Molecular Control of Fruit Development in *Pisum sativum*

Mark Alexander Bal

University of East Anglia

John Innes Centre

Submitted to the University of East Anglia in fulfilment of the requirements for the degree of Doctor of Philosophy, September 2024

Experimental Work Statement

All presented experimental work in this thesis is my own work, with the exceptions of **Figure 2.4** and **Figure 2.9**. For **Figure 2.4**, all work was carried out by postdoctoral research fellow Zhe Ji at the University of Oxford.

For **Figure 2.9**, though I myself extracted the tissues and ground them under liquid nitrogen, LCMS measurement of auxin metabolites was conducted in the Laboratory of Growth Regulators, Palacký University Olomouc, Czechia. Measurements were performed by Aleš Pěnčík and Ondrej Novak.

Abstract

This thesis concerns the molecular mechanisms underpinning the development of the fruit (pod) in the crop legume *Pisum sativum* (pea), with a split focus on both hormonal influences on seed-to-pod communication, and genetic variation in a yield-relevant pod trait (pod tip pointedness). Legume research outside of nodulation has been substantially neglected since the advent of plant molecular genetics, but the development of increasingly sophisticated tools for managing genomic data and the urgent, climate-driven needs for increases in sustainable plant protein production are poised to fuel a resurgence in crop legume research. Pea is unusual in that it produces both the most common plant auxin (indole-acetic acid, IAA) but also a second, chlorinated auxin known as 4-Cl-IAA. Here, a previously reported role for this dual-auxin system in governing pea pod development is revisited and expanded upon to elucidate how the evolutionary duplication of small-molecule growth regulators might affect signalling. Additionally, the biosynthesis of the second auxin of pea is explored *in vitro*. Finally, a pea mutation (*acutilegumen*) which affects whether the pod tip is blunt (as in wild-type) or pointed (in the mutant) is mapped to an LTR-retrotransposon, so providing a candidate sequence underlying the *acutilegumen* pod phenotype. This research lays the groundwork for the identification of a pharmaceutically relevant pea enzyme and a crop trait which could have impacts on agricultural yield. This thesis informs our empirical understanding of how the duplication of signalling ligands is shaping eudicot evolution.

Acknowledgements

I express my gratitude to all who have made the following research possible, including Professor Ian Clark and Professor Cathie Martin for their directorship of the EDESIA PhD programme, and to the Wellcome Trust for funding.

I would like to express particular thanks to my primary supervisor, Professor Lars Østergaard, for his assistance throughout the duration of the PhD. His contributions to my understanding of developmental molecular genetics, awareness of available techniques for experimentation, and clerical assistance in the substantial administrative aspect of the degree have proven invaluable, and the research presented here could not have been completed without his expertise and knowledge. I once again congratulate him on his well-deserved appointment to the position of Sherardian Professor of Botany at the University of Oxford.

I would like to also express thanks to my secondary supervisor Dr Cathrina Edwards and her research team, for experience in *in vitro* digestion assays and cell wall purification. I wish to also thank two informal, tertiary supervisors Professor Noel Ellis and Dr Julie Hofer. Noel and Julie's input to the research has been critical and based on longstanding experience in pea genetics, and a recombinant inbred population that they originally produced formed a core part of this research. They proved especially important in framing my understanding of the history of the *ACUTILEGUMEN* gene, as well as aiding me through training in pea crossing. I thank Julie for countless insightful scientific discussions and for her assistance in independently phenotyping a series of recombinant inbred lines for comparison with my own results. I thank Noel especially for his aid in locating markers within the ever-changing permutations of the pea genome. I would like also to thank Dr Andy Chen, with whom I phenotyped a larger RIL population for a genome-wide association study.

I offer sincere thanks to the members of my lab who have assisted and supported me during the course of my PhD: Dr Surbhi Rana, Dr Aaron Ang Chun Hou, Dr Bhavani Natarjan, Nicola Trozzi, William Bezodis and Yurun Jin, who all have assisted me on both technical and conceptual levels at multiple points during my research. I would also like to express special thanks to Neil McKenzie, who was instrumental in nucleic acid purification (particularly genomic DNA) and quality control prior to sequencing, and for his diligent and tireless lab management. I would like to also express thanks to Nicola Stacey for assistance with cloning and phenotyping of a pea diversity panel.

I would like to thank Dr Martin Vickers, Dr Pirita Paajanen, Dr Marielle Vigouroux and Dr Burkhard Steuernagel of the JIC bioinformatics team for training and assistance.

Contents

I. Introduction

I.I. Preface: Legume Crops, Human Health and the Future of our Planet

At the time of writing this thesis, atmospheric carbon dioxide has reached 420 parts per million (ppm), exceeding the internationally designated "safe level" (350ppm)[1]. More people are alive today than at any other point in the history of our species, at over 8 billion. Over the 4-year course of this PhD, plant-poor, highly processed diets will have caused approximately 44 million preventable deaths globally[2]. The 21st-century threats of anthropogenic climate change, food insecurity among an increasing population, and mortality resulting from global adoption of ultra-processed Western-pattern diets are complex, interwoven problems and will require coordinated, interdisciplinary solutions developed by social and natural scientists and effected by forward-thinking policy makers. While there is no single, effortless solution that can simultaneously resolve all of these challenges, cultivation and consumption of legume crops will play an unprecedented role in bringing us into a healthier, more sustainable future[3].

Legumes are a clade of eudicot plants descended from a single common ancestor which (as with many other extant angiosperm families) originated shortly after the K-Pg mass extinction event 65 million years ago[4]. Forming the family Fabaceae, legumes represent the third-largest family of plants[5], and this speciosity is also reflected in the many members of this family which are commonly found in farms and kitchens. Most legumes (including all major food crop legumes) are facultatively symbiotic with diazotrophic (nitrogen-fixing) bacteria; this trait is widespread in the family and is the subject of enthusiastic molecular research[6]. Beyond its innately fascinating nature, this nitrogen-fixing ability strongly differentiates legumes from other plants along both agronomic and nutritional dimensions. With access to nitrogen fixed directly from the atmosphere, legumes can grow in N-depleted substrates, and this makes them an important group for crop rotations; they can capitalise on fields which have been depleted in nutrients from other crops and can fix nitrogen on such impoverished soils for later, non-legume crops (so-called "green manure"). Nutritionally, legumes accumulate unusually high levels of plant protein as compared to other crops, likely due to their nitrogen-fixing symbioses. This protein-richness has made legume domestication a recurring feature of the development of early civilisations on Earth[7].

In antiquity, legumes were widely and routinely consumed as part of the diet due to the ease of cultivation and storage of legume crops as dry seeds[8,9]. Across the ancient world, legume grains were boiled to create countless cultural variations on pea (or lentil, bean,

etc.) soup, ubiquitously served in conjunction with a cultivated monocot cereal (wheat, rice, maize). As technology has advanced, however, we have come to largely neglect legumes as a protein food source and are now paying the price for this oversight in both environmental collapse and compromised health[10].

Though the work described in this thesis is primarily motivated by fascination with the fundamental biology of this most critical clade of plants, the importance of legume research for the future of food in society cannot be overstated. By pushing the limits of what we know about this family further outward, the goals of future food security, human wellbeing, and agronomic sustainability can be pursued.

I.II. Auxin Signalling in Plants – an Overview

Plants form a multicellular kingdom of eukaryotes which, along with animals and fungi, exhibit widespread and highly specific specialisation of cell types and tissues into functional niches across their bodies. As largely sessile, indeterminate organisms, plants have evolved an array of complex developmental programs to coordinate the growth of their different tissues to maximise evolutionary fitness. Unlike in animals where the majority of processes that cause gross morphological change tend to be limited to short bursts of developmental activity (embryogenesis, metamorphosis, puberty), most plants continuously emit new shoot and root system organs throughout their entire lives and exhibit a relatively more uniform distribution of developmental activity. The programs that underpin plant development can essentially be thought of as encoders, dispersers, integrators and interpreters of both endogenous signals such as positional information, age or metabolic status and exogenous signals such as environmental conditions or pest and pathogen attack. Plant developmental programs underpin the common observation that two genetically identical plants can end up looking different to one-another (particularly if grown in different environments) – this phenomenon is known as *phenotypic plasticity*[11].

One important class of developmental coordinators is plant hormones, which are small molecules that are produced and emitted by one cell or tissue and can elicit developmental responses in both near and distant cells. In plants, several such phytohormones have been discovered and extensively studied; none more so than indole-3-acetic acid (IAA), the most naturally abundant representative of a group of related hormones known as auxins[12].

Outwardly, auxin appears to be a near-omni-functional hormone. Virtually all aspects of plant development involve auxin at some level, including both shoot and root tropisms, embryogenesis, organogenesis, lateral root initiation, fruit growth, vascular differentiation, meristem maintenance and others[13] . This seemingly limitless role of auxin is mediated primarily, but not exclusively, by the canonical auxin signal transduction pathway.

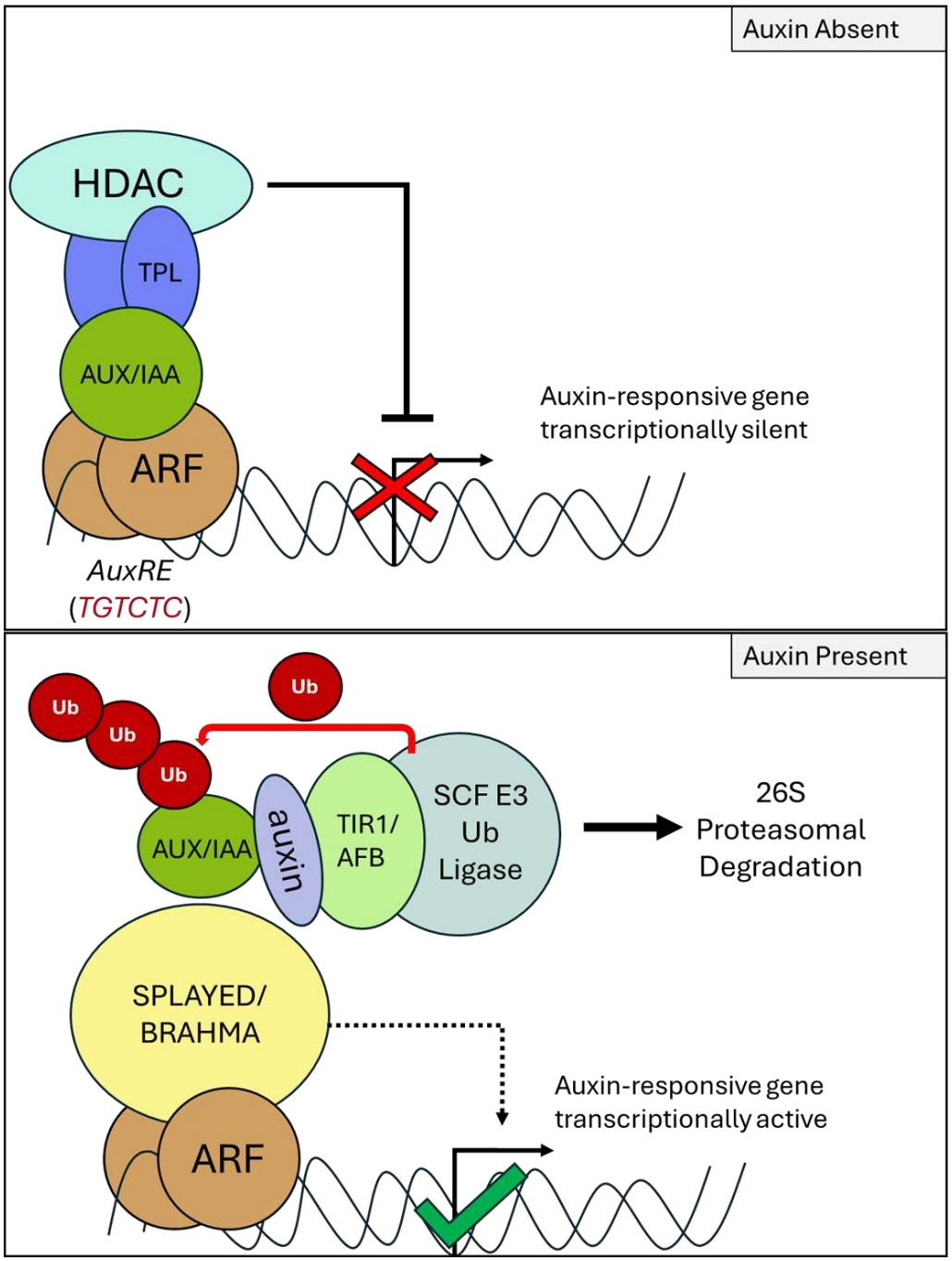


Figure 1.1: Schematic representation of auxin signal transduction pathway. ARF: AUXIN RESPONSE FACTOR, AUX/IAA: AUX/IAA repressor, TPL: TOPLESS, HDAC: Histone Deacetylase enzyme, AuxRE: Auxin-responsive element, TIR1/AFB: Auxin-binding F-Box, Ub: Ubiquitin, SCF E3 Ub Ligase: SCF E3 ubiquitin ligase complex, SPLAYED/BRAHMA: Plant SWI/SNF-like ATP-dependent chromatin remodelling complex.

The canonical auxin signalling pathway is a transcriptional response, identified from protracted research on auxin-insensitive and developmentally aberrant mutants in the model plant *Arabidopsis thaliana* (reviewed in [12]). Genes which respond directly to auxin exhibit a repeated cis-regulatory element in their promoter sequences known as the auxin-responsive element (AuxRE, consensus sequence TGTCTC). These AuxREs are bound by transcriptional regulator proteins (transcription factors) known as auxin response factors (ARFs), which are dimeric DNA-binding proteins. In the absence of auxin, ARFs are bound by transcriptional repressor proteins known as Aux/IAA repressors (Aux/IAAs). Aux/IAAs, in turn, recruit the corepressor protein TOPLESS (TPL) and, through TPL, ATP-dependent chromatin remodelling complex proteins. Together, these proteins form a multimeric inhibitory complex which renders the downstream gene transcriptionally inactive. Auxin functions by promoting interaction between the Aux/IAA repressor and the F-Box component (TRANSPORT INHIBITOR RESISTANT 1/ AUXIN RESPONSE F-BOX (TIR1/AFB)) of the SCF E3 ubiquitin ligase complex, which leads to polyubiquitination of the Aux/IAA repressor, marking it for degradation. The subsequent proteasomal degradation of Aux/IAAs is sufficient to liberate the ARFs to interact with transcription-promoting chromatin remodelling factors, such as SPLAYED and BRAHMA[14] so leading to an increase in gene expression. Additionally, TIR1/AFB auxin receptors have been demonstrated to have adenylate cyclase (cAMP producing) activity when bound with auxin and Aux/IAAs, and the ability of AFBs to generate cAMP is required for auxin-responsive gene expression[15].

The diversity of these signalling components can partly explain how different cells respond to auxin in different ways. Different tissues express different complements of ARFs, Aux/IAAs, TPL-related (TPR) corepressors and TIR1/AFB receptors, which have different dissociation constants and activatory/repressive strengths in relation to gene transcription. Furthermore, ARFs can be sub-categorised into three clades: classes A, B and C. Only class A ARFs are activatory (possessing a glutamine-rich middle region which allows them to interact with activatory factors) and so competition for AuxRE binding between activatory A class ARFs and repressive class B and C ARFs adds an additional layer of complexity to the auxin response[16].

Other, non-canonical mechanisms of transcriptional auxin response have been identified, such as through the highly derived and unusual ARF known as ETTIN (AtARF3). ETTIN, unlike other ARFs, is capable of interacting directly with corepressor protein TPL via an intrinsically disordered, ETTIN-SPECIFIC (ES) domain[17]. This interaction is

disrupted by auxin which, again unusually, is capable of binding directly to ETTIN and promoting TPL dissociation and alteration of gene expression.

More recently, a non-transcriptional auxin response has also been described. AUXIN-BINDING PROTEIN1 (ABP1) is an extracellular auxin receptor which is capable of binding auxin in the apoplast and, when bound, activates the cell membrane TRANSMEMBRANE KINASE 1 (TMK1) to effect an intracellular phosphorelay[18]. Downstream, this leads to altered phosphorylation of a range of target proteins and is thought to trigger highly rapid auxin responses such as membrane depolarisation.

Hence, auxin signalling can produce both transcriptomic and phosphoproteomic changes in receptive cells, and much exciting work remains to be done to further elucidate these pathways in *Arabidopsis* and newer model species such as *Marchantia polymorpha*. It is likely that the multi-layered structure of the auxin signalling network, and in particular tissue-specific expression of signalling protein paralogs, underpins not only intra-organismal diversity in developmental programmes, but may even explain some fraction of the developmental/ morphological diversity observed across plant species.

I.III. Unanswered Questions and Thesis Scope

Connecting the incredible power of molecular genetics to crop science is an ongoing challenge for biology, as the experimental tractability of model plant species is often absent from their larger, slower-growing, and sometimes polyploid counterparts. To this pea is no exception, as its large, 4.5 Gb genome was published only as recently as 2019[19]. Despite its diploid nature, the sheer size of the pea genome makes it bioinformatically challenging to work with and demanding of specialised computing infrastructure for exploration and analysis.

This thesis aims to explore the signalling and biosynthesis of pea auxins in the context of fruit development, and to map the gene responsible for the shape of the pod apex (blunt or pointed), using a range of genetic and biochemical techniques.

II. 4-Cl-IAA Signalling across Fabaceae

II.I. Introduction

Pea (*Pisum sativum*) is unusual among legumes in that it produces both the most common auxin indole-3-acetic-acid (IAA) alongside a secondary, more recently evolved chlorinated auxin, 4-Cl-IAA (**Figure 2.1**). 4-Cl-IAA was originally discovered as its methyl ester conjugate in immature pea seeds[20], and subsequent work that quantified its presence across the plant body has revealed that it is low across all above ground vegetative tissues excepting reproductive structures[21]. This unusual phenomenon, in which a recently evolved hormone "duplicate" accumulates to high levels in the yield-relevant developing reproductive tissues of a major legume crop has motivated the research reported here.

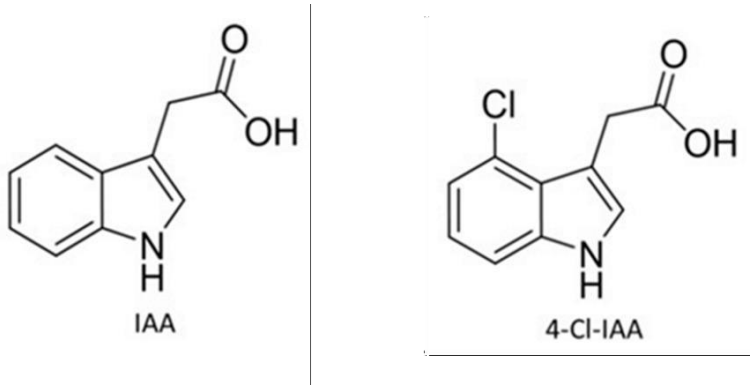


Figure 2.1: IAA and 4-Cl-IAA molecules, two naturally occurring auxins in developing pea fruit and seeds.

Additional work by the same research group also explored whether 4-Cl-IAA was produced in other legumes and discovered that while 4-Cl-IAA could also be detected in *Vicia amurensis*, it was absent from clade Phaseoleae, which began to indicate it may be restricted to close relatives of pea such as *Vicia*[2]. Unexpectedly, however, there were also reports of 4-Cl-IAA from the seeds of a conifer, *Pinus sylvestris*[23]. In 2015, extremely detailed follow-up work by HK Lam and colleagues tested a wide range of legume species and also re-tested several species of *Pinus.* They found 4-Cl-IAA to be restricted to Tribes Fabeae and Trifolieae[24] and to be absent from pines altogether. Specifically, they extended the presence of 4-Cl-IAA to *Medicago truncatula, Melilotus indicus*, and three *Trifolium* species. They also confirmed the presence of 4-Cl-IAA in *Vicia faba* and confirmed the absence of 4-Cl-IAA from *Cicer arietinum*. This provided our currently up-to-date view on the distribution of 4-Cl-IAA among legumes, in which

it is restricted to the Fabeae/Trifoleae (which together form a clade known as the F/T clade) and the absence of 4-Cl-IAA from sister clade Cicerae. However, a more recent preprint has also reported the presence of 4-Cl-IAA and 6-Cl-IAA in the unrelated tropical vine *Anredera cordifolia*[25]. Additionally, endogenous 5-Cl-IAA has been reported in *Rheum* (rhubarb)[26]. Thus, our current understanding of the distribution of 4-Cl-IAA is that it evolved independently in families Fabaceae (pea and relatives) and Basellaceae (*Anredera*) and that 5-Cl-IAA evolved in Polygonaceae (*Rheum*). While all these species are eudicots, their families aren't closely related, which suggests that evolution of halogenated auxins has occurred multiple times across plant evolution. Whether this represents a case of true convergent evolution or a parallelism remains to be determined, but it is presently speculated that all halogenating species evolved halogenation by a similar enzymatic mechanism (see Chapter 3 for a more in-depth discussion). Moreover, the true distribution of halogenated auxins across the plant kingdom remains unknown, though they are likely truly absent from well-studied species in which they have not so far been detected (e.g. *Arabidopsis*, tomato, wheat).

A pioneering historical observation related to this dual-auxin system in pea fruit was that immature pea fruit, when cut along their abaxial suture and left otherwise alone, will continue to grow and develop[27]. By contrast, seed removal causes pod growth arrest and eventual senescence and abscission. Pod growth can be rescued through the exogenous application of 4-Cl-IAA, but not IAA[28,39]. This observation was found to correlate with concentration of exogenously applied auxin, i.e. application of higher concentrations of IAA led to less growth and applications of higher concentrations of 4-Cl-IAA led to more growth[29]. Given both hormones are synthesised endogenously *in planta* but have distinct effects on the development of deseeded fruit, it was hypothesised that IAA and 4-Cl-IAA have diverged in function and may have even become antagonistic in their mechanism of signalling. This divergence in 4-Cl-IAA and IAA signalling will be referred to as the **differential growth response**.

Ozga *et al* were both the original observers of the differential growth response and so far, the only advancers of an explanatory model. According to their findings across multiple papers, IAA and 4-Cl-IAA exert their distinct effects on pod growth primarily by upregulating different sets of target genes (identified by qPCR). These can be broadly categorised into 3 sets of growth-critical differentially expressed genes: auxin receptors, gibberellic acid-related genes and ethylene-related genes.

As described above, auxin is primarily detected within plant cells by TIR1/AFB F-Box proteins which are competent to weakly bind auxin and, when bound to auxin, develop a

high affinity for Aux/IAA transcriptional repressors and mark the latter for proteasomal degradation. Ozga et al[29] have reported that of the three major TIR1/AFBs to be expressed in immature pea fruit (PsTIR1a, PsTIR1b, and PsAFB1) that PsTIR1b is differentially expressed by exogenous application of IAA or 4-Cl-IAA to deseeded pea fruit. Deseeding itself triggers heightened expression of PsTIR1b, and the authors therefore conclude that TIR1b is associated with growth arrest. 4-Cl-IAA, but not IAA, can attenuate this spike in TIR1b transcript abundance, and this is presented as one mechanism for the differential growth response. This model is not without its questions though, as differential expression of TIR1b would itself require some kind of differential perception mechanism of IAA and 4-Cl-IAA; i.e., it is unclear in this explanatory model which protein is binding to IAA and 4-Cl-IAA and effecting this differential gene expression.

Pod growth after fertilisation is extremely fast and is hypothesised (as is typical of other fruits such as *Arabidopsis* siliques) to be mediated primarily by rapid cell expansion, driven by gibberellic acid signalling. Gibberellic acid metabolism has been proposed by Ozga et al to be under differential control between IAA and 4-Cl-IAA. GA can exist in a series of active and inactive forms in plant cells, and interconversion between these forms is a primary mechanism for regulating the abundance of free, active GA. 4-Cl-IAA treatment was shown by qPCR to lower the transcript abundance of GA inactivating enzyme *GA2ox1*[30], but to increase the expression of GA anabolic enzyme *GA20ox1*[30]. As such, it has been hypothesised that 4-Cl-IAA triggers the production of high levels of active GA in the pea fruit and simultaneously inhibits the inactivation of GA, which would explain the rapid growth response of deseeded fruit to 4-Cl-IAA treatment. This proposed model is supported by the observation that treatment of deseeded pea fruit with GA alone can stimulate pod growth[30], and that 4-Cl-IAA/GA co-treatment have a synergistic, positive effect on pod growth[30]. By contrast, IAA/GA combinatorial treatment was reported to nullify GA-mediated pod growth[31]. As with the TIR1b-mediated model, however, this GA-based explanation does not account for the mechanism by which 4-Cl-IAA, but not IAA, increases the expression of genes encoding GA anabolic enzymes.

Another secondary explanation provided by Ozga et al is that 4-Cl-IAA and IAA have different effects on ethylene signalling in developing fruit. In plants, ethylene acts as a mobile signalling ligand which can alter the expression of ethylene-responsive genes[32]. When ethylene is absent, ER-localised transmembrane ethylene receptors (ETR1, ETR2, ERS1, ERS2, and EIN4) constitutively activate the kinase CTR1. CTR1 then phosphorylates EIN2, rendering EIN2 inactive. This prevents EIN2 from inhibiting the

SCF E3 ubiquitin ligase-mediated degradation of transcription factors EIN3 and EIL1, which occurs via binding to the F-box proteins EBF1 and EBF2. When ethylene binds to its ER-localised transmembrane receptors, it prevents them from activating CTR1 and in turn prevents CTR1 from phosphorylating EIN2. This leads to the cleavage of EIN2′s C-terminal domain (EIN2-C). EIN2-C inhibits the ubiquitination of EIN3/EIL1 by increasing the degradation of the mRNAs encoding EBF1 and EBF2, which allows for EIN3/EIL1-mediated gene expression. EIN2-C also upregulates EIN3/EIL1 activity by direct interaction in the nucleus. CTR1 also translocates to the nucleus in response to ethylene, where it stabilises EIN3 by binding to and inhibiting EBF proteins[33]. Ozga et al found that 4-Cl-IAA, but not IAA, increased the transcript abundances of ethylene receptor genes *PsERS1* and *PsETR2* when exogenously applied to deseeded fruit[34]. Conceivably, this could lead to an increase in the abundance of the ethylene receptors at the protein level, which in turn could increase the threshold level of ethylene required to overcome the constitutive activation of CTR1. Furthermore, 4-Cl-IAA treatment increased *PsEBF1* and *PsEBF2* transcript abundance, which may lead to an accelerated degradation of EIN3/EIL1. Hence, it is possible that 4-Cl-IAA partly acts by decreasing the sensitivity of young fruits to ethylene, and this explains why 4-Cl-IAA-stimulated growth is resistant to ethylene produced both from the application of 4-Cl-IAA itself and from exogenous ethephon. IAA appears to lack this ethylene desensitising activity relative to 4-Cl-IAA, thus rendering the pods more susceptible to ethylene-mediated senescence.

Taken together, these results create a model in which 4-Cl-IAA promotes pod elongation through regulating expression of a range of growth-critical genes. As these results have only ever been produced by one lab and only a specific series of genes have been described as being differentially expressed from qPCR experiments, the work reported here began by trying to recapitulate these same results and to obtain a broader view of the mechanism of the 4-Cl-IAA/IAA differential growth response. Additionally, expansion of this work into other legumes beyond the *Pisum* genus was considered to potentially improve the evolutionary insights into this dual-auxin system.

Secondly, a role in regulating auxin biosynthesis has been reported for the phosphorylated disaccharide signalling molecule trehalose-6-phosphate (T6P). T6P is synthesised *in planta* by trehalose phosphate synthase (TPS), and trehalose phosphate phosphatase (TPP) dephosphorylates T6P to produce trehalose. In plants, mutations in TPS have severe, typically embryo-lethal phenotypes, and so Meitzel et al.[35] used a heterologous system to explore the role of T6P in pea seed development. Specifically, they generated transgenic pea lines with an elevated embryonic expression of

either *TPS* or *TPP*, by expression of heterologous *Escherichia coli* genes *otsA* and *otsB*, in conjunction with an embryo-specific promoter *USP*. *proUSP::TPP* pea lines (with reduced levels of T6P) showed up to 70% reductions in 4-Cl-IAA level in embryos, suggesting a possible role for T6P in upregulating 4-Cl-IAA biosynthesis[35]. This motivated an experiment in which deseeded pea fruit were treated with T6P to see if this could potentially rescue pod growth.

II.II. Materials and Methods

1. Plant Materials

All seeds were obtained from the publicly accessible JIC Germplasm Resources Unit (GRU), in the JIC *Pisum* Germplasm sub-collection. Cane-supported pea plants (lines JI2822 and JI3253, both *Pisum sativum*) were grown in glasshouse conditions in 9 cm pots. A 16-hour photoperiod was ensured using supplementary lighting outside of summertime. In autumn and winter, supplemental heating provided the glasshouse with a day/ night temperature of 18 ^{o}C/12 ^{o}C. In summer, side and roof vents and air handling units were used to increase circulation and cool the glasshouse interior. Tarps and internal thermal screens were also used in summer to reduce glare from the sun, but no electrically powered supplemental cooling was used. Seeds were sown at 2 cm depth in institute "Arabidopsis mix", which is a Peat/Loam/Grit mix (65:25:10) supplemented with 3 kg/m^3 limestone. Pots were saturated twice each day by the automated watering system.

2. Deseeding and hormone treatments

To carry out deseeding experiments, pods were measured to a developmental stage of 15-20 mm in length, which corresponds to 2 days after anthesis (DAA) where anthesis is defined as full reflex of petals as is convention in *Pisum*. Pea plants self-pollinate cleistogamously around 24-48 hours before anthesis. An incision was made with a clean scalpel along the abaxial (also known as "dorsal") suture of the immature pod, which is the edge to which the seeds do not attach. This incision was then carefully widened through the use of fine tweezers and, if seeds were to be removed, they were plucked out with the tweezers. Funiculi were left intact inside the pod wherever possible.

Hormone treatments were applied either through daily pipetting (for all numerical data acquisition) or through lanolin (for timelapse photography). For lanolin treatment, lanolin (The Soap Kitchen, South Yorkshire, UK) was melted in an Eppendorf tube using a heat block at 60 ^{o}C, with hormones IAA (Merck Life Science UK Limited, Dorset, UK) and 4-Cl-IAA (Cayman Chemical, Cambridge Bioscience, UK) added to a final concentration of 100 µM from a 10 mM stock (1% v/v auxin/lanolin). For lanolin, hormones stocks were dissolved in DMSO (Merck), and so for mock treatment DMSO was added to 1% v/v in lanolin.

For daily pipetting treatments, a 50ml solution of 0.1% v/v Silwet L-77 (BHGS Ltd, Evesham, UK) was prepared and used either alone for mock treatment, or with 100 µM IAA (Merck) or 100 µM 4-Cl-IAA (Cayman Chemical), both diluted to 1% v/v from 10

mM stocks. Hormone solution was added to the point of pod saturation, 50-200 µl. For kakeimide (KKI) treatment, the same Silwet-based solution was prepared: solutions of 20 µM KKI (ProbeChem, Shanghai, China), 100 µM IAA (Merck) and a combined 20 µM KKI/100 µM IAA solution were prepared by dilution from stocks. KKI was diluted to 20 µM from a 10 mM stock by 0.2% v/v dilution in 0.1% v/v Silwet L-77.

Pods lengths were recorded every day at the same time as the length along their adaxial (uncut) suture – this allowed control of variable pod curvature during deseeding treatments. Pod length decreases during pod senescence/death were not recorded to give a record of cumulative growth.

3. RNA-Sequencing

Immature pea pods of length 15-20 mm were deseeded and treated with either mock solution (0.1 % v/v Silwet, BHGS Ltd), IAA (100 µM IAA + 0.1 % v/v Silwet) or 4-Cl-IAA (100 µM 4-Cl-IAA + 0.1 % v/v Silwet). Pods were snap frozen in liquid nitrogen and ground using a pre-chilled pestle and mortar, tissue collection took place between 10-11AM in all experiments to minimise circadian variations in the pod transcriptome or auxin level. RNA was then extracted from frozen tissue powder using the RNEasy Plant Mini-kit as per manufacturer's instructions, using optional buffer RLC (as opposed to RLT) for extraction to improve RNA concentrations from sugar-rich pea tissues. RNA quantification and quality (260/280 nm and 260/230 nm absorptions) were assessed via nanodrop, with a minimum concentration of 1 µg/µl RNA used for further steps. RNA samples were then sent to Novogene (Cambridge, UK) for RNA-Sequencing to 40x depth enriched for mRNAs.

RNA-sequencing data was downloaded and analysed using the slurm job scheduling system to interface with the Norwich Bioscience Institute's High-Performance Computing Cluster (HPC). Hisat2 was used to map RNA reads to the reference pea genome JI2822 v1.2, and Samtools/Stringtie used to generate tsv files for each biological replicate. These were then uploaded to Degust for DEG analysis and visualisation. Degust employs a pairwise Student's t-test (a linear model) with a Benjamini-Hochenberg correction for multiple testing. Because significant differences in expression are being searched for across a large number of genes, use of a conventional p-value of 0.05 would lead to a 5% false positive rate. This can be reduced (though not eliminated) by using the Benjamini-Hochenberg correction, a transform which converts p-values to q-values which are more stringent and less likely to yield false positives. Degust automatically deploys this correction.

4. Auxin Metabolite Quantification preparation

Auxin metabolites were quantified with collaborators at the Laboratory of Growth Regulators, Palacký University Olomouc, Czechia. Pod tissues were prepared at JIC through snap freezing of immature JI3253 pods at stages 2 days after anthesis, 5 days after anthesis and 7 days after anthesis. These pods were then ground into a powder on pre-chilled mortars, and the powder weighed to 50 mg before being sent to Olomouc on dry ice. Collaborators carried out quantification via small molecule extraction and LC-MS as has been previously published[36].

5. Data Visualisation and Statistics

With the exception of RNA-Seq analyses (see Section II.II., 3) all plotted data was visualised using R, using the packages 'ggplot2', 'dplyr', 'ggthemes' and 'extrafont'. Statistical significance was tested using a pairwise Student's t-test (a linear model), with a p-value of <0.05 declared to be significant.

II.III. Results

1. 4-Cl-IAA, but not IAA, is capable of promoting the growth of deseeded immature pea fruit

In order to begin exploring the role of 4-Cl-IAA in pea fruit development, previously reported experiments on immature pericarps were recapitulated. Immature pea fruit (15-20 mm in length, 2 days after anthesis) were deseeded and treated with 100 μM of IAA or 4-Cl-IAA. Only 4-Cl-IAA treatment, but not IAA treatment, resulted in growth comparable to a fruit with seeds retained (**Figure 2.2**) in accord with previous reports. This suggested that, indeed, 4-Cl-IAA and IAA may have subfunctionalised relative to one-another during legume evolution.

It has been previously reported that IAA is itself inhibitory to growth when exogenously applied to deseeded pea fruit (relative to mock/ water control treatments). However, in this result a statistically significant difference in pod growth between mock and IAA treatments was not observed, which suggested the pod is "IAA-blind" rather than IAA having played an active role in inhibiting pod growth.

The plant's apparent ability to "discriminate" between IAA and 4-Cl-IAA (which are both functional auxins *in planta* and are structurally near-identical) was deemed a point of major interest. The ability of 4-Cl-IAA, but not IAA, to stimulate pod growth in the absence of developing seeds suggested that the two auxins could be altering the expression of different suites of target genes. It has been previously reported that some specific genes are differentially expressed upon IAA vs. 4-Cl-IAA treatment (from quantitative PCR analyses of extracted RNAs); hereunder a whole-transcriptome view was pursued by RNA-sequencing. It was hypothesised that the genes which were expressed under 4-Cl-IAA treatment would favour pod growth, and that IAA would have either a more limited effect on pod gene expression or that it would upregulate a totally different set of growth-inhibiting target genes.

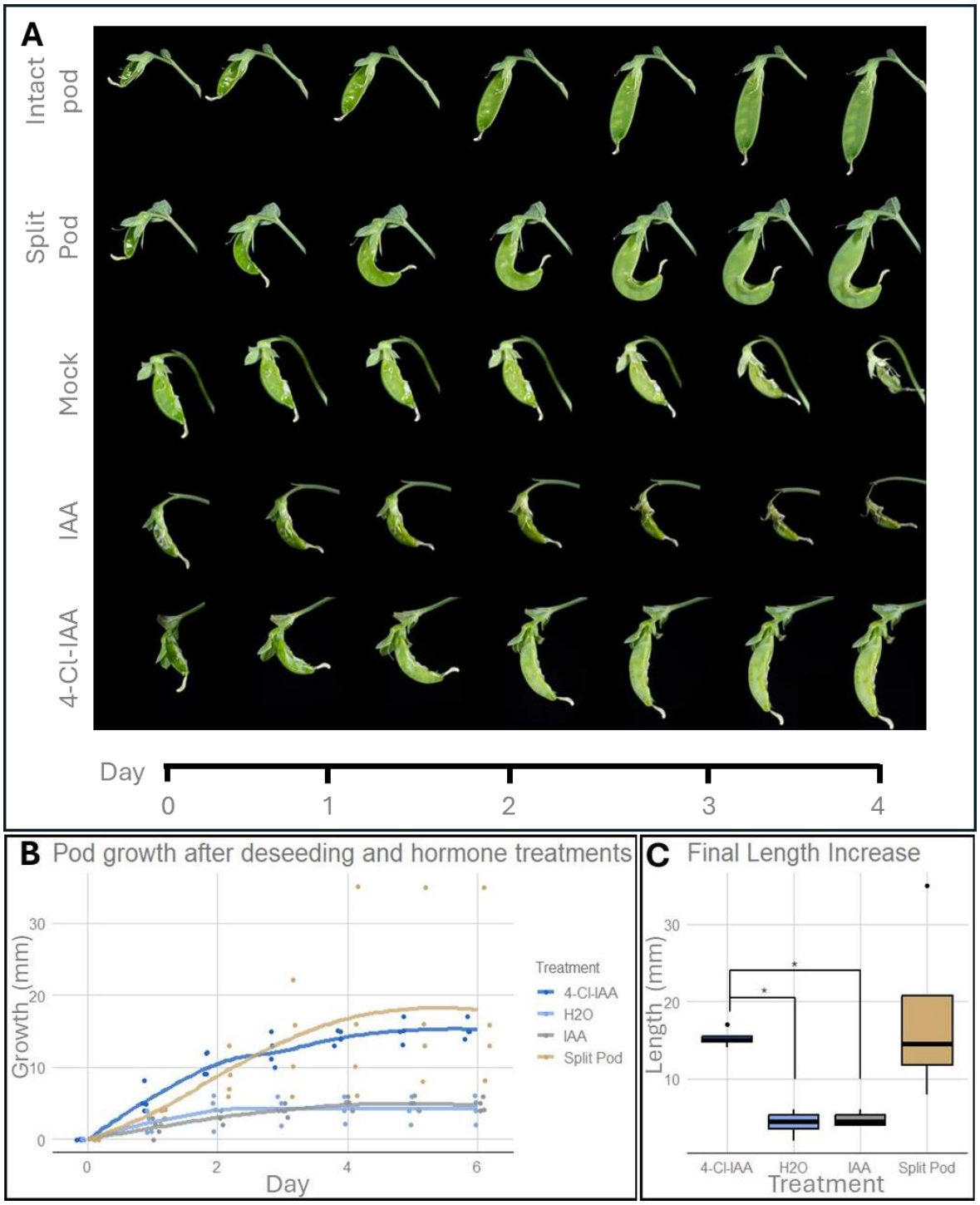


Figure 2.2: Responses of pea fruit to deseeding and hormone treatments. A - Representative 4-day morphological responses of immature pea fruit to abaxial suture cutting, deseeding and hormone treatments, seeds removed and replaced with H_2O (mock), 100 µM IAA or 100 µM 4-Cl-IAA treatment. B - Pod growth per day by treatment (split pod seeds retained, seeds removed and replaced with water, IAA or 4-Cl-IAA treatment). C – Final length increase by treatment, 4-Cl-IAA significantly longer than IAA or mock (H_2O) treatment (Student's t-test, n=4, $p<0.05$).

2. 4-Cl-IAA and IAA do not induce the expression of different target genes in deseeded pea fruit, and pod responses to each hormone are transcriptomically similar

It was hypothesised that the different growth responses of deseeded pea fruit to IAA and 4-Cl-IAA may be underpinned by the upregulation of different sets of target genes by each of the two hormones. RNA-sequencing (RNA-Seq) was conducted to explore this possibility, in the search for differentially expressed genes (DEGs) between 4-Cl-IAA and IAA treatment of deseeded pods.

Immature (15-20mm, 2 days after anthesis) pea pericarps were cut along their abaxial suture and seeds removed. The interior surfaces of the pod were then treated with either mock solution, 100 μM IAA or 100 μM 4-Cl-IAA for 2 hours. After 2 hours, pods were snap frozen and RNA extracted for RNA-Seq.

Despite being performed in octuplicate, no statistically significant DEGs were observed between IAA and 4-Cl-IAA treatment at $q<0.05$ (**Figure 2.3**). Both IAA and 4-Cl-IAA did show differential expression of genes relative to the mock (H_2O-treated) control (279 DEGs between H_2O and IAA and 356 DEGs between H_2O and 4-Cl-IAA treatment at $q<0.05$).

Canonical auxin-responsive genes were observed to be upregulated in both IAA and 4-Cl-IAA treatment relative to control treatment, indicating that the deseeded pods were responding to auxin as expected. These included genes encoding GRETCHEN-HAGEN3 (GH3) auxin conjugation enzymes (*e.g. PSAT2G169400*), SMALL AUXIN UPREGULATED RNAs (SAUR, *PSAT7G181440*) and AUX/IAA repressors (*PSAT5G299400*). These gene families have been shown to be upregulated in early responses to auxin treatment across plant species in the literature.

As mentioned above, 356 genes were observed to be upregulated by 4-Cl-IAA treatment relative to the water control while 279 genes were observed to be upregulated by IAA treatment relative to water (**Figure 2.4**). 209 of these DEGs were shared between IAA and 4-Cl-IAA treatment while 70 only appeared to be differentially expressed when comparing IAA to the control. Another 147 of these DEGs only appeared significant when comparing 4-Cl-IAA to the control. In no case did any of these genes show a statistically significantly different level of expression between IAA and 4-Cl-IAA treatments when compared directly. Quantitative PCR validation of this outcome likewise was unable to reproducibly detect any statistically significant difference in the level of expression of these genes between IAA and 4-Cl-IAA treatment (experiments performed by Master's student, data unpublished).

There have been previously published reports that claim that 4-Cl-IAA and IAA alter the expression of specific genes in immature pea fruit and that these genes' differential expression can explain the ability of 4-Cl-IAA, but not IAA, to rescue growth. Previous reports detail that 4-Cl-IAA alone attenuates the expression of auxin receptor *TIR1b* and GA catabolic enzyme *GA20ox1*, while 4-Cl-IAA alone increases the expression of the GA anabolic enzyme-encoding gene, *GA20ox1*. *PsTIR1b* (*PSAT3G024480*), *PsGA20ox1* (*PSAT1G113960*), and *PsGA2ox1* (*PSAT4G173520*) were not found to be differentially expressed across any pairwise comparison of treatments (**Figure 2.4**).

Additionally, it has also been published that 4-Cl-IAA, but not IAA, significantly increases the transcript abundance of ethylene receptor genes *PsERS1* and *PsETR2* and also of ethylene-responsive F-Box genes *PsEBF1* and *PsEBF2*. This is thought to reduce the sensitivity of the immature fruit to ethylene-mediated growth arrest and senescence, thus conferring on 4-Cl-IAA a protective role in preventing immature fruit abortion. As previously, no such differences in the expression of *PsERS1, PsETR2, PsEBF1* or *PsEBF2* were detected in the RNA-Seq dataset.

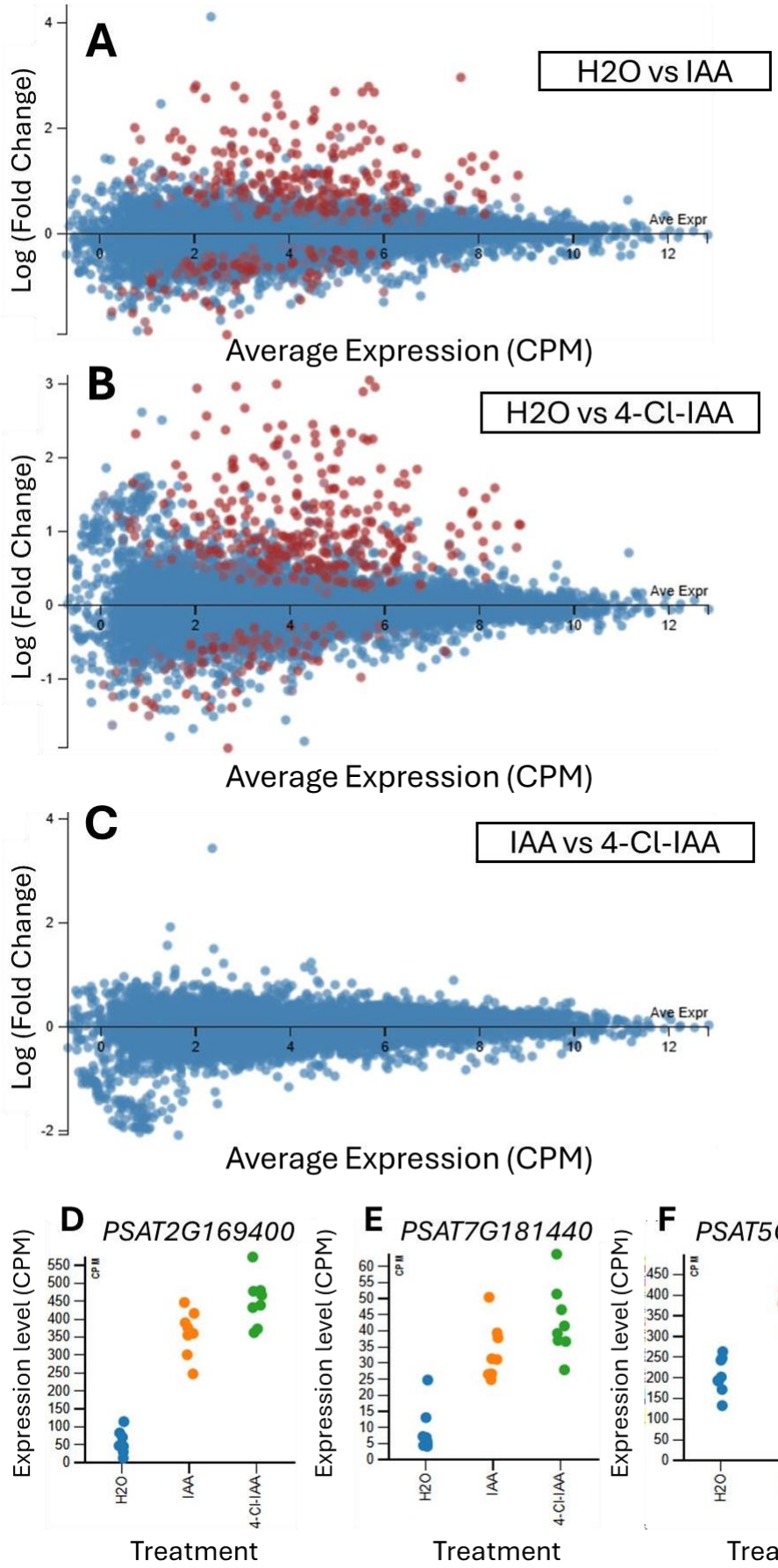

A
H2O vs IAA
Log (Fold Change)
Ave Expr
Average Expression (CPM)
B
H2O vs 4-Cl-IAA
Log (Fold Change)
Ave Expr
Average Expression (CPM)
C
IAA vs 4-Cl-IAA
Log (Fold Change)
Ave Expr
Average Expression (CPM)
D
PSAT2G169400
E
PSAT7G181440
F
PSAT5G299400
Expression level (CPM)
H2O
IAA
4-Cl-IAA
Treatment

Figure 2.3 (Page 25): Overview of pod transcriptomic responses to deseeding and 2-hour treatment with 100 μM exogenous IAA or 4-Cl-IAA relative to H_2O (mock) control, biological reps in octuplicate. A – MA plot of pea gene expression in response to IAA relative to H_2O treatment, B – MA plot of pea gene expression in response to 4-Cl-IAA relative to H_2O treatment, C – MA plot of pea gene expression in response to 4-Cl-IAA relative to IAA treatment, y-axis is log(fold change) and x-axis is average level of expression for the given gene (mean transcripts per million). Red points indicate genes that are significantly differentially expressed between the two treatments (pairwise Student's t-test with Benjamini-Hochberg correction, n = 8, $q<0.05$), blue points indicate genes where no significant difference was detected between the two treatments. Transcript abundance (counts per million) by treatment of: D - GH3 gene *PSAT2G169400*, E –SAUR gene *PSAT7G181440*, F –AUX/IAA gene *PSAT5G299400*, blue points are water (control) treatment, orange points are IAA treatment, green points are 4-Cl-IAA treatment.

Note: After completion of this PhD, work by postdoc Zhe Ji (ORCiD: 0000-0002-0651-2598) in the same lab repeated the above experiment (Figure 2.3) but at a 4-hour timepoint instead of 2 hours. This led to the detection of a number of DEGs between IAA and 4-Cl-IAA, indicating that the transcriptomic differences elicited between these two hormones arise over longer timeframes than 2 hours.

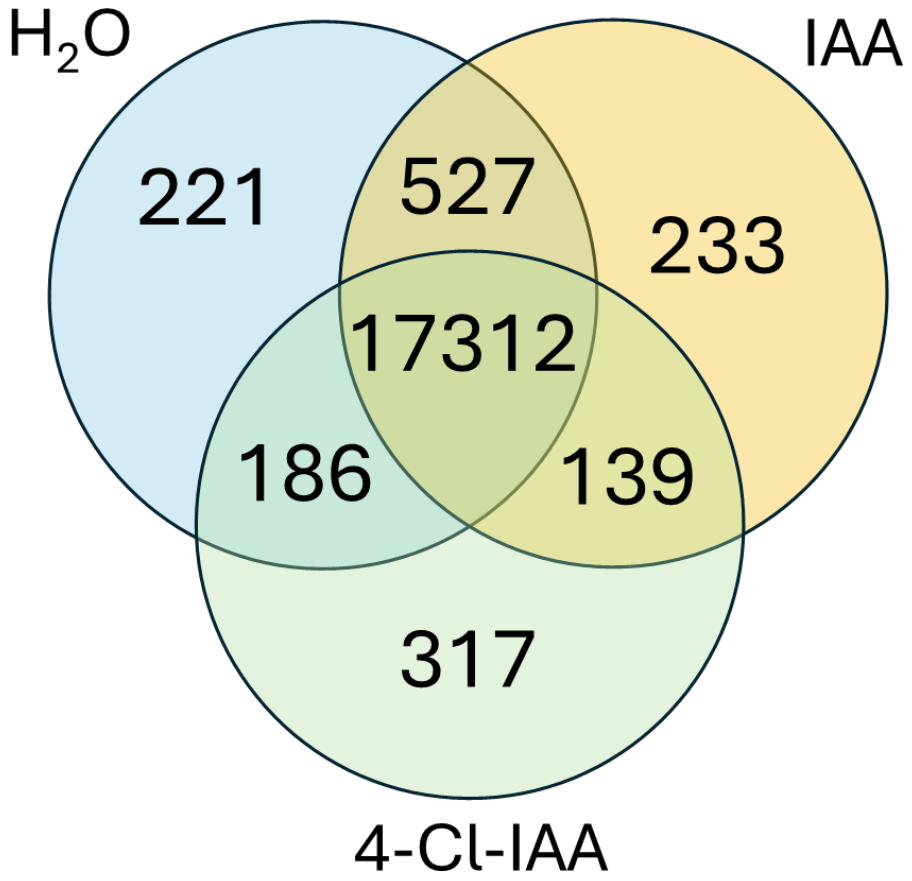


Figure 2.4: Numbers of upregulated differentially expressed genes (DEGs, q<0.05) between 4-Cl-IAA, water and IAA after deseeding and 4-hour treatment with 50 µM exogenous auxins, 5 biological reps. 317 genes were identified as being uniquely upregulated by 4-Cl-IAA, while 233 genes were identified as uniquely upregulated by IAA. RNA-Seq experiment performed by Zhe Ji at University of Oxford. Statistical analysis was conducted using FPKM values for normalisation of reads and a negative binomial test with Benjamini-Hochberg correction for multiple testing (by Novogene Cambridge Sequencing Centre).

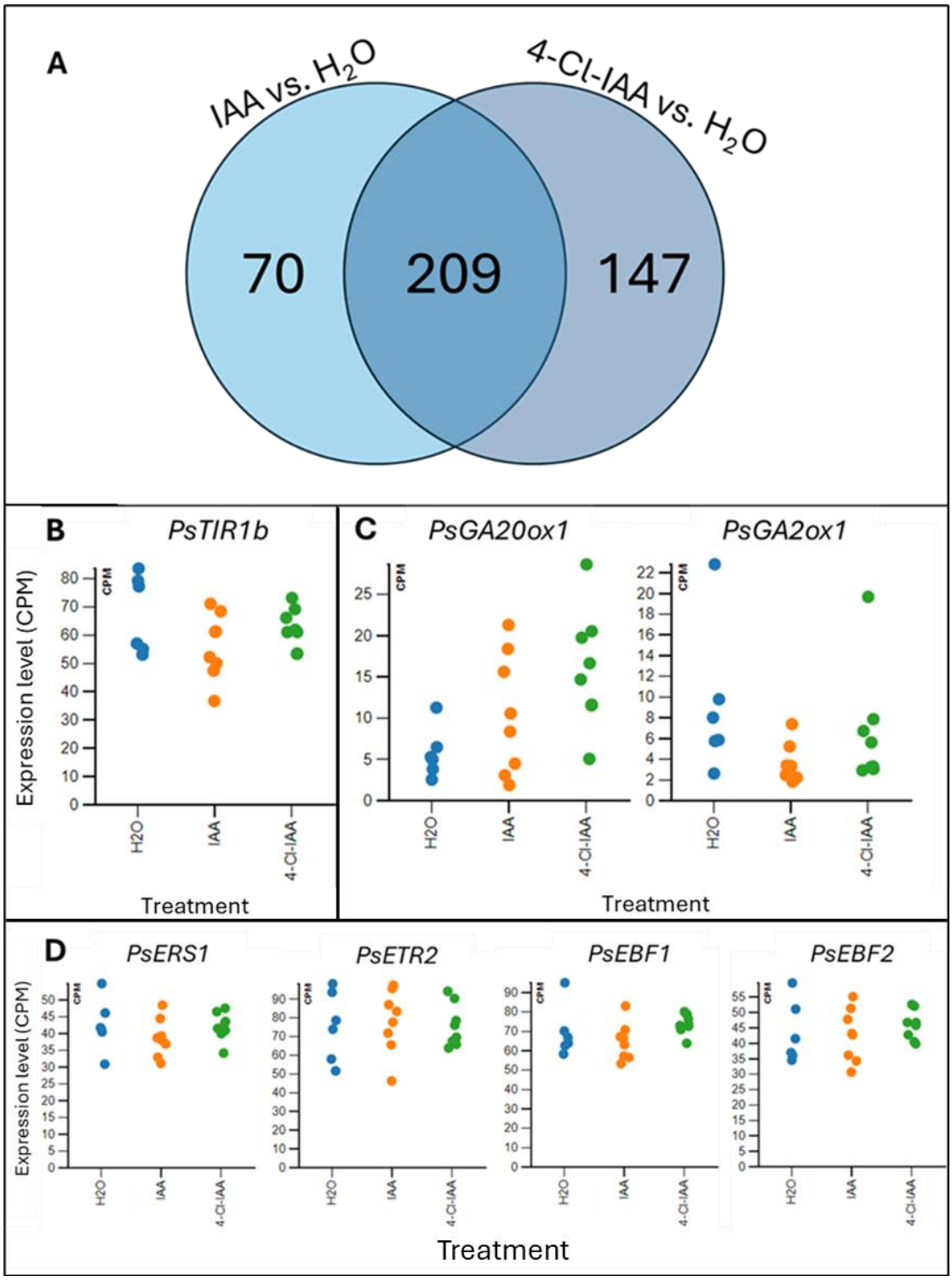


Figure 2.5: A – Numbers of differentially expressed genes (DEGs) between 4-Cl-IAA and water and IAA and water after deseeding and 2-hour treatment with 100 µM exogenous auxins, biological reps in octuplicate. B - Transcript abundance (counts per million) of *PsTIR1b* by treatment, C - Transcript abundance (counts per million) of *PsGA20OX1* and *PsGA2OX1* by treatment, D - Transcript abundance (counts per million) of *PsERS1*, *PsETR2*, *PsEBF1* and *PsEBF2* by treatment, blue points are water (control) treatment, orange points are IAA treatment, green points are 4-Cl-IAA treatment.

3. Trehalose-6-phosphate is not capable of independently rescuing the growth of deseeded pea pericarps

Within immature pea seeds, trehalose-6-phosphate (T6P) has been demonstrated to alter the expression of genes encoding members of the TRYPTOPHAN AMINOTRANSFERASE (TAR) family of auxin biosynthetic enzymes[40]. This led to the hypothesis that T6P may be capable of promoting increased synthesis of 4-Cl-IAA across reproductive tissues and, thereby, may stimulate the growth of deseeded pods when applied alone. However, no significant growth increase was observed when treating deseeded pods with T6P relative to water controls (**Figure 2.6**). As such, it was concluded that T6P's role in promoting auxin biosynthesis is likely limited to the developmental context of seed starch accumulation.

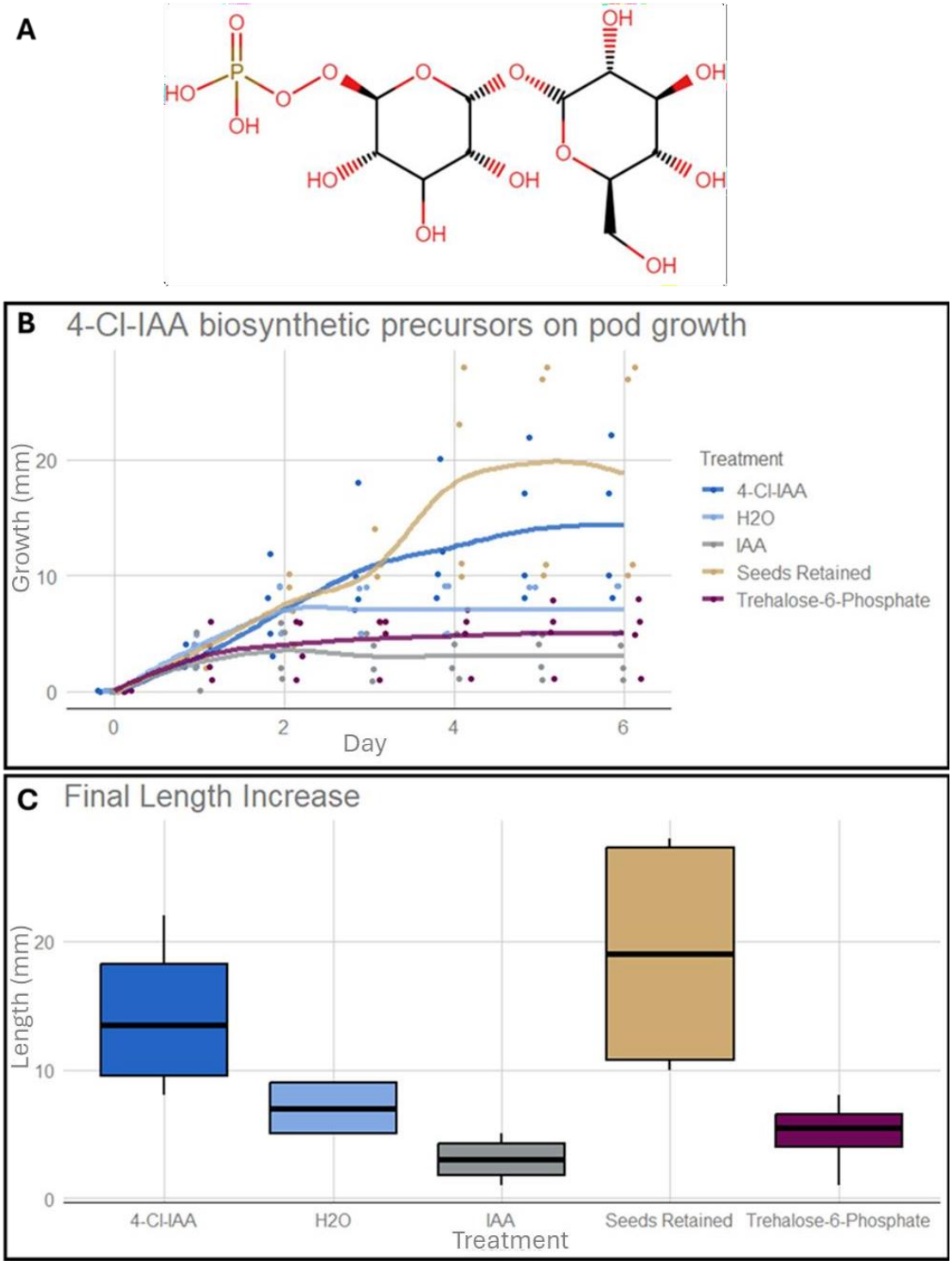


Figure 2.6: Responses of pea fruit to deseeding and trehalose-6-phosphate treatment. A – Trehalose-6-phosphate molecule. B - Pod growth per day by treatment (split pod seeds retained, seeds removed and replaced with water, IAA or 4-Cl-IAA treatment). C – Final length increase by treatment, T6P not significantly longer than mock (H_2O) treatment (n=4, p>0.5).

4. 4-Cl-IAA/ IAA discrimination is limited to a subset of legume species

The reproducibility of the observation that 4-Cl-IAA, but not IAA, could rescue immature developing fruit raised the question of whether this apparent "auxin discrimination" was a universal feature of the legume family or was limited to *Pisum* as a unique case. Deseeding work was thus replicated across a selection of papilionoid legume species, most of which do not produce 4-Cl-IAA (those outside of the Fabeae/Trifoleae (F/T) clade). Several species did not produce useful data, as the fruit were too delicate and died in response to the mechanical damage associated with deseeding regardless of the internal treatment (*Lotus* species were noticeably poor responders). However, to the extent that each fruit was experimentally tractable, both 4-Cl-IAA *and* IAA were able to promote pod elongation for all non-*Pisum* species tested and showed no significant difference in their effects on pod growth (**Figure 2.7**). This indicates that the ability of the fruit to distinguish between 4-Cl-IAA and IAA may be an evolutionarily recent phenomenon. In particular, the 4-Cl-IAA-producing *Trigonella* (fenugreek), which is the most closely related of tested species to *Pisum*, responded positively to both IAA and 4-Cl-IAA treatment, indicating that the differential growth response to these two hormones arose within the F/T clade.

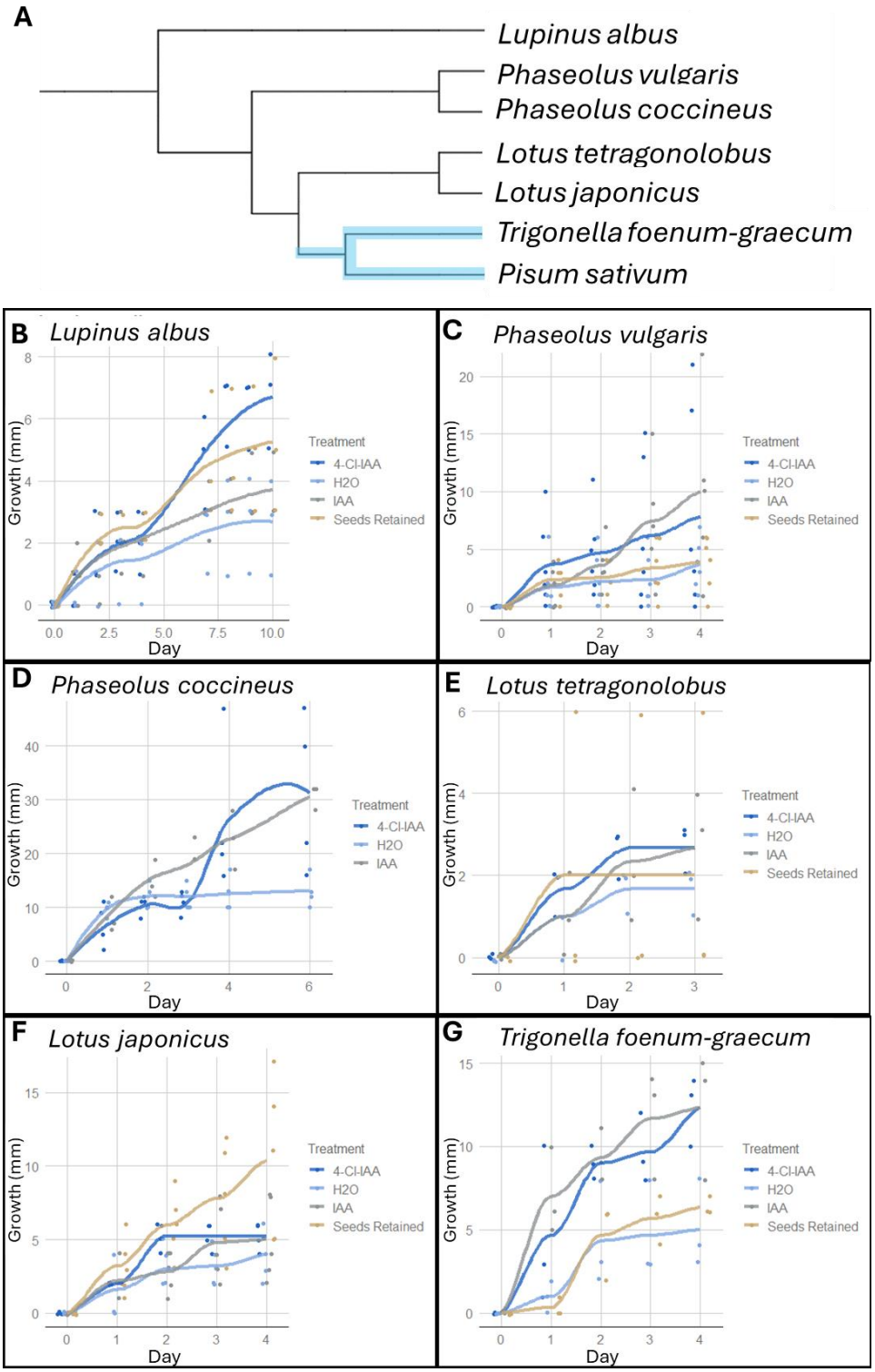


Figure 2.7: A - Phylogeny of tested papilionoid legumes, with members of the F/T clade highlighted in blue. B-G – Growth of split pods per day by treatment (split pod seeds retained, seeds removed and replaced with water, IAA or 4-Cl-IAA) across species. B - *Lupinus albus*, lupin; C - Phaseolus vulgaris, *common bean*; D - *Phaseolus coccineus*, scarlet runner bean; E - *Lotus tetragonolobus*, asparagus pea; F - *Lotus japonicus*; G - *Trigonella foenum-graecum*, fenugreek.

5. GH3 activity is not required for auxin discrimination during pea fruit growth

The RNA-Seq experiment described in Section 2 above, unexpectedly showed no obvious transcriptomic differences between IAA and 4-Cl-IAA treatment (**Figure 2.3**). This suggested that the growth differences observed were not the result of different affinities of either auxin to the TIR1/AFB auxin receptors present in the pea fruit. As such, it was hypothesised that, while 4-Cl-IAA and IAA may induce the same transcriptional changes by the 2-hour timepoint, the 4-Cl-IAA-induced response may be more "durable" than the IAA-induced response over a longer period of time. This was hypothesised to be the result of different rates of inactivation of each auxin by the auxin-inactivating GH3 enzymes, which may have had a stronger affinity for IAA than 4-Cl-IAA.

In order to test this hypothesis, the selective GH3 inhibitor kakeimide (KKI)[37] was used to see if it could convert IAA into a growth-promoting treatment in immature pea fruit. However, treatment with either KKI alone or with IAA and KKI together did not increase pod growth relative to controls (**Figure 2.8**). As such, it was concluded that the mechanism by which IAA inhibits fruit growth must be GH3-independent.

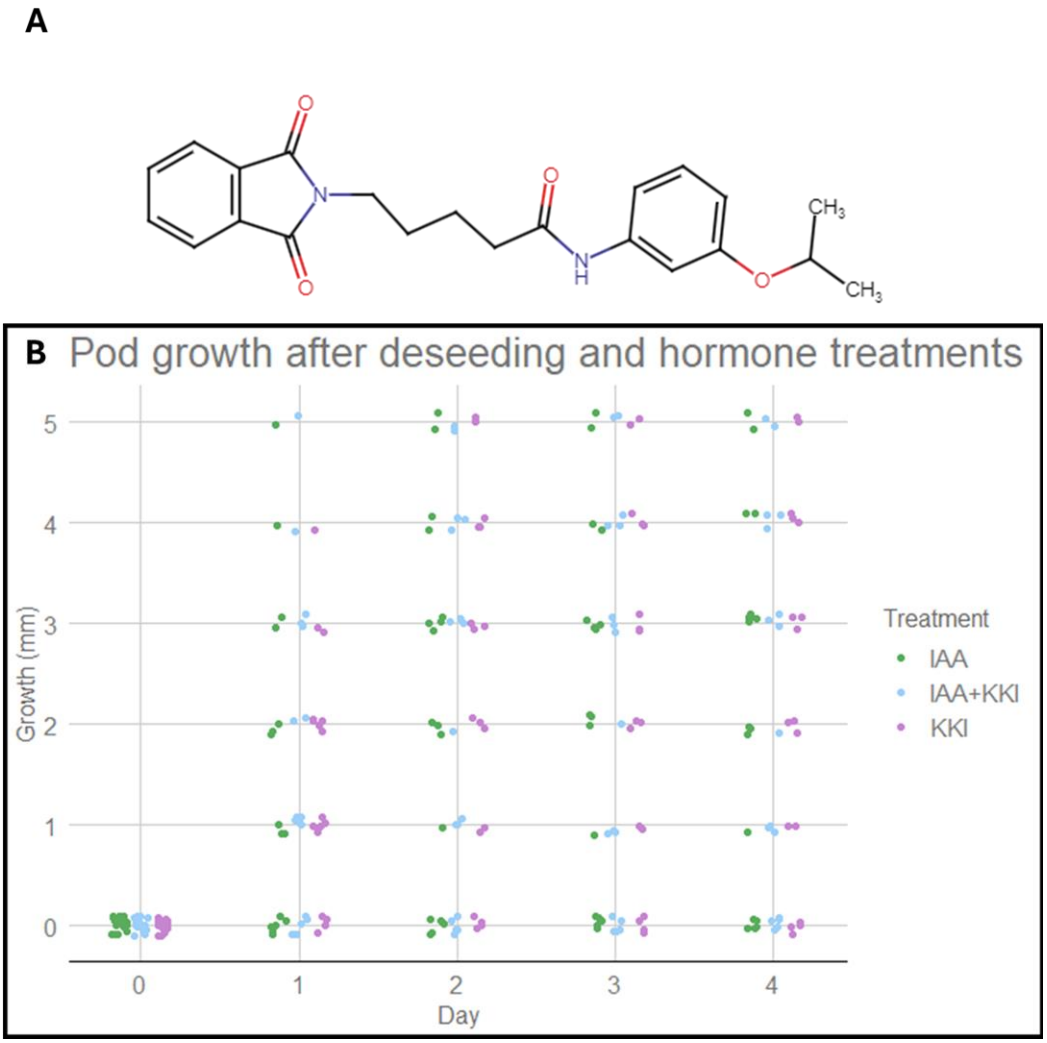


Figure 2.8: A – Kakeimide (KKI) molecule, a synthetic inhibitor of auxin-inactivating GH3 enzymes B - Pod growth per day by treatment (IAA treatment 100 µM, KKI treatment 20 µM, and combined IAA 100 µM + KKI 20 µM treatment) in deseeded pea fruit. No statistically significant difference was observed between treatments (Student's t-test, n=17, $p>0.05$).

6. IAA, rather than 4-Cl-IAA, is more abundant in immature whole pea fruit, but the most abundant auxin species is aspartate-conjugated IAA

The ability of exogenous 4-Cl-IAA but not IAA to rescue pod growth motivated an investigation into the endogenous 4-Cl-IAA and IAA of immature pea fruit. Immature whole pods (with seeds still inside) were harvested at 2, 5 or 7 days after anthesis (2DAA, 5DAA, 7DAA) and snap frozen and weighed for quantification of both unconjugated and conjugated auxin species. Aspartate-conjugated IAA dominated the auxin species observed in the fruit (**Figure 2.9**), with all other forms of auxin occurring at minor concentrations. In terms of free auxin, IAA was more abundant than 4-Cl-IAA across all observed developmental stages, though 4-Cl-IAA began to approach IAA levels at 7DAA.

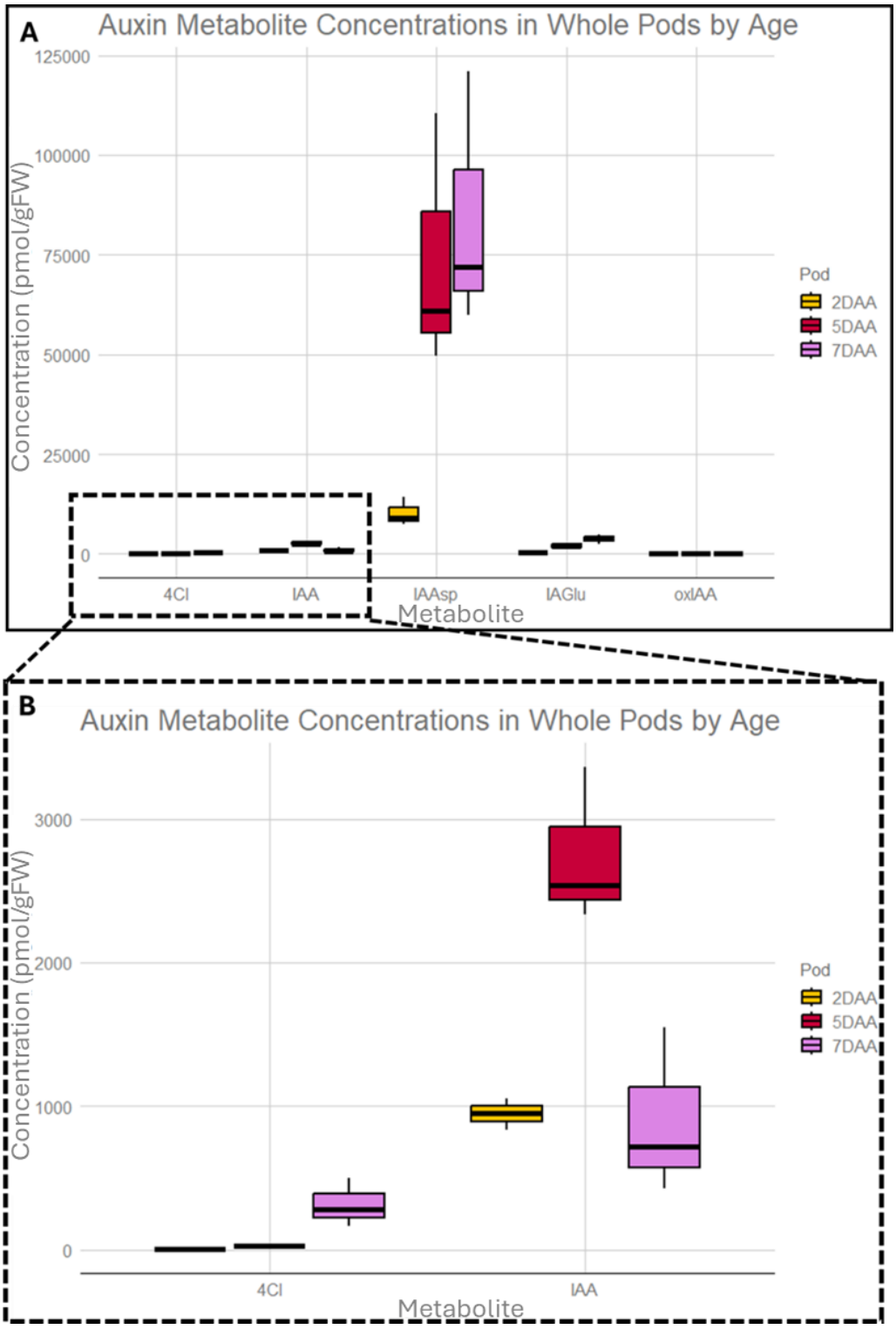


Figure 2.9: A – Auxin metabolite (4Cl – 4-Cl-IAA, IAA, IAAsp – IAA-aspartate, IAGlu – IAA-glutamate, oxIAA - oxidised IAA) levels in picomoles per gram of fresh weight (pmol/gFW) across immature whole pods at 2, 5 and 7 days after anthesis (2DAA, 5DAA, 7DAA). B – 4-Cl-IAA and IAA concentrations at 2, 5 and 7 days after anthesis. LCMS measurement of auxin metabolites was conducted in the Laboratory of Growth Regulators, Palacký University Olomouc, Czechia, by Aleš Pěnčík and Ondrej Novak.

II.IV. Discussion

Broadly speaking, it has been established across angiosperm species that exogenous auxin can substitute for developing seeds in deseeded fruit. This has previously demonstrated in non-legume species such as strawberry[38] and, when emasculating the flower and treating with IAA rather than deseeding the fruit, *Arabidopsis*[39]. Based on these results, it was hypothesised that auxin is synthesised in developing seeds and exported into the maternal fruit tissue to promote growth. This model has found support across experimental investigations which demonstrate that auxin is indeed synthesised in immature seeds[40]. Auxin reception in the fruit is mediated by the canonical auxin signalling pathway and, indeed, mutants in components of this pathway (e.g. *Solanum lycopersicum auxin response factor 8a/8b* double mutants) have been shown to exhibit parthenocarpy[41] (*i.e.* fruit growth without fertilisation). Hence, this creates a model in which ARFs (bound by AUX/IAAs and TPL) create a pre-existing transcriptional "lock" on fruit development which can be "unlocked" by IAA-mediated degradation of AUX/IAA repressors. In this case, the IAA is generated by fertilised seeds and so, in wild type plants, fruit do not begin to develop until fertilisation and seed IAA production begin. This is important from the perspective of organismic resource allocation, as the limited photosynthate, amino acids and micronutrients that are available to the plant are partitioned across multiple seeds and multiple potential offspring and limited only to fruit which are successfully fertilised.

Pisum sativum is unusual in that it not only produces 4-Cl-IAA as well as IAA, but in that across above-ground vegetative tissues 4-Cl-IAA levels remain low but increase markedly in developing reproductive structures[1]. As such, it appears that 4-Cl-IAA has subfunctionalised relative to IAA as the fruit's response to IAA has been lost. The presence and apparently divergent functions of these two auxins in immature pea fruit represents an experimentally tractable and yield-relevant example of evolutionary hormone duplication for study. The previously reported observation that 4-Cl-IAA, but not IAA, is capable of promoting growth of deseeded fruit was supported across multiple experiments, but the molecular underpinnings of the differential growth response have been difficult to clarify.

The strongest hypothesis to explain the observation that exogenous 4-Cl-IAA can rescue the growth of immature pea fruit, while IAA cannot, was that each hormone upregulates different target genes which are respectively conducive or neutral/inhibitory to fruit growth. Conceptually, this could be the result of each auxin having different affinities for the auxin-binding components of the auxin signalling pathway, and thus different

probabilities to form the AFB/auxin/Aux/IAA trimolecular complex. If this were the case, it could be argued that 4-Cl-IAA leads to different degradation dynamics of Aux/IAA transcriptional repressors and thus different effects on global auxin-responsive gene expression. Probing this question directly by experimentation would require using AFB/auxin affinity assays such as yeast-2-hybrid (Y2H) or surface plasmon resonance (SPR) methods. For Y2H, interaction between two proteins can be tested by expressing the two interaction partners as fusion proteins. The first partner (known as the "prey") is fused to the activation domain (AD) of a particular transcription factor while the second partner (the "bait") is fused to the same transcription factor's DNA-binding domain (DBD). Interaction between the two proteins is determined if the transcription factor assembles into a functional unit (DBD-bait/prey-AD), determined by visual identification of expression of a particular reporter gene to which the DBD binds (often encoding an enzyme which yields a coloured product on the testing media). For testing auxin-dependent AFB/AUX/IAA interaction, Prigge et al[42] demonstrated that, when grown on auxin-containing media, the AFB/**auxin**/AUX/IAA trimolecular complex could assemble in yeast, using a *lacZ* reporter system. Separately testing *Pisum* AFBs and AUX/IAAs with IAA and 4-Cl-IAA in the media may allow identification of unique AFB/**4-Cl-IAA$_{auxin}$**/AUX/IAA (or AFB/**IAA$_{auxin}$**/AUX/IAA) interactions which may explain the ability of immature fruit to discriminate between the two auxins.

However, before committing to this experimental line of enquiry it was deemed important to check whether the natural conclusion of this model could actually be observed in developing pea fruit, i.e. that 4-Cl-IAA and IAA do indeed cause different changes in the fruit transcriptional profile.

Extensive RNA-Seq analysis involving two independent experiments (the latter of which is shown above) did not find any significant or reproducible difference in the expression of any genes between 4-Cl-IAA-treated and IAA-treated immature fruit, at the 2-hour timepoint. This observation was unexpected, particularly in light of previous reports that at least some mechanistically important genes related to GA/ethylene synthesis and signalling were differentially regulated by the two auxins[29,31]. After repeated qPCR assays and an inability to reproducibly detect a differentially expressed gene between IAA and 4-Cl-IAA treated pods, it was concluded that IAA and 4-Cl-IAA do indeed elicit the same transcriptional responses from pea fruit at the 2-hour timepoint. However, given that IAA-treated pods undergo growth arrest and eventual senescence, there must be a point after the 2-hour-mark at which the transcriptomes of the alternately treated pods do indeed diverge. This was shown by postdoc Zhe Ji at the University of Oxford (**Figure 2.4**), as taking place at 4 hours, indicating a narrow temporal window in which

transcriptomes diverge between IAA-treated and 4-Cl-IAA-treated pods. As such, it remains unresolved as to the extent to which the TIR1-signalling pathway provides the initial discriminatory step in the pod's response to each auxin.

One possible mechanism for differential growth is that the auxin-induced transcriptional effect of 4-Cl-IAA may last longer or could be "more durable" than that elicited by IAA. The 4-Cl-IAA response could perhaps last longer than the IAA response if IAA is more susceptible to the auxin inactivation/degradation pathway than 4-Cl-IAA. The auxin degradation pathway begins with the reversible conjugation of free IAA to amino acids such as aspartate and glutamate, mediated by the GRETCHEN-HAGEN3 (GH3) auxin-conjugating enzymes[37]. This conjugated IAA is then susceptible to irreversible oxidation by DIOXYGENASE FOR AUXIN OXIDATION (DAO) which will render the IAA molecule permanently inactive[37]. GH3 enzymes are a diverse family and have a broad substrate affinity which goes beyond both auxins and amino acid conjugates, but there is evidence that among auxinic molecules individual GH3s do indeed have substrate preferences[43]. Indeed, there are also reports on the relative inactivity of some *Arabidopsis* GH3 enzymes on halogenated auxins including 4-Cl-IAA and synthetic 2,4-D[44], though importantly these were tested *in vitro* and may not necessarily reflect the conditions *in planta.* As such, it was hypothesised that IAA may have a higher affinity for the pod-expressed GH3s than 4-Cl-IAA, and that the IAA-induced transcriptomic response is therefore relatively ephemeral compared to that induced by 4-Cl-IAA. There is a further element to this model: both IAA and 4-Cl-IAA rapidly induce the expression of GH3-encoding genes as part of a negative feedback loop, which could suggest that there is an asymmetric cross-antagonism between the two auxins.

The synthetic GH3 inhibitor kakeimide (KKI)[37], when applied in combination with IAA, does not rescue the growth of deseeded fruit. This suggests that the inability of IAA to stimulate pod elongation is not the result of GH3-mediated IAA inactivation. However, in this experiment it is also conceivable that the KKI was unable to fully permeate into the tissues of the pod and may thus have been unable to exert its effect.

Alternative hypotheses about the mechanism by which the pea plant discriminates between IAA and 4-Cl-IAA yet remain. Noncanonical components of the auxin signalling pathway include ABP1, an apoplast-localised auxin receptor that connects with its plasma membrane-localised partner, TMK1 to link apoplastic auxin levels to an internal kinase cascade[19], which may exhibit a different affinity for IAA and 4-Cl-IAA. This has never been explored experimentally but was rejected as a possible line of enquiry for this PhD for two reasons. One, the auxin and zinc-binding moieties of ABP1

are totally conserved across all legumes which were tested for deseeding experiments, which makes it unlikely that ABP1 could be the factor that has allowed pea to (evolutionarily) begin discriminating against IAA (as non-discrimination legumes have essentially the same ABP1 homologue). Secondly, the phosphoproteomic changes induced by ABP1/auxin/TMK1 interaction are thought to be critical for highly rapid responses to auxin, whereas the observed auxin discrimination between IAA and 4-Cl-IAA takes place over the timescale of days.

Transport mechanisms are also a possible explanatory variable, as IAA and 4-Cl-IAA may have different affinities for the PIN-FORMED (PIN) auxin transporters[45]. Though generally poorly explored, what little evidence exists indicates that 4-Cl-IAA, perhaps counterintuitively, likely has a higher affinity for PIN transporters than IAA. This has been tested in the context of regeneration of somatic embryos from *Arabidopsis*[46]. Namely, 4-Cl-IAA was found to be poor at stimulating regeneration in *Arabidopsis* explants until those explants were treated with NPA, an auxin transport inhibitor. How exactly a differential rate of transport might factor into an explanatory model for auxin discrimination remains unknown.

Identifying which element of pea fruit molecular biology causes this differential growth response has proven challenging, but comparative studies across legume species have provided a potential insight. Deseeding and hormone treatment of non-*Pisum* legumes did not result in different responses to IAA and 4-Cl-IAA, as both auxins are generally competent to promote the growth of deseeded fruit outside of the *Pisum* genus. This is not unexpected, as IAA's ancestral role in fruit development is to promote immature fruit growth in response to successful fertilisation[41]. Indeed, this result suggests that the subfunctionalisation of IAA and 4-Cl-IAA has involved a loss of fruit responsiveness to IAA more so than any specific gain-of-function for 4-Cl-IAA. This loss of IAA's ability to promote fruit growth appears to be very evolutionarily recent, as of the tested species only *Pisum sativum* has IAA-unresponsive fruit. Attempts were made to deseed a wider range of species (including *Cicer arietinum, Vicia faba* and *Lens culinaris*) to improve the resolution for when auxin discrimination evolved, but the ability of IAA to promote fruit growth of the F/T clade legume *Trigonella foenum-graecum*, suggests that auxin discrimination evolved *after* auxin duplication (~25M years ago)[24]. Phrased another way, whilst an ancestor of pea evolved the ability to discriminate between 4-Cl-IAA and IAA, early diverging 4-Cl-IAA-producing species have not. The consequences of auxin duplication on organismic fitness or, indeed, whether any selective advantage exists to 4-Cl-IAA production and/or discrimination remains unknown. This research also raises questions as to the extent to which ligand chemical evolution in plant signalling systems

has been responsible for the morphological diversity observed across the kingdom, though so far there is no conclusive evidence that auxin duplication has expanded F/T clade morphospace.

III. Halogenation of Tryptophan during 4-Cl-IAA Biosynthesis

III.I. Introduction

As described in the previous chapter, pea and related legumes within the Fabeae/Trifolieae (F/T) clade produce both the universal plant auxin indole-3-acetic acid (IAA) and the chlorinated auxin variant 4-chloroindole-3-acetic acid (4-Cl-IAA)[25]. While the *in planta* biological activities of these auxins were explored first, it quickly became clear that the biosynthetic pathway for 4-Cl-IAA was not fully resolved.

The observation that a specific hormone in immature seeds of pea (4-Cl-IAA) is capable of rescuing the growth of deseeded pods led us to hypothesise that the endosperm may be a major site of auxin biosynthesis. Endosperm is a unique feature of angiosperm plants: during fertilisation, the pollen tube brings two pollen nuclei into the megagametophyte to effect **double fertilisation** (reviewed in [69]). This entails the expected fusion of one of the sperm nuclei with the egg cell to generate a diploid (2n) zygote, but concurrently the other sperm nucleus fuses to the two haploid polar nuclei of the central cell. This yields a 3n sister cell known as the endosperm, which contains two maternal and one paternal genome in its triploid nucleus. This 3n endosperm nucleus then undergoes repeated mitosis without cell division to yield an enormous, polynucleate cell which later cellularises and fuels the growth of the developing embryo. This endosperm is mostly retained in the mature seeds of Poaceous monocots such as wheat, but across dicots is almost always consumed during the development of the seed and is thus absent at seed maturity and abscission. Seeds which, at maturity, have already consumed all of their endosperm (including pea and *Arabidopsis*) are referred to as exalbuminous. Previous studies in strawberry (*Fragaria* spp.)[40] and *Arabidopsis thaliana*[40] have demonstrated high expression of auxin biosynthetic genes within immature endosperm, but the status of auxin biosynthesis in pea endosperm remains unknown.

IAA can be synthesised in plants by a variety of different biosynthetic pathways (reviewed in [47]). These include both tryptophan-dependent (Trp-dependent) pathways, in which the core proteinogenic amino acid tryptophan is used as the biosynthetic precursor for IAA, or tryptophan-independent pathways, in which IAA is synthesised from non-tryptophan indolic precursors in the cytosol. Within Trp-dependent pathways, the dominant route for IAA synthesis in *Arabidopsis* is a linear, two-step conversion of tryptophan into indole pyruvic acid (IPyA) via the activity of TRYPTOPHAN

AMINOTRANSFERASE RELATED (TAR), followed by a conversion of IPyA into IAA by the activity of YUCCA (YUC). This pathway also predominates in pea auxin biosynthesis[48].

Tivendale et al. 2012[48] offer the most recently published progress on understanding the biosynthesis of 4-Cl-IAA. They heterologously expressed the pea TAR enzyme-encoding genes *PsTAR1* and *PsTAR2* in *E. coli*. By LCMS, Tivendale et al. were able to demonstrate that incubating the transformed bacteria with tryptophan led, as expected, to indole-pyruvic acid (IPyA) synthesis, but feeding of chlorinated tryptophan led to the synthesis of chlorinated indole pyruvic acid (4-Cl-IPyA). Furthermore, they detected both 4-Cl-Trp and, at low levels, 4-Cl-IPyA from pea seed extracts[48], which led them to the model that in 4-Cl-IAA biosynthesis, tryptophan is first converted to 4-Cl-tryptophan, followed by the two-step conversion of 4-Cl-Tryptophan into 4-Cl-IPyA and then 4-Cl-IAA by TAR and YUC. This model is presented in **Figure 3.1**.

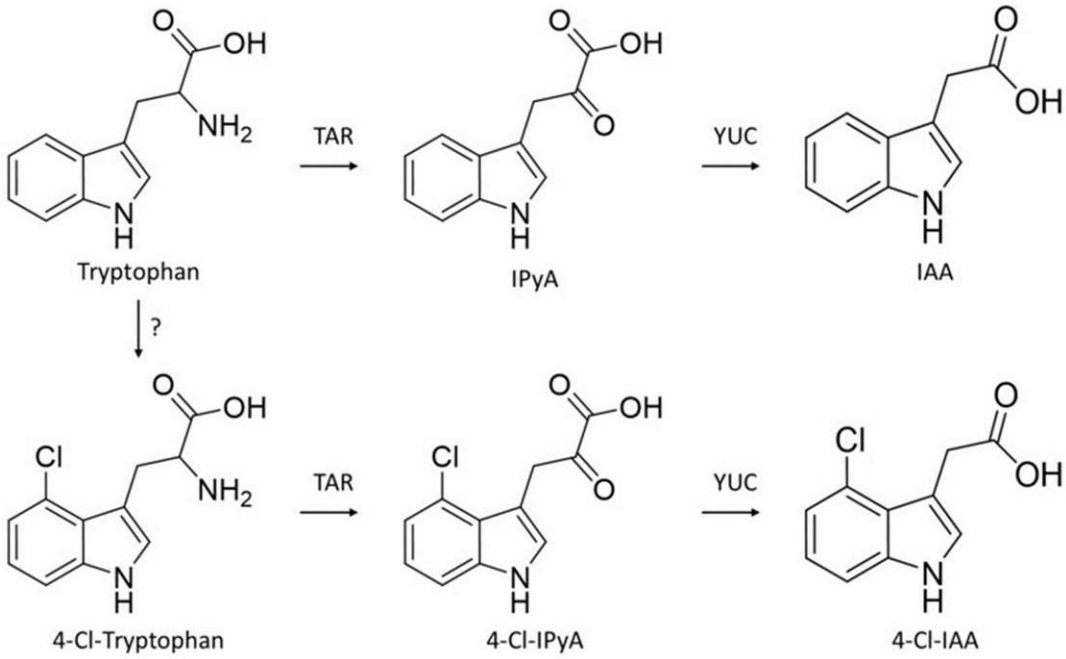


Figure 3.1: Parallel biosynthetic pathway for the synthesis of IAA and 4-Cl-IAA in pea. Uknown halogenase enzyme is indicated with '?'. TAR - TRYPTOPHAN AMINOTRANSFERASE-RELATED, YUC - YUCCA

Halogen atoms such as chlorine are conspicuously absent from core metabolites (e.g. fatty acids, nucleic acids, amino acids, carbohydrates) but are common additions to secondary metabolites across domains of life[49,50]. All nonradioactive halogens (fluorine, chlorine, bromine and iodine) have been observed to be incorporated into biological molecules. These include bacterial metabolites such as (chlorinated) syringomycin E[51],

plant metabolites such as fluoroacetic acid[52] and even human thyroid hormones triiodothyronine and thyroxine, which are iodinated[53]. Perhaps unsurprisingly, a concomitant diversity in halogenases is also observed in nature; halogenases have originated within several families of unrelated enzymes multiple times across the tree of life[57].

Halogenases have evolved from several families, but some halogenases have specifically evolved from mutations in hydroxylases/oxygenases, as halide ions are roughly similar to pseudohalides such as hydroxide anions[54]. Indeed, as few as just one amino acid change can convert a hydroxylase into a halogenase, as demonstrated by Papadopoulou et al. 2021[55] when they converted L-proline cis-4-hydroxylase into a proline halogenase via a targeted D108G mutation. As such, there is an ongoing explosion of scientific interest in biohalogenation both as a naturally occurring process and as a tool for synthetic biochemistry for high-value pharmaceutical[56], agrochemical[57] and antibiotic[58] organohalide products.

Despite a vigorous investigation of microbial halogenation which has led to the identification of large numbers of microbial halogenases, comparatively little progress has been made in studying halogenation within higher eukaryotes including plants, which produce a range of halogenated molecules. At time of writing, only a single plant halogenase has so far been identified and biochemically characterised, DECHLOROACUTUMINE HALOGENASE (DAH) from the early diverging eudicot family Menispermaceae[50]. The plants in this family (including *Menispermum canadense* and *Sinomenium acutum*) produce the extremely toxic chlorinated alkaloid (−)-acutumine via DAH and other enzymes. DAH itself is a 2-oxoglutarate-dependent halogenase, having evolved from a 2-oxoglutarate-dependent dioxygenase and concordant with the evolutionary trend of oxygenases giving rise to halogenases. Though a variety of other halogenated metabolites are synthesised by plants, no other plant halogenase has yet been identified, though one chlorinated plant metabolite (microtubule assembly inhibitor maytansine[59] from *Maytenus serrata*) has been demonstrated to be produced by a bacterial symbiont. Much still remains to be understood about halogenation in plants, and it is likely that several high-value halogenated metabolites of the future may be identified by focused research in this area.

Tryptophan is a core amino acid, and much as other core amino acids are targets of halogenases (e.g. threonine halogenase *SyrB2*[51] and tyrosine halogenase *TPO*[53]) so too is tryptophan halogenated by a range of bacterial tryptophan halogenases. So far, all discovered indole halogenases belong to the flavin-dependent halogenase (FDH)

family[60], and the tryptophan halogenases of this family include PyrH[61], SttH[62], ThHal[63], PrnA[64] and RebH[65]. The defining characteristic of these enzymes is their regioselectivity: they halogenate tryptophan only at specific positions on the indole ring at either carbon 5, 6 or 7 (**Figure 3.2**). However, a recent discovery concerning the activity of AetF[66,67] expands this model, as AetF is an unusual tryptophan brominase which di-halogenates tryptophan at both positions 5 and 7 (but not 6). Regardless, it is notable that a position-4 halogenase has yet to be identified among FDHs, or indeed any other enzyme family.

In order to determine whether the pea tryptophan-4-halogenase was an FDH enzyme, known FDHs were aligned to the publicly available Caméor pea genome via BLAST. No hits were observed however, and furthermore no pea proteins were identified to contain both of the widely conserved GxGxxG and WxWxIP[60] motifs (which together permit flavin binding) that characterise this enzyme family.

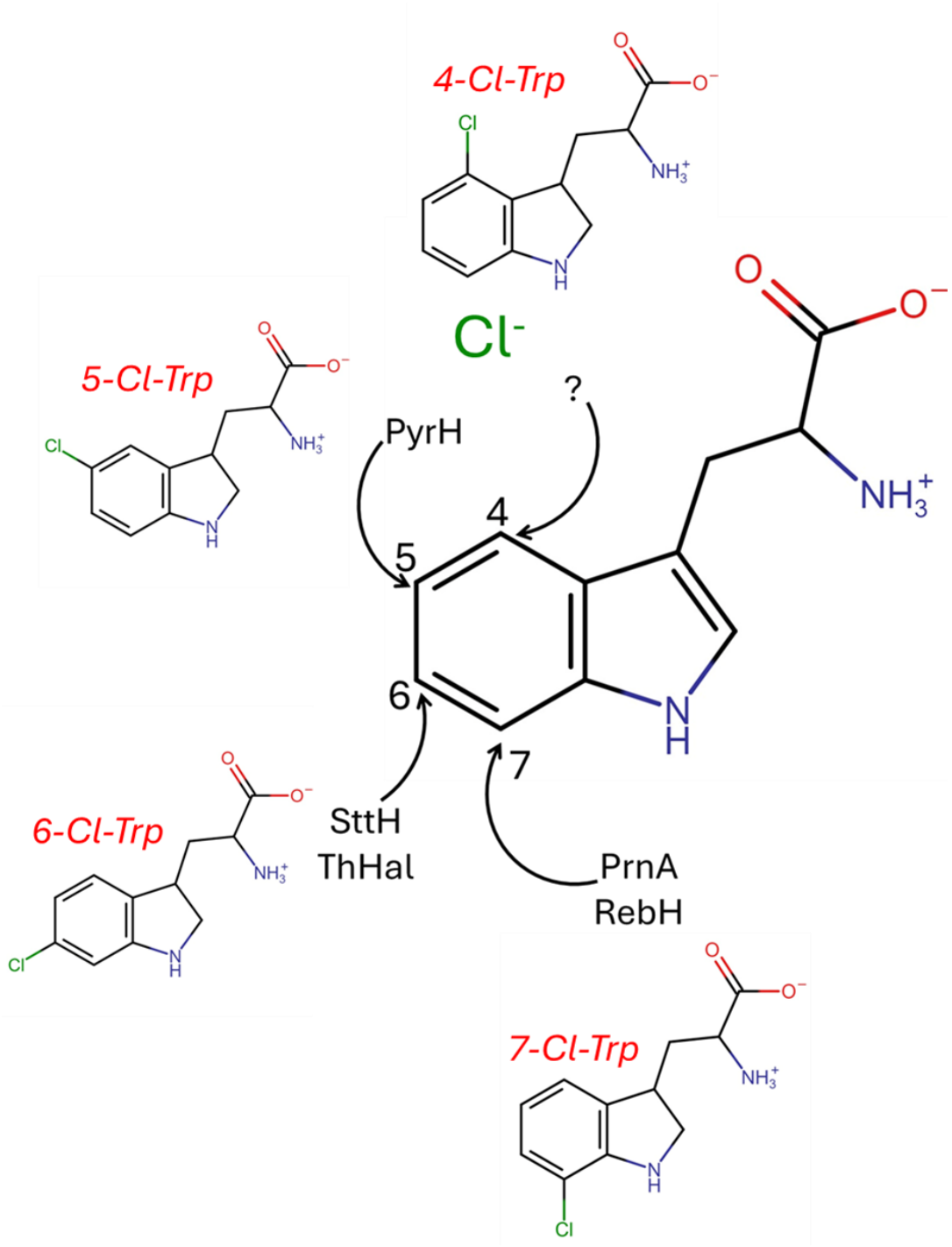


Figure 3.2: Schematic representation of possible chlorinated tryptophan isomers found in nature: 4-Cl-Trp, 5-Cl-Trp, 6-Cl-Trp, 7-Cl-Trp. Numbers indicate positions on indole ring of tryptophan molecule, and black text indicates well-described FDH enzymes capable of chlorinating at the given position. Uknown tryptophan-4-halogenase of pea remains unknown ('?').

Given the non-homology of any pea genome-encoded proteins to flavin-dependent halogenases, and the absence of any FDHs that are capable of halogenating at position 4, it was concluded that the halogenase upstream of 4-Cl-IAA biosynthesis was likely not related to bacterial indole halogenases. This then motivated a less targeted approach to identification of the halogenase enz1yme, by attempting to isolate it directly from the reproductive structures themselves.

III.II. Materials and Methods

1. Plant Materials

JI2822 pea plants were cultivated as described above (Materials and Methods, Section II.II.). Endosperm was harvested from pods by the harvesting and opening of immature pods, and piercing seeds with the tip of a sharp clean scalpel or tweezers. The liquid endosperm was then pipetted out and pooled in a 5 ml tube on ice. Endosperm was collected to a volume of between 2 and 4 ml, depending on availability. Approximately 100 JI2822 pods provide 2 ml of endosperm for assaying, but this is highly variable with the specific developmental stage of each seed (which varies within and between pods). Harvested endosperm was not stored prior to the enzyme assay, and new endosperm was freshly harvested prior to each assay.

2. Deseeding and hormone treatments

JI2822 plants were deseeded as described above (Section II.II., 2.) and then treated with 100 µM of either IAA (Merck Life Science UK Limited, Dorset, UK), 4-Cl-IAA (Cayman Chemical, Cambridge Bioscience, UK), 4-Cl-tryptophan (Biosynth Ltd., Berkshire, UK), tryptophan (Merck), or mock solution (0.1% v/v Silwet), by daily pipetting to the point of pod saturation. Opened, but not deseeded pods were used as a control (“Split Pod”).

3. Halogenase assay

Pooled endosperm was divided into aliquots of equal volume (50-120 µl, depending on endosperm availability) across replicates and kept on ice. To these aliquots, 80 µl of assay solution was added to give a final assay volume of 130-200 µl. Assay solution was based on 1M Tris-HCl buffer (Fisher Scientific, Leicestershire, UK) diluted to 100 mM with deionised water (10% v/v) and adjusted to pH 5.5 by the gradual addition of concentrated hydrochloric acid with a pH meter. 80 µl of this buffer was either used alone in the “0 µM tryptophan” replicates or, if to be used for testing the addition of exogenous tryptophan to the endosperm, adjusted with a 5 mM stock of tryptophan to the desired concentration. 80 µl of either tryptophan-containing or tryptophan-absent

assay solution was then added to the 50-120 µl endosperm and incubated overnight at 5 ^{o}C.

4. Analysis of tryptophan and chlorinated tryptophan levels

After incubation, 180 µl of 20% v/v methanol was added to the assay samples followed by centrifugation at 8000 rpm at room temperature. The supernatant was used for LC-MS. LC-MS analysis was conducted using a Waters Xevo TQ Absolute Triple Quadrupole Mass Spectrometer. The MS detector was set to positive ion mode using electrospray ionisation (ESI+). For chromatography, a Kinetex 2.6 µm EVO C18 100 Å column (50 x 2.1 mm, Phenomenex) with a flow rate of 0.5 ml/min, injecting a volume of 10 µl per sample. 0.01% v/v formic acid (solution A) and pure acetonitrile (solution B) were used for the mobile phase. These began 99% A/1% B for 4 minutes, then at 40% A/60% B for 0.5 minutes, followed by 100% B for 1.10 minutes, then 99% A/1% B for 2 minutes as a wash. All sample runs were interspersed with blanks (20% v/v methanol solution) to limit carryover. Integration of total ion current was used to relatively quantify chlorinated tryptophan across samples. Relative abundance was determined by automatic integration of peaks using MassLynx/TargetLynx.

5. RNA-Seq of pea endosperm and pod

200 µl of endosperm was frozen on dry ice prior to RNA extraction, while 3 pods per biological replicate (all at 3 DAA) were pooled for the pod RNA samples. RNA was extracted from pea endosperm and pod tissues using the RNEasy Plant Mini-Kit as described above (Section II.II.), and data were analysed according to the same pipeline, and visualised in Degust.

6. Sucrose gradient centrifugation

A sucrose gradient from 20-60% (w/v) was prepared with steps of 10% in a 15 ml tube, with 1 ml per step. 2 ml of endosperm was pipetted onto the top layer and centrifuged at 8000 rpm for 4 minutes, which led to fractionation of subcellular components. Fractions were pipetted off from top to bottom and separated into individual tubes for assaying.

7. Proteomics analysis

The pellet generated by sucrose centrifugation was washed several times in 100 mM pH 5.5 Tris-HCl buffer, and provided dried to the JIC proteomics platform. The JIC proteomics platform performed proteomics analysis on this pellet using a Thermo Scientific Orbitrap Eclipse Tribri mass spectrometer LCMS with Thermo UltiMate 3000

RSLCnano LC system. Data were analysed by the platform using Proteome Discoverer software (Thermo) with Percolator algorithm.

8. Agrobacterial infiltration of *Nicotiana benthamiana* leaves for heterologous enzyme expression

The genes encoding the 33 candidate enzymes were synthesised to order from Integrated DNA Technologies (IDT) and cloned into a pDONR vector via a BP gateway cloning reaction. These pDONR constructs were then verified by gel electrophoresis and by sequencing to bear the correct inserts and were used in LR reactions with a *pEAQ-HT-DEST1*[68] vector to produce a construct for heterologous expression in tobacco. *pEAQ-HT-DEST1* drives expression by combining the Cauliflower mosaic virus *35S* promoter (CaMV *35S*) with the 5'- and 3'UTR of Cowpea mosaic virus (CPMV) RNA-2. The nopaline synthase (nos) terminator was used to terminate transcription[68]. Constructs were transformed into *Agrobacterium* line GV3101. Preparation and syringe infiltration of Agrobacterium suspensions was carried out according to previously published methods (see Sainsbury et al. 2012, "3.3. Agroinfiltration of *N. benthamiana*"[68]). After 5 days, tobacco samples were snap frozen and ground in 20% methanol for LCMS-based analysis. Methanol samples were centrifuged at maximum speed for 5 minutes prior to LCMS.

III.III. Results

1. 4-Cl-IAA's biosynthetic precursors are incapable of promoting pod growth

To better spatially localise the exact site of 4-Cl-IAA biosynthesis, solutions of the biosynthetic precursors to 4-Cl-IAA were prepared and used to treat deseeded fruit. It was hypothesised that, should the pod express the enzymes involved in 4-Cl-IAA biosynthesis, that it would be stimulated by the exogenous application of 4-Cl-IAA's precursor molecules. However, the pod did not show a significant growth response to tryptophan nor chlorinated tryptophan (**Figure 3.3**), which suggests that the pod is not capable of tissue-autonomous production of 4-Cl-IAA from amino acid precursors. This is in accord with the prior observation (See Chapter 2 above) that deseeded fruit undergo growth arrest. Overall, this supports a model in which 4-Cl-IAA's entire biosynthetic pathway occurs in the developing seed, prior to auxin export into the nascent fruit. Additionally, this indicated that the halogenase upstream of 4-Cl-IAA biosynthesis would be expressed in the developing seeds rather than the immature fruit.

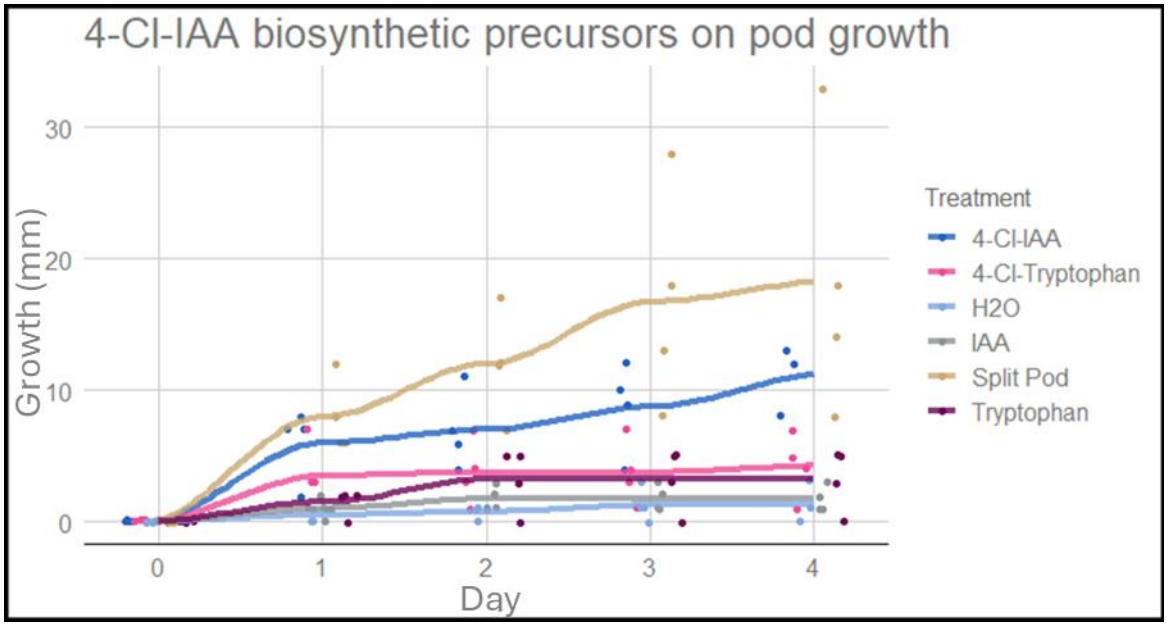


Figure 3.3: Pod growth per day by treatment (100 μM 4-Cl-IAA, 4-Cl-Tryptophan, H_2O, IAA, Split Pod (seeds retained) and Tryptophan all in 0.1% v/v Silwet). Tryptophan and 4-Cl-Tryptophan are biosynthetic precursors of 4-Cl-IAA.

2. An *in vitro* assay suggests the presence of a tryptophan-4-halogenase in immature pea endosperm, but high background levels of 4-Cl-tryptophan prevent conclusive assaying.

After it was established that pea fruit do not express the tryptophan halogenase, attention was redirected towards the developing seed. Pea endosperm is large enough to be easily pipetted out of developing seeds, and so it was extracted and pooled into chilled tubes for the assaying for halogenase activity. Endosperm of equal volume and concentration was fed with variable concentrations of exogenous tryptophan (added in equal volumes) and refrigerated overnight to allow all enzymatic reactions to proceed to their end state. The final solution was 60% endosperm by volume.

After overnight incubation the analytes were centrifuged at max speed and the supernatant diluted 10x with 20% methanol. LC-MS of these resultant samples revealed that as exogenous tryptophan was added in larger concentrations, so too did the detected quantity of 4-Cl-tryptophan increase (**Figure 3.4**). This suggested the presence of the tryptophan halogenase in the early endosperm, however a very high, native background value for 4-Cl-tryptophan in endosperm, and low replicate number (n=3), made conclusive determination impossible. Based on the size of peaks of standards of known concentrations, both 4-Cl-tryptophan and tryptophan occur natively in pea endosperm in the micromolar range (**Table 3.1**).

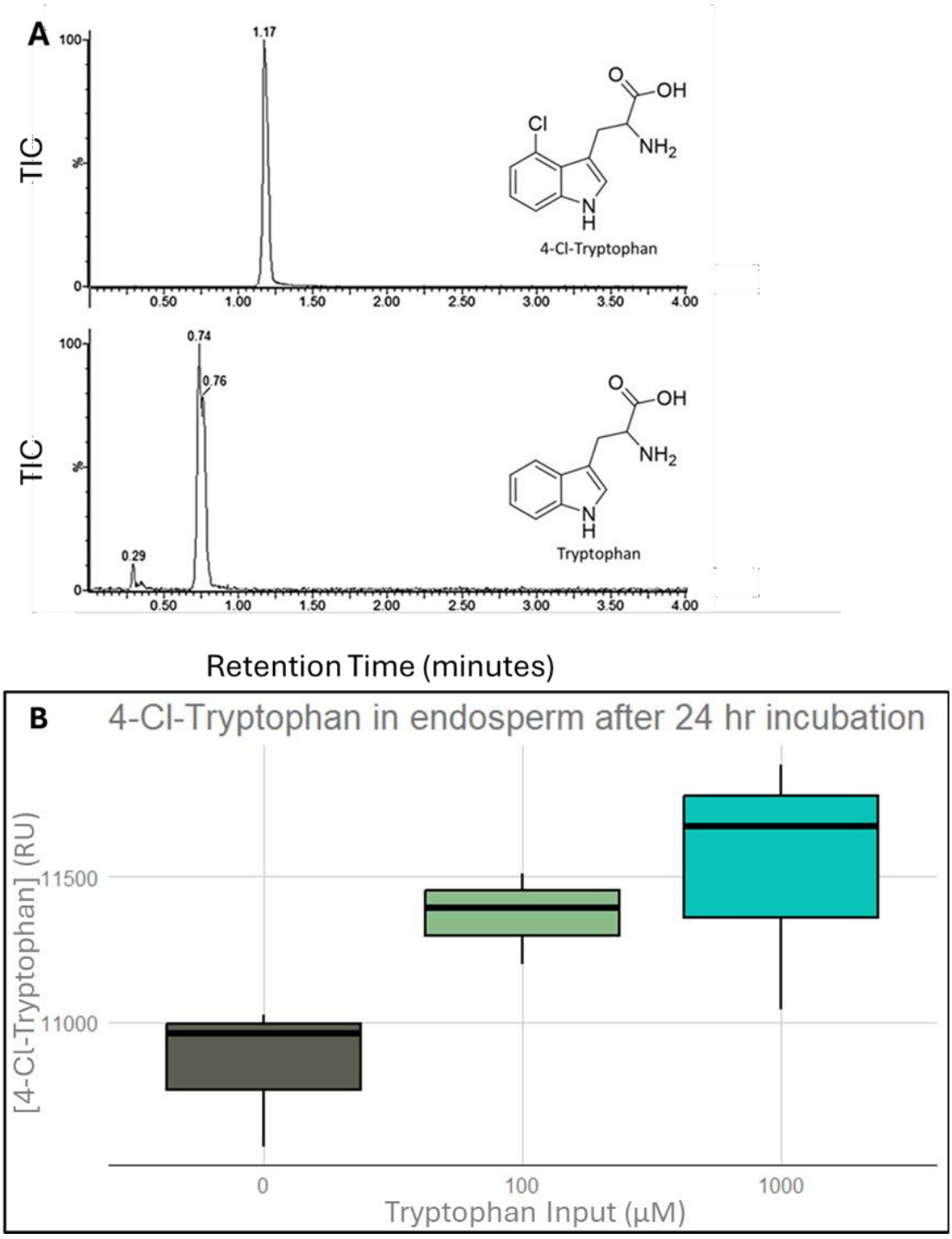


Figure 3.4: Possible halogenation of tryptophan *in vitro* using pea endosperm. A – Total Ion Current (TIC) against retention time in minutes of 4-Cl-tryptophan and tryptophan, injected in 20% methanol (v/v) at 10 µM. B – Concentration of 4-Cl-tryptophan after overnight incubation of pea endosperm with exogenous tryptophan at varying input concentrations (n=3 per concentration).

3. *Phaseolus coccineus* endosperm cannot halogenate tryptophan *in vitro*

To ensure the observed halogenation reaction was truly enzymatic, the experiment was repeated in the exact same conditions but substituting pea endosperm with endosperm from a legume species outside of the 4-Cl-IAA-producing F/T clade (scarlet runner bean, *Phaseolus coccineus*). As expected, *Phaseolus* showed no *in vitro* halogenation and no presence of 4-Cl-tryptophan in its endosperm (**Table 3.1**). This suggested that the halogenation observed in pea endosperm was the product of the F/T clade-specific halogenase enzyme.

4. The halogenase localises to the densest cytoplasmic material

In order to understand the subcellular localisation of the halogenase, centrifugation at maximum speed for 5 minutes was used to produce a fractionated endosperm with a supernatant and pellet. The supernatant was subjected to the halogenation assay (with whole endosperm as a positive control) and it was found that the supernatant was not capable of halogenating tryptophan (**Table 3.1**). Fresh supernatant was then re-fractionated on a 20-60% (w/v) sucrose gradient. This second fractionation yielded a supernatant, a "cloud" of slowly sedimenting, lighter material and a pellet of dense material. Of these three fractions, only the dense pellet was found to increase in 4-Cl-tryptophan content when fed exogenous tryptophan (**Table 3.1**). As such, it was concluded that the halogenase was likely associated with the densest material in the endosperm, speculated to possibly be the nuclei and associated endoplasmic reticulum.

Table 3.1: Halogenation assays according to varying input conditions. **Data are preliminary and require further replication**, with mean 4-Cl-tryptophan concentrations [4-Cl-Trp] in relative units (area under curve) indicated after 24 hour incubation at 5°C with 0 μM or 20 μM tryptophan. NOTE: Due to small quantities of endosperm available, replicates are <3 in supernatant and sucrose pellet experiments.

Experimental condition	[4-Cl-Trp] RU at 0 μM Tryptophan	[4-Cl-Trp] RU at 20 μM Tryptophan
Whole pea endosperm	921,536	1,223,180
Phaseolus coccineus endosperm	0	0
Pea endosperm supernatant	790,404	799,321
Pea endosperm sucrose pellet	1,538,505	1,732,743

5. The endosperm endomembrane proteome reveals candidate potential halogenases

As the halogenase potentially appeared to be associated with the heaviest fraction of the centrifugate, a pellet was generated on a sucrose gradient and washed several times with Tris buffer. This pellet was taken for proteomic analysis by the JIC proteomics platform, which performed mass spectrometry to identify the peptides present in the pellet. This was then cross-referenced to an in-house pea protein database to yield 8096 detected proteins at $q<0.05$, albeit only 3756 proteins with >5 detected peptides.

Of these 8096 proteins, all cytochrome P450s (CYPs), flavin-binding monooxygenases (FMOs), 2-oxoglutarate-dependent dioxygenases (2OGDs) and peroxidases were identified and collected into a list of 90 proteins. This list was then annotated manually using the nearest *Arabidopsis* homologue of each protein. This generated a range of possible candidates, and the list was revised using RNA-Seq to compare the transcriptomes of pea endosperm against the transcriptome of the deseeded pea pod (**Figure 3.5)**. Any enzymes that were expressed in the endosperm at a level higher than that in the pod ($q<0.05$) were taken as priority candidates. This generated a list of 33 enzymes for testing (**Table 3.2**).

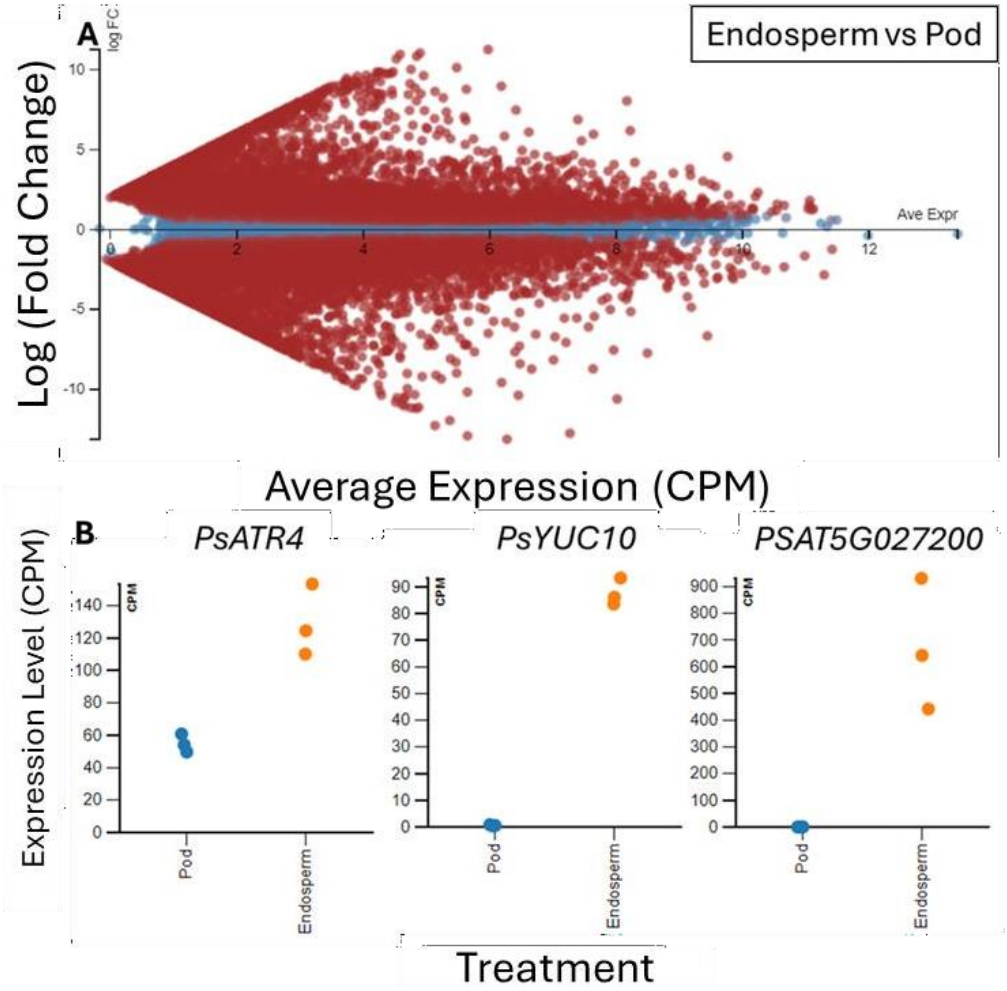


Figure 3.5: Overview of endosperm transcriptome vs deseeded pod transcriptome. A – MA plot of genes expressed in endosperm vs pod, y-axis is log(fold change) and x-axis is average level of expression for the given gene (mean transcripts per million). Red points indicate DEGs between the two treatments (pairwise Student's t-test with Benjamini-Hochberg correction, n = 3, $q<0.05$), blue points indicate genes where no significant difference was detected between the two treatments. Red points above the x-axis are significantly more expressed in endosperm than in pod tissues (n=6719), and red points below the x-axis are significantly more expressed in pod tissues than endosperm (n=6864). B –Three examples of DEGs of interest with transcript abundance (counts per million) of CYP450 enzyme *PSATR4* (*PSAT3G200960*), flavin-binding monooxygenase enzyme *PsYUC10* (*PSAT3G024760*) and 2-oxoglutarate-dependent dioxygenase *PSAT5G027200.* All were significantly more expressed in the endosperm relative to the pod ($q<0.05$). Blue points are pod expression levels in each biological replicate, orange points are endosperm expression levels in each biological replicate.

Table 3.2: List of 33 candidate enzymes derived from pea endosperm proteomics + endosperm vs pod RNA-Seq refinement pipeline. Genes are classified as cytochrome P450 (CYP), peroxidase, flavin-binding monooxygenase (FMO) or 2-oxoglutarate-dependent dioxygenase (2OGD). Annotation is manual and based on annotation from The Arabidopsis Information Resource (TAIR)[70], which was also used to perform BLAST of the amino acid sequences of each '*PSAT*' gene against *Arabidopsis thaliana*. Genes highlighted in red have yet to be heterologously expressed in tobacco leaves, all others have been heterologously expressed in tobacco leaf vis *Agrobacterium*-mediated infiltration for transient expression.

Gene	Class	Annotation
PSAT0S697G0040	CYP	CYP94D2-related
PSAT1G127560	CYP	ELONGATED UPPERMOST INTERNODE CYP714A1/CYP714A2, epoxidises (inactivates) gibberellins, double mutant has bigger organs
PSAT1G168200	Peroxidase	PRX53, encodes a protein with sequence similarity to peroxidases that is involved in lignin biosynthesis. Loss of function mutations show abnormal development of xylem fibers and reduced levels of lignin biosynthetic enzymes.
PSAT1G218560	Peroxidase	PRX25, encodes a cationic cell-wall-bound peroxidase homolog that is involved in the lignification of cell walls. Regulated by COG1, involved in seed longevity.
PSAT2G127280	Peroxidase	CM3 Chorismate mutase 3
PSAT2G165600	Peroxidase	Peroxidase enzyme, no further annotation
PSAT2G169200	CYP	CYP81D8-related, glucosinolate biosynthesis
PSAT3G024720	FMO	YUCCA Enzyme (YUC10)-related
PSAT3G127800	2OGD	2-oxoglutarate/Fe(II)-dependent oxygenases that hydroxylates JA to 12-OH-JA
PSAT3G148400	2OGD	No meaningful annotation, but weakly related to

		aromatic (but not indolic) glucosinolate biosynthesis gene in *Arabidopsis*
PSAT3G167760	CYP	CYP81D8-related, glucosinolate biosynthesis
PSAT3G200960	CYP	ALTERED TRYPTOPHAN REGULATION 4, ATR4, CYP83B1, - Encodes an oxime-metabolizing enzyme in the biosynthetic pathway of indole glucosinolates. Is required for phytochrome signal transduction in red light. Mutation confers auxin overproduction.
PSAT3G178480	Peroxidase	Peroxidase enzyme, no further annotation
PSAT4G017880	FMO	OPR3 encodes a 12-oxophytodienoate reductase that is required for jasmonate biosynthesis. Mutants are male sterile and defective in pollen dehiscence.
PSAT4G202960	CYP	LACERATA, CYP86A8, Encodes a member of the CYP86A subfamily of cytochrome p450 genes. Expressed at moderate levels in flowers, leaves, roots and stems.Mutant seeds have reduced seed longevity, higher tetrazolium salt uptake and reduction, and reduced lipid polyester barriers
PSAT5G027200	2OGD	No *Arabidopsis* homolog, no annotation
PSAT5G058720	Peroxidase	BETA GALACTOSIDASE10-related
PSAT5G060240	FMO	YUCCA Enzyme (YUC10)-related
PSAT5G082680	FMO	YUCCA Enzyme YUC11-related
PSAT5G111120	CYP	CYP94B3 is a jasmonoyl-isoleucine-12-hydroxylase that catalyzes the formation of 12-OH-JA-Ile from JA-Ile. By reducing

		the levels of this the biologically active phytohormone, CYP94B3 attenuates the jasmonic acid signaling cascade. CYP94B3 transcript levels rise in response to wounding
PSAT5G145880	CYP	CYP79B2 is involved in tryptophan metabolism. Converts Trp to indo-3-acetaldoxime (IAOx), a precursor to IAA and indole glucosinolates.
PSAT5G169320	CYP	CYP71B36, most closely related to flavonoid 3' hydroxylases
PSAT5G201640	CYP	CYP75B1 is required for flavonoid 3' hydroxylase activity. Enzyme abundance determines Quercetin/Kaempferol metabolite ratio.
PSAT5G250480	Peroxidase	PRX25, encodes a cationic cell-wall-bound peroxidase homolog that is involved in the lignification of cell walls. Regulated by COG1, involved in seed longevity.
PSAT5G294920	2OGD	STRONG HOMOLOGY encodes a member of the six Arabidopsis IAA-amino acid conjugate hydrolase subfamily and conjugates and conjugates IAA-Ala in vitro. Gene is expressed most strongly in roots, stems, and flowers. The mRNA is cell-to-cell mobile.
PSAT6G030800	2OGD	No *Arabidopsis* homolog, no annotation
PSAT6G133400	CYP	CYP96A1-related
PSAT7G044720	2OGD	gibberellin 20-oxidase, weak homology
PSAT7G140600	CYP	CYP84A1 encodes ferulate 5-hydroxylase (F5H). Involved in lignin biosynthesis.
PSAT7G189120	CYP	CYP94C1 has highest omega-hydroxylase

		activity with 9,10-epoxystearic acid, while also metabolized lauric acid (C12:0) and C18 unsaturated fatty acids.
PSAT7G189240	CYP	CYP94C1 has highest omega-hydroxylase activity with 9,10-epoxystearic acid, while also metabolized lauric acid (C12:0) and C18 unsaturated fatty acids.
PSAT7G220280	CYP	CYP93D1-related
PSAT7G227920	Peroxidase	PRX17, a cell wall-localized class III peroxidase that is directly regulated by the MADS-box transcription factor AGL15 and is involved in lignified tissue formation.

6. Heterologous expression of 26 candidate enzymes has yet to reveal the halogenase

After proteomics and transcriptomics were used to narrow the list, candidate enzymes were prioritised for individual testing by heterologous expression in *Nicotiana benthamiana* (tobacco) leaf according to a well-established prior protocol (Sainsbury et al. 2009). The genes were synthesised to order with a 3xFLAG tag at their C-terminus (connected by a GGG linker) and cloned into a pEAQ-HT-DEST1 expression vector. These constructs were transformed into *Agrobacterium* line GV3101. Tobacco leaves were syringe infiltrated with 3-5ml of *Agrobacterium* suspension, and after 5 days, tobacco samples were snap frozen and ground in 20% methanol for LCMS-based analysis. An empty-vector control was added as a negative control. Of the 26 candidate enzymes tested, none led to the accumulation of detectable levels of 4-Cl-tryptophan (see Appendix **Supplementary Figure S1**) indicating that they are not the tryptophan-4-halogenase. However, non-chlorinated tryptophan was detected in all tested samples. Protein quantification by Western blotting has yet to be carried out at time of writing.

III.IV. Discussion

It was found that addition of the biosynthetic precursors of 4-Cl-IAA to deseeded, immature pea fruit was not able to rescue growth and concluded that the entire biosynthetic pathway of 4-Cl-IAA takes place within the endosperm itself. This is in accord with prior reports that auxin (albeit non-chlorinated IAA) is synthesised in the endosperm of other flowering plants, prior to export into the developing fruit via PIN-FORMED (PIN) auxin efflux carriers[40]. The spatial limitation of auxin synthesis to the developing endosperm can be speculated to ensure evolutionary fitness – it is hypothesised that, as endosperm cells only arise from successful fertilisation events, the highly auxin-responsive fruit tissues rely on endosperm-origin auxin to detect successful fertilisation events and, potentially, to monitor the development of seeds throughout the fruit enlargement (pod elongation) phase. This is supported indirectly by the observation that, across species, mutations in auxin-related genes can lead to parthenocarpy, i.e. a disruption of the parent plant's ability to detect and respond to successful vs unsuccessful floral fertilisation[41]. Local auxin synthesis within fruits is not unheard of, however, but so far has primarily been reported in cases of fruit **metamorphosis** (in which fruit change shape after fertilisation relative to pre-fertilisation ovary shape). This can be seen through comparison of *Arabidopsis thaliana* siliques, which do not locally synthesise auxin and do not undergo post-fertilisation shape change, with *Capsella rubella,* in which the characteristic "heart-shaped" fruit bear shoulders that arise from local synthesis of auxin (Dong et al., accepted, currently in preprint [71]). Whether pea fruit exhibit any fruit-local auxin synthesis remains unknown, but the absence of a post-fertilisation shape change in pea along with the total dependence of deseeded fruit on exogenous 4-Cl-IAA for continued growth suggests that fruit-local auxin synthesis is minor and incapable of substituting for endosperm-origin 4-Cl-IAA.

In pea, it has also been reported that auxin synthesis in seeds during the later, **seed filling**, phase of reproductive growth is mediated by parent plant-derived trehalose-6-phosphate[36], and that this seed-origin auxin acts tissue autonomously to promote the expression of starch biosynthetic enzymes[72]. As such, the role of auxin in pea fruit development can be summarised as beginning by synthesis in the endosperm and likely export into the developing fruit to stimulate fruit growth during pod elongation. Subsequently, at the point of the pod reaching its maximal length (and possibly synchronously with endosperm cellularisation), seed-localised auxin synthesis and signalling result in starch synthesis and the development of the seed's large cotyledons[36,72].

When tested *ex planta* by feeding exogenous tryptophan, pea endosperm converts a detectable quantity of added tryptophan into 4-Cl-tryptophan, preliminarily suggesting the presence of a tryptophan-4-halogenase. So far, no tryptophan-4-halogenase has yet been identified in nature, and all indole halogenases have so far been identified in microorganisms[72]. Further work on the pea endosperm as a possible site for tryptophan halogenation will need to find a way of overcoming the high background levels of 4-Cl-tryptophan in the assay, which prevented robust detection of the halogenation reaction by LCMS. Possible solutions include using an alternative source of endosperm from a non-4-Cl-producing legume and using fractions of pea endosperm protein in this, lower-4-Cl background. It is also important to further refine the LCMS with regular standard curves of 4-Cl-tryptophan and tryptophan, for precise molar quantification of assay outcomes. It would also be essential to ensure that the observed, preliminary halogenation activity can be reproduced at room temperature with a protease inhibitor cocktail, as the lowered temperature of the overnight incubation (at 5^{o}C) may compromise the halogenase's activity. This may also allow for a time-course study of halogenation kinetics.

Of the 26 candidate oxygenases and peroxidases tested by heterologous expression in tobacco, none yielded a detectable level of 4-Cl-tryptophan, which would indicate that none are the endosperm-resident tryptophan-4-halogenase. However, it is also possible that there was an issue with the heterologous expression system, which may require troubleshooting and refinement. Specifically, it is not yet clear if the enzymes were successfully expressed. Protein samples were taken for each infiltration and await western blotting against the 3xFLAG-tagged candidate enzymes to ensure that each of them was expressed (and in-frame). Should this prove to be the case, it is further possible that even if one of the candidate enzymes was the halogenase, that 4-Cl-tryptophan was not detected due to rapid chemical conversion by the tobacco's endogenous enzyme repertoire, though this seems unlikely. Finally, unpublished reports from Kondhare and colleagues at the National Chemical Laboratory, Pune, India, indicate that *Nicotiana benthamiana* may synthesise 4-Cl-IAA endogenously in its root tips. The heterologous expression may therefore be interfered with by the rapid conversion of 4-Cl-tryptophan into 4-Cl-IAA. It is unlikely, however, that even a highly efficient halogenase would erase all detectable traces of 4-Cl-tryptophan, and 4-Cl-tryptophan was not detected in empty-vector control leaves.

Future work on identifying the pea halogenase would be well-served by continued testing of candidate enzymes in tandem with more sophisticated attempts (e.g. affinity chromatography) to isolate the halogenase from pea endosperm.

Identification of the pea halogenase will have at least two important implications for pure and applied science. First, by completing the biosynthetic pathway of 4-Cl-IAA, it would be possible to begin using mutagenic approaches to identify mutants in the halogenase gene for phenotyping. All present evidence that 4-Cl-IAA is important for fruit growth in pea is based on highly manipulated, exogenous auxin-feeding experiments which are likely prone to artefacts. To analyse the phenotype of a plant which is mutated in the 4-Cl-tryptophan halogenase gene would allow for assessment of the pea's inability to produce 4-Cl-IAA, whilst still producing the predominant auxin, IAA. This would allow for conclusive investigation on 4-Cl-IAA's overall significance in pea reproductive development. Should this mutant phenotype compromise fitness (e.g. by rendering a plant infertile) it may reflect an evolutionary trend in which duplicated auxins (or perhaps even phytohormones generally) develop essential functions and become unlikely to be secondarily lost. This may offer a glimpse into the role of signalling molecule diversification in driving morphological complexity in plants, though at present there is no unifying morphological trait that distinguishes 4-Cl-IAA-producing legumes relative to non-4-Cl-IAA-producing legumes.

Furthermore, identification of the halogenase could provide useful insight into the evolution of halogenases in higher organisms in general. Other plant species also produce chlorinated auxins including 5-Cl-IAA (*Rheum*)[26] and 4-Cl-IAA and 6-Cl-IAA (*Anredera*)[25]. The identification of the halogenase in pea may assist in identifying the halogenases of these other plant species too and, thereafter, may facilitate a bioinformatic investigation into how widespread halogenated auxins truly are among plants. Whether the evolution of chlorinated auxin biosynthesis is occurring by true convergent evolution (i.e., in which plants are evolving the biosynthesis via totally different mechanisms) or whether it reflects a parallelism (in which a proto-halogenase is repeatedly evolving halogenase function) remains unknown. Investigation into the chemical diversity of plant signalling metabolites in nature will be an important next step in extending our understanding of plant evolution and may in the longer term prove informative for addressing very fundamental questions about what makes particular signalling molecules evolutionarily advantageous. This may also prove important for the future of synthetic biology, i.e. engineering of new-to-nature or heterologous signalling systems for applied sciences.

Secondly, many prescription drugs, agrochemicals and antibiotics are singly or multiply halogenated. 30% of drugs approved by the FDA in 2021 were halogenated[74], and 96% of pesticides developed between 2010 and 2017 have been halogenated[57]. This is because the addition of halogens can alter the lipophilicity and steric arrangement of

atoms within a given molecule, and thus alter its shape and ability to interact with its molecular target (e.g. increasing affinity for a target of interest or increasing stability by decreasing affinity for degradation enzymes). The discovery of a tryptophan-4-halogenase, therefore, could be an important contribution to the chemical synthesis sector and may affect the ease with which high-value chemicals can be produced. It could perhaps also expand the range of molecular structures available for chemical development.

IV. ACUTILEGUMEN

IV.I. Introduction

In recent years, there has been a stagnation of pea yields in the UK[75]. Despite exhibiting tremendous diversity in pod phenotype[76], little is known about the genetic factors governing pea fruit diversity. We hypothesise that exploration of the genetic basis of pod shape may allow for breeding of new pea lines with specific pod morphologies. Though the relationship between pod shape and yield remains poorly explored, it is possible that breeding pea cultivars with altered pod dimensions may represent an avenue for crop improvement in the future.

ACUTILEGUMEN (from "acute" and "legume", i.e. pointed pod, also known as *Bt*, which is short for "blunt") is a pea pod character that has been of interest to geneticists for over a century. Wild type *ACUTILEGUMEN* results in blunt-tipped pods, while a recessive, loss-of function mutation (*acutilegumen*) causes pointed tips in pods. It was reported in 1904 by William Bateson, Edith Saunders and Reginald Punnett in their report to the Royal Society, entitled "Experimental Studies in the Physiology of Heredity"[77]. In it, Bateson (later Director of the John Innes Horticultural Institution) writes:

"There is great diversity in the shapes characterising the pods of various types, but they may be classified into those known as "stumpy", which, when the pod is full, have a blunt rounded end, and those which are *pointed*. We have found the former always dominant over the latter, and F_2 segregates normally. This observation of the dominance of the stumpy over the pointed types is in agreement with the similar results obtained by Tschermak and will almost certainly prove to be of wide general application, and of value to the practical breeder."

(Note: Erich Tschermak was an Austrian plant geneticist and one of the three independent "rediscoverers" of Gregor Mendel's principles of heredity from the latter's 1865 paper 'Versuche über Pflanzen-Hybriden', Experiments on Plant Hybridisation.)

Bateson believed that "stumpy" and "pointed" pod characters would prove valuable to the practical breeder, but it is unclear if this has proven to be the case thus far. The lack of identification of the *ACUTILEGUMEN* gene itself is likely one reason why pod apex is poorly characterised in the agricultural context, though previous work has mapped this gene to the end of chromosome 3 (linkage group V) [76,78]. During the course of the PhD

corresponding to this thesis, interactions with plant breeders and agronomy researchers have revealed two unpublished, unverified observations related to pod apex shape. First, that processing of pea pods for the production of frozen fresh (garden) peas involves large-scale opening of harvested pods by mechanical means. During this processing, pointed pod apices have been reported to be more likely to break off the pod than blunt apices. The resulting pod shard is then of roughly similar dimensions to an individual pea and becomes impossible to separate from the frozen peas thereafter, lowering the quality of the resultant food product. Secondly, it has been reported that abscission of petals is hindered by pointed pod apices, as the pointed tip pierces and retains the abscised petal on the growing pod. This results in the petal remaining on the tip as the pod grows, increasing humidity around the pod apex and increasing the risk of fungal infection and subsequent tip rot. Neither of these observations have been studied in rigorous field experiments, likely due to the absence of near-isogenic blunt and pointed podded lines and the subtlety of the phenotype.

Here, on the 120th anniversary of its reporting to the Royal Society, an investigation into the characterisation of *ACUTILEGUMEN* is described.

IV.II. Materials and Methods

1. Plant Materials

Pea plants were cultivated as described above (Materials and Methods, Section II.II.).

2. Recombinant Inbred Line (RIL) Mapping

An initial list of 21 RILs with recombination events close to the end of chromosome 3 was generated by Noel Ellis and sown for cultivation. Pods at the end of pod elongation were phenotyped by visual examination and direct comparison with developmentally equivalent pods from JI3253 and JI281. Phenotyping was carried out by two independent researchers (myself and Julie Hofer) and results compared for confirmation of *ACUTILEGUMEN/ acutilegumen* allocation.

For phenotyping of the entire RIL population, phenotyping was carried out alone.

3. RNA-Sequencing

For RNA-seq of JI281 and JI3253, pods were harvested at different developmental stages. “Young” pods were harvested 2 days after anthesis (DAA), “Middle” pods were harvested at 5 DAA and “Old” pods were harvested at 7 DAA, this latter timeframe corresponding to the completion of pod elongation. Two to three pods were pooled for each biological replicate. Pods were placed into liquid nitrogen to snap freeze.

For RNA-seq of JI2822 and FN3241/187, early stage, developing ovaries were excised with tweezers and placed into pre-chilled, open Eppendorf tubes on dry ice to snap freeze. These were harvested 3 days prior to anthesis (-3 DAA) and 20-25 ovaries used for each biological replicate.

RNA extraction, RNA-Sequencing and analysis was carried out as described above (Section II.II.).

4. Genomic DNA (gDNA) sequencing

Whole seedlings of JI2822 and FN3241/187 were grown to the 2-week stage and then placed into an opaque black bag for 48 hours to induce chlorosis and reduce tissue sugar content. Whole DNA was then extracted by lab assistant Neil McKenzie and subjected to QC protocols. Passing DNA preps were sent to Novogene for sequencing in line with their concentration and quality requirements.

DNA reads were mapped to JI2822 v1.0 and JI2822 v1.2 using Burrows-Wheeler Alignment (*bwa*) and viewed with Integrative Genomics Viewer (IGV). Manual analysis of *ACUTILEGUMEN* sequences was performed using Microsoft Word.

IV.III. Results

1. JI3253 and JI281 exhibit different pod shapes due to *acutilegumen* mutation

JI3253 and JI281 are pea lines available from the John Innes Centre Germplasm Resource Unit (GRU). JI3253 is the French Cultivar Caméor, while JI281 is a *Pisum sativum* line from Ethiopia with a substantial genomic difference relative to modern European cultivars. JI3253 has previously been crossed to the type line for *acutilegumen* JI799 (data unpublished) and all F1 exhibited pointed pod tips, indicating that JI3253 is a true *acutilegumen* mutant. Direct comparison of JI3253 and JI281 revealed that they exhibited substantially different pod-tip shapes (**Figure 4.1**).

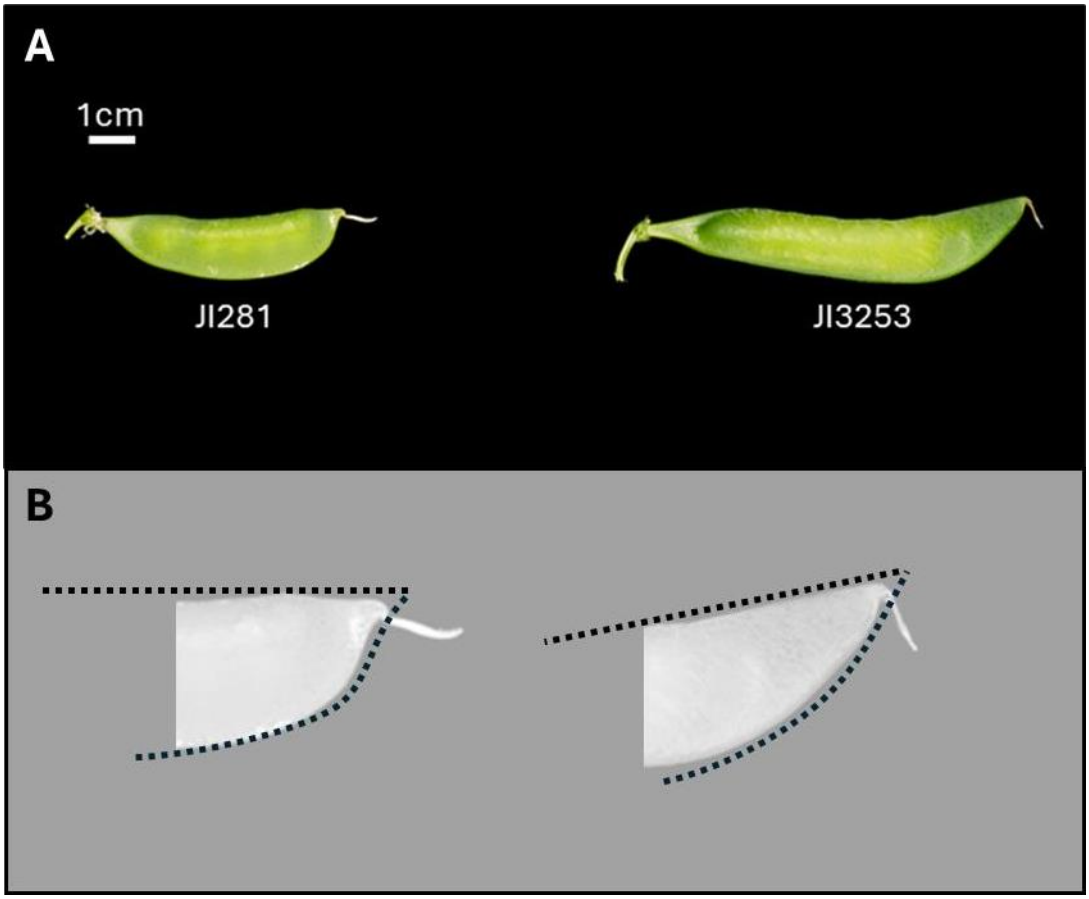


Figure 4.1: Typical pod shapes of JI281 and JI3253 at the completion of seed filling. A - Whole pods of each line, B - Zoom-in of pod apex shape from JI281 (left) and JI3253 (right).

As JI281's pods were blunt-tipped relative to JI3253, it was decided that JI281 was phenotypically wild-type (WT) with respect to pod apex and the *ACUTILEGUMEN* gene.

2. Genetic mapping reveals the presence of a 3Mb region for *ACULTILEGUMEN* at the end of chromosome 3

As described in a previously published paper[79] a recombinant inbred line (RIL) population was established between JI281 and JI3253 for genetic mapping of traits of interest. This population had been densely genotyped at the F5 level though has now been updated more recently with newer (F7) genotyping data.

As it was known from previous publications that acutilegumen mapped to the end of chromosome 3 (also known as linkage group V)[76,78], a list of 21 RILs with crossover events close to the end of chromosome 3 was generated with the assistance of Ellis et al. These RILs were phenotyped in the F6 generation for blunt vs. pointed-tipped pods by independent visual examination by two different researchers, which agreed. This revealed two blunt-podded RILs, RIL450 and RIL442, whose crossover events bound a limited region of 6 Axiom markers (**Figure 4.2**).

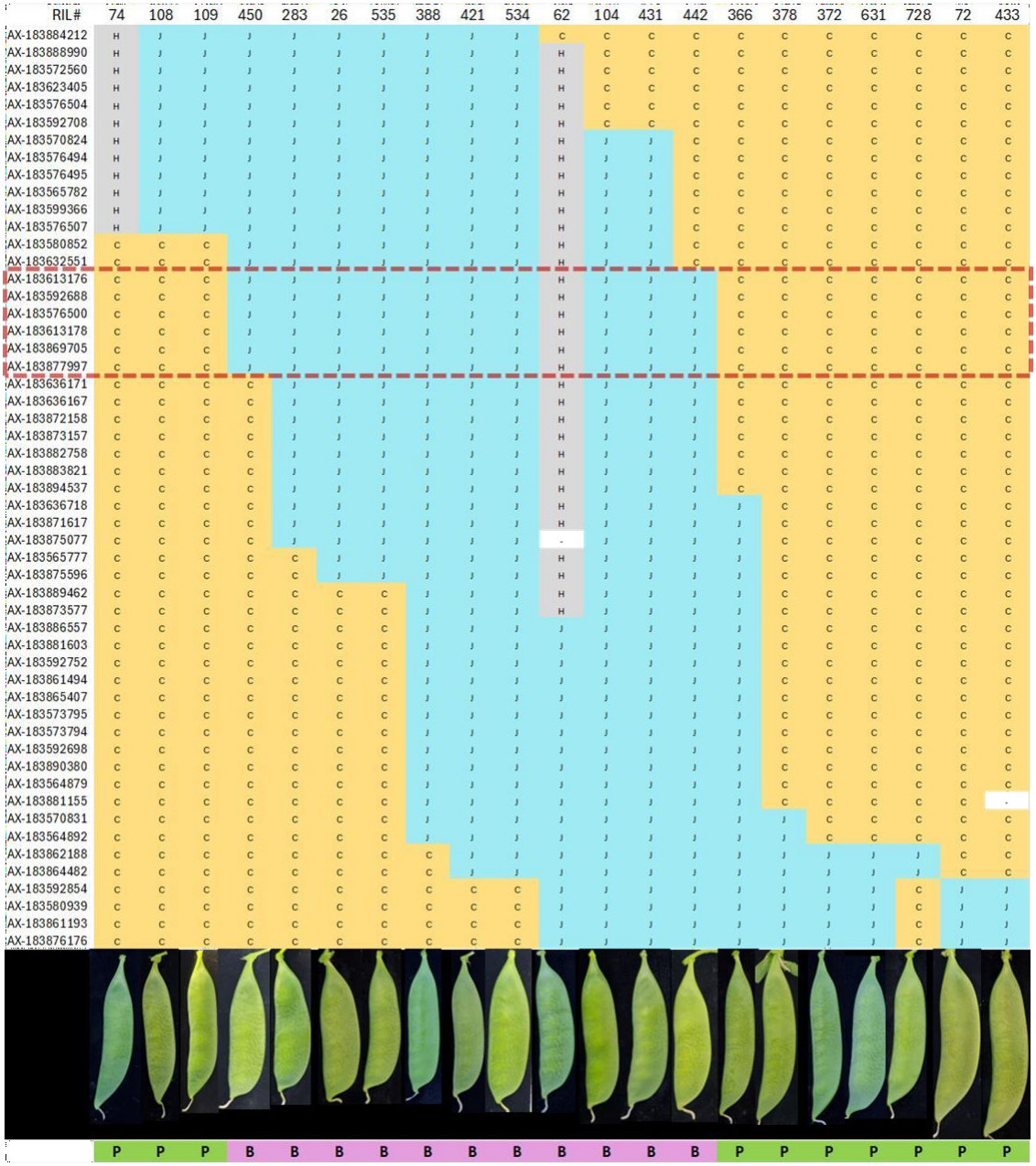


Figure 4.2: Mapping of *acutilegumen* via phenotyping of 21 JI3253xJI281 recombinant inbred lines (RILs), with representative photos of pods from each RIL. Each coloured column represents one RIL, with RIL number in the top row. AX – Axiom Marker number, 'P' – pointed pod tip, 'B' – blunt pod tip. Uppermost axiom marker (AX-183884212) is at the end of chromosome 3 (at position 1004647bp in JI2822v1.0). Subsequent rows indicate marker positions that are deeper into chromosome 3. Red box highlights markers within crossover events in RIL450 and RIL442, which are the RILs at which the phenotype shifts (bottom two rows). Blue cells 'J' are positions in the chromosome where the Axiom marker is JI281-type, while orange cells 'C' are positions where the marker is JI3253-type.

Due to the subtlety of the phenotype, an additional 486 RILs (with recombination events randomly distributed across the entire genome) were scored in the F7 generation for blunt vs. pointed pods (**Figure 4.3**). These were designated as either confidently blunt (223 RILs) or pointed-tipped (241 RILs) by visual examination, or otherwise as too ambiguous to score (22 RILs). After phenotyping, the recorded pod shapes of each RIL were compared to the genetic markers within the genomic location established for *ACUTILEGUMEN*. Of the 464 RILs which were confidently phenotyped as either blunt or pointed-tipped, 448 agreed with the established genomic location of *ACUTILEGUMEN* as described in **Figure 4. 2**, while only 16 RILs had a genotype that did not match the expectation. Were the established genomic location unlinked to *ACUTILEGUMEN*, it would be expected that 50% of all RILs would have a genotype that did not match their predicted genotype, and as such the observed 97% agreement rate supported the established genomic position of *ACUTILEGUMEN*.

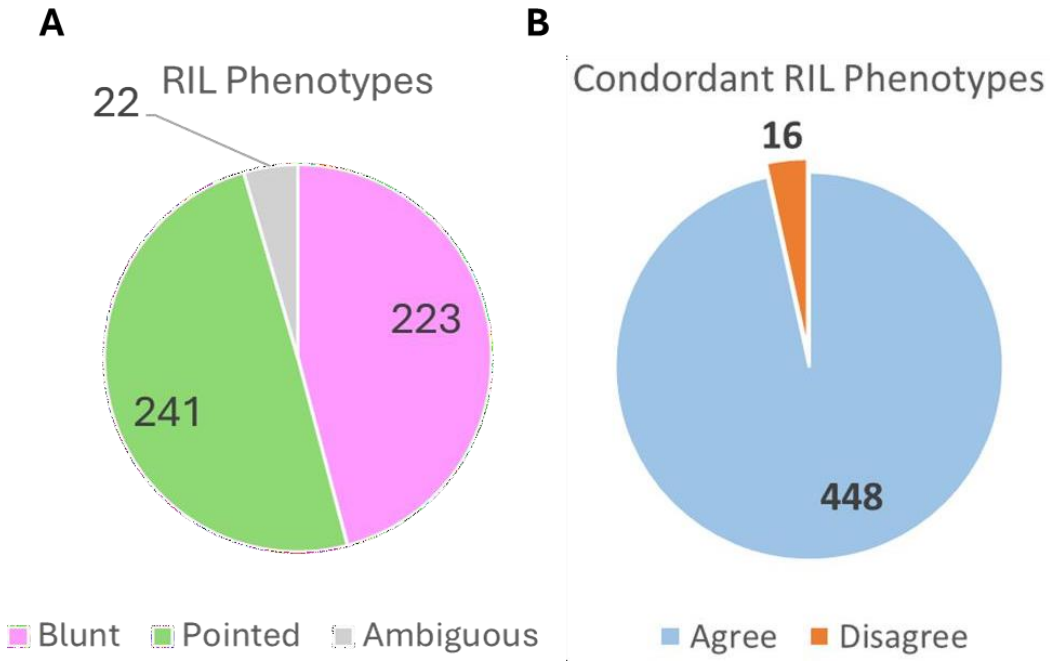


Figure 4.3: Scoring of larger, random RIL set for blunt vs pointed pods. A – number of RILs designated as blunt-tipped, pointed-tipped, or ambiguous. B – number of RILs whose phenotype agreed with the previously established genomic position of *ACUTILEGUMEN* as described in **Figure 4.2**.

Re-examination of RILs which disagreed with the established genomic location (RILs 6, 64, 116, 292, 344, 393, 445, 460, 3, 334, 347, 422, 424, 425, 681, 704) revealed pod shapes that were sufficiently ambiguous that they could be designated according to either phenotype, or categorically mis-phenotyped lines.

RIL450 and RIL442's bound region is highlighted by a red box in **Figure 4.2** and is flanked on the outside by Axiom markers AX-183632551 and AX-183636171. In JI2822v1.0, these mapped to positions 3,880,891bp-6,926,081bp on chromosome 3, giving a physical region of 3,045,190bp for *ACUTILEGUMEN*'s genomic position. In JI2822v1.2, the chromosome was flipped and so the axiom markers instead mapped to 531,728,047bp-534,767,378 though the genic content of the region did not change.

Between positions 531,728,047-534,767,378 there were 19 genes that mapped from the publicly available Caméor v1a annotation[20]. These genes are listed in **Table 4.1**. These genes thus became candidates for *ACUTILEGUMEN*.

Table 4.1: List of 19 candidate *ACUTILEGUMEN* genes derived from mapping of pod apex shape in the RIL population, with description from Caméor v1a gene annotation.

Gene	Description
PSAT3G000600	Domain of unknown function
PSAT3G000640	Inorganic pyrophosphatase
PSAT3G000680	Ribonuclease III domain
PSAT3G000720	Myb-like DNA-binding domain
PSAT0S2325G0040	O-methyltransferase
PSAT0S1614G0120	O-methyltransferase
PSAT0S1614G0080	Uknown gene
PSAT4G105240	Uknown gene
PSAT0S3721G0160	Hydrolase activity + acting on acid anhydrides + in phosphorus-containing anhydrides
PSAT0S1908G0040	Intracellular membrane-bounded organelle
PSAT5G200280	Late embryogenesis abundant protein
PSAT5G200320	Response regulator receiver domain
PSAT7G187960	Uknown gene
PSAT0S362G0040	RNA recognition motif. (a.k.a. RRM + RBD + or RNP domain)
PSAT0S3525G0040	COG4 transport protein
PSAT0S3525G0080	Dehydrogenase E1 component
PSAT0S3535G0120	60S acidic ribosomal protein
PSAT0S3535G0160	Unknown gene
PSAT0S359G0040	B3 DNA binding domain

3. RNA-sequencing reveals differences in expression of *ACUTILEGUMEN* candidates genes between JI281 and JI3253 (Caméor)

As the *acutilegumen* mutation was recessive, it was speculated that it was most likely caused by the loss-of-function mutation of a gene with a functional role in reproductive development. It was therefore hypothesised that one mechanism by which *acutilegumen* could have been rendered nonfunctional could have been loss of its expression relative to wild-type. As such, pods at different ages (young, mid and old) were harvested from JI3253 and JI281 for RNA-seq. None of the 19 candidate genes showed total absence of expression between JI3253 and JI281, though there were some significant differences (**Figure 4.4**). Only two genes, *PSAT3G000600* and *PSAT0S362G0040*, showed different expression levels across all pod developmental stages, though most genes showed significantly different levels of expression within at least one developmental stage.

Interestingly, there were large numbers of genes across developmental stages which showed total absence of expression in one of the two genotypes (visible as “arms” of

MA plot in **Figure 4.5**). These "arms" include genes which are only expressed in JI281 (pink) which would be the most natural home for a very strong *ACUTILEGUMEN* candidate, but none of the 19 candidate genes exhibit a genotype-exclusive expression pattern. JI281-exclusivity seems to break down as pods age (**Figure 4.5 C**) but across all stages there are at least some genes which exhibit genotype-exclusive expression.

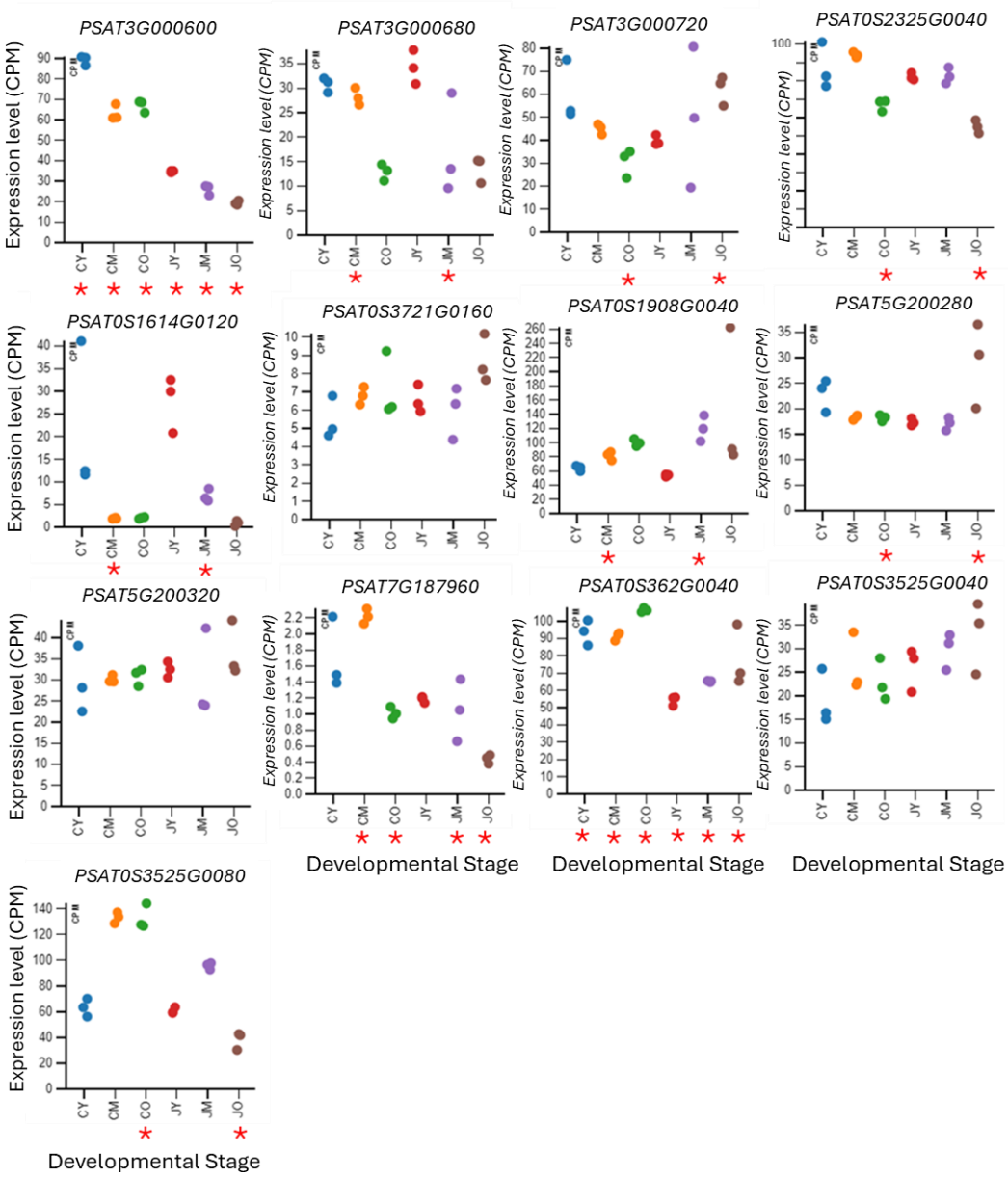


Figure 4.4: Transcript abundance (counts per million) for candidate *ACUTILEGUMEN* gene expression profiles when comparing mutant pod transcriptomes (Caméor, JI3253) to wild-type pod transcriptomes (JI281) across pod development. 3 biological replicates. 'C' – Caméor/JI3253, 'J' – JI281, 'Y' – young, 'M' – mid, 'O' – old, i.e. 'JM' indicates expression values belonging to mid-elongation stage JI281 pods. Caméor v1a gene names are presented above each plot. Expression levels were compared pairwise between genotypes but within developmental stages (i.e. CY vs. JY, CM vs. JM and CO vs. JO), red star indicates that the given expression level is significantly different from the expression level in the same developmental stage in the other genotype (pairwise Student's t test with Benjamini–Hochberg procedure, n=3, $q<0.05$). Of the 19 candidate *ACUTILEGUMEN* genes described in **Table 4.1**, 6 genes which showed no expression across any replicate are excluded from this figure.

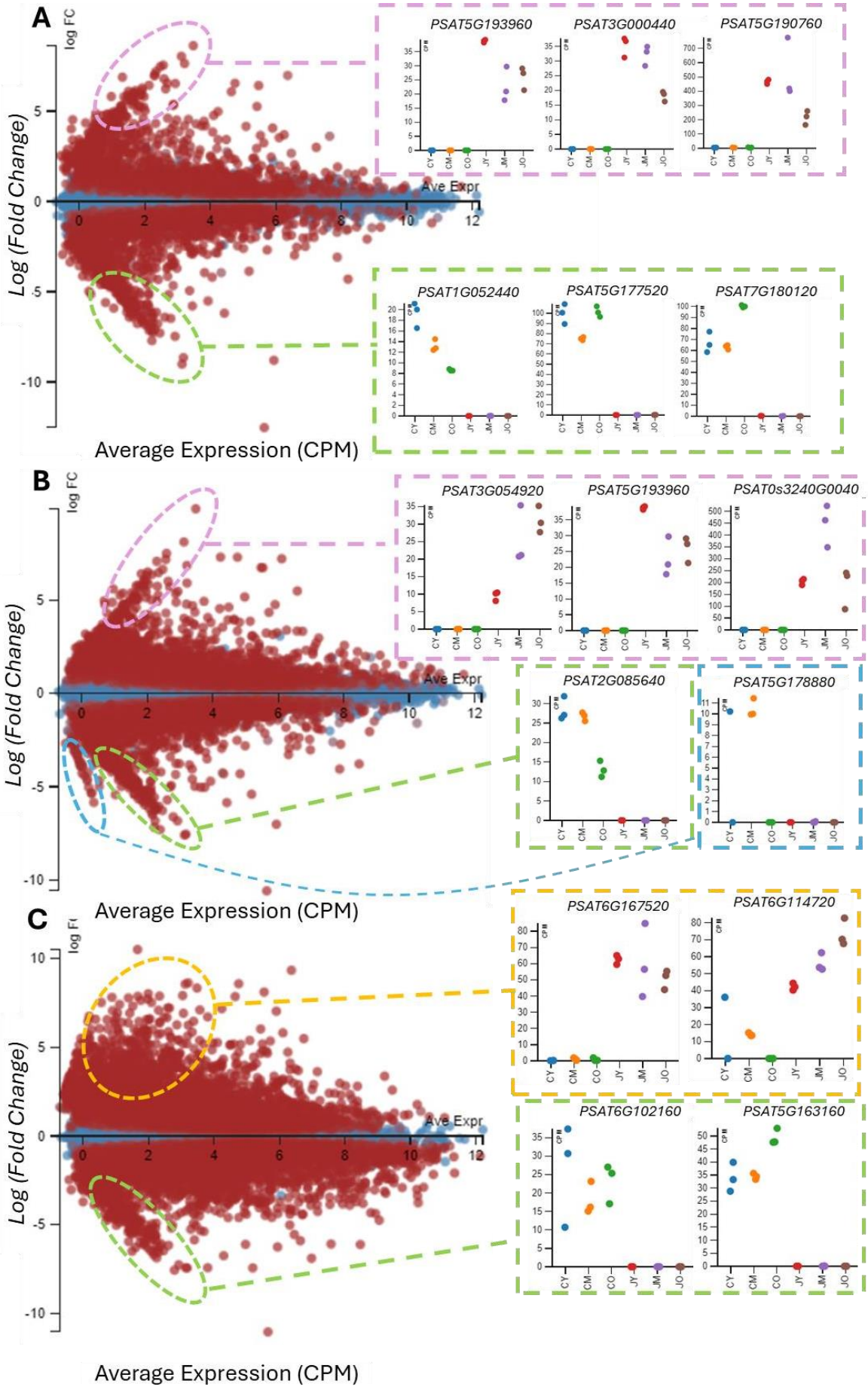

A
log FC
Log (Fold Change)
Ave Expr
Average Expression (CPM)
PSAT5G193960
PSAT3G000440
PSAT5G190760
PSAT1G052440
PSAT5G177520
PSAT7G180120
B
log FC
Log (Fold Change)
Ave Expr
Average Expression (CPM)
PSAT3G054920
PSAT5G193960
PSAT0s3240G0040
PSAT2G085640
PSAT5G178880
C
Log (Fold Change)
Ave Expr
Average Expression (CPM)
PSAT6G167520
PSAT6G114720
PSAT6G102160
PSAT5G163160
CY
CM
CO
JY
JM
JO

Figure 4.5 (page 76): Overview of JI3253 and JI281 pod transcriptomes over developmental time. A – MA plot of pod gene expression in young pods, B – MA plot of pod gene expression in middle-stage pods, C – MA plot of pod gene expression in old pods, y-axis is log(fold change) and x-axis is average level of expression for the given gene (mean transcripts per million). Red points indicate genes that are significantly differentially expressed between JI281 and JI3253 (pairwise Student's t-test with Benjamini-Hochberg correction, n=3, $q<0.05$), blue points indicate genes where no significant difference was detected between the two genotypes. Pink boxes indicate typical expression profiles for genes which are expressed only in JI281, green boxes indicate typical expression profiles for genes which are expressed only in JI3253, blue box indicates genes which are primarily expressed in mid-stage JI3253, and orange box indicates general zone in which genes are expressed in JI281 and expressed significantly less (but at a non-zero level) in JI3253. Individual gene transcript abundance (counts per million) is presented.

As none of the *ACUTILEGUMEN* candidate genes showed total loss of expression in either line (and several were not detected at all) it was difficult to conclusively rule on which candidate was the most likely. At the time of writing, the only TARGETED INDUCED LOCAL LESIONS IN GENOMES (TILLING) population available for *Pisum* was itself in a JI3253 background, making testing phenotypes for induced mutation in each of the *ACUTILEGUMEN* candidate genes impossible (as the progenitor Caméor is already an *acutilegumen* mutant).

4. Identification and verification of a *de novo acultilegumen* mutant in a JI2822 background

During the mapping in JI281 and JI3253, a fast-neutron (FN) population in a JI2822 background was sown at John Innes and grown to maturity. JI2822 is wild-type with respect to pod apex shape, however visual phenotyping of this population led to the identification of a pointed-tipped mutant, designated FN3241/187. FN3241/187 pods were compared against wild-type JI2822 (**Figure 4.6**) and their pointedness observed to originate from as early as 3 days before anthesis. As such, it was speculated that the *acutilegumen* phenotype may originate from defects during gynoecium primordium formation.

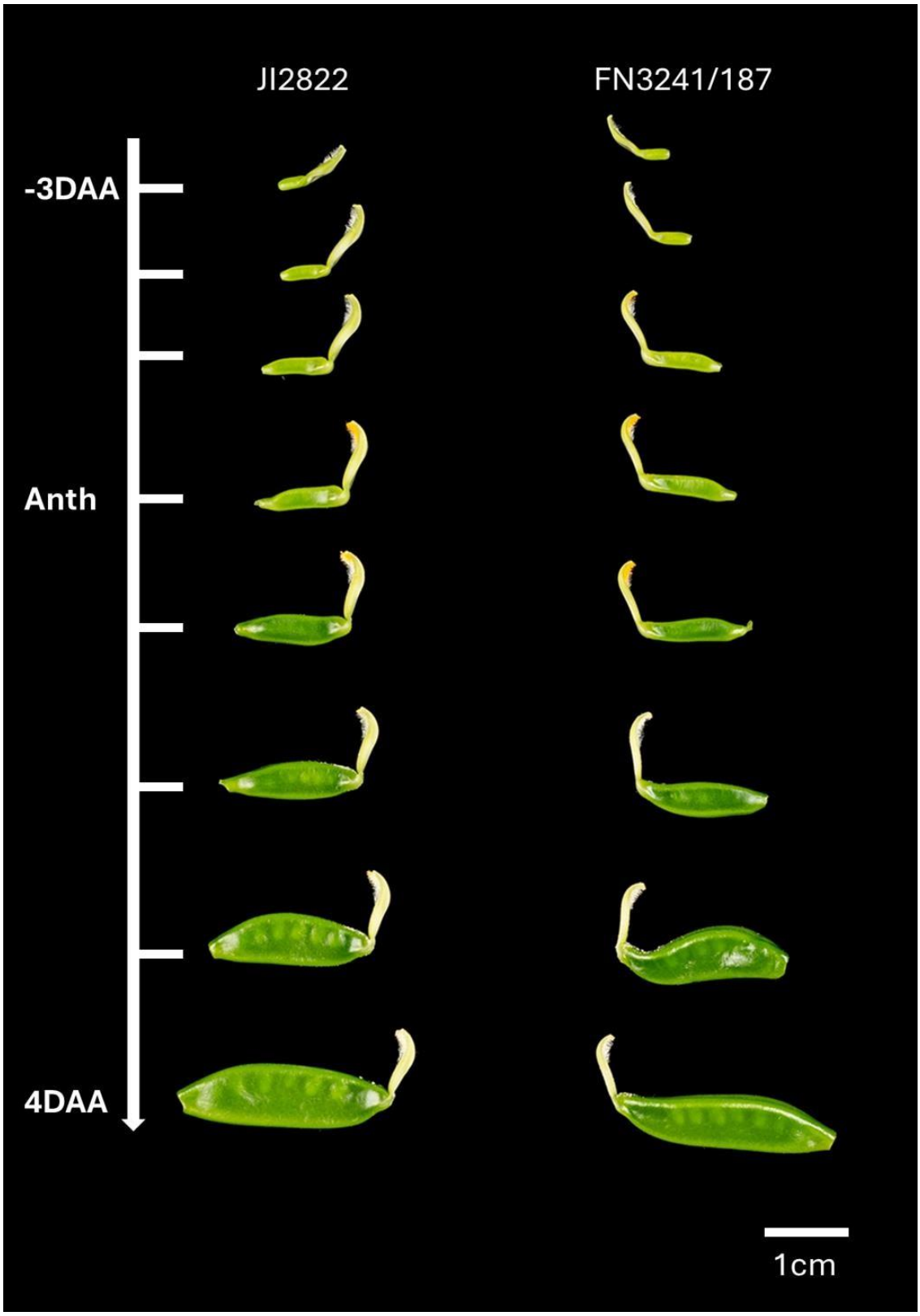


Figure 4.6: Developmental series of JI2822 (left) and FN3241/187 (right) fruit from 3 days prior to anthesis (-3DAA) to 4 days after anthesis (4DAA). Anthesis (Anth) is defined as full reflex of petals.

To ensure that FN3241/187 was a true, *de novo acutilegumen* mutant, an allelism test was performed using JI3253 (**Figure 4.7**). Crossing JI3253 to JI2822 led to F1 plants with blunt-tipped pods (as would be expected for a heterozygous plant *ACUTILEGUMEN/acutilegumen*) but, by contrast, when JI3253 was crossed to FN3241/187, the F1 plants exhibited pointed-tipped pods, which would only be possible if they exhibited the genotype *acutilegumen/acutilegumen*). As such, JI3253 and FN3241/187 were concluded to be **allelic** to one-another, and FN3241/187 was identified as a mutant of *acutilegumen* relative to JI2822.

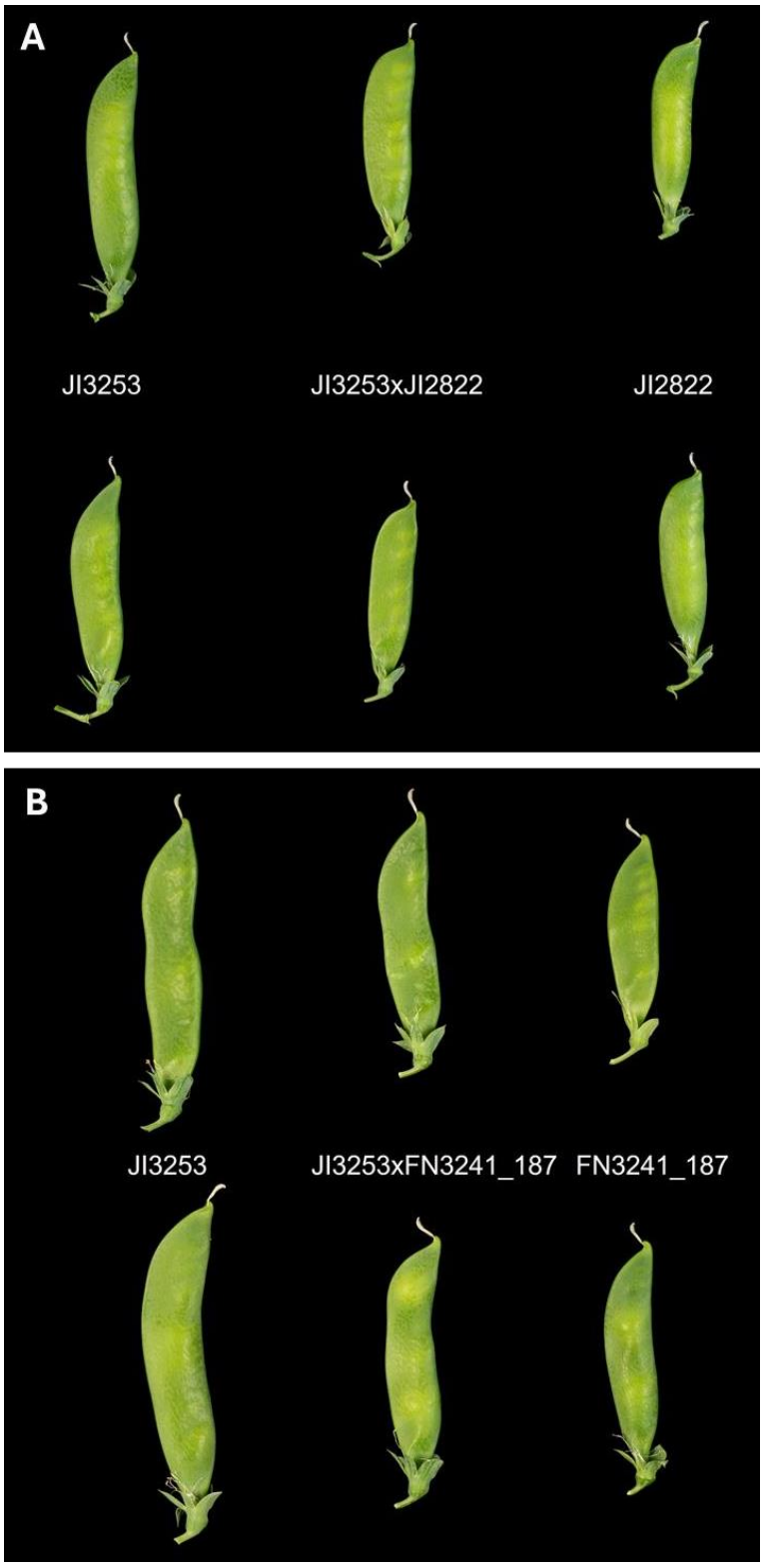


Figure 4.7: Allelism test of putative *acutilegumen* mutant FN3241/187 by examination of pod shape. Leftmost column of two pods are JI3253, rightmost column of two pods are either JI2822 (A) or FN3241/187 (B). Middle two pods are F1 progeny of cross. JI3253xFN3241/187 F1 pods are pointed-tipped, in contrast to JI3253xJI2822 F1 blunt-tipped pods.

5. Mapping of whole-genome data identifies a single base substitution mutation as a strong candidate for *acutilegumen*

As a genomic location was established for *ACUTILEGUMEN* (from 531,728,047-534,767,378 in JI2822 v1.2) and a pair of lines with *acutilegumen* vs. wild-type (FN3241/187 vs JI2822) was identified, it was determined that genomic DNA-sequencing should allow for fine-mapping of the *acutilegumen* mutation in FN3241/187. This was pursued despite the caveat that FN3241/187 was not backcrossed to JI2822, and so likely exhibited a number of other background mutations in its genome which were unrelated to the *acutilegumen* phenotype. This was facilitated by the pre-established region for *ACUTILEGUMEN* from mapping in Cameor and JI281. Both wild-type JI2822 and mutant FN3241/187 genomic DNA samples were sent for paired-end sequencing. The sequencing data was mapped to the JI2822 v1.2 reference genome using software *bwa* for alignment of reads, and the stretch of chromosome 3 defined by the RIL mapping was explored for genomic differences between JI2822 and FN3241/187.

A single C→T single base substitution (SBS) mutation was identified at 533,554,597, and no other mutations were identified in the defined region of chromosome 3 (see **Figure 4.8** for view of mutation in IGV). Therefore, it was concluded that this SBS mutation was the cause of the observed *acutilegumen* phenotype, and was either in, or was close to, the *acutilegumen* gene.

Examination of the genomic context for this mutation revealed that there were no annotated genes in close proximity. The nearest gene upstream of the mutation was *PSAT3G000600*, at position 533,053,303 (501,294bp upstream) whilst the nearest gene downstream was *PSAT0S359G0040*, which was located near the end of the feasible region at 534,764,000 (1,209,403bp downstream, **Figure 4.8**). Both genes are far beyond the range of any known promoter or enhancer element, and so it was inferred that the observed SBS likely wasn't affecting pod shape by cis-regulatory changes on the expression of *PSAT3G000600* or *PSAT0S359G0040*. With no annotated genes located in proximity to the SBS, it was decided that transcriptomics might yield novel transcripts that map near to the SBS.

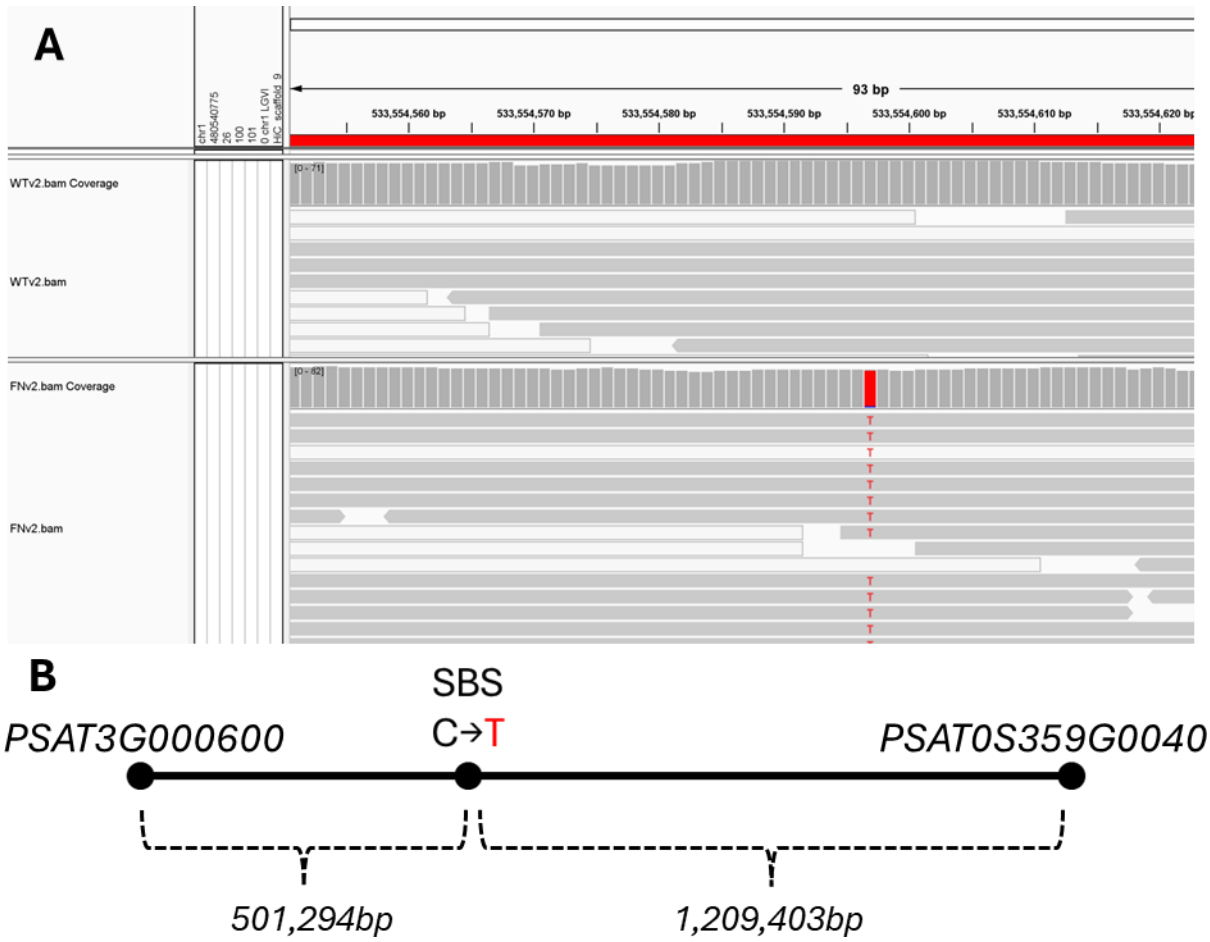


Figure 4.8: Detection of a single-base substitution mutation at position 533,554,597 on chromosome 3 in JI2822 v1.2. A – IGV view of mutation. Upper set of gDNA reads belong to wild-type JI2822 (WTv2) while lower reads belong to FN3241/187 (FNv2). Red 'T' indicates thymine base which mismatches from reference JI2822 genome (cytosine), found only in FN3241/187 DNA reads. B – Schematic representation of the mutation's location (SBS) relative to nearest annotated flanking genes.

6. RNA-sequencing does not identify transcripts mapping to the causative mutation, but suggests an auxin-related mechanism for pod apex change

In order to more properly characterise the SBS mutation differentiating FN3241/187, RNA-seq was performed to look for transcripts that might map to the SBS. As the phenotypic differences between JI2822 and FN3241/187 appear very early during gynoecium development (**Figure 4.6**), it was decided to carry out RNA-seq using very early-stage ovaries (-3DAA). Ovaries were collected and pooled together to form each biological replicate, and RNA-seq data were analysed using hisat2 for mapping.

No transcripts were detected mapping closely to the SBS mutation, though expression was detected at several of the annotated genes within the defined region of chromosome 3. A pair of unannotated, near-perfect repeat transcripts was, however, detected far away from the SBS (1.6Mb upstream, mapping to positions 531,925,000 and 531,937,000). Curiously, these unannotated genes showed expression in wild-type JI2822 ovaries, but almost total loss of expression in FN3241/187 (**Figure 4.9** for IGV view). Re-mapping in both JI2822 v1.0 and v1.2 did not lead to any structural rearrangement, and so it was concluded that the unannotated DEG was too far from the SBS to be affected in *cis*. However, the presence of a DEG between wild-type and FN3241/187 was encouraging. BLASTing (blastn) of these transcript sequences against the NCBI Viridiplantae database yielded only uncharacterised *Pisum* sequences described as "PREDICTED: *Pisum sativum* uncharacterized LOC […] ncRNA", a computationally predicted annotation from the sequenced pea line ZW6[80]. The lack of proximity of the transcripts to the SBS and lack of functionally relevant annotation led to abandonment of these paired transcripts for further study.

The presence of at least one DEG motivated a global search for DEGs across the genome. Analysis of transcript abundances of annotated genes revealed 125 DEGs at $q<0.05$. Three major clusters of DEGs (**Figure 4.10**) were observed. These included:

1- Genes which were substantially more expressed in JI2822
2- Genes which were substantially more expressed in FN3241/197
3- Genes which were mildly less expressed in FN3241/187

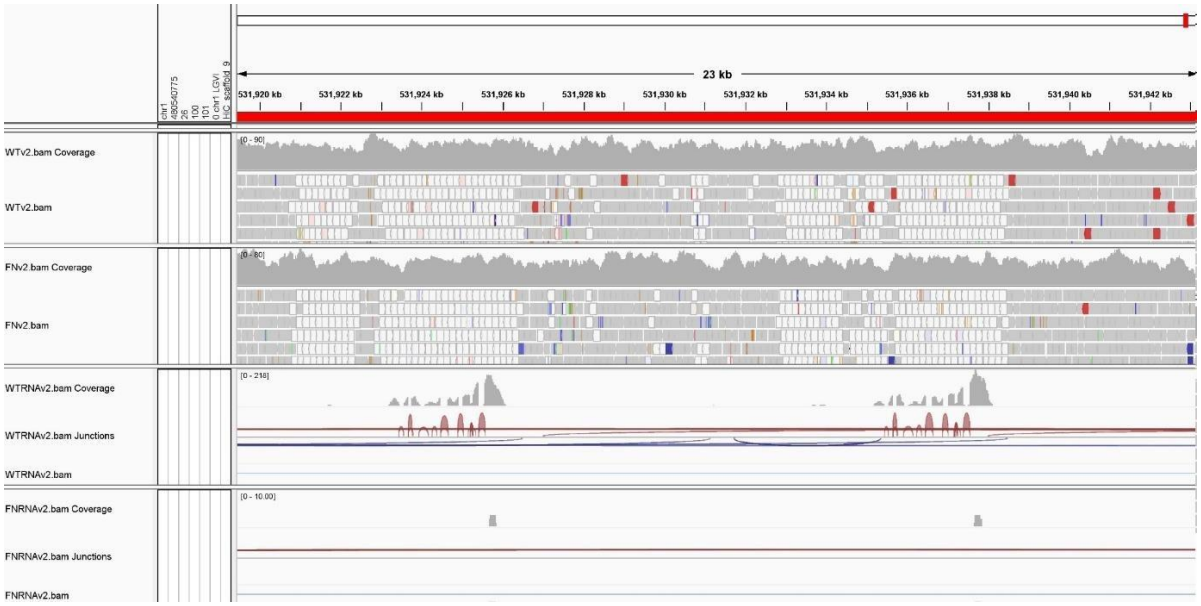


Figure 4.9: Detection of directly repeating, differentially expressed transcripts at positions 531,925,000 and 531,937,000 on chromosome 3 in JI2822 v1.2. Transcripts are visible in “WTRNAv2.bam Coverage” and “FNRNAv2.bam Coverage” as grey peaks. Note reduced peak height in FN.

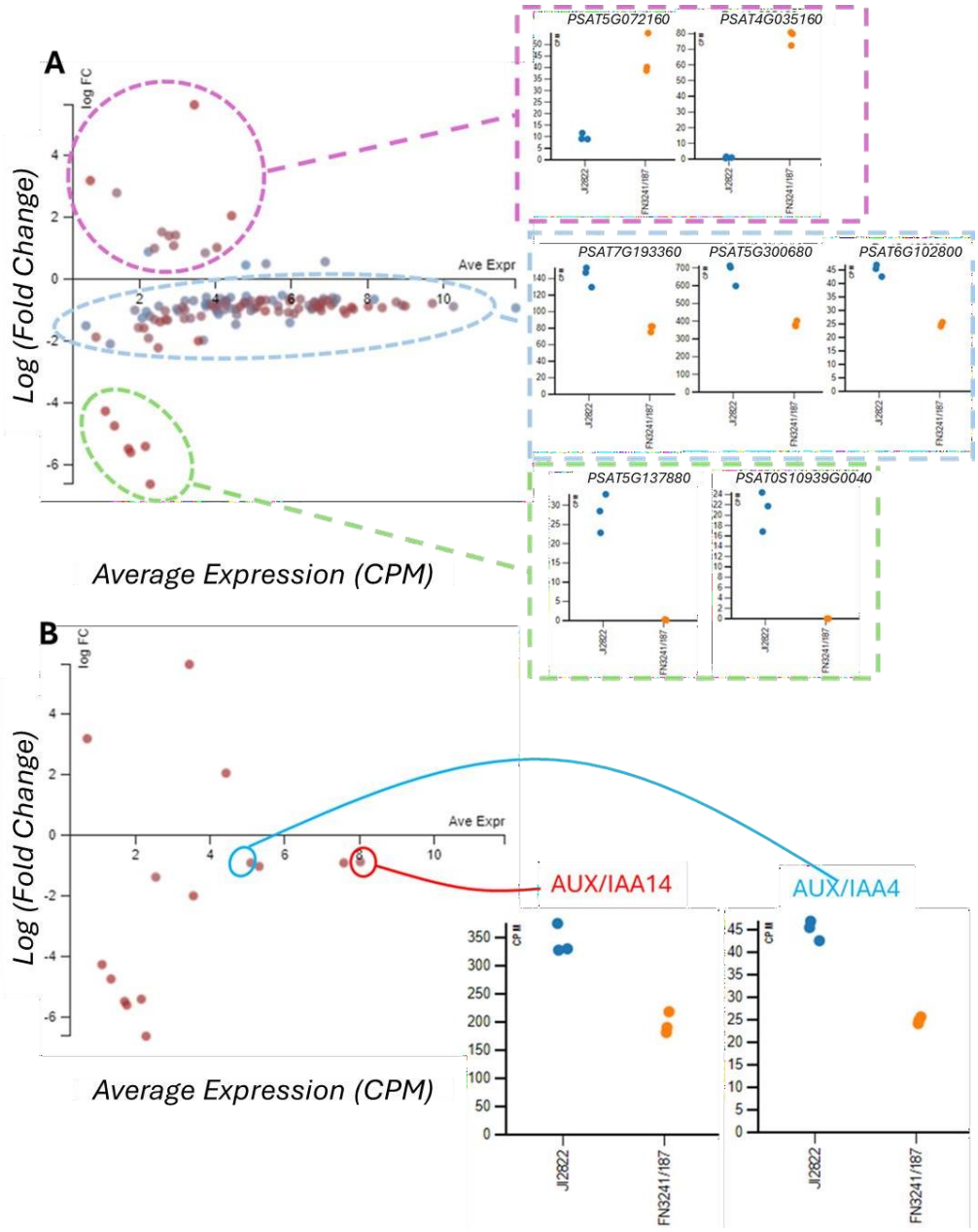


Figure 4.10: Overview of JI2822 and FN3241/187 ovary transcriptomes. DEG analysis at A – MA plot of ovary gene expression at q<0.05, B – MA plot of ovary gene expression at q<0.01 with highlighted expression profiles of *AUX/IAA14* (red) and *AUX/IAA4* (blue), y-axis is log(fold change) and x-axis is average level of expression for the given gene (mean transcripts per million). All points indicate genes that are significantly differentially expressed between JI2822 and FN3241/187 (pairwise Student's t-test with Benjamini-Hochberg correction, n=3, q<0.05 or q<0.01). Pink box indicates genes which are more expressed in FN3241/187, blue box indicates genes which are slightly less expressed in FN3241/187, green box indicates genes which are substantially less expressed in FN3241/187. Individual gene transcript abundance (counts per million) is presented.

Manual functional annotation of highly significant DEGs (at q<0.01) is represented in **Table 4.2**. Notably, transcriptional repressors *AUX/IAA4* and *AUX/IAA14* were significantly less expressed in FN3241/187 than in JI2822. Given AUX/IAAs are central repressors of the canonical auxin signal transduction pathway, and have a well-established role in plant developmental processes, this led to the hypothesis that the reduction in expression of these AUX/IAAs may be directly upstream of the altered shape of the gynoecium apex. This does assume that there is a correlation of an inferred, lowered protein abundance for the AUX/IAAs in accord with the lowered expression in FN3241/187. Further, the 125 DEGs at q<0.05 were assumed to be downstream of the SBS mutation, though the mechanism by which the SBS site could be regulating these genes *in trans* remained elusive.

Table 4.2: List of 15 DEGs at q<0.01 when comparing transcriptomes of JI2822 and FN3241/187, with manual functional annotation. Mapping in JI2822v1.2.

Gene ID	Function	Chromosome	Change in FN rel to WT
PSAT4G003960	unknown	2	down
PSAT0S10939G0040	snakin-2-like? (antimicrobial peptide)	7	down
PSAT4G035160	ESCRT-related	4	up
PSAT5G137880	GLR3.6-like	5	down
PSAT5G137920	GLR3.6-like	5	down
PSAT5G269800	DEK domain "MYB-like"	5	up
PSAT0S2524G0040	unknown "ncRNA"	6	down
PSAT5G086800	No entries on NCBI	5	down
PSAT3G116440	Vestitone reductase (phytoalexin biosynthesis)	3	down
PSAT5G072160	unknown	5	up
PSAT7G061200	basic blue copper protein	7	down
PSAT4G218720	CER1-like (wax biosynthesis)	4	down
PSAT5G308080	ACT-domain (glutamine synthetase feedback regulation)	5	down
PSAT6G057880	**Aux/IAA14**	6	down
PSAT6G102800	**Aux/IAA4**	6	down

7. *ACUTILEGUMEN* is descended from an LTR retrotransposon

Manual annotation of the sequence peripheral to the SBS revealed it to be located in one of two long, repeated sequences. BLASTing against the NCBI database revealed the sequence surrounding the SBS to be related to "*Pisum sativum* peabody/Ty3-type retrotransposon gag-pol precursor (gag-pol) pseudogene" (annotation from *P. sativum* cultivar 'Alaska'). As such, the raw sequence surrounding the SBS was manually searched for repeating sequences, which revealed two near-identical direct repeat sequences of 1993bp and 1990bp (a 3bp insertion within a microsatellite extending the longer of the two). These repeats were located at the termini of a stretch of sequence, and as such, it was hypothesised that they may have been long-terminal repeats (LTRs) of an LTR retrotransposon-like sequence. Translation of the sequence between the two LTRs yielded 21 short (2 - 129 amino acid) open reading frames in the 5' to 3' direction, but translation in the 3' to 5' direction yielded 3 open reading frames (in the first frame) of 311, 703 and 388 amino acids in length. As such, it was hypothesised that the coding sequence for this LTR-like sequence ran in the 3' to 5' sequence.

Overall, the entire LTR element-like sequence (including the LTRs) was 8519bp in length, with the initial LTR 1993bp long. Within this first LTR, at position 1065bp, was the location of the G→A SBS mutation in FN3241/187 (**Figure 4.11, now viewed from opposite strand**). Additionally, alignment of the upper and lower LTRs revealed several other positions at which the first LTR mis-matched the second LTR. As such, FN3241/187's *acutilegumen* differed from JI2822's *ACUTILEGUMEN* in that the former carries one extra mis-matched base between its two LTRs. Alignment of the upper and lower LTRs from JI3253 (an independent *acutilegumen* mutant) revealed them to share all the mismatches as JI2822 but with one, unique mismatch at position 1532. This base is a cytosine in the lower LTR but is an adenine in the upper LTR. It was therefore hypothesised that the upper base was likely originally a cytosine and underwent a C→T SBS mutation, albeit the JI3253 mutation is 467 bp downstream of the FN3241/187 mutation.

Though successfully finely mapped to an LTR-like sequence, it remained unclear if this sequence was indeed still functionally capable of transposition, or even if it was expressed at all. Regardless, it was hypothesised that the LTR-like sequence itself was *ACUTILEGUMEN*, and G→A and C→T SBS mutations led to loss of the gene's function in FN3241/187 and JI3253, respectively.

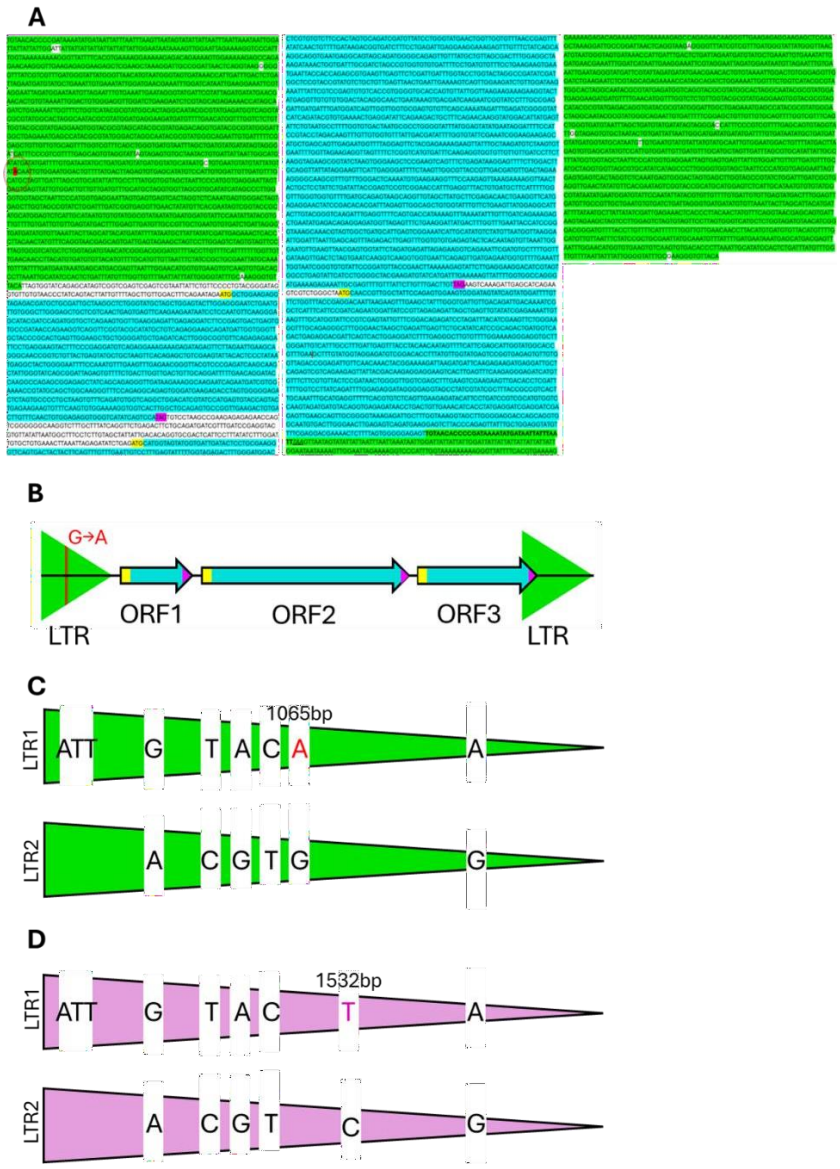


Figure 4.11: Structure of *acutilegumen*. A – Raw sequence of *acutilegumen* in FN3241/187, with LTRs highlighted in green. Red, circled base is the G→A SBS mutation, and unhighlighted bases are non-identical between the two LTRs. 3 ORFs are highlighted in blue, START codons are highlighted in yellow, with STOP codons highlighted in purple, final ORF leaks into second LTR and is highlighted in bold with STOP underlined. B – Schematic representation of *acutilegumen*. Green arrows are LTRs, blue arrows are ORFs. Yellow highlights are START codons; purple highlights are STOP codons. Red highlight is G→A SBS mutation in FN3241/187. C – Schematic representation of alignment of upper (LTR1) and lower (LTR2) LTRs from FN3241/187. Individual mismatches are indicated, with a single ATT insertion also represented. Red 'A' represents G→A SBS mutation, with position in the LTR indicated. D - Schematic representation of alignment of upper and lower LTRs from JI3253, with purple 'T' indicating C→T substitution at position 1532 bp.

The observed similarities between the *ACUTILEGUMEN* alleles of JI2822, FN3241/197 and JI3253 allowed for relatively easy comparison. However, comparison of these alleles to the JI281 allele of *ACUTILEGUMEN* reflects the enormous genomic distance that separates JI281 from other *Pisum sativum* lines. JI281 is an early-diverging *Pisum sativum* landrace from Ethiopia. JI281 exhibits a much larger number of mismatches between the upper and lower LTRs of its *ACUTILEGUMEN* allele than JI3253 or FN3241/187 (22 as compared to 6), none of which are shared (**see Table 4.3 for a comprehensive list of LTR mismatches in each line**). Additionally, JI281 lacks the 3bp microsatellite ATT insertion found in the upper LTR of JI3253 and JI2822, but JI281 does exhibit a unique 1 bp insertion in its lower LTR at position 854. The G→A and C→T SBS mutations that characterise FN3241/187 and JI3253, respectively, are absent in JI281 (i.e. the aligned bases are the ancestral G at position 1062 and C at position 1524).

Table 4.3: List of mismatches/ insertions between upper and lower *ACUTILEGUMEN* LTRs in FN3241/187 (JI2822 background), JI3253 and JI281.

Position	LTR1	LTR2
FN3241/187		
90, 91, 92	ATT	-
283	G	A
917	T	C
950	A	G
1042	C	T
1065	G	A
1981	A	G
JI3253		
90, 91, 92	ATT	-
283	G	A
917	T	C
950	A	G
1042	C	T
1532	C	T
1981	A	G
JI281		
219	G	A
240	C	T
322	A	G
575	C	T
581	G	T
614	T	C
760	G	A
782	G	A
853	A	G
853	-	T
862	G	T
863	G	T
918	C	T
934	T	G
993	C	G
1051	T	C
1063	A	C
1194	A	G
1370	G	T
1469	T	G
1481	A	C
1631	T	C
1653	C	T

IV. Discussion

The reported mapping of *ACUTILEGUMEN* to the specified LTR-like sequence is not uncontentious. First, it was unexpected that the mutation in FN3241/187 would map to as gene-poor a genomic context as has been observed, and that all available candidate genes (which were derived from the publicly available Cameor v1a annotation) would be rejected as *ACUTILEGUMEN*. The Cameor v1a annotation was derived from multiple independent RNA-seq datasets which were combined and used to annotate for the presence of genes[20]. This was concerning, as it suggested the above-reported inability to detect any transcripts mapping to the SBS was perhaps also true across multiple other independent RNA-seq experiments, casting doubt on the hypothesis that *ACUTILEGUMEN* is expressed at all. Additionally, it was speculated that due to the unconventional nature of *ACUTILEGUMEN* as an LTR-like gene, and the omnipresence of LTR-like sequences in the non-genic portion of the *Pisum* genome[20], that *ACUTILEGUMEN* was missed by computational predictions (Eugene 4.3). Finally, while transposons have been reported previously to be critical to plant phenotypes (and, indeed, were discovered by focused work on another crop plant: maize[81]), there is no evidence from the gDNA sequencing of JI2822 and FN3241/187 that a transposition event has taken place, as insertion scars would show up as mis-matching base stretches on either end of the LTR-like sequence. As such, it is unclear if *ACUTILEGUMEN* retains a presumably ancestral ability to undergo copy-and-paste transposition within the pea genome.

In favour of the proposed sequence as the identity of *ACUTILEGUMEN*: the mapped region defined by the Axiom markers in JI3253 and JI281, when checked in FN3241/187, contained only the single G→A SBS mutation which, by inference, must be causative for the observed pointed pod tip. In support of this conclusion, the observed SBS is located in the first LTR, which is where the promoter of LTR-retroelements is located. In FN3241/187, this G→A substitution can be observed *relative to JI2822, the wild-type*, but importantly it can also be observed by comparison *of the first LTR to the second LTR*, which bears the ancestral G base at the corresponding location. Comparing the upper and lower LTRs of the wild-type, JI2822 allele of *ACUTILEGUMEN* reveals a total of 5 mismatches and one 3 bp microsatellite insertion. As such, *acutilegumen* has gained one additional mismatch between its LTRs. LTRs are generally identical upon transposon insertion[82], but the exact ancestral sequence cannot be determined by comparison of the upper and lower LTRs alone, as mutations may have taken place in either during the course of molecular evolution. The loss of identity across the LTRs has likely compromised *ACUTILEGUMEN's* transposition capability[82].

JI3253 (Caméor) is an independent *acutilegumen* mutant to FN3241/187, though the two mutants are allelic. JI3253 is also allelic to JI0799, the *acutilegumen* type line, as has been found in previous unpublished work by Hofer and Ellis (data unpublished). Though the above mapping of *ACUTILEGUMEN* began by investigating JI3253, the lack of a corresponding wild type prevented conclusive identification of the causative allele (JI281 being highly divergent from JI3253 across its entire genome). However, due to its nature as an LTR-like sequence, *acutilegumen* bears an internal control by which mutations can be inferred to have taken place as the mismatches between the two LTRs accumulate over evolutionary time. Importantly, JI3253 has 5 shared mismatches (and one shared 3bp microsatellite insertion) to JI2822 (**Figure 4.11**), indicating that in the last common ancestor of these two pea lines, the *ACUTILEGUMEN* LTR sequences were identical to those found presently in (wild type) JI2822. However, the one unshared mismatch which discriminates the JI3253 *acutilegumen* allele from JI2822's allele, is an additional C→T substitution mutation. This mutation is located relatively near to the SBS in FN3241/187 (467 bp downstream) and the co-occurrence of these similar mutations in two pea lines with pointed pods motivated the conclusion that the identified sequence was indeed *ACUTILEGUMEN*. This does not represent absolute proof, however, as it may yet be that *ACUTILEGUMEN* is in fact a separate sequence to the LTR-like locus reported here, and that the similarities between FN3241/187 and JI3253's alleles are coincidental. Though this seems an unlikely possibility, conclusive evidence for the proposed location of *ACUTILEGUMEN* will not be reached until either an additional, independent mutant is obtained from a pre-sequenced wild type line or *ACUTILEGUMEN* function is knocked out or knocked down (e.g. via VIRUS-INDUCED GENE SILENCEING, VIGS[83]) and a blunt-to-pointed phenotypic change simultaneously observed.

The total lack of shared mismatches between JI2822 and JI281 indicates that, in the last common ancestor of these two pea lines, there may have been no mismatches at all between the upper and lower LTR sequences, which may suggest that the two lines evolutionarily split shortly after ACUTILEGUMEN's birth-by-insertion event. Though containing abundantly more mismatches between its LTRs, JI281's ACUTILEGUMEN allele is believed to be functional, as it yields the wild type, blunt-tipped pods characterising JI281 (and half of the JI281xCameor RIL population). This motivates the hypothesis that total number of mismatches between the upper and lower LTRs bears little impact on ACUTILEGUMEN's function but rather that a small number of critical bases are essential, or alternatively that the LTR mutations that evolutionarily preceded those in JI3253 and FN3241/187 'primed' the gene for loss-of-function. Future work

should go on to compare the observed *ACUTILEGUMEN* sequences from JI3253, JI2822 and JI281 with other exotic, early diverging pea lines, including *Pisum abyssinicum*[80], an independent domesticate and *Pisum fulvum*[84], a near-wild landrace. This may allow for bioinformatic reconstruction of the ancestral allele, which would clarify the presence of mutations in sequenced lines in the future.

In the absence of a mapping transcript, it is very difficult to imagine a convincing model for *ACUTILEGUMEN's in planta* molecular mechanism, but possible explanations linking this gene to pod shape can be categorised into three broad suites of possibilities. First, it is possible (and, arguably, suggested by both our RNA-Seq data and the Cameor v1a annotation) that the *ACUTILEGUMEN* locus is not expressed at all, and therefore technically does not constitute a gene. To allow for an effect on pod shape, however, this would require the observed mutation in *acutilegumen* lines to affect the expression of nearby genes *in cis*, e.g. as a cis-regulatory element: an enhancer or repressor. In this "structural" model, mutation within the upper LTR compromises the ability of *acutilegumen* to regulate *in cis*, thus altering the expression of its targets. This is unsupported, however, by the above reported genomics and transcriptomics. Genomically, there are no annotated genes located within half a megabase of the SBS mutation (**Figure 4.8**) which places all annotated genes far beyond the influence of the SBS as a cis-regulatory element. Additionally, there were no unannotated transcripts (from the RNA-seq data) mapping any closer to the SBS mutation within this range, indicating *ACUTILEGUMEN* as being located in a true gene desert. As such, this structural, cis-regulatory model for *ACUTILEGUMEN's* function is rejected, and unlikely to be causal for the observed fruit phenotype.

Second, it is possible that *ACUTILEGUMEN* is regulating its targets *in trans.* This hypothesis would posit that *ACUTILEGUMEN* is a true gene, i.e. it is expressed and yields a gene product (either a regulatory RNA or an mRNA which encodes a regulatory protein) and that it thus alters the expression of a number of downstream targets. This is weakly supported by comparing the transcriptomes of JI2822 and FN3241/187, which includes 125 DEGs at $q<0.05$. It is possible that these DEGs include direct and/or indirect targets of *ACUTILEGUMEN*'s gene product, and that alteration in the expression of these target genes leads to alteration in the protein abundances of these targets. For genes that belong to families with well-established roles in plant development, e.g. *AUX/IAA*s, this could provide a plausible explanation for how *acutilegumen* mutation may lead to altered gynoecium development and, ultimately, altered fruit shape at maturity. Additionally, the relatively small number of affected target genes (125) would indicate that a *trans-regulatory ACUTILEGUMEN* gene would

sit quite low in the transcriptional hierarchy. The primary (and not trivial) objection to this hypothesis is the lack of detected transcripts mapping to the *ACUTILEGUMEN* locus itself: it is difficult to suggest that *ACUTILEGUMEN* is expressed, and to speculate on the role of its expressed product, before expression at this locus has been detected. In support of this model, however, is the proximity of the JI3253 and FN3241/187 *acutilegumen* mutations to the LTR retroelement's ancestral promoter sequence (located, as in all LTR retrotransposons, in its upstream LTR). It remains possible that the *ACUTILEGUMEN* transcript is sufficiently rare that it is not detected in RNA-seq and/or that it is not protein-coding, and therefore would have been missed in our transcriptomics experiment which enriched for polyadenylated RNA species (i.e. protein-coding mRNAs). In favour of the former of these two explanations is the observation that the developmental difference between JI2822 and FN3241/187 occurs extremely early (at least before -3DAA) and therefore may emerge at the primordium stage. A transcript which affects development in such an ephemeral and physically small precursor structure could very conceivably go undetected in any RNA-seq experiment that isn't aimed specifically at gathering RNAs from highly specialised (and low in number) primordium cells. This is because, even with high sequencing depths, unrelated and more abundant RNAs may outcompete and dilute the desired transcript to below detectable levels during the sequencing phase.

Finally, similarly to the last hypothesis, it is also conceivable that *ACUTILEGUMEN* may encode a protein, which is expressed early in gynoecium development, but which is not regulatory in nature. This hypothesis is not essentially different from the previous one, except that it would indicate that the DEGs that differentiate JI2822 and FN3241/187 ovaries are byproducts of a mutation in a nonregulatory biological process, i.e. that the transcriptome is adapting to an altered function of a particular protein, rather than the genes being targets of *ACUTILEGUMEN* itself *per se*. This hypothesis seems unlikely too, however, as the gag-pol ORFs that lie between the two LTRs of *ACUTILEGUMEN* share little homology with any non-retroelement proteins and are poorly conserved across pea lines. As such, hypothesis 2, in which *ACUTILEGUMEN* is a regulatory factor that operates *in trans*, is the current working hypothesis.

A final caveat to the reported DEGs is that FN3241/187 does bear more than one SBS mutation within its genome, and that background mutations may also be influencing the expression of some of the observed DEGs. Backcrossing to JI2822 could clean the FN genome of these background mutations and allow for a more powerful comparison of the transcriptome.

The mapping of *ACUTILEGUMEN* to such an unusual genic sequence, which is sufficiently rarely expressed to have evaded RNA-seq-based methods of pea genome annotation thus far, is illustrative of the power of pea genomics[19] and RIL-based mapping methods[79]. With sufficiently wide crosses, it seems likely that many further genes in the future may yet be identified by this combined approach of classical genetics with the only recently accessible pea genome sequence[24], which will play an important role for securing future yield of this plant protein source to feed a growing global population under a changing climate.

V. Conclusion

The work presented above provides evidence that the ability of the pea fruit to 'discriminate' between IAA and 4-Cl-IAA is an evolutionarily recent phenomenon, having only arisen *after* the evolution of the 4-Cl-producing halogenase 25 million years ago when the Fabeae/Trifolieae clade split from its sister clade Cicerae[24]. Identification of the halogenase enzyme itself remains an outstanding challenge for the future in this area, though its potential spatial localisation to pea endosperm (and possible detection of its activity *in vitro*) paves the way for further enzymatic and protein purification studies.

A great deal remains to be understood about the evolution of plant signalling ligands. The IAA/ 4-Cl-IAA system represents a useful and evolutionarily recent example of how duplication in a signalling molecule can go on to shape differential growth responses, though conclusive assessment of 4-Cl-IAA's role in pea fruit development will still require a halogenase mutant plant. Additionally, IAA/ 4-Cl-IAA provide a recent and auxinic counterpart to the established strigolactone/ karrikin system. Strigolactones are butenolide-containing, polycyclic signalling molecules derived from the carotenoid biosynthesis pathway and are important in regulating a range of processes across plant development[85]. Karrikins, by contrast, are molecules produced by burning plant matter which stimulate germination and photomorphogenesis across a wide range of plant species including *Arabidopsis*[86] and tomato[87]. Karrikins primarily act through the karrikin receptor KARRIKIN-INSENSITIVE2 (KAI2), though developmental phenotypes (such as increased seed dormancy) in *kai2* mutants grown in the absence of karrikins[88] indicate the existence of an endogenous ligand of the KAI2 receptor which remains unidentified. The strigolactone receptor D14 is homologous to KAI2, indicating an ancient duplication of a signalling protein giving rise to multiple signalling systems (reviewed in [89]). Though the evolution of KAI2-ligands vs. strigolactones remains unclear (as the KAI2-ligand has yet to be identified) their evolution represents a comparable signalling molecule divergence as the IAA/4-Cl-IAA system, albeit

strigolactone and KAI2-ligand signalling pathways are hypothesised to have diverged early in land plant evolution[89]. IAA and 4-Cl-IAA sit at the opposite end of this history and provide an opportunity for exploration of how recent signalling molecule diversification can drive evolution in signalling networks.

An additional evolutionary question pertains to the tryptophan-4-halogenase: i.e. whether it was ancestrally expressed in endosperm or whether its localisation to endosperm occurred after its halogenase function evolved. Once the halogenase-encoding gene is identified, close comparison of spatial expression profiles of homologous genes from ancestrally diverging, non-4-Cl-IAA producing legumes may prove informative in this regard. Identification of proto-halogenase homologues may also lead to identification of the causative amino acid changes that led to gain of halogenase function and, in due course, may eventually permit CRISPR/Cas9[90,91] mediated conversion of proto-halogenase enzymes into tryptophan halogenases across non-4-Cl-producing crop legumes such as chickpea and soy. This could be interesting from the perspective of pure research, as it would allow for observation of early developmental responses to immediate evolutionary gain of 4-Cl-IAA. Such an experiment would provide a model for how the halogenase may have developmentally affected the original ancestor of the F/T clade and whether it provides any obvious benefit to organismic fitness. If not, then it might motivate the hypothesis that the halogenating allele of the proto-halogenase enzyme originally reached fixation in the ancestor of the F/T clade via genetic drift, i.e. in a selectively neutral manner. Besides these more abstract research purposes, it is also possible (though remains to be demonstrated) that transformation of non-4-Cl-producing legume crops with the halogenase may lead to alteration in yield or crop performance.

Importantly, work of this kind has already begun via heterologous expression of bacterial tryptophan halogenases (*pyrH, ThdH, prnA*) in *Arabidopsis*[92]. Due to the absence of a tryptophan-4-halogenase among bacteria, however, this allowed only for 5,6,7-Cl-IAA-producing *Arabidopsis* plants to be generated. Unexpectedly, this did not result in altered development relative to wild-type controls[92]. It is difficult to abstract results from *Arabidopsis* heterologous expression systems into the context of the legume fruit.

4-Cl-IAA is also potentially important to aiding our development of herbicides. Many auxinic herbicides are synthetically halogenated auxin analogues[93]. Auxinic herbicides cause rapid growth responses followed by death[94]. The lethality of synthetic auxin treatment has been speculated to partially result from crosstalk with other signalling

systems, with auxins upregulating the expression of ABA biosynthetic *NCED* genes and accumulation of leaf ABA concentrations 3-fold higher than would be typically observed in drought-stressed plants[95]. This large synthesis of ABA has been associated with mass downregulation of a range of photosynthesis-critical genes, leading eventually to plant cell death, though a more robust demonstration of this hypothesis (via testing of *nced* mutants for resistance) is outstanding. Auxinic herbicides are generally active only against dicots and are resisted by monocots, this feature has made auxinic herbicides important for the cultivation of cereal crops[94]. It is broadly unknown why herbicides exhibit species selectivity – i.e. proving toxic to some plant species and not to others, but resistance to auxinic herbicides has both evolved naturally amongst dicot agricultural weeds (e.g. *Brassica scoparia aux/iaa16* resistance to 2,4-D[96]) and has been induced by mutations in *Arabidopsis* (e.g. *afb5* resistance to synthetic auxin picloram[97]). IAA, interestingly, does not kill plants at any reported concentration, though applications of large quantities of IAA lead to developmentally aberrant plants. 4-Cl-IAA unexpectedly differs in this regard in that it has been used as a herbicide previously[98], despite being naturally produced by F/T clade legumes. Indeed, at sufficiently high concentrations (I_{50} against pea plants is 15mg/kg fresh weight) 4-Cl-IAA can even inhibit the growth of pea plants and was further found to be active against *Sinapis alba* and *Hordeum vulgare*[98]. Whether 4-Cl-IAA toxicity is species-selective, and why, remains an open question with substantial economic relevance.

Though known mostly as a plant hormone, IAA is a ubiquitous molecule across the biosphere and is produced from tryptophan by mammalian gut symbionts[99]. Microbiome-origin IAA is not inconsequential for host physiology and has been implicated in reducing spinal inflammation (Ankylosing Spondylitis)[100], maintaining the integrity of the intestinal epithelium[101], IAA gavage slows progression of myopia in mice models by stimulating collagen synthesis (via *COL1A1* expression)[102], and exerts anti-depressive effects in mouse models of psychological stress[103]. Most excitingly, recent work has demonstrated that enteric IAA concentrations correlate with efficacy of chemotherapeutic treatment of pancreatic ductal adenocarcinoma (PDAC) in humans[99], and IAA was successfully used as a chemotherapy adjuvant in mice models of PDAC[99]. This was due to an increase in reactive oxygen species (ROS) within the cancer cells, leading to reduced tumour cell proliferation (and likely thus aiding the FIRINOX chemotherapeutic treatment)[99]. IAA was found to increase ROS by reducing the expression of H_2O_2-reducing enzymes *GPX7* and *GPX3*, whose activities are required for oxidative stress prevention[99]. Exactly how IAA causes a reduction in the expression of *GPX7* and *GPX3* was not explored, but other research has found IAA (and indolic and

aromatic metabolites generally) to interact with and alter the activity of at least one bHLH transcriptional factor (ARYL HYDROCARBON RECEPTOR, AhR)[104]. However, IAA-mediated AhR activation has also been implicated in promoting pancreatic tumour growth by activating AhR in tumour-associated macrophages, which may stimulate the expression and secretion of tumour-promoting proteins such as VASCULAR ENDOTHELIAL GOWTH FACTOR (VEGF)[104]. In the absence of IAA, AhR is sequestered into the cytosol in a complex of inhibitory chaperones (HSP90, XAP2, p23 and SRC) but upon ligand binding AhR is liberated from these partners to translocate to the nucleus and alter expression of a range of target genes. This is mediated by direct binding of AhR to xenobiotic responsive elements (XREs) within the promoters of target genes (reviewed in [105]). Whether 4-Cl-IAA or 4-Cl-tryptophan likewise have AhR-interacting and/or health-promoting potential remains to be explored, but the additional chlorine may stabilise or alter the activity of these indole metabolites in therapeutically relevant ways.

This thesis also presents *ACUTILEGUMEN*, an LTR-transposon-derived sequence which likely encodes a trans-regulatory factor and influences the expression of gynoecium developmental genes including *AUX/IAA4* and *AUX/IAA14*. BLASTing of the *ACUTILEGUMEN* sequence against NCBI reveals large numbers of homologous legume sequences, though very few have any functional annotation. One result, with 60% coverage and 94% identity to the part of *ACUTILEGUMEN* to which it matches, is listed as "*Pisum sativum* peabody/Ty3-type retrotransposon gag-pol precursor (gag-pol) pseudogene". 'Ty' elements ("Transposons of yeast")[106] are sequences which encode the characteristic gag ("group antigens") and pol (reverse transcriptase) genes of retroviral origin, which allow them to transpose by transcription into RNA, and reverse transcription into cDNA for reintegration into the genome at a new site. All Ty genes lack the "env" (envelope) proteins characteristic of retroviruses, and so do not assemble their nucleic acids into capsids as true viruses do and, consequently, are not infectious[106]. Ty genes are classified according to the order in which the protein modules of their gag and pol genes are arranged and the precise molecular mechanism by which they effect duplication and transposition. Ty3 elements, which in yeast are defined by their exploitation of $tRNA^{MET}$ for priming of reverse transcription and proclivity for insertion into RNA polymerase III promoter sequences, are also capable of translational frameshifting[110]. If *ACUTILEGUMEN* is indeed a Ty3-like element, this complicates bioinformatic translation of its LTR-flanked ORFs, and more detailed study into the range of possible gene products across wild type and *acutilegumen* pea lines is warranted. Though there was no evidence from the gDNA sequencing and mapping that

acutilegumen plants had undergone a transposition event, transposons have been widely implicated in altering fruit shape across a range of plant species.

Harbinger, for example, is a transposon which, in some lines of melon, has transposed into the promoter region of *EIN2* and reduced its expression, thus impairing ethylene signalling[111]. This was found to be due to the spreading of DNA methylation from the newly inserted *Harbinger* element into the promoter of *EIN2*. In melon, this led to the conversion of hermaphroditic flowers to male flowers (androecy) due to the requirement of ethylene signalling for carpel development. *EIN2* silencing also altered fruit shape when combined into a *wip1* background, indicating a role for *EIN2* in fruit elongation[107].

Transposons are ubiquitous within the pea genome, and 75-97% of the genome is composed of repetitive elements[20]. This is likely due to recent mass transposition of elements across the pea genome, and partly explains the large size of the pea genome compared to most legumes (4.5Gb). *Vicia faba*, a close relative of pea, also exhibits one of the largest genomes of any diploid plant at 11.9Gb[108], also replete with repetitive sequences. Although a fantastic effort has been made towards annotation of the protein-coding genes in pea, *ACUTILEGUMEN*'s status as apparent "junk" DNA is a reminder of the potential phenotypic relevance even of highly repetitive elements within the genome. The key next steps for *ACUTILEGUMEN* research are to attempt targeted disruption of the gene's function by VIRUS-INDUCED GENE SILENCING (VIGS)[83] or, perhaps eventually, by CRISPR/Cas9-mediated knockout[90,91]. Concomitant with this molecular work, whole-plant phenotypic work pertaining to yield, thousand-seed weight, ease of processing and risk of pod tip disease could now be carried out using a near-isogenic pair which could be generated through repeated backcrossing of FN3241/187 onto JI2822. This latter work may allow for better understanding of the agricultural detriment (if any) of *acutilegumen* genotype and will provide a marker for breeders to use when considering pea fruit morphology.

Their large genomes and difficulty of transformation have prevented legumes from enjoying as strong an interest as model plants during the golden age of plant molecular genetics. Fortunately, the ever-declining cost of sequencing and vastly improved bioinformatic software are slowly facilitating a renaissance for crop legumes among amongst a wider audience of plant biologists. Under the mounting pressures of climate change and a growing world population, demand for plant protein sources is increasing and legumes are well-poised for greater cultivation in the future[5].

However, it seems likely that legume research is to continue bifurcating into distinct sub-fields. These can generally be categorised into model species-driven research into

the molecular detail of nodulation (usually *Medicago truncatula* and *Lotus japonicus*)[109], novel attempts to develop model-like, dwarfed, rapid-cycling and eventually transformable genotypes of crop legumes (e.g. JI2822 in *Pisum sativum*), and finally large-scale, bioinformatics-heavy analyses of diversity across genotypes including landraces and ancient progenitor lines[79]. The success of legume research in the future will require close communication between these sub-fields to allow for rapid dissemination of new information and deployment into crop lines, combined with even further interdisciplinary research branching into nutrition and human health.

VI. Bibliography

1. Rockström, J., Steffen, W., Noone, K. et al. A Safe Operating Space for Humanity. *Nature* 461, 472–475 (2009).

2. Willett, W. *et al.* Food in the Anthropocene: the EAT–Lancet Commission on Healthy Diets from Sustainable Food Systems. *Lancet* 393, 447–492 (2019).

3. Semba, R. D., *et al.* Legumes as a Sustainable Source of Protein in Human Diets. *Glob Food Sec* 28, 100520 (2021).

4. Koenen, E. J. M. *et al.* The Origin of the Legumes is a Complex Paleopolyploid Phylogenomic Tangle Closely Associated with the Cretaceous-Paleogene (K-Pg) Mass Extinction Event. *Syst Biol* 70, 508–526 (2021).

5. Smýkal, P. *et al.* Legume Crops Phylogeny and Genetic Diversity for Science and Breeding. *CRC Crit Rev Plant Sci* 34, 43–104 (2015).

6. Mathesius, U. Are legumes different? Origins and Consequences of Evolving Nitrogen Fixing Symbioses. *J. Plant Physiol* vol. 276, Article 153765 (2022).

7. Bohra, A. *et al.* The Key to the Future Lies in the Past: Insights from Grain Legume Domestication and Improvement Should Inform Future Breeding Strategies. *Plant Cell Physiol* 63, 1554–1572 (2022).

8. Fuller, D. Q. & Harvey, E. L. The Archaeobotany of Indian pulses: Identification, Processing and Evidence for Cultivation. *Environ Archaeol* 11, 219–246 (2006).

9. Hymowitz, T. On the Domestication of the Soybean. *Econ Bot* 24, 408–421 (1970).

10. Foyer, C. H. *et al.* Neglecting Legumes has Compromised Human Health and Sustainable Food Production. *Nat Plants* 2, Article 16112 (2016).

11. Stotz, G. C., Salgado-Luarte, C., Escobedo, V. M., Valladares, F. & Gianoli, E. Global Trends in Phenotypic Plasticity of Plants. *Ecol Lett* 24, 2267–2281 (2021).

12. Yu, Z., Zhang, F., Friml, J., and Ding, Z. Auxin Signaling: Research Advances Over the Past 30 Years. J. Integr. *Plant Biol* 64, 371–392 (2022).

13. Gomes, G. L. B. & Scortecci, K. C. Auxin and its Role in Plant Development: Structure, Signalling, Regulation and Response Mechanisms. *Plant Biol* 23, 894–904 (2021).

14. Miin-Feng, W. *et al.* Auxin-Regulated Chromatin Switch Directs Acquisition of Flower Primordium Founder Fate *eLife* 4, e09269 (2015).

15. Qi, L. *et al.* Adenylate Cyclase Activity of TIR1/AFB Auxin Receptors in Plants. *Nature* 611, 133–138 (2022).

16. Rienstra, J., Hernández-García, J. & Weijers, D. To bind or not to bind: how AUXIN RESPONSE FACTORs select their target genes. *J Exp Bot* 74, 6922–6932 (2023).

17. Kuhn, A. *et al.* Direct ETTIN-Auxin Interaction Controls Chromatin States in Gynoecium Development. *eLife* 9, e51787 (2020).

18. Friml, J. *et al.* ABP1–TMK Auxin Perception for Global Phosphorylation and Auxin Canalization. *Nature* 609, 575–581 (2022).

19. Kreplak, J. *et al.* A Reference Genome for Pea Provides Insight into Legume Genome Evolution. *Nat Genet* 51, 1411–1422 (2019).

20. Marumo, S., Abe, H., Hattori, H. & Munakata, K. Isolation of a Novel Auxin, Methyl 4-Chloroindoleacetate from Immature Seeds of *Pisum sativum*. *Agric Biol Chem* 32, 117–118 (1968).

21. Katayama, M., Thiruvikraman, S. V & Marumo, S. Localization of 4-Chloroindole-3-Acetic Acid in Seeds of *Pisum sativum* and Its Absence from All Other Organs. *Plant Cell Physiol* 29, 889-891 (1988).

22. Katayama, M., Thiruvikraman, S. V & Marumo, S. Identification of 4-Chloroindole-3-Acetic Acid and Its Methyl Ester in Immature Seeds of *Vicia amurensis* (the Tribe *Vicieae*), and Their Absence from Three Species of *Phaseoleae*. *Plant Cell Physiol* 28, 383-386 (1987).

23. Ernstsen, A., Sandberg, G. & Lundstrom, K. Identification of Oxindole-3-Acetic Acid, and Metabolic Conversion of Indole-3-Acetic Acid to Oxindole-3-Acetic Acid in Pinus Sylvestris Seeds. *Planta* 172, 47-52 (1987).

24. Lam, H. K., McAdam, S. A. M., McAdam, E. L. & Ross, J. J. Evidence that Chlorinated Auxin is Restricted to the Fabaceae but not to the Fabeae. *Plant Physiol* 168, 798–803 (2015).

25. Ma, M.-H. *et al.* Enhanced Expression of 4-Cl-IAA and 6-Cl-IAA by Touch Stimulus for Rapid and Differential Growths of the Madeira Vine. *bioRxiv* https://doi.org/10.1101/2022.04.13.488269 (2022).

26. Wojtania, A., Markiewicz, M. & Waligórski, P. Regulation of the Bud Dormancy Development and Release in Micropropagated Rhubarb 'Malinowy'. *Int J Mol Sci* 23 (3) 1480 (2022).

27. Reinecke, D. M., *et al.* Molecular Properties of 4-Substituted Indole-3-Acetic Acids Affecting Pea Pericarp Elongation. *Plant Growth Regul* 27, 39–48 (1999).

28. Jayasinghege, C. P. A., *et al.* TIR1 Auxin Receptors are Implicated in the Differential Response to 4-Cl-IAA and IAA in Developing Pea Fruit. *J Exp Bot* 70, 1239–1253 (2019).

29. Ozga, J. A. *et al.* Developmental and Hormonal Regulation of Gibberellin Biosynthesis and Catabolism in Pea Fruit. *Plant Physiol* 150, 448–462 (2009).

30. Van Huizen, R., Ozga, J. A. & Reinecke, D. M. Seed and Hormonal Regulation of Gibberellin 20-Oxidase Expression in Pea Pericarp. *Plant Physiol* 115, 123–128 (1997).

31. Johnstone, M. M. G., Reinecke, D. M. & Ozga, J. A. The Auxins IAA and 4-Cl-IAA Differentially Modify Gibberellin Action via Ethylene Response in Developing Pea Fruit. *J Plant Growth Regul* 24, 214–225 (2005).

32. Binder, B. M. Ethylene Signaling in Plants. *J Biol Chem* 295, 7710–7725 (2020).

33. Park, H. L. *et al.* Ethylene-Triggered Subcellular Trafficking of CTR1 Enhances the Response to Ethylene Gas. *Nat Commun* 14, 365 (2023).

34. Jayasinghege, C. P. A., *et al.* Regulation of Ethylene-Related Gene Expression by Indole-3-Acetic Acid and 4-Chloroindole-3-Acetic Acid in Relation to Pea Fruit and Seed Development. *J Exp Bot* 68, 4137–4151 (2017).

35. Meitzel, T. *et al.* Trehalose 6-Phosphate Promotes Seed Filling by Activating Auxin Biosynthesis. *New Phytol* 229, 1553–1565 (2021).

36. Pencík, A., *et al.* Ultra-rapid auxin metabolite profiling for high-throughput mutant screening in Arabidopsis. *J Exp Bot*, 69 (10) 2569–2579 (2018).

37. Fukui, K. *et al.* Chemical Inhibition of the Auxin Inactivation Pathway Uncovers the Roles of Metabolic Turnover in Auxin Homeostasis, *Proc Natl Acad Sci USA* 119 (32) e2206869119 (2022).

38. Nitsch, J. P. Growth and Morphogenesis of the Strawberry as Related to Auxin. *Am J Bot* 37, 3, 211-215 (1950).

39. Vivian-Smith A, Koltunow AM. Genetic Analysis of Growth-Regulator-Induced Parthenocarpy in Arabidopsis. *Plant Physiol* 121 (2) 437-51 (1999).

40. Guo, L. *et al.* Mechanism of Fertilization-Induced Auxin Synthesis in the Endosperm for Seed and Fruit Development. *Nat Commun* 13, Article 3985 (2022).

41. Hu, J., Li, X. & Sun, T. ping. Four Class A AUXIN RESPONSE FACTORs Promote Tomato Fruit Growth Despite Suppressing Fruit Set. *Nat Plants* 9, 706–719 (2023).

42. Prigge, M. J., *et al. Physcomitrella patens* Auxin-Resistant Mutants Affect Conserved Elements of an Auxin-Signaling Pathway. *Curr Biol*, 20, (21) 1907-1912 (2010).

43. Mackelprang, R., Okrent, R. A. & Wildermuth, M. C. Preference of *Arabidopsis thaliana* GH3.5 Acyl Amido Synthetase for Growth Versus Defense Hormone Acyl Substrates is Dictated by Concentration of Amino Acid Substrate Aspartate. *Phytochem* 143, 19–28 (2017).

44. Staswick, P. E. *et al*. Characterization of an *Arabidopsis* Enzyme Family that Conjugates Amino Acids to Indole-3-Acetic Acid. *Plant Cell* 17, 616–627 (2005).

45. Adamowski, M. & Friml, J. PIN-dependent Auxin Transport: Action, Regulation, and Evolution. *Plant Cell* 27, 20–32 (2015).

46. Balcerowicz, M. Stop the flow: Improving Somatic Embryogenesis using Auxin Transport Inhibitors. *Plant J* 118, 293–294 (2024).

47. Di, D. W., *et al*. The Biosynthesis of Auxin: How Many Paths Truly Lead to IAA? *J Plant Growth Regul* 78, 275–285 (2024)

48. Tivendale, N. D. *et al.* Biosynthesis of the Halogenated Auxin, 4-chloroindole-3-acetic acid. *Plant Physiol* 159, 1055–1063 (2012).

49. Cochereau, B., *et al.* Halogenation in Fungi: What Do We Know and What Remains to Be Discovered? *Molecules* 27, Article 3157 (2022).

50. Kim, C. Y. *et al.* The Chloroalkaloid (−)-Acutumine is Biosynthesized via a Fe(II)- and 2-oxoglutarate-Dependent Halogenase in Menispermaceae Plants. *Nat Commun* 11, 1–7 (2020).

51. Dé, F., Vaillancourt, R. H., Yin, J. & Walsh, C. T. SyrB2 in Syringomycin E Biosynthesis Is a Nonheme Fe II-Ketoglutarate-and O_2-Dependent Halogenase. *Proc Natl Acad Sci USA* 102 (29) 10111-10116, (2005).

52. Mead RJ & Segal W. Fluoroacetic Acid Biosynthesis: a Proposed Mechanism. *Aust J Biol Sci.* 25 (2) 327-333 (1972).

53. Godlewska, M. & Banga, P. J. Thyroid Peroxidase as a Dual Active Site Enzyme: Focus on Biosynthesis, Hormonogenesis and Thyroid Disorders of Autoimmunity and Cancer. *Biochimie* 160, 34–45 (2019).

54. Xu, G. & Wang, B. G. Independent Evolution of Six Families of Halogenating Enzymes. *PLoS One* 11 (5) e0154619 (2016).

55. Papadopoulou, A. *et al.* Re-Programming and Optimization of a L-Proline cis-4-Hydroxylase for the cis-3-Halogenation of its Native Substrate. *ChemCatChem* 13, 3914–3919 (2021).

56. Fraley, A. E. & Sherman, D. H. Halogenase Engineering and its Utility in Medicinal Chemistry. *Bioorg Med Chem Lett* 28, 1992–1999 (2018).

57. Jeschke, P. Latest Generation of Halogen-Containing Pesticides. *Pest Manag Sci* 73, 1053–1066 (2017).

58. Faleye, O. S., Boya, B. R., Lee, J. H., Choi, I. & Lee, J. Halogenated Antimicrobial Agents to Combat Drug-Resistant Pathogens. *Pharmacol Rev* 76, 90–141 (2024).

59. Kusari, P. *et al.* Cross-Species Biosynthesis of Maytansine in *Maytenus serrata*. *RSC Adv* 6, 10011–10016 (2016).

60. Phintha, A., Prakinee, K. & Chaiyen, P. Structures, Mechanisms and Applications of Flavin-Dependent Halogenases. *Enzymes* 47, 327–364 (2020).

61. Zehner, S. *et al.* A Regioselective Tryptophan 5-Halogenase is Involved in Pyrroindomycin Biosynthesis in *Streptomyces rugosporus* LL-42D005. *Chem Biol* 12, 445–452 (2005).

62. Lee, J. *et al.* Characterization of a Tryptophan 6-Halogenase from Streptomyces albus and Its Regioselectivity Determinants. *ChemBioChem* 21, 1446–1452 (2020).

63. Menon, B. R. K. *et al.* Structure and Biocatalytic Scope of Thermophilic Flavin-Dependent Halogenase and Flavin Reductase Enzymes. *Org Biomol Chem* 14, 9354–9361 (2016).

64. Dong, C. *et al.* Structural Biology: Tryptophan 7-Halogenase (PrnA) Structure Suggests a Mechanism for Regioselective Chlorination. *Science (1979)* 309, 2216–2219 (2005).

65. Bitto, E. *et al.* The Structure of Flavin-Dependent Tryptophan 7-Halogenase RebH. *Proteins* 70, 289–293 (2008).

66. Breinlinger, S. *et al.* Hunting the Eagle killer: A Cyanobacterial Neurotoxin Causes Vacuolar Myelinopathy. *Science (1979)* 371, eaax9050 (2021).

67. Jiang, Y. *et al.* The Single-Component Flavin Reductase/Flavin-Dependent Halogenase AetF is a Versatile Catalyst for Selective Bromination and Iodination of Arenes and Olefins. *Angew. Chem. Int. Ed.* 2022, 61, e202214610 (2022).

68. Sainsbury, F., Thuenemann, E. C. & Lomonossoff, G. P. PEAQ: Versatile Expression Vectors for Easy and Quick Transient Expression of Heterologous Proteins in Plants. *Plant Biotechnol J* 7, 682–693 (2009).

69. Berger, F., Hamamura, Y., Ingouff, M. & Higashiyama, T. Double Fertilization - Caught in the Act. *Trends Plant Sci* 13, 437–443 (2008).

70. www.arabidopsis.org. The Arabidopsis Information Resource (TAIR). (2000).

71. Dong, Y. *et al. SHOOTMERISTEMLESS* Autoactivation: A Prerequisite for Fruit Metamorphosis *biorXiv* doi:https://doi.org/10.1101/2024.02.23.581830 (2024)

72. Mcadam, E. L. *et al.* Evidence that Auxin is Required for Normal Seed Size and Starch Synthesis in Pea. *New Phytol* 193–204 (2017).

73. Van Pée, K. H. & Patallo, E. P. Flavin-Dependent Halogenases Involved in Secondary Metabolism in Bacteria. *Appl Microbiol Biotechnol* 70, 631–641 (2006).

74. Tiz, D. B. *et al.* New Halogen-Containing Drugs Approved by FDA in 2021: An Overview on Their Syntheses and Pharmaceutical Use. *Molecules* 27 (5) 1643 (2022).

75. Ritchie, H., Rosado, P. & Roser, M. Agricultural Production. *Our World in Data* Retrieved from https://ourworldindata.org/grapher/pea-yields [online resource] (2023).

76. Ellis, T. H. N. *et al.* Diversity of Pod Shape in *Pisum*. *Diversity* 13, 203 (2021).

77. Bateson, W., Saunders, E. R. & Punnett R C. Experimental Studies on the Physiology of Heredity. *Reports to the Evolution Commitee of the Royal Society*, Report 4 (1904).

78. Feng, C. *et al.* Genomic and Genetic Insights into Mendel's Pea Genes. *bioRXiV*, doi:10.1101/2024.05.31.596837 (2024).

79. Ellis, N. *et al.* Recombinant Inbred Lines Derived from Wide Crosses in *Pisum*. *Sci Rep* 13, Article 20408 (2023).

80. Yang, T. *et al.* Improved Pea Reference Genome and Pan-Genome Highlight Genomic Features and Evolutionary Characteristics. *Nat Genet* 54, 1553–1563 (2022).

81. Barbara Mcclintock, B. The Origin and Behavior of Mutable Loci in Maize. *Proc Natl Acad Sci USA* 36 (6) 344-355 (1950).

82. Kijima, T. E. & Innan, H. On the Estimation of the Insertion Time of LTR Retrotransposable Elements. *Mol Biol Evol* 27, 896–904 (2010).

83. Zulfiqar, S. *et al.* Virus-Induced Gene Silencing (VIGS): A Powerful Tool for Crop Improvement and Its Advancement towards Epigenetics. *Int J Mol Sci* 24, Article 5608 (2023).

84. Hellwig, T. *et al.* Environmental and Genetic Determinants of Amphicarpy in *Pisum fulvum*, a Wild Relative of Domesticated Pea. *Plant Sci* 298, Article 110566 (2020).

85. Bonhomme, S. & Guillory, A. Synthesis and Signalling of Strigolactone and KAI2-ligand Signals in Bryophytes. *J Exp Bot* 73, 4487–4495 (2022).

86. Nelson, D. C. *et al.* Karrikins Discovered in Smoke Trigger *Arabidopsis* Seed Germination by a Mechanism Requiring Gibberellic Acid Synthesis and Light. *Plant Physiol* 149, 863–873 (2009).

87. Jain, N., Kulkarni, M. G. & Van Staden, J. A Butenolide, Isolated from Smoke, can Overcome the Detrimental Effects of Extreme Temperatures during Tomato Seed Germination. *Plant Growth Regul* 49, 263–267 (2006).

88. Conn, C. E. & Nelson, D. C. Evidence that KARRIKIN-INSENSITIVE2 (KAI2) Receptors may Perceive an Unknown Signal that is not Karrikin or Strigolactone. *Front Plant Sci* 6, 1219 (2016).

89. Wang, Q., Smith, S. M. & Huang, J. Origins of Strigolactone and Karrikin Signaling in Plants. *Trends Plant Sci* 27, 450–459 (2022).

90. Song, G. *et al.* CRISPR/Cas9: A Powerful Tool for Crop Genome Editing. *Crop J* 4, 75–82 (2016).

91. Li, G. *et al.* Development of an *Agrobacterium*-mediated CRISPR/Cas9 system in Pea (*Pisum sativum* L.). *Crop J* 11, 132–139 (2023).

92. Walter, A., *et al.* Chlorinated Auxins - How does *Arabidopsis thaliana* deal with them? *Int J Mol Sci* 21 (7) 2567 (2020).

93. Mithila, J., *et al.* Evolution of Resistance to Auxinic Herbicides: Historical Perspectives, Mechanisms of Resistance, and Implications for Broadleaf Weed Management in Agronomic Crops. *Weed Sci* 59, 445–457 (2011).

94. Uzunova, V. V., *et al.* Tomographic docking suggests the mechanism of auxin receptor TIR1 selectivity. *Open Biol* 6 (10) 160139 (2016).

95. McCauley, C. L. *et al.* Transcriptomics in *Erigeron canadensis* reveals rapid photosynthetic and hormonal responses to auxin herbicide application. *J Exp Bot* 71, 3701–3709 (2020).

96. LeClere, S., Wu, C., Westra, P. & Sammons, R. D. Cross-Resistance to Dicamba, 2,4-D, and Fluroxypyr in *Kochia scoparia* is endowed by a mutation in an *AUX/IAA* gene. *Proc Natl Acad Sci USA* 115 (13) E2911–E2920 (2018).

97. Walsh, T. A. *et al.* Mutations in an auxin receptor homolog *AFB5* and in *SGT1b* confer resistance to synthetic picolinate auxins and not to 2,4-dichlorophenoxyacetic acid or indole-3-acetic acid in *Arabidopsis. Plant Physiol* 142, 542–552 (2006).

98. Engvild, K. C. Herbicidal Activity of 4-Chloroindoleacetic Acid and other Auxins on Pea, Barley and Mustard. *Physiol Plant* 96, 333–337 (1996).

99. Tintelnot, J. *et al.* Microbiota-derived 3-IAA Influences Chemotherapy Efficacy in Pancreatic Cancer. *Nature* 615, 168–174 (2023).

100. Shen, J. *et al.* Indole-3-Acetic Acid Alters Intestinal Microbiota and Alleviates Ankylosing Spondylitis in Mice. *Front Immunol* 13, 762580 (2022).

101. Li, M. *et al.* Gut Microbiota Metabolite Indole-3-Acetic Acid Maintains Intestinal Epithelial Homeostasis through Mucin Sulfation. *Gut Microbes* 16 (1) 2377576 (2024).

102. Li, H. *et al.* Gut Microbiota-Derived Indole-3-Acetic Acid Suppresses High Myopia Progression by Promoting Type I Collagen Synthesis. *Cell Discov* 10, 89 (2024).

103. Chen, Y. *et al.* Indole Acetic Acid Exerts Anti-Depressive Effects on an Animal Model of Chronic Mild Stress. *Nutrients* 14 (23) 5019 (2022).

104. Hezaveh, K. *et al.* Tryptophan-Derived Microbial Metabolites Activate the Aryl Hydrocarbon Receptor in Tumor-Associated Macrophages to Suppress Anti-Tumor Immunity. *Immunity* 55, 324-340 (2022).

105. Hu, J., Ding, Y., Liu, W. & Liu, S. When AHR Signaling Pathways Meet Viral Infections. *Cell Commun Signal* 21 (1) 42 (2023).

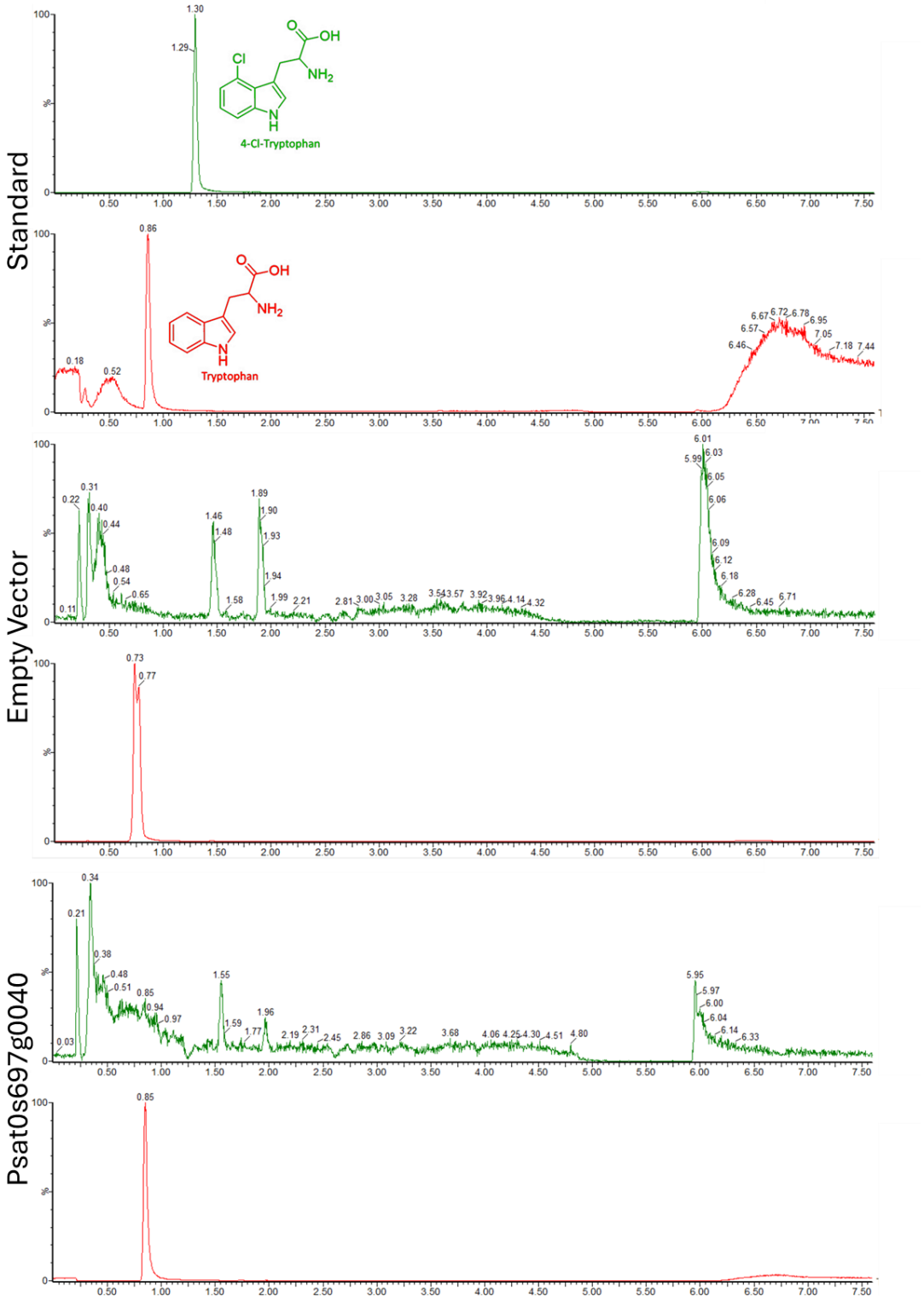
106. Krastanova, O., Hadzhitodorov, M. & Pesheva, M. Ty Elements of the Yeast *Saccharomyces cerevisiae*. *Biotechnol Biotechnol Equip* 19, 19–26 (2005).

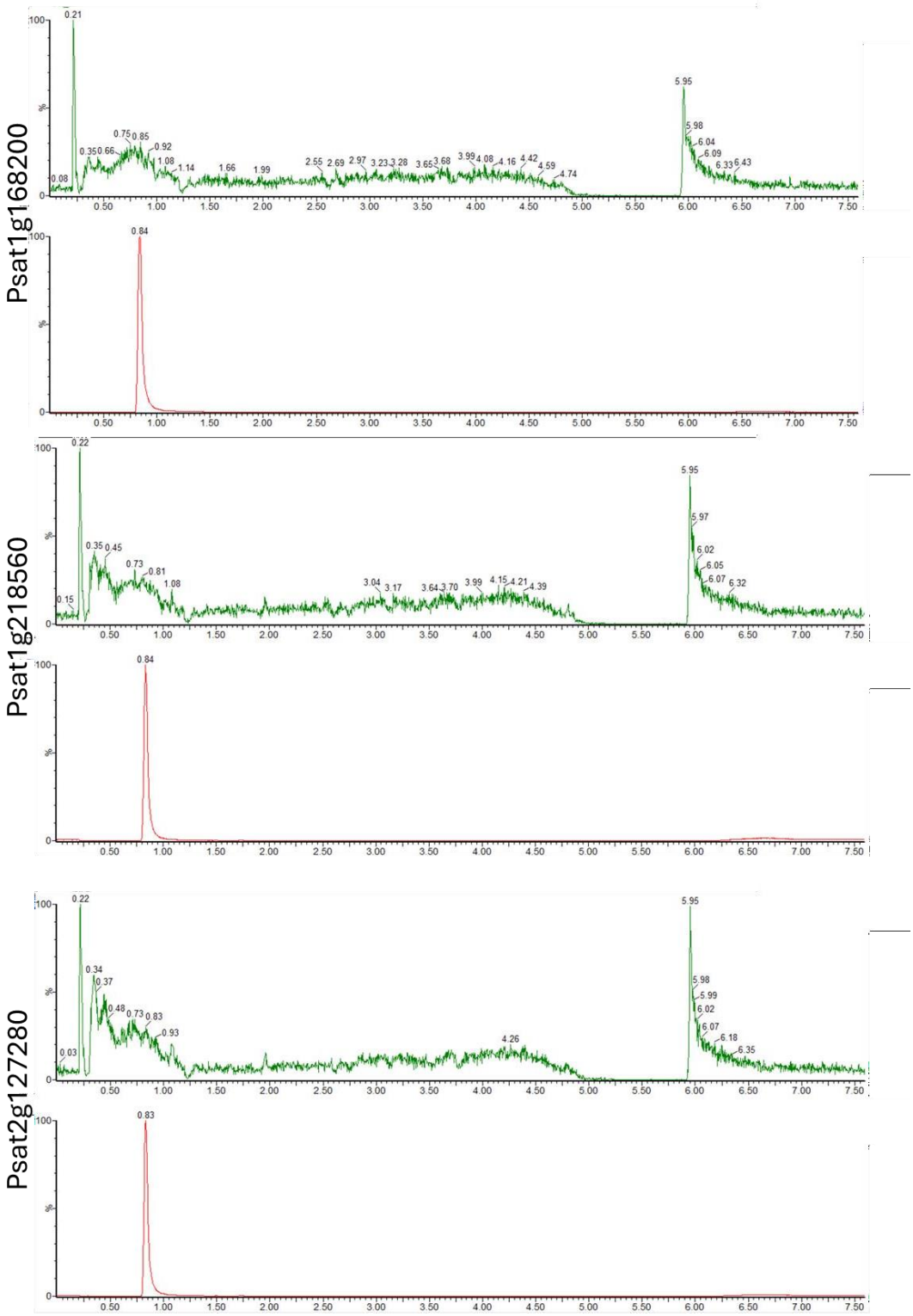
107. Huang, H. Y. *et al.* Harbinger Transposon Insertion in Ethylene Signaling Gene Leads to Emergence of New Sexual Forms in Cucurbits. *Nat Commun* 15, 4877 (2024).

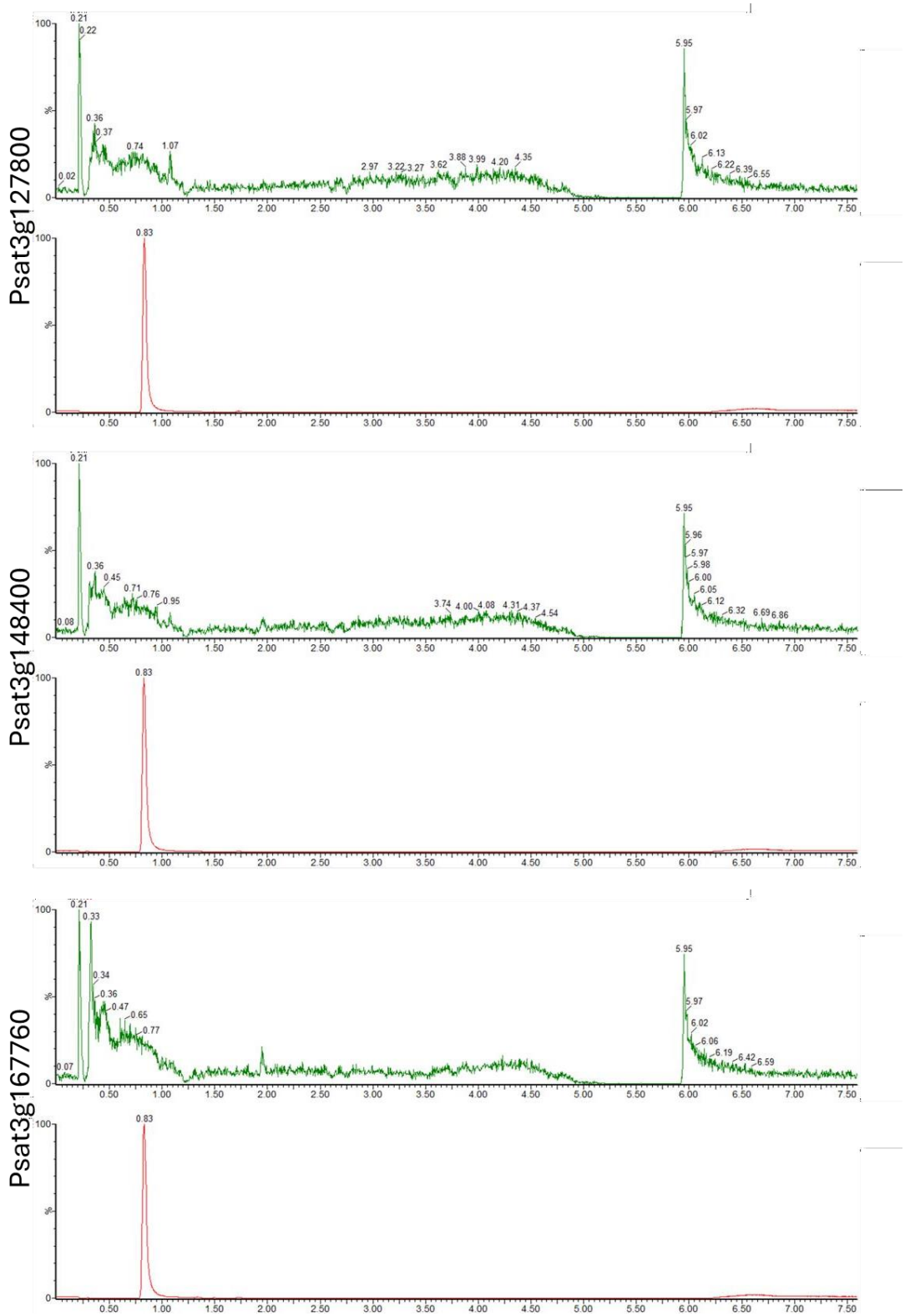
108. Jayakodi, M. *et al.* The Giant Diploid Faba Genome Unlocks Variation in a Global Protein Crop. *Nature* 615, 652–659 (2023).

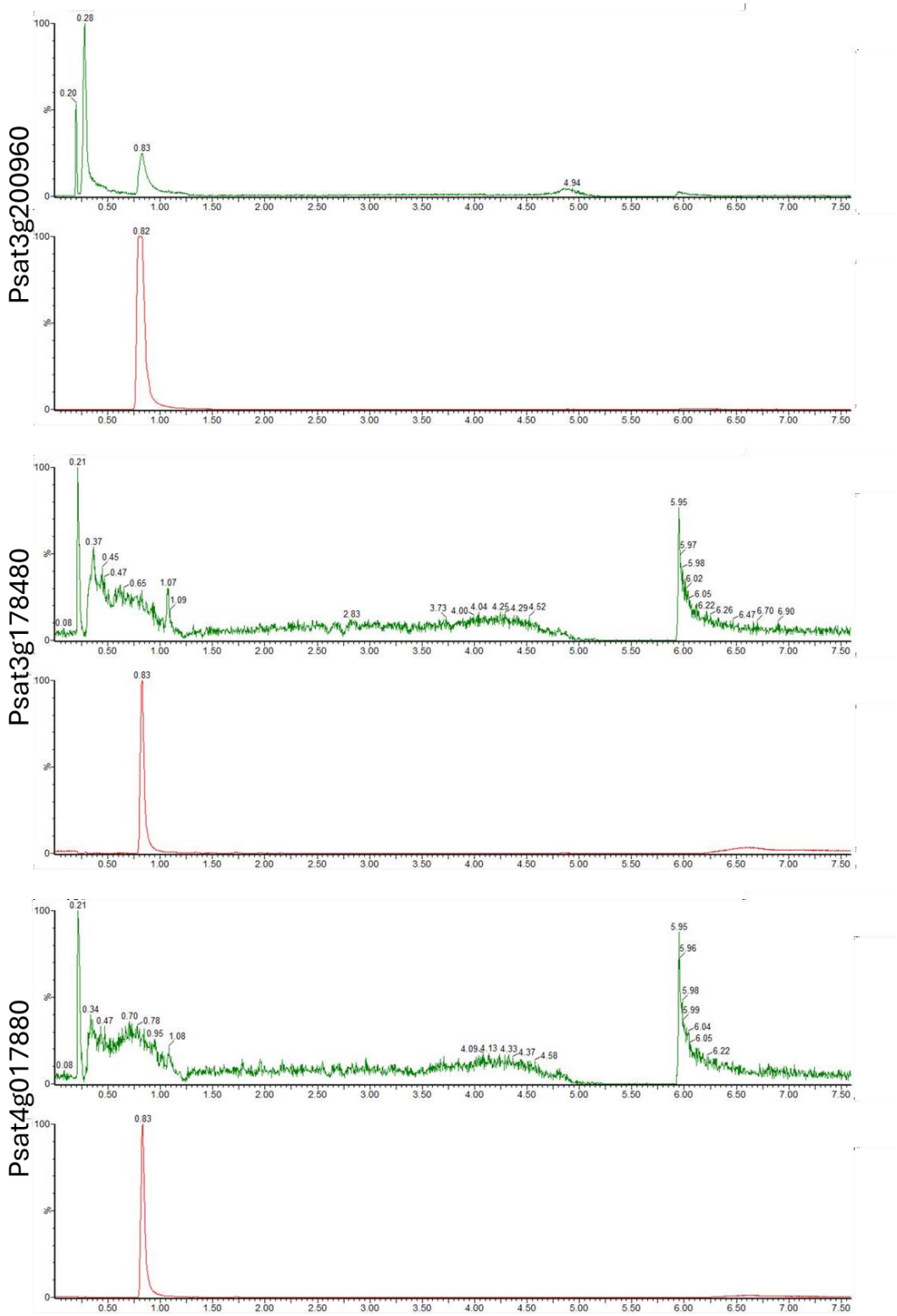
109. Roy, S. *et al.* Celebrating 20 Years of Genetic Discoveries in Legume Nodulation and Symbiotic Nitrogen Fixation. *Plant Cell* 32, 15–41 (2020).

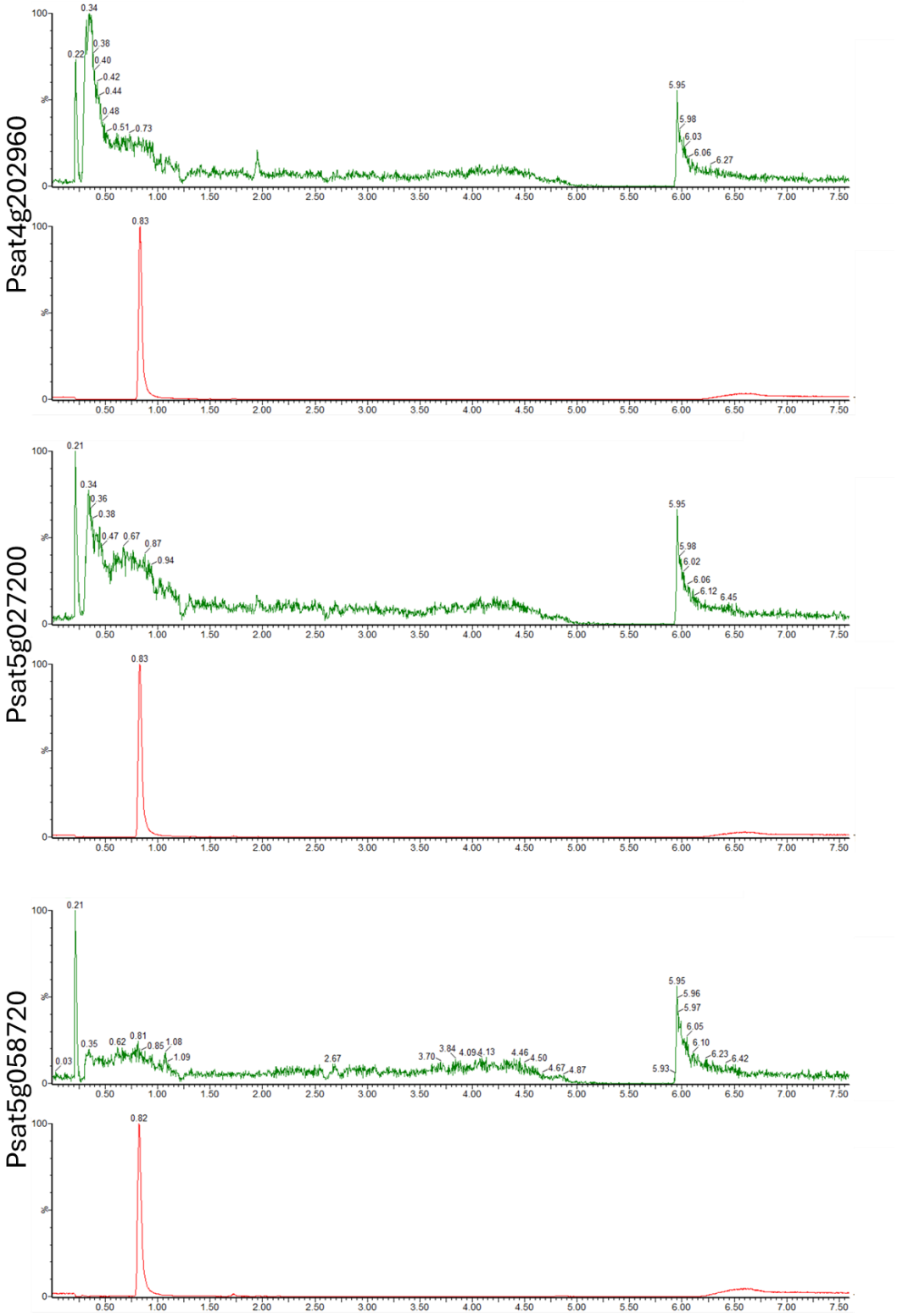
VII. Appendix

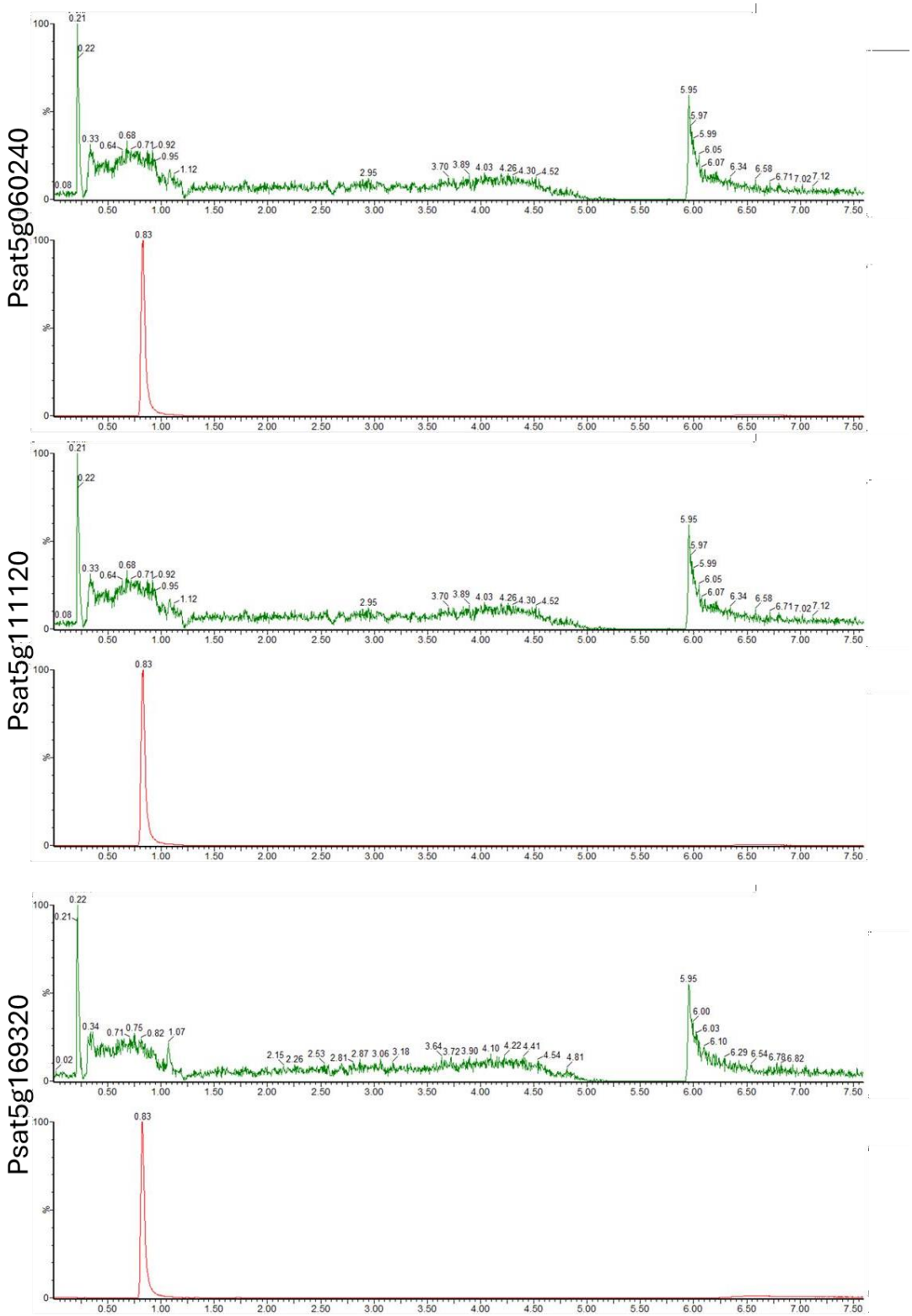


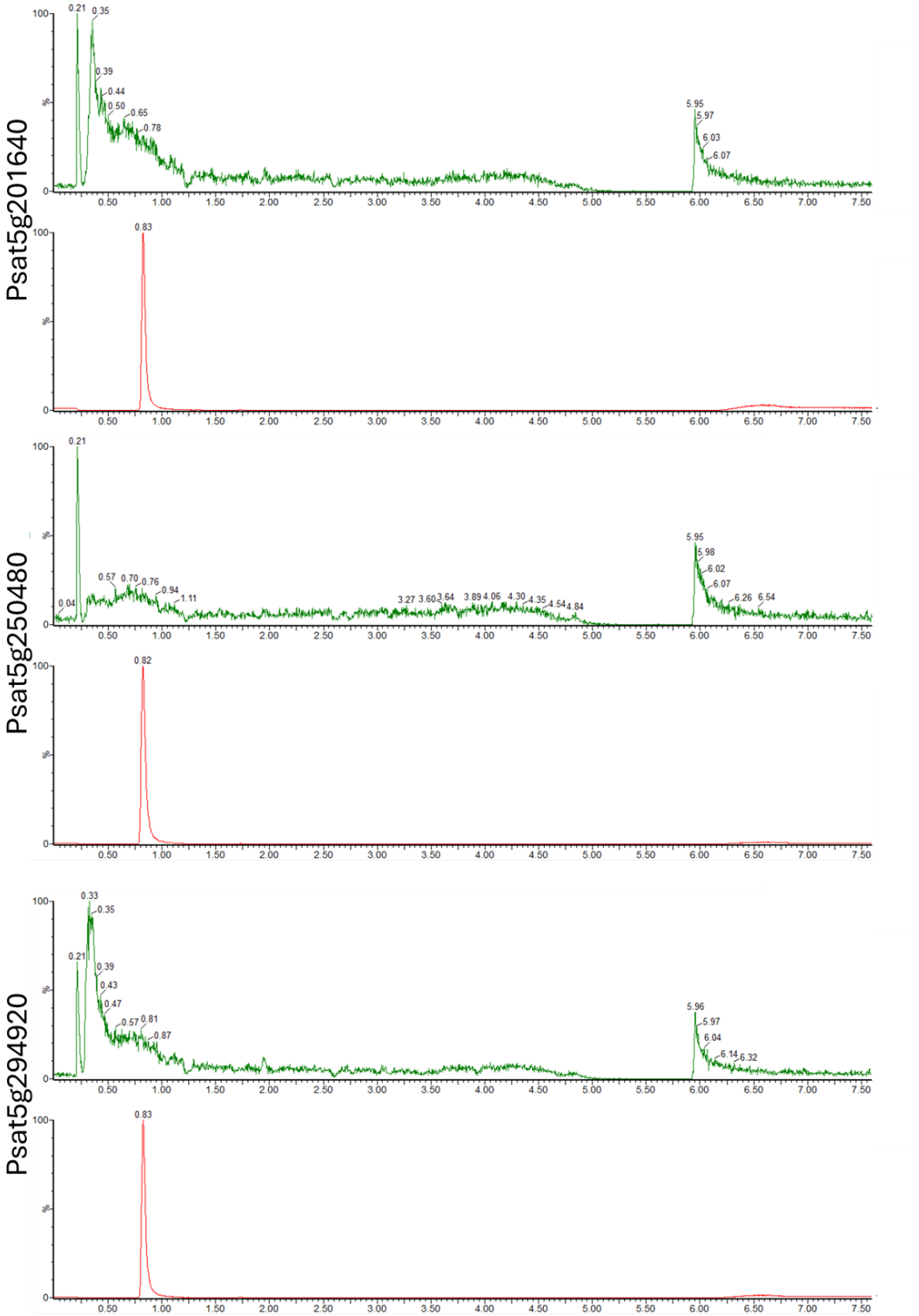

Psat1g168200
Psat1g218560
Psat2g127280

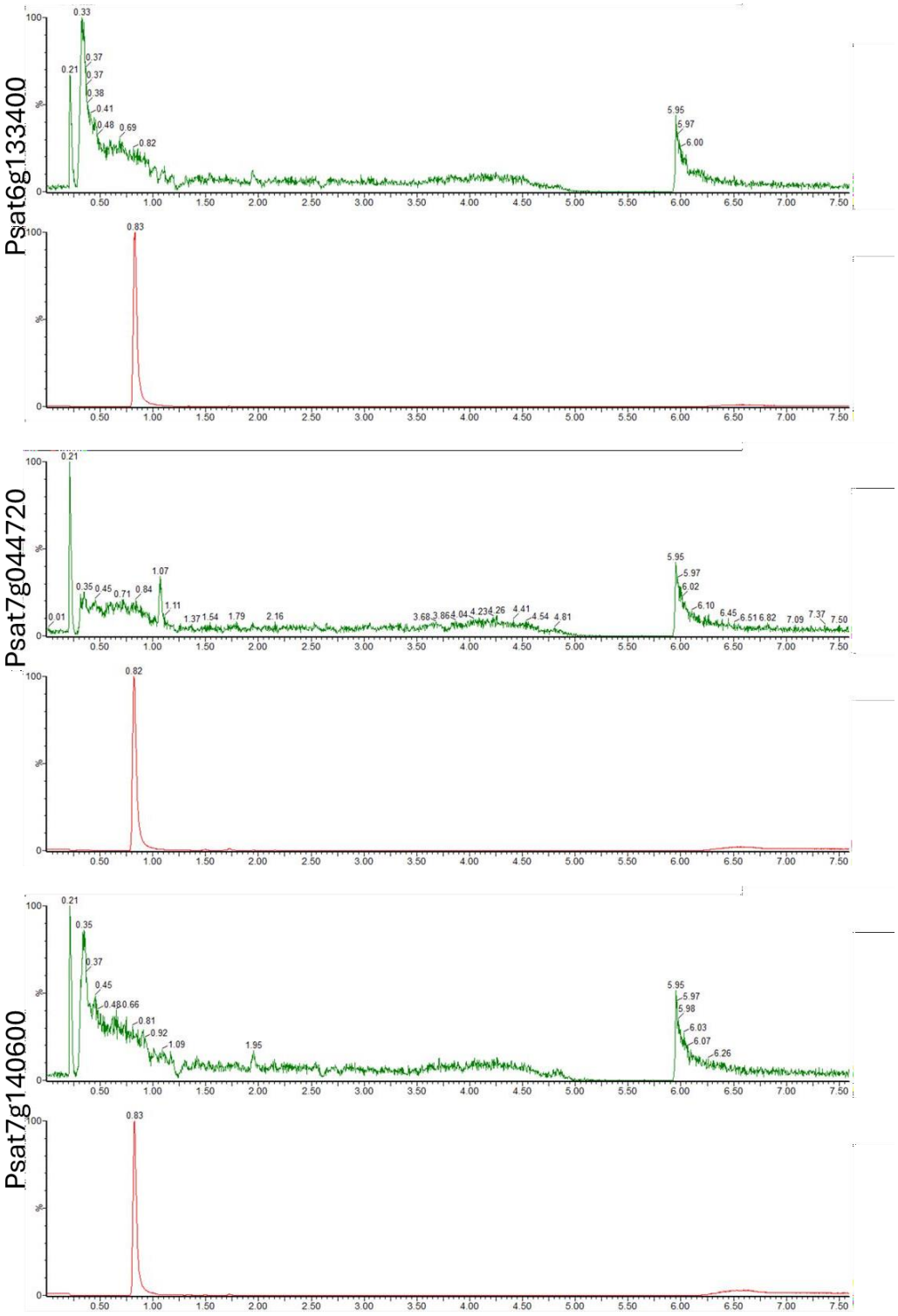

Psat3g127800
Psat3g148400
Psat3g167760

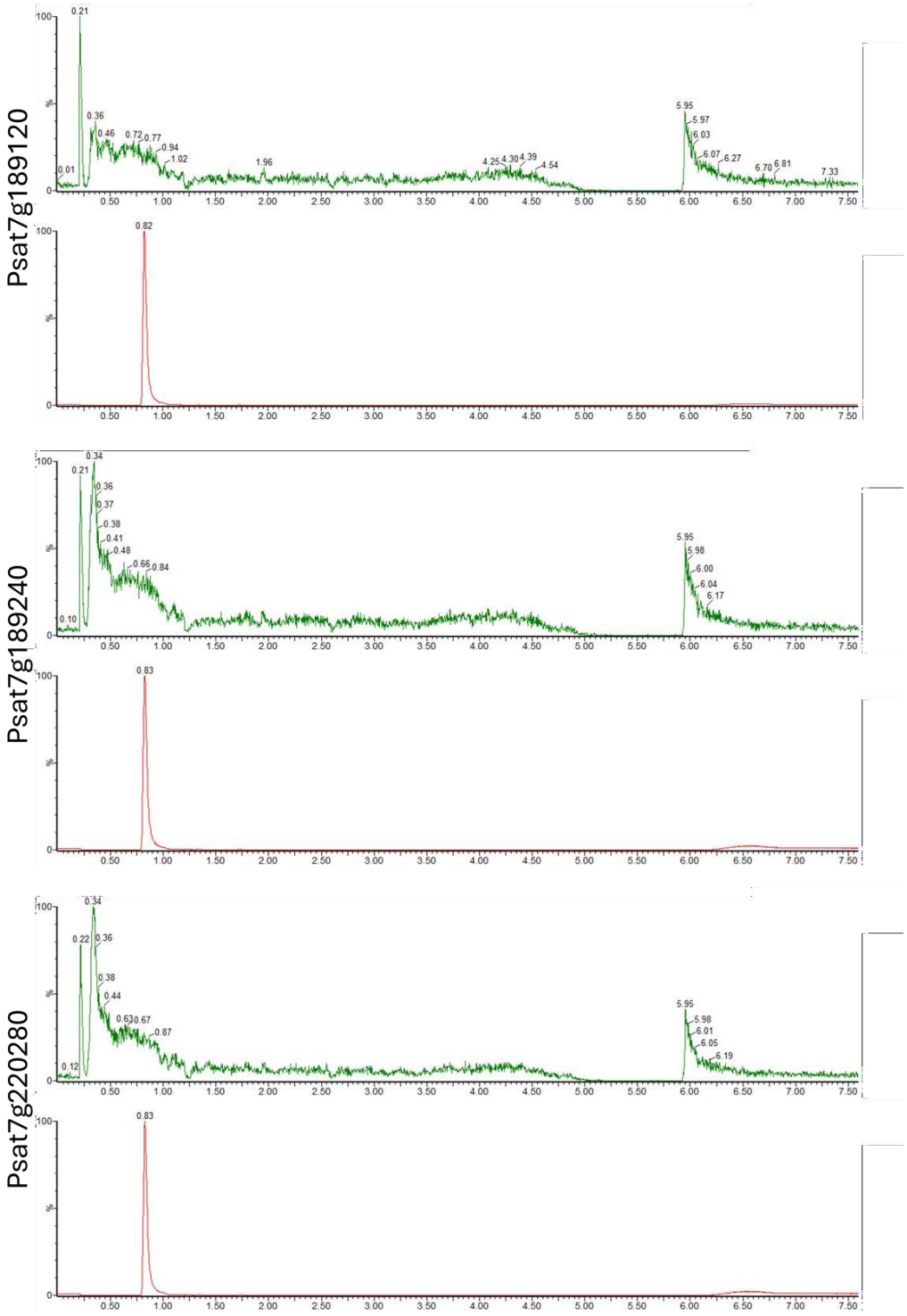

Psat3g200960
Psat3g178480
Psat4g017880

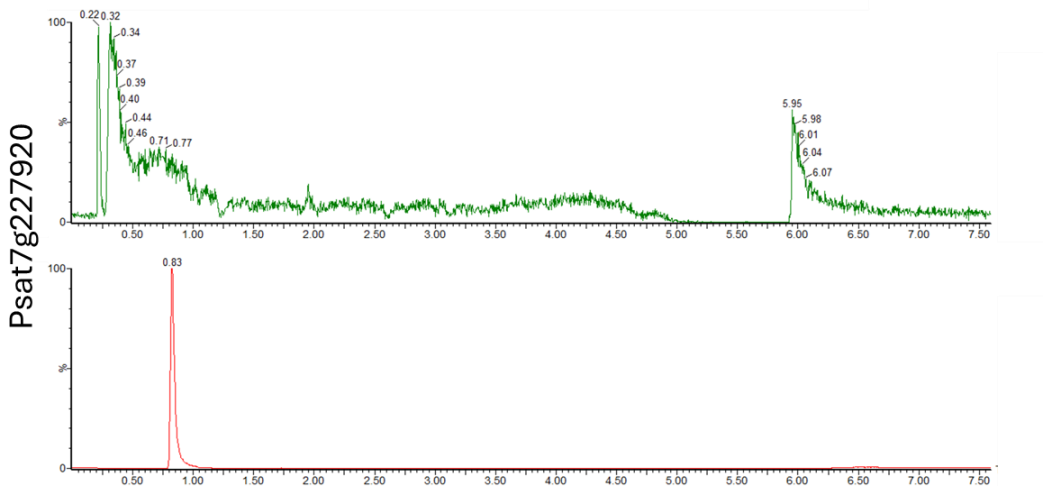

Psat4g202960
Psat5g027200
Psat5g058720


Psat5g060240
Psat5g111120
Psat5g169320


Psat5g201640
Psat5g250480
Psat5g294920


Psat6g133400
Psat7g044720
Psat7g140600


Psat7g189120
Psat7g189240
Psat7g220280



Supplementary Figure S1 – Total Ion Current (TIC) against retention time in minutes of 4-Cl-tryptophan (green) and tryptophan (red) across tobacco extracts after transient expression of specific enzyme candidates. 'Standard' – Typical peaks generated by 4-Cl-tryptophan and tryptophan from solution of dissolved pure compound. 'Empty Vector' – Peaks from tobacco leaf extracts transformed with pEAQ-HT-DEST1 without a cloned enzyme candidate.

VIII. List of Abbreviations

2OGD: 2-oxoglutarate dependent dioxygenase, a class of plant oxygenase enzymes.

4-Cl-IAA: 4-chloroindole-3-acetic acid, an auxin variant and plant hormone.

4-Cl-Trp: 4-chloro-tryptophan, a chlorinated variant of amino acid tryptophan.

ABP1: AUXIN BINDING PROTEIN1, an apoplastic auxin coreceptor.

ARF: AUXIN RESPONSE FACTOR, an auxin-responsive family of transcription factors, which may be activatory or inhibitory to gene expression.

AuxRE: Auxin-responsive element, a cis-regulatory element (TGTCTC consensus) which responds to auxin by acting as a binding site for ARFs.

BLAST: Basic Local Alignment Tool, a bioinformatic algorithm for alignment of nucleic acid or protein sequences.

CRISPR/Cas9: Clustered regularly interspaced short palindromic repeats/ CRISPR-associated protein 9, a bacterial immune signalling system and technology for targeted genetic manipulation across organisms.

CTR1: CONSTITUTIVE TRIPLE RESPONSE 1, a RAF family kinase which constitutively inhibits EIN2 via phosphorylation and is inactivated by ethylene perception.

CYP: Cytochrome p450, originally identified as pigments, CYPs are an enormous superfamily of oxygenase enzymes with considerable diversity across the tree of life.

DAH: DECHLOROACUTUMINE HALOGENASE the first and thus far only halogenase enzyme identified in plants and required for synthesis of highly toxic alkaloid acutumine by Menispermaceae family plants.

DAO: DIOXYGENASE FOR AUXIN OXIDATION1, an oxidase which irreversibly inactivates auxin molecules and prefers to attack conjugated auxins.

DEG: Differentially expressed gene, a gene whose transcript abundance (measured in counts per million) significantly differs between two different

treatments in a transcriptomics dataset, when statistical tests are corrected for multiple testing (i.e. q value is below a designated threshold e.g. $q<0.05$, $q<0.01$).

EIL1/2: ETHYLENE INSENSITIVE-LIKE1/2, ethylene-responsive transcriptional activators.

EIN2/4: ETHYLENE INSENSITIVE2 and 4, EIN4 is an ethylene-inactivated receptor kinase, and EIN2 is a signalling protein which undergoes cleavage when released from CTR1-mediated inhibitory phosphorylation. EIN2-C is C-terminal EIN2 which translocates to nucleus to stabilise ethylene-responsive transcription factors and to destabilise F-box protein-encoding mRNAs.

ERS1/ERS2: ETHYLENE RESPONSE SENSOR1/2, ethylene-inactivated receptor kinases.

ETR1/ETR2: ETHYLENE RESPONSE1/2, ethylene-inactivated receptor kinases.

F/T Clade: Fabeae/Trifolieae clade, a clade of papilionoid legumes including all species in defunct Tribes Fabeae and Trifolieae, inc. pea, fava bean, lentil, *Medicago, Trifolium, Trigonella* and others.

FDA: United States Food and Drug Administration

FDH: Flavin-dependent halogenase, a family of bacterial halogenases capable of halogenating a range of aromatic substrates, and the only halogenase family to include indole halogenases.

FMO: Flavin-binding monooxygenase, a family of oxygenase enzymes.

FN3241/187: Fast Neutron Mutant 3241/187, a *de novo acutilegumen* mutant pea line in genetic background JI2822.

GA: Gibberellic acid, a plant hormone universally associated with cell and tissue elongation.

GA20ox1: GIBBERELLIN 20 OXIDASE 1, a GA biosynthetic enzyme.

GA2ox1: GIBBERELLIN 2 OXIDASE 1, a GA inactivating enzyme.

gDNA: genomic DNA.

GH3: GRETCHEN-HAGEN3, a family of proteins capable of reversibly inactivating auxin by conjugating it to a range of substrates, principally amino acids.

GPX: GLUTATHIONE PEROXIDASE, a family of enzymes capable of reducing hydrogen peroxide and protecting cells from oxidative damage.

IAA: Indole-3-acetic acid, the most abundant auxin in the plant kingdom and a central plant growth regulator.

IGV: Integrated Genome Viewer, a software for visualisation and analysis of genomes and sequencing data.

IPyA: Indole-pyruvic acid, an intermediate in tryptophan-dependent auxin biosynthesis.

JI281: A genetically distinct *Pisum sativum* line obtained from Ethiopia.

JI2822: An experimentally tractable, dwarfed, early flowering recombinant inbred line of *Pisum sativum*. The third pea line to have an assembled genome.

JI3253: Cameor, the first *Pisum sativum* line to be sequenced and have its genome assembled.

KAI2: KARRIKIN-INSENSITIVE2, a receptor for both karrikins (xenobiotic molecules produced from burning plant matter) and for an as-yet unidentified endogenous ligand in plants.

KKI: Kakeimide, a synthetic GH3 inhibitor.

K-Pg: Cretaceous-Paleogene mass extinction event, triggered by the arrival of the Chicxulub impactor into the Earth, killing all non-avian dinosaurs and precipitating the transition of the biosphere from gymnosperm-dominated to angiosperm-dominated.

LCMS: Liquid chromatography-mass spectrometry, a technique for separation and identification of molecules.

LTR: Long terminal repeat, identical, usually direct repeats that demarcate the ends of LTR-retrotransposons.

NCBI: National Center for Biotechnology Information, an National Institutes of Health database collection of biological data.

PCR: Polymerase chain reaction, a technique for *in vitro* sequence-specific DNA synthesis.

PDAC: Pancreatic ductal adenocarcinoma, a highly deadly form of exocrine pancreatic cancer.

PIN: PIN-FORMED1, a family of auxin efflux proteins which mediate polar auxin transport across cells and tissues.

RIL: Recombinant inbred line, a highly homozygous descendent of a cross between two parents generated by repeated self-fertilisation and single seed descent, used for genetic mapping of traits that genetically differ between original parents.

RNA-Seq: RNA-sequencing.

ROS: Reactive Oxygen Species.

SBS: Single-base substitution mutation.

T6P: Trehalose-6-phosphate, a disaccharide signalling molecule in plants.

TAIR: The Arabidopsis Information Resource, an online set of databases containing data generated on model Brassicaceae species *Arabidopsis thaliana*.

TAR: TRYPTOPHAN AMINOTRANSFERASE-RELATED, an enzyme which converts tryptophan to IPyA during auxin biosynthesis.

TILLING: Targeted Induced Local Lesions In Genomes, a large-scale reverse genetics approach involving treatment of a plant species with a mutagen followed by sequencing to generate and identify mutants in sequences of interest.

TIR1/AFB: TRANSPORT INHIBITOR RESISTANT1/AUXIN-BINDING F-BOX, an auxin co-receptor and component of the ubiquitin ligase complex, required for alleviation of ARF repression by promoting ubiquitination and degradation of Aux/IAA repressors.

TMK1: TRANSMEMBRANE KINASE1, a cell surface receptor kinase which is activated by auxin-bound ABP1.

TPL/TPR: TOPLESS/TOPLESS-RELATED, a family of generic transcriptional corepressors, involved in but not exclusive to auxin signalling.

TPP: TREHALOSE-6-PHOSPHATE PHOSPHATASE, an enzyme which removes the phosphate group from T6P to generate trehalose.

TPS: TREHALOSE-6-PHOSPHATE SYNTHASE, an enzyme which synthesises T6P, mutant is embryo lethal in *Arabidopsis*.

Trp: Tryptophan, a core proteinogenic amino acid.

Ty: Transposon of yeast, a family of LTR transposons in *Saccharomyces cerevisiae*

VIGS: VIRUS-INDUCED GENE SILENCING, a technique for targeted silencing of specific transcripts by exploiting endogenous defence systems against RNA viruses.

Y2H: Yeast-2-hybrid, a technique for detecting protein-protein interactions by fusing a transcriptional activation domain to one protein and a DNA-binding domain to another and detecting expression of a marker gene in transformed yeast.

YUC: YUCCA, an auxin biosynthetic enzyme responsible for conversion of IPyA into IAA.