAGTTTACGACATCCTGCAC
	GGTCGTATTTGGGCGGGCCACAAC	GGTCGTATTTGGGCGGGCCACAAC
	ATTAAGCGTTTCGACTGCGTTCGT	ATTAAGCGTTTCGACTGCGTTCGT
	ATCCGTGATGCGTTTGCGGAGATT	ATCCGTGATGCGTTTGCGGGCATT
	GGTCTGCCGCCGCCGGAACCGAAG	GGTCTGAGCCCGCCGGAACCGAAG
	GCGACCATCGACAGCCTGAGCCTG	GCGACCATCGACAGCCTGAGCCTG
	CTGAGCCAGAAGTTCGGTAAACGT	CTGAGCCAGAAGTTCGGTAAACGT
	GCGGGCGACATGAAAATGGCGAGC	GCGGGCGACATGAAAATGGCGAGC
	CTGGCGACCTATTTTGGTCTGGGC	CTGGCGACCTATTTTGGTCTGGGC
	GATCAAGCGCACCGTAGCCTGGAC	GATCAAGCGCACCGTAGCCTGGAC
	GATGTGCGTATGAACCTGGAAGTG	GATGTGCGTATGAACCTGGAAGTG
	ATCAAACACTGCAGCACCGTGCTG	GTGAAATATTGCGCGACCGTGCTG
	TTCCTGGAAAGCAGCGTTCCGGAC	TTCCTGGAAAGCAGCGTTCCGGAC
	ATTCTGACCGATATGAGCTGGCTG	ATTCTGAAAGATATGAGCTGGTTT
	TTTCCGCGTAAGAGCCCGCGTACC	AGCCCGCGTAAGAGCCCGCGTACC
	CGTAGCAACGAGAAAAGCCTGCCG	CGTAGCAACGAGAAAAGCCTGCCG
	AACGGTGTGCGTGAAAGCCCGACC	AACGGTGTGCGTGAAAGCCCGACC
	AGCAGCAGCAGCAGCCCGAAGACC	AGCAGCAGCAGCAGCCCGAAGACC
	GACCCGAGCAGCAGCAGCGTGGAC	GACCCGAGCAGCAGCAGCGTGGAC
	GCGACCGCGGTTAAAAACCACCCG	GCGACCACCGTTAAAAACCACCCG
	ATCATTAGCCTGCTGACCGAGTGC	ATCATTAGCCTGCTGACCGAGTGC
	AGCGAAAGCGACACCAGCAGCTGC	AGCGAAAGCGACACCAGCAGCTGC
	GAGGACCCGAGCGACATCACCACC	GAGGACCCGAGCGACATCACCACC
	CTGATTAGCAAGCTGCACATTGGC	CTGATTAGCAAGCTGCACATTGGC
	ACCCTGACCGATGCGGCGGATGAA	ACCCTGCGTGATGCGGCGGATGAA
	GCGAAGACCGTGCGTCAGCAAGGT	GCGAAGACCGTGCGTCAGCAAGGT
	AGCACCGACCCGAACGCGAAAGAT	AGCACCGACCCGAACGCGAAAGAT
	GAGAGCTTCCTGGGCGTTAACGAA	GAGAGCTTCCTGGGCGTTAACGAA
	GTGAGCGTTAGCAGCATCCGTGCG	GTGAGCGTTAGCAGCATCCGTGCG
	AGCCTGATTCCGCTGTACCGTCGT	AGCCTGATTCCGCTGTACCGTGGC
	AGCCTGCGTATGGAGCTGTTCCAC	GGCCTGCGTATGGAGCTGTTCCAC
	AACGACACCCCGCTGCACCTGTGC	AACGACACCCCGCTGCACCTGTGC
	TGGTACAGCCTGAAGATCCGTTTT	TGGTACAGCCTGAAGATCCGTTTT
	GGTATTAGCCGTAAATATGTGGAT	GGTATTAGCCGTAAATATGTGGAT
	CACGTTGGCCGTCCGAAGATGAAC	CACGTTGGCCGTCCGAAGATGAAC
	ATCGTGGTTGACATTCCGCCGGAT	ATCGTGGTTGACATTCCGCCGGAT
	CTGTGCAAAATCCTGGACGCGAGC	CTGTGCAAAATCCTGGACGCGAGC
	GATGCTGCGGCGCACAACCTGCTG	GATGCTGCGGCGCACAACCTGCTG
	ATTGACAGCAGCACCAGCAGCGAT	ATTGACAGCAGCACCAGCAGCGAT
	TGGCGTCCGACCGTGATGCGTAAG	TGGCGTCCGACCGTGATGCGTAAG
	AAAGGTTTTGCGAACTACCCGACC	GAAGGTTTTGCGAACTACCCGACC
	GCGCGTCTGCAGATCAGCAGCGAG	GCGCGTCTGCAGATCAGCAGCGAG
	AGCGGCACCCAGGTTTATCAAAAA	AGCGGCACCCAGGTTCATCAAAAA
	GAGGAGCCGCTGGGCACCAACCAA	GAGGAGCCGCTGGGCACCAACCAA
	AAACTGGACTTCAGCAGCGATAAC	AAACTGGACTTCAGCAGCGATAAC
	TTTGAGAAACTGGAAAGCGCGCTG	TTTGAGAAACTGGAAAGCGCGCTG
	CTGCCGGGCACCCTGGTGGACGCG	CTGCCGGGCACCCTGGTGGACGCG
	TTCTTTAGCGTTGAACCGTACGAT	TTCTTTAGCCTGGAACCGTACGAT
	TATAAGAAAATGGTGGGCATCCGT	TATAAGAAAATGGTGGGCATCCGT
	CTGGCGGCGCGTAAACTGGTTATC	CTGGCGGCGCGTAAACTGGTTATC
	CAACTGAAGAAGTAA	CATCTGAAGAAGTAA

References

Abd, N., Boon, A., Tan, C., Ardiyana, N., Othman, R.Y., and Khalid, N. 2020. A new plant expression

system for producing pharmaceutical proteins. Mol. Biotechnol. **62**: 240–251. Springer US. doi:10.1007/s12033-020-00242-2.

Adams, K.L. 2007. Evolution of duplicate gene expression in polyploidy and hybrid plants. J. Hered. **98**: 136–141.

Aguilera, P.M., Galdeano, F., Quarin, C.L., Amelio Ortiz, J.P., and Espinoza, F. 2015. Inheritance of aposporous apomixis in interspecific hybrids derived from sexual *Paspalum plicatulum* and apomictic *Paspalum guenoarum*. Crop Sci. **55**: 1947–1956. doi:10.2135/cropsci2014.11.0770.

Aliyu, O.M., Seifert, M., Fuchs, J., and Sharbel, T.F. 2013. Copy number variation in transcriptionally active regions of sexual and apomictic *Boechera* demonstrates independently derived apomictic lineages. Plant Cell **25**: 3808–3823. doi:10.1105/tpc.113.113860.

Arkhangelsky, E., Duek, A., and Gitis, V. 2012. Maximal pore size in UF membranes. J. Memb. Sci. **394**–**395**: 89–97. doi:10.1016/j.memsci.2011.12.031.

Aslanidis, C., and de Jong, P.J. 1990. Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res. **18**: 6069–6074. doi:10.1093/nar/18.20.6069.

Audagnotto, M., and Peraro, M.D. 2017. Protein post-translational modifications : *in silico* prediction tools and molecular modeling. Comput. Struct. Biotechnol. J. **15**: 307–319. doi:10.1016/j.csbj.2017.03.004.

Barcaccia, G., and Albertini, E. 2013. Apomixis in plant reproduction : a novel perspective on an old dilemma. Plant Reprod. **26**: 159–179. doi:10.1007/s00497-013-0222-y.

Barke, H., Daubert, M., and Horandl, E. 2018. Establishment of apomixis indiploid F2 hybrids and inheritance of apospory from F1 to F2 hybrids of the *Ranunculus auricomus* complex. Front. Plant Sci. **9**: 1111. doi:/10.3389/fpls.2018.01111.

Beck, J.B., Alexander, P.J., Allphin, L., Al-Shehbaz, I.A., Rushworth, C., Bailey, C.D., and Windham, M.D. 2012. Does hybridization drive the transition to asexuality in diploid *Boechera*? Evolution (N. Y). **66**: 985–995. doi:10.1111/j.1558-5646.2011.01507.x.

Bicknell, R., and Koltunow, A. 2004. Understanding apomixis : recent advances and remaining conundrums. Plant Cell Online **16**: 228–246. doi:10.1105/tpc.017921.Apomixis.

Blom, N., Gammeltoft, S., and Brunak, S. 1999. Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. J. Mol. Biol. **294**: 1351–1362. [Online] Available: http://www.cbs.dtu.dk/services/NetPhos-3.1/.

Bonecki, M., Rotkiewicz, P., Skolnick, J., and Kolinski, A. 2003. Protein fragment reconstruction using various modeling techniques. J. Comput. Aided. Mol. Des. **17**: 725–738.

Booth, J.G., Eilertson, K.E., Olinares, P.D.B., and Yu, H. 2011. A bayesian mixture model for comparative spectral count data in shotgun proteomics. Mol. Cell. Proteomics **10**: M110.007203. doi:10.1074/mcp.M110.007203.

Boutilier, K., Offringa, R., Sharma, V.K., Kieft, H., Ouellet, T., Zhang, L., Hattori, J., Liu, C., van Lammeren, A., Miki, B., Custers, J., and van Lookeren Campagne, M. 2002. Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. Plant Cell **14**: 1737–1749.

Breyer, W.A., and Matthews, B.W. 2000. Structure of *Escherichia coli* exonuclease I suggests how processivity is achieved. Nat. Struct. Biol. **7**: 1125–1128.

Brucet, M., Querol-Audi, J., Bertlik, K., Lloberas, J., Fita, I., and Celada, A. 2008. Structural and biochemical studies of TREX1 inhibition by metals. Identification of a new active histidine conserved in DEDDh exonucleases. Protein Sci **17**: 2059–2069. doi:10.1110/ps.036426.108.

Brucet, M., Querol-Audí, J., Serra, M., Ramirez-Espain, X., Bertlik, K., Ruiz, L., Lloberas, J., Macias, M.J., Fita, I., and Celada, A. 2007. Structure of the dimeric exonuclease TREX1 in complex with DNA displays a proline-rich binding site for WW domains. J. Biol. Chem. **282**: 14547–14557. doi:10.1074/jbc.M700236200.

Carlson, E.D., Gan, R., Hodgman, C.E., and Jewett, M.C. 2012. Cell-free protein synthesis : applications come of age. Biotechnol. Adv. **30**: 1185–1194. Elsevier Inc. doi:10.1016/j.biotechadv.2011.09.016.

Carman, J.G. 1997. Asynchronous expression of duplicate genes in angiosperms may cause apomixis, bispory, tetraspory and polyembryony. Biol. J. Linn. Soc. **61**: 51–94.

Caston, J.R. 2013. Conventional electron microscopy, cryo-electron microscopy and cryo-electron tomography of viruses. *Edited By*M. Mateu Structure. Springer, Dordrecht. doi:https://doi.org/10.1007/978-94-007-6552-8_3.

Chen, J., Lui, L., You, C., Gu, J., Ruan, W., Zhang, L., Gan, J., Cao, C., Huang, Y., Chen, X., and Ma, J. 2018. Structural and biochemical insights into small RNA 3′ end trimming by *Arabidopsis* SDN1. Nat. Commun. **9**: 3585. doi:10.1038/s41467-018-05942-7.

Conner, J.A., Goel, S., Gunawan, G., Cordonnier-Pratt, M., Johnson, V., Liang, C., Wang, H., Pratt, L., Mullet, J., and DeBarry, J. 2008. Sequence analysis of bacterial artificial chromosome clones from the apospory-specific genomic region of *Pennisetum* and *Cenchrus.* Plant Physiol **147**: 1396–1411.

Conner, J.A., and Ozias-Akins, P. 2017. Apomixis : engineering the ability to harness hybrid vigor in crop plants. Pages 17–34 *in* A. Schmidt, ed. Plant Germline Development: Methods and Protocols. Humana Press, Heidelberg Germany. doi:10.1007/978-1-61779-436-0.

Corral, J.M., Vogel, H., Aliyu, O.M., Hensel, G., Thiel, T., Kumlehn, J., and Sharbel, T.F. 2013. A conserved apomixis-specific polymorphism is correlated with exclusive exonuclease expression in premeiotic ovules of apomictic *Boechera* species. Plant Physiol **163**: 1660–1672. doi:10.1104/pp.113.222430.

Coskun, O. 2016. Separation techniques : chromatography. North. Clin. Istanbul **3**: 156–160. doi:10.14744/nci.2016.32757.

Coughlan, J., Sefanovie, S., and Dickinson, T. 2014. Relative resource allocation to dispersal and competition demonstrates the putative role of hybridity in geographical parthenogenesis. J. Bio **41**: 1603–1613. doi:10.1111/jbi.12316.

d'Erfurth, I., Jolivet, S., Froger, N., Catrice, O., Novatchkova, M., and Raphael, M. 2009. Turning meiosis into mitosis. PLoS Biol. **7**. doi:e1000124.

van Dijk, P.J. 2003. Ecological and evolutionary opportunities of apomixis : insights from *Taraxacum* and *Chondrilla*. Philos. Trans. R. Soc. B Biol. Sci. **358**: 1113–1121. doi:10.1098/rstb.2003.1302.

van Dijk, P.J., and Van Damme, J. 2000. Apomixis technology and the paradox of sex. Trends Plant Sci. **5**: 81–84.

van Dijk, P.J., and Vijverberg, K. 2005. The significance of apomixis in the evolution of the angiosperms : a reappraisal. Pages 101–116 *in* F. Bakker, L. Chatrou, B. Gravendeel, and P. Pelzer, eds. In Plant Species- Level Systematics: New Perspectives on Pattern and Process. Koeltz Scientific Books., Liechtenstein: A.R.G. Gantner Verlag. doi:10.13140/2.1.4567.9042.

Donnelly, M.I., Wilkens Stevens, P., Stols, L., Xiaoyin Su, S., Tollaksen, S., Giometti, C., and Joachimiak, A. 2001. Expression of a highly toxic protein, Bax, in *Escherichia coli* by attachment of a leader peptide derived from the GroES cochaperone. Protein Expr. Purif. **22**: 422–429. doi:10.1006/prep.2001.1442.

Earl, L.A., Falconieri, V., Milne, J.L., and Subramaniam, S. 2017. Cryo-EM : beyond the microscope. Curr. Opin. Struct. Biol. **46**: 71–78. Elsevier Ltd. doi:10.1016/j.sbi.2017.06.002.

FAO 2009. 2050: A third more mouths to feed. [Online] Available: http://www.fao.org/news/story/en/item/35571/icode/ [2018 Feb. 16].

Fei, X., Shi, J., Liu, Y., Niu, J., and Wei, A. 2019. The steps from sexual reproduction to apomixis. Planta **249**: 1715–1730. Springer Berlin Heidelberg. doi:10.1007/s00425-019-03113-6.

Fernandez, J.M., and Hoeffler, J.P. 1999. Introduction : so many possibilities : how to choose a system to achieve your specific goal. Pages 1–5 *in* J.M. Fernandez and J.P. Hoeffler, eds. Gene expression systems. Academic Press, San Diego. [Online] Available: https://www.embl.de/pepcore/pepcore_services/cloning/choice_expression_systems/comparison_expression_systems/index.html.

Fox, B.G., and Blommel, P.G. 2009. Autoinduction of protein expression. Curr. Protoc. Protein Sci. **Apr**: 5.23.1-5.23.18. doi:10.1002/0471140864.ps0523s56.

Gallegos, J.E., and Rose, Alan, B. 2019. An intron-derived motif strongly increases gene expression from transcribed sequences through a splicing independent mechanism in *Arabidopsis thaliana*. Sci. Rep. **9**: 13777.

Gibbons, J. 2014. Western blot : protein transfer overview. North Am. J. Med. Sci. **6**: 158–159.

Giri, D. 2015. High performance liquid chromatography (HPLC) : principle, types, instrumentation and applications. [Online] Available: https://laboratoryinfo.com/hplc/ [2018 Jul. 14].

GRAIN 2001.September. Apomixis : the plant breeder's dream. Seedling **18**. [Online] Available: https://www.grain.org/article/entries/218-apomixis-the-plant-breeder-s-dream.

Gregor, T. 2013. Apomicts in the vegetation of central Europe. Tuexenia **33**: 233–257.

Gupta, R., Jung, E., and Brunak, S. 2004. Prediction of N-glycosylation sites in human proteins. **4**: 203–206. [Online] Available: http://www.cbs.dtu.dk/services/NetNGlyc/.

Gutierrez-Gonzalez, M., Farias, C., Tello, S., Perez-Etcheverry, D., Romero, A., Zuniga, R., Riberiro, Caroline, H., Lorenzo-Ferreiro, C., and Molina, M.C. 2019. Optimization of culture conditions for the expression of three different insoluble proteins in *Escherichia coli*. Sci. Rep. **9**: 16850. doi:10.1038/s41598-019-53200-7.

Hadi, Masood, Z., and Wilson, David, M.I. 2000. Second human protein with homology to the *Escherichia coli* abasic endonuclease exonuclease III. Environ. Mol. Mutagen. **36**: 312–324.

Hage, D.S., Anguizola, J.A., Bi, C., Li, R., Matsuda, R., Papastavros, E., Pfaunmiller, E., Vargas, J., and Zheng, X. 2012. Pharmaceutical and biomedical applications of affinity chromatography : recent trends and developments. J. Pharm. Biomed. Anal. **69**: 93–105. Elsevier B.V. doi:10.1016/j.jpba.2012.01.004.

Hand, M.L., and Koltunow, A.M.G. 2014. The genetic control of apomixis : asexual seed formation. Genetics **197**: 441–450. doi:10.1534/genetics.114.163105.

Hao, J.L., Qiang, S., Chronbock, T., van Kleunen, M., and Liu, Q.Q. 2011. A test of Baker's law :

breeding systems of invasive species of *Asteraceae* in China. Biol. Invasions **13**: 571–580.

Hartmann, M., Štefánek, M., Ák, P.Z.Ř., Man, P.H.E.Ř., Chrtek, J.Ř.I.C.H., and Mráz, P. 2017. The Red Queen hypothesis and geographical parthenogenesis in the alpine hawkweed *Hieracium alpinum* (Asteraceae). Biol. J. Linn. Soc. **122**: 681–696.

Hasan, M.M., and Khatun, M.S. 2018. Prediction of protein post-translational modification sites : an overview. Proteom Bioinform **2**: 049–057. doi:10.29328.

Heng, S., Wei, C., Jing, B., Wan, Z., Wen, J., Yin, B., Ma, C., Tu, J., Fu, T., and Shen, J. 2014. Comparative analysis of mitochondrial genomes between the hau cytoplasmic male sterility (CMS) line and its iso-nuclear maintainer line in *Brassica juncea* to reveal the origin of the CMS-associated. BMC Genomics **15**: 1–12. doi:10.1186/1471-2164-15-322.

Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**: 351–359.

Hojsgaard, D., Greilhuber, J., Pellino, M., Paun, O., Sharbel, T.F., and Horandl, E. 2014a. Emergence of apospory and bypass of meiosis via apomixis after sexual hybridisation and polyploidisation. New Phytol. **204**: 1000–1012. doi:10.1111/nph.12954.

Hojsgaard, D., and Horandl, E. 2019. The rise of apomixis in natural plant populations. Front. Plant Sci. **10**: 358. doi:10.3389/fpls.2019.00358.

Hojsgaard, D., Klatt, S., Baier, R., Carman, J.G., and Hörandl, E. 2014b. Taxonomy and biogeography of apomixis in angiosperms and associated biodiversity characteristics. CRC. Crit. Rev. Plant Sci. **33**: 414–427. doi:10.1080/07352689.2014.898488.

Hojsgaard, D., Martínez, E.J., Acuña, C.A., Quarin, C.L., and Pupilli, F. 2011. A molecular map of the apomixis-control locus in *Paspalum procurrens* and its comparative analysis with other species of *Paspalum*. Theor. Appl. Genet. **123**: 959–971. doi:10.1007/s00122-011-1639-z.

Horandl, E. 2006. The complex casuality of geographical parthenogenesis. New Phytol. **17**: 525–538.

Huang, K.W., Hsu, K.C., Chu, L.Y., Yang, J.M., Yuan, H.S., and Hsiao, Y.Y. 2016. Identification of inhibitors for the DEDDh family of exonucleases and a unique inhibition mechanism by crystal structure analysis of CRN-4 bound with 2-Morpholin-4-ylethanesulfonate (MES). J. Med. Chem. **59**: 8019–8029. doi:10.1021/acs.jmedchem.6b00794.

Iakoucheva, L.M., Radivojac, P., Brown, C.J., O'Connor, T.R., Sikes, J.G., Obradovic, Z., and Dunker, A.K. 2004. The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res. **32**: 1037–1049. doi:10.1093/nar/gkh253.

invitrogen (2010). EasySelect Pichia expression kit : for expression of recombinant proteins using pPICZ and pPICZα in *Pichia pastoris*.

Jha, P., and Kumar, V. 2018. BABY BOOM (BBM) : a candidate transcription factor gene in plant biotechnology. Biotechnol. Lett. **40**: 1467–1475.

Kantama, L., Sharbel, T.F., Schranz, M.E., Mitchell-Olds, T., de Vries, S., and de Jong, H. 2007. Diploid apomicts of the *Boechera holboellii* complex display large-scale chromosome substitutions and aberrant chromosomes. Proc. Natl. Acad. Sci. U. S. A. **104**: 14026–14031. doi:10.1073/pnas.0706647104.

Karunarathne, P., Schedler, M., Martínez, E.J., Honfi, A.I., Novichkova, A., and Hojsgaard, D. 2018. Intraspecific ecological niche divergence and reproductive shifts foster cytotype displacement and

provide ecological opportunity to polyploids. Annu. Rev. Plant Biol. **121**: 1183–1196. doi:10.1093/aob/ mcy004.

Kijas, A.W., Lim, Y.C., Bolderson, E., Cerosaletti, K., Gatei, M., Jakcob, B., Tobias, F., Taucher-Scholz, G., Oakley, G., Concannon, P., Wolvetang, E., Khanna, K.K., Wiesmuller, L., and Lavin, M.F. 2015. ATM-dependent phosphorylation of MRE11 controls extent of resection during homology directed repair by signalling through Exonuclease 1. Nucleic **43**: 8352–8367. doi:/10.1093/nar/gkv754.

Kirchheimer, B., Wessely, J., Gattringer, A., Dietmar Moser, K.H., Schinkel, C., Simmone Klatt, M., Cacciangia, M., Dellinger, A., Guisan, A., Kuttner, M., Luigi Mariorano, J., Nieto-Lugilde, D., Plutzar, C., Svenning, J.-C., Willner, W., Horandle, E., and Dullinger, S. 2018. Reconstructing geographical parthenogenesis : effects of niche differentiation and reproductive mode on Holocene range expansion of an alpine plant. Ecol. Lett. **21**: 392–401. doi:10.1111/ele.12908.

Kliver, S., Rayko, M., Komissarov, A., Bakin, E., Zhernakova, D., Prasad, K., Rushworth, C., Baskar, R., Smetanin, D., Schmutz, J., Rokhsar, D.S., Mitchell-olds, T., and Grossniklaus, U. 2018. Assembly of the *Boechera retrofracta* genome and evolutionary analysis of apomixis-associated genes. Genes (Basel). **9**: 1–16. doi:10.3390/genes9040185.

Knorre, D.G., Kudryashova, N. V., and Godovikova, T.S. 2009. Chemical and functional aspects of posttranslational modification of proteins. Acta Naturae **1**: 29–51. [Online] Available: https://pubmed.ncbi.nlm.nih.gov/22649613/.

Koch, M.A. 2015. A new chromosome was born : comparative chromosome painting in *Boechera*. Trends Plant Sci. **20**: 533–535. doi:10.1016/j.tplants.2015.07.001.

Koltunow, A.M., Bicknell, R.A., and Chaudhury, A.M. 1995. Apomixis : molecular strategies for the generation of genetically identical seeds without fertilization. Plant Physiol **108**: 1345–1352.

Koltunow, A.M., and Grossniklaus, U. 2003. Apomixis : a developmental perspective. Annu. Rev. Plant Biol. **54**: 547–574. doi:10.1146/annurev.arplant.54.110901.160842.

Kramer, D., and Pellenz, S. 2013. Western blotting (immunoblot) : gel electrophoresis for proteins. [Online] Available: https://www.antibodies-online.com/resources/17/1224/western-blotting-immunoblot-gel-electrophoresis-for-proteins/ [2020 Dec. 1].

Kucej, M., Fermaintt, C.S., Yang, K., Irizarry-Caro, R.A., and Yan, N. 2017. Mitotic phosphorylation of TREX1 C terminus disrupts TREX1 regulation of the oligosaccharyltransferase complex. Cell Rep. **18**: 2600–2607. doi:10.1016/j.celrep.2017.02.051.

Kukal, M.S., and Irmak, S. 2018. Climate-driven crop yield and yield variability and climate change impacts on the U.S. great plains agricultural production. Sci. Rep. **8**: 1–18. doi:10.1038/s41598-018-21848-2.

Kumar, P., Joy, J., PAndey, A., and Gupta, D. 2017. PRmePRed : a protein arginine methylation prediction tool. PLoS One **12**: e0183318. doi:10.1371/journal.pone.0183318.

Kurien, B.T., Dorri, Y., Dillon, S., Dsouza, A., and Scofield, R.H. 2011. An overview of western blotting for determining antibody specificities for immunohistochemistry. Methods Mol. Biol. **717**: 55–67.

Liu, Y., Wang, M., Xi, J. Luo, F., and Li, A. 2018. PTM-ssMP : a web server for predicting different types of post-translational modification sites using novel site-specific modification profile. Int. J. Biol. Sci. **14**: 946–956. [Online] Available: http://bioinformatics.ustc.edu.cn/PTM-ssMP/index/.

LMC International 2016. The economic impact of canola on the Canadian economy. Winnipeg, Canada.

[Online] Available: https://www.canolacouncil.org/news/canola-now-worth-$267-billion-to-canadian-economy/.

Lovell, J.T., Aliyu, O.M., Mau, M., Schranz, M.E., Koch, M., Kiefer, C., Song, B.H., Mitchell-Olds, T., and Sharbel, T.F. 2013. On the origin and evolution of apomixis in *Boechera*. Plant Reprod. **26**: 309–315. doi:10.1007/s00497-013-0218-7.

Lovell, J.T., Grogan, K., Sharbel, T.F., and McKay, J.K. 2014. Mating system and environmental variation drive patterns of adaptation in *Boechera spatifolia* (Brassicaceae). Mol. Ecol. **23**: 4486–4497. doi:10.1111/mec.12879.

Lovell, J.T., Williamson, R.J., Wright, S.I., Mckay, J.K., and Sharbel, F. 2017. Mutation accumulation in an asexual relative of *Arabidopsis*. PLOS Genet. **13**: 1–14. doi:10.1371/journal.pgen.1006550.

MacPhee, D.J. 2010. Methodological considerations for improving Western blot analysis. J. Pharmacol Toxicol Methods **61**: 171–177.

Mahmood, T., and Yang, P.-C. 2012. Western Blot: Technique, Theory, and Trouble Shooting. North Am. J. Med. Sci. **4**: 429–434. doi:10.4103/1947-2714.100998.

Mandakova, T., Schranz, M.E., Sharbel, T.F., de Jong, H., and Lysak, M.A. 2015. Karyotype evolution in apomictic *Boechera* and the origin of the aberrant chromosomes. Plant J. **82**: 785–793. doi:10.1111/tpj.12849.

Mattanovich, D., Branduardi, P., Dato, L., Gasser, B., Sauer, M., and Porro, D. 2012. Recombinant protein production in yeasts. Pages 329–358 *in* A. Lorence, ed. Recombinant Gene Expression: Reviews and Protocols, third. Arkansas Biosciences Institute, Jonesboro, Ar, USA. doi:10.1007/978-1-61779-433-9.

Mau, M., Lovell, J.T., Corral, J.M., Kiefer, C., Koch, M.A., Aliyu, O.M., and Sharbel, T.F. 2015. Hybrid apomicts trapped in the ecological niches of their sexual ancestors. Proc. Natl. Acad. Sci. U. S. A. **112**: E2357-65. doi:10.1073/pnas.1423447112.

McBride, A., and Silver, P. 2001. State of the arg : protein methylation at arginine comes of age. Cell **106**: 5–8.

Mol, C.D., Kuo, C.F., Thayer, M.M., Cunningham, R.P., and Tainer, J.A. 1995. Structure and function of the multifunctional DNA-repair enzyme exonuclease III. Nature **374**: 381–386.

Montelione, G.T., Zheng, D., Huang, Y.J., Gunsalus, K.C., and Szyperski, T. 2000. Protein NMR spectroscopy in structural genomics. Nat. Struct. Biol. **7**: 982–985. doi:10.1038/80768.

Morafraile, E.C., Bugallo, A., Carreira, R., Fernandez, M., Martin-Castellanos, C., Blanco, M., and Segurado, M. 2020. Exo1 phosphorylation inhibits exonuclease activity and prevents fork collapse in rad53 mutants independently of the 14-3-3 proteins. Nucleic Acids Res. **48**: 3053–3070. doi:10.1093/nar/gkaa054.

Moser, M.J., Holley, W.R., Chatterjee, A., and Mian, I.S. 1997. The proofreading domain of *Escherichia coli* DNA polymerase I and other DNA and/Or RNA exonuclease domains. Nucleic Acids Res. **25**: 5110–5118.

Muller, H.J. 1964. The relation of recombination to mutational advance. Mutat. Res. - Fundam. Mol. Mech. Mutagen. **1**: 2–9. doi:10.1016/0027-5107(64)90047-8.

Oldfield, C.J., Xue, B., Van, Y.-Y., Ulrich, E.L., Markley, J.L., Dunker, A.K., and Uversky, V. 2013. Utilization of protein intrinsic disorder knowledge in structural proteomics. Biochim Biophys Acta

1834: 487–498.

Otter, D. 2003. PROTEIN : determination and characterization. Academic Press. doi:10.1016/B0-12-227055-X/00980-9.

Ozias-Akins, P. 2006. Apomixis : developmental characteristics and genetics. CRC. Crit. Rev. Plant Sci. **25**: 199–214. doi:10.1080/07352680600563926.

Paule, J., Dunkel, F.G., Schmidt, M., and Gregor, T. 2018. Climatic differentiation in polyploid apomictic *Ranunculus auricomus* complex in Europe. BMC Ecol. **18**: 1–12. BioMed Central. doi:10.1186/s12898-018-0172-1.

Paun, O., Stuessy, T.F., and Horandl, E. 2006. The role of hybridization, polyploidization and glaciation in the origin and evolution of the apomictic *Ranunculus cassubicus* complex. New Phytol. **171**: 223–236. doi:10.1111/j.1469-8137.2006.01738.x.

Peng, K., Radivojac, P., Vucetic, S., Dunker, A.K., and Obradovic, Z. 2006. Length-dependent prediction of protein intrinsic disorder. BMC Bioinformatics **7**.

Pinheiro, A., Pozzobon, M., Do Valle, C., Penteado, M., and Carneiro, A. 2000. Duplication of the chromosome number of diploid *Brachiaria brizantha* plants using colchicine. Plant Cell Rep **19**: 274–278.

Polegri, L., Calderini, O., Arcioni, S., and Pupilli, F. 2010. Specific expression of apomixis-linked alleles revealed by comparative transcriptomic analysis of sexual and apomictic Paspalum simplex Morong flowers. J. Exp. Bot. **61**: 1869–1883.

Pupilli, F., and Barcaccia, G. 2012. Cloning plants by seeds : inheritance models and candidate genes to increase fundamental knowledge for engineering apomixis in sexual crops. J. Biotechnol. **159**: 291–311. Elsevier B.V. doi:10.1016/j.jbiotec.2011.08.028.

Radivojac, P., Iakoucheva, L.M., Oldfield, Chistopher, J., Obradovic, Z., Uversky, V., and Dunker, A.K. 2007. Intrinsic disorder and functional proteomics. Biophys. J. **92**: 1439–1456.

Ramachandran, V., Matzikies, M., Dienemann, A., and Sprenger, F. 2007. Cyclin A degradation employs preferentially used lysines and a cyclin box function other than Cdk1 binding. Cell Cycle **6**: 171–181. doi:10.4161/cc.6.2.3716.

Ramsey, J., and Schemske, D.W. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annu. Rev. Ecol. Syst. **29**: 467–501.

Ramulu, K.S., Sharma, V.K., Naumova, T.N., Dijkhuis, P., and Lookeren Campagne, M.M. 1999. Apomixis for crop improvement. Protoplasma **208**: 196–205. doi:10.1007/BF01279090.

Ravi, M., Marimuthu, M., and Siddiqi, I. 2008. Gamete formation without meiosis in *Arabidopsis*. Nature **451**: 1121–1124.

Rawlings, A.E. 2016. Membrane proteins: always an insoluble problem? Biochem. Soc. Trans. **44**: 790–795. doi:10.1042/BST20160025.

Razzaq, A., Shamsi, S., Ali, A., Ali, Q., Sajjad, M., Malik, A., and Ashraf, M. 2019. Microbial proteases applications. Front. Bioeng. Biotechnol. **June**. doi:10.3389/fbioe.2019.00110.

Redon, R., Ishikawa, S., Fitch, K., Feuk, L., Perry, G., Andrews, T., Fiegler, H., Shapero, M., Carson, A., Chen, W., Cho, E., Dallaire, S., Freeman, J., Gonzalez, J., Gratacos, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J., Marshall, C., Mei, R., Montogomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, J., Zerjal,

T., Zhang, J., Amengol, L., Conrad, D., Estivill, X., Typer-Smith, C., Carter, N., Aburatani, H., Lee, C., Jones, K., Scherer, S., and Hurles, M. 2006. Global variation in copy number in the human genome. Nature **Nov 23**: 444–54.

Rushworth, C.A., Song, B.H., Lee, C.R., and Mitchell-Olds, T. 2011. *Boechera*, a model system for ecological genomics. Mol. Ecol. **20**: 4843–4857. doi:10.1111/j.1365-294X.2011.05340.x.

Sarramegna, V., Talmont, F., Demange, P., and Milon, A. 2003. Heterologous expression of G-protein-coupled receptors : comparison of expression systems from the standpoint of large-scale production and purification. Cell. Mol. Life Sci. **60**: 1529–1546. doi:10.1007/s00018-003-3168-7.

Schinkel, C.C.F., Kirchheimer, B., Dellinger, A.S., Klatt, S., Winkler, M., Dullinger, S., and Horandl, E. 2016. Correlations of polyploidy and apomixis with elevation and associated environmental gradients in an alpine plant. AoB Plants **8**. doi:10.1093/aobpla/plw064.

Schranz, M.E., Dobes, C., A, K.M., and Mitchell-Olds, T. 2005. Sexual reproduction, hybridization, apomixis, and polyploidization in the genus *Boechera* (Brassicaceae). Am. J. Bot. **92**: 1797–1810.

Schranz, M.E., Kantama, L., De Jong, H., and Mitchell-Olds, T. 2006. Asexual reproduction in a close relative of *Arabidopsis* : a genetic investigation of apomixis in *Boechera* (Brassicaceae). New Phytol. **171**: 425–438. doi:10.1111/j.1469-8137.2006.01765.x.

Sharbel, T.F., Voigt, M.L., Corral, J.M., Galla, G., Kumlehn, J., Klukas, C., Schreiber, F., Vogel, H., and Rotter, B. 2010. Apomictic and sexual ovules of *Boechera* display heterochronic global gene expression patterns. Plant Cell **22**: 655–671. doi:10.1105/tpc.109.072223.

Sharbel, T.F., Voigt, M.L., Corral, J.M., Thiel, T., Varshney, A., Kumlehn, J., Vogel, H., and Rotter, B. 2009. Molecular signatures of apomictic and sexual ovules in the *Boechera holboellii* complex. Plant J. **58**: 870–882. doi:10.1111/j.1365-313X.2009.03826.x.

Shaul, O. 2017. How introns enhance gene expression. Int. J. Biochem. Cell Biol. **91**: 145–155. doi:10.1016/j.biocel.2017.06.016.

Shevelev, I. V., and Hübscher, U. 2002. The 3′–5′ exonucleases. Nat. Rev. Mol. Cell Biol. **3**: 364–376. doi:10.1038/nrm804.

Siena, L., Sartor, M., Espinoza, F., Quarin, C.L., and Ortiz, J. 2008. Genetic and embryological evidences of apomixis at the diploid level in *Paspalum rufum* support recurrent auto-polyploidization in the species. Sex. Plant Reprod. **21**: 205–215.

de Silva, U., Choudhury, S., Bailey, S.L., Harvey, S., Perrino, F.W., and Hollis, T. 2007. The crystal structure of TREX1 explains the 3′ nucleotide specificity and reveals a polyproline II helix for protein partnering. J. Biol. Chem. **282**: 10537–10543. doi:10.1074/jbc.M700039200.

Simioni, C., and do Valle, C. 2009. Chromosome duplication in *Brachiaria* (A. Rich.) Stapf allows intraspecific crosses. Crop Breed Appl Biotechnol **9**: 328–333.

Singh, M., Conner, J., Zeng, Y.-L., Hanna, W., Johnson, V., and Ozias-Akins, P. 2010. Characterization of apomictic BC7 and BC8 pearl millet : meiotic chromosome behavior and construction of an ASGR-carrier chromosome-specific library. Crop Sci. **50**: 892–902.

Smyth, M.S., and Martin, J.H. 2000. X ray crystallography. J. Clin. Pathol. Mol. Pathol. **53**: 8–14. doi:10.1136/mp.53.1.8.

Sochor, M., Šarhanová, P., Pfanzelt, S., and Trávníček, B. 2017. Is evolution of apomicts driven by the phylogeography of the sexual ancestor? Insights from European and *Caucasian brambles* (Rubus,

Rosaceae). J. Biogeogr. **44**: 2717–2728. doi:10.1111/jbi.13084.

Spillane, C., Curtis, M.D., and Grossniklaus, U. 2004. Apomixis technology development-virgin births in farmers' fields? Nat. Biotechnol. **22**: 687–691. doi:10.1038/nbt976.

Stender, E.G., O'Shea, C., and Skriver, K. 2015. Subgroup-specific intrinsic disorder profiles of *Arabidopsis* NAC transcription factors : identification of functional hotspots. Plant Signal. Behav. **10**: e1010967.

Stöcklin, J., and Armbruster, G.F.J. 2016. Environmental filtering, not local adaptation of established plants, determines the occurrence of seed- and bulbil-producing *Poa alpina* in a local flora. Basic Appl. Ecol. **17**: 586–595. Gesellschaft für Ökologie. doi:10.1016/j.baae.2016.06.004.

Stols, L., Gu, M., Dieckman, L., Raffen, R., Collart, F.R., and Donnelly, M.I. 2002. A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. Protein Expr. Purif. **25**: 8–15. doi:10.1006/prep.2001.1603.

Studier, F.W. 2005. Protein production by auto-induction in high-density shaking cultures. Protein Expr. Purif. **41**: 207–234.

Taskin, K.M., Ozbilen, A., Sezer, F., Hurkan, K., and Gunes, S. 2017. Structure and expression of DNA methyltransferase genes from apomictic and sexual *Boechera* species. Comput. Biol. Chem. **67**: 15–21. doi:10.1016/j.compbiolchem.2016.12.002.

Terpe, K. 2006. Overview of bacterial expression systems for heterologous protein production : from molecular and biochemical fundamentals to commercial systems. Appl. Microbiol. Biotechnol. **72**: 211–222. doi:10.1007/s00253-006-0465-8.

ThermoFisher Scientific 2020a. Gateway cloning protocols. [Online] Available: https://www.thermofisher.com/ca/en/home/life-science/cloning/gateway-cloning/protocols.html#bp [2020 Sep. 5].

ThermoFisher Scientific 2020b. Antibodies for western blotting. [Online] Available: https://www.thermofisher.com/ca/en/home/life-science/protein-biology/protein-assays-analysis/western-blotting/detect-proteins-western-blot/western-blot-detection-reagents/western-blot-antibodies.html [2020 Nov. 3].

Toenniessea, G.H. 2001. Potential role of apomixis. Pages 6–7 *in* Y. Savidan, J.G. Carmen, and T. Dresselhaus, eds. The Flowering of Apomixis: From Mechanisms to Genetic Engineering. CIMMYT, IRD, European Commission DG VA (FAIR), El Batan, Mexico.

United Nations 2015. World population projected to reach 9.7 billion by 2050. [Online] Available: http://www.un.org/en/development/desa/news/population/2015-report.html [2018 Feb. 16].

Verma, R., Boleti, E., and George, A.J.T. 1998. Antibody engineering : comparison of bacterial, yeast, insect and mammalian expression systems. J. Immunol. Methods **216**: 165–181.

Voigt-Zielinski, M.L., Piwczyński, M., and Sharbel, T.F. 2012. Differential effects of polyploidy and diploidy on fitness of apomictic *Boechera*. Sex. Plant Reprod. **25**: 97–109. doi:10.1007/s00497-012-0181-8.

Vonck, J., and Mills, D.J. 2017. Advances in high-resolution cryo-EM of oligomeric enzymes. Curr. Opin. Struct. Biol. **46**: 48–54. The Authors. doi:10.1016/j.sbi.2017.05.016.

Wang, X., Hunter, A.K., and Mozier, N.M. 2009. Host cell proteins in biologics development : identification, quantitation and risk assessment. Biotechnol. Bioeng. **103**: 446–458.

doi:10.1002/bit.22304.

Wilchek, M., and Chaiken, I. 2000. An overview of affinity chromatography in affinity chromatography : methods and protocols. Humana Press: 1–6.

Willems, P., Horne, A., Van Parys, T., Goormachtig, S., De Smet, I., Botzki, A., Van Breusegem, F., and Gevaert, K. 2019. The Plant PTM Viewer, a central resource for exploring plant protein modifications. Plant J. **99**: 752–762.

Wingfield, P.T., Palmer, I., and Liang, S.M. 2014. Folding and purification of insoluble (inclusion body) proteins from *Escherichia coli*. Curr. Protoc. Protein Sci. **2014**: 6.5.1-6.5.30. doi:10.1002/0471140864.ps0605s78.

Womack, C. 2011. Bypassing common obstacles in protein expression. NEB Expressions. Austin, TX.

Xie, L., Lui, W., Li, Q., Chen, S., Xu, M., Huang, Q., Zeng, J., Zhou, M., and Xie, J. 2015. First succinyl-proteome profiling of extensively drug-resistant mycobacterium tuberculosis revealed involvement of succinylation in cellular physiology. J. Proteome Res. **14**: 107–119. doi:10.1021/pr500859a.

Xu, J., Sun, H., Huang, G., Liu, G., Li, Z., Yang, H., Jin, L., Cui, X., Shi, L., Ma, T., Kameyama, A., and Dong, W. 2019. A fixation method for the optimisation of western blotting. Sci. Rep. **9**: 6649. doi:10.1038/s41598-019-43039-3.

Yin, L., Chen, H., Cao, B., Lei, J., and Chen, G. 2017. Molecular characterization of MYB28 involved in aliphatic glucosinolate biosynthesis in Chinese kale (*Brassica oleracea var. alboglabra Bailey*). Front. Plant Sci. **8**: 1–10. doi:10.3389/fpls.2017.01083.

Yokoyama, S. 2003. Protein expression systems for structural genomics and proteomics. Curr. Opin. Chem. Biol. **7**: 39–43. doi:10.1016/S1367-5931(02)00019-4.

Zayas, J.F. 1997. Solubility of proteins. Pages 6–7 *in* Functionality of Proteins in Food. Springer Berlin Heidelberg. doi:10.1007/978-3-642-59116-7_2.