Impact of growth-related genes on petal size in *Arabidopsis thaliana* and the formation of two distinct floral morphs in *Amsinckia spectabilis*

Dissertation

zur Erlangung des akademischen Grades

„doctor rerum naturalium“

(Dr. rer. nat.)

in der Wissenschaftsdisziplin „Genetik“

eingereicht an der

Mathematisch-Naturwissenschaftlichen Fakultät der Universität Potsdam

Institut für Biochemie und Biologie

von

Lisa Bartholomäus

Disputation: 27. August 2021, Potsdam

Hauptbetreuer:	Prof. Michael Lenhard
Gutachter:	Prof. Dr. Mark Johnston
	Prof. Dr. Héctor Candela Antón

„Science is simply common sense at its best, that is, rigidly accurate in observation, and merciless to fallacy in logic."
Thomas Henry Huxley

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-51986
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-519861

Contents

Summary

The life cycle of higher plants is based on recurring phases of growth and development based on repetitive sequences of cell division, cell expansion and cell differentiation. This dissertation deals with two projects, each of them investigating two different topics that are related to cell expansion. The first project is examining an *Arabidopsis thaliana* mutant exhibiting overall cell enlargement and the second project is analysing two naturally occurring floral morphs of *Amsinckia spectabilis* (*Boraginaceae*) differing (amongst others) in style length and anther heights due to differences in longitudinal cell elongation. The EMS-mutant *eop1* was shown to exhibit a petal size increase of 26% caused by cell enlargement. Further phenotypes were detected, such as cotyledon size increase (based on larger cells) as well as increased carpel, sepal, leaf and pollen sizes. Plant height was shown to be increased and more highly branched trichomes explained the hairy *eop1* phenotype. Fine mapping revealed the causal SNP to be a C to T transition at the last nucleotide of intron 7 of the *INCURVATA11* (*ICU11*) gene, a 2-oxoglutarate /Fe(II)-dependant dioxygenase, and thus causing missplicing of the mRNA. Two *T-DNA* insertion lines (*icu11-2* & *icu11-4*) confirmed *ICU11* as causal gene by exhibiting increased petal size. A comparison of three *icu11* alleles, which possessed different mutation-related changes, either overexpressing *ICU11* or modified mRNAs, was the base for investigating the molecular mechanism that underlies the observed phenotype. Different approaches revealed contradictory results regarding ICU11 protein functionality in the *icu11* mutants. A complementation assay proved the three mutants to be exchangeable and *ICU11* overexpression in the wild-type led to an *icu11*-like phenotype, arguing for all three *icu11* mutants to be gain-of-function (GOF) mutants. Contradicting this conclusion, the *icu11-4* line could be rescued by a genomic *ICU11* transgene. A model, based on the assumption that an overexpression of *ICU11* is inhibiting the function of the protein, and thus causing the same effect as a loss-of-function (LOF) protein was proposed. Further, *icu11-3* (*eop1*) mutants were shown to have an increased resistance towards paclobutrazol, a gibberellin (GA) inhibitor and an upregulation of *At*GA20ox2, a main GA biosynthesis gene. Additionally, ICU11 subcellular localization was discovered to be cytoplasmic, supporting the assumption, that *ICU11* affects GA biosynthesis and overall GA level, possibly explaining the observed (GA-overdose) phenotype.

The second project aimed to identify the genetic base of the *S*-locus in *Amsinckia spectabilis*, as the *Amsinckia* genus represents untypical characteristics for a heterostylous species, such as no obvious self-incompatibility (SI) and the repeated transition towards homostylous and fully selfing variants. The work was based on three *Amsinckia spectabilis* forms: a heterostylous form, consisting of two floral morphs with reciprocal positioning of sexual organs (S-morph: high anthers and a short style and L-morph: low anthers and a long style), and two homostylous forms, one large-flowered and partially selfing and the other small-flowered and fully selfing. The maintenance of the two floral morphs is genetically based on the *S*-locus region, containing genes that encode for the morph-specific traits, which are marked by a tight linkage due to suppressed recombination. Natural populations are found to possess a 1:1 S:L morph ratio, that can be explained by predominant disassortative mating of the two morphs, causing the occurrence of the dominant *S*-allele only in the heterozygous state (heterozygous (*Ss*) for the S-morph and homozygous recessive (*ss*) for the L-morph). Investigation of morph-specific phenotypes detected 56% elongated L-morph styles and 58% higher positioned S-morph anthers. Approximately 50% of the observed size differences were explained by an increase in cell elongation. Moreover, additional phenotypes were found, such as 21% enlarged S-morph pollen and no obvious SI, confirmed by hand pollinated seed counts, *in vivo* pollen tube growth and the development of homozygous dominant *SS* individuals via selfing. The *Amsinckia spec. S*-locus was assumed to at least consist of the *G*- (style length), the *A*- (anther height) and the *P*- (pollen size) locus. Comparative Transcriptomics of the two morphs revealed 22 differentially expressed markers that were found to be located within two contigs of a *SS* individual PacBio genome assembly, allowing the localization of the *S*-locus to be delimited to a region of approximately 23 Mb. Contradictory to revealed *S*-loci within the plant kingdom, no strong argument for a present hemizygous region was found to be causal for the suppressed recombination of the *S*-locus, so that an inversion was assumed to be the causal mechanism.

Zusammenfassung

Der Lebenszyklus von Pflanzen ist geprägt von sich wiederholenden Wachstums- und Entwicklungsphasen, die auf wiederkehrenden Abläufen, bestehend aus Zellteilung, Zellvergrößerung und Zelldifferenzierung, basieren. Diese Dissertation ist aus zwei Projekten aufgebaut, die sich beide mit unterschiedlichen Blickwinkeln des Zellwachstums beschäftigen. Im ersten steht die Charakterisierung einer *Arabidopsis thaliana* Mutante, die eine generelle Zellvergrößerung aufweist, im Vordergrund. Das zweite fokussiert sich auf zwei natürlich vorkommende Blütenmorphologien in *Amsinckia spectabilis* (*Boraginaceae*), die sich, aufgrund von Zelllängenunterschieden, in Griffellänge und Höhe der Staubblattposition unterscheiden. Es wurde gezeigt, dass die EMS-Mutante *eop1* durch größere Zellen 26% größere Blütenblätter aufweist. Außerdem wurden weitere Phänotypen beschrieben, wie zum Beispiel, vergrößerte Kotyledonen, (ebenfalls aufgrund von Zellvergrößerung), Fruchtblätter, Kelchblätter, Rosettenblätter und Pollen. Die Gesamtwuchshöhe der Mutante zeigte sich ebenfalls erhöht und zusätzliche Trichomäste erklärten den haarigen Phänotyp. Feinkartierung enthüllte eine C zu T Transition des letzten Nukleotids des Introns 7 des *INCURVATA11* (*ICU11*) Gens, einer 2-oxoglutarat/Fe(II)-abhängigen Dioxygenase, als ursächlichen SNP, welcher missgespleißte mRNA verursacht. Zwei *T-DNA* Insertionslinien (*icu11-2* & *icu11-4*), ebenfalls mit vergrößerten Blütenblättern, bestätigten *ICU11* als kausales Gen, und erlaubten somit die Analyse von drei verschiedenen *icu11* Allelen. Ein Vergleich der verursachten molekularen Veränderung durch die jeweiligen Mutationen ermittelte Unterschiede in den drei Mutanten, wie zum Beispiel Überexpression von *ICU11*, als auch die Modifikation von *ICU11* mRNA. Zusammen bildete das die Grundlage für die Untersuchung des molekularen Mechanismus, der für den beobachteten Phänotyp verantwortlich ist. Verschiedene Ansätze ermittelten widersprüchliche Ergebnisse hinsichtlich der Proteinfunktion von ICU11 in den drei Mutanten. So zeigte eine Komplementierungsanalyse, dass alle drei Mutationen austauschbar sind, was, zusammen mit der Beobachtung, dass eine *ICU11* Überexpression im Wildtyp zu einem *icu11*-ähnlichen Phänotyp zeigte, dazu führte, dass die *icu11* Mutanten als gain-of-function Mutationen eingeordnet wurden. Im Widerspruch dazu stand die Entdeckung, dass sich *icu11-4* durch ein genomisches *ICU11* Transgen retten ließ. So wurde ein Model, basierend auf der Annahme, dass eine *ICU11* Überexpression die Proteinfunktion ebenso hemmt wie ein nichtfunktionales Protein,

vorgeschlagen. Außerdem wurde eine erhöhte Resistenz der *icu11-3* (*eop1*) gegenüber Paclobutrazol, einem Gibberellin (GA)-Inhibitor, und die Aktivierung der Expression von *At*GA20ox2, einem Haupt-GA-Biosynthese-Gen, festgestellt. Zusätzlich wurde eine zytoplasmatische Lokalisation von ICU11 detektiert, sodass ein Einfluss von *ICU11* auf die GA-Biosynthese und somit auf das Gesamt-GA-Level angenommen wird, der den beobachteten (GA-überdosierten) Phänotyp erklären könnte.

Das zweite Projekt strebte die Identifizierung der genetischen Grundlage des *S*-Locus in *Amsinckia spectabilis* an, da die Gattung *Amsinckia* einige untypische Charakteristiken für eine heterostyle Art, wie zum Beispiel das Fehlen einer offensichtlichen Selbstinkompatibilität (SI), sowie die mehrmalige Entwicklung zu Homostyly und 100% autonomem Selbsten, aufweist. Die Analyse basierte auf drei *Amsinckia spectabilis* Varianten: einer heterostylen Form, bestehend aus zwei Blütenmorphologien mit gegensätzlich positionierten Sexualorganen (S-Morph: hohe Staubblattposition und kurzer Griffel und L-Morph: niedrige Staubblattansätze und langer Griffel), und zwei homostylen Formen, einer großblütigen teilweise selbstenden und einer kleinblütigen voll selbstenden. Natürliche Populationen weisen ungefähr ein 1:1 S:L Morph-Verhältnis auf, welches sich durch vorherrschend disassortative Paarung beider Morphs erklären lasst. Dadurch kann das dominante *S*-Allel ausschließlich heterozygot auftreten (heterozygot (*Ss*) im S-morph und homozygot rezessiv (*ss*) im L-morph). Die Suche nach Morph-spezifischen Phänotypen offenbarte 56% längere L-Morph Griffel und 58% höhere S-Morph Staubblattansätze. Zusätzlich wurden 21% größere S-Morph Pollen, sowie das Fehlen einer offensichtlichen SI gefunden. Dies war die Grundlage für die Annahme, dass der *Amsinckia spec.* *S*-Locus mindestens aus *G*- (Griffel), *A*- (Staubblatt) und *P*- (Pollen) Locus besteht. Vergleichende Transkriptom-Analyse beider Morphs offenbarte 22 unterschiedlich exprimierte Marker, die in 2 Contigs der PacBio Genom-Assemblierung eines *SS*-Individuums lokalisiert werden konnten. Dies erlaubte die genetische Einengung des *S*-Locus auf einen Bereich von circa 23 Mb. Gegensätzlich zu bisher aufgeklärten *S*-Loci in anderen Pflanzenarten konnte kein Hinweis auf eine hemizygote Region gefunden werden, die die supprimierte Rekombination am *S*-Locus erklären könnte, sodass eine Inversion als Ursache dieser vermutet wurde.

Abbreviations

2OGDs	2-oxoglutarate /FeII-dependant dioxygenases
A.	*Amsinckia*
A-locus	encodes for anther height within the *S*-locus
At	*Arabidopsis thaliana*
BR	Brassinosteroides
CDK	CYCLIN-DEPENDANT KINASE
cDNA	complementary DNA
Col-0	Columbia-0
CPS	*ent*-copalyl diphosphate *ent*-CDP synthase
CTAB buffer	Cetyltrimethylammoniumbromid
CYCs	CYCLINs
ddH2O	double distilled water
DNA	deoxyribonucleic acid
Dnase	deoxyribonuclease
EDTA	ethylendiaminetetraacetic acetate
eop1	*enhancer of paps1*
EtOH	Ethanol
GA	gibberellic acids
GA1	*CPS*
GA2	*KS*
GA20ox	GA 20-oxidase
GA3	KO
GA5	GA20ox
Gannox	GA *nn*-oxidase
GFP	green fluorescent protein
G-locus	encodes for style length within the *S*-locus
GOF	gain-of-function
ICU11	INCURVATA11
KAO	*ent*-kaurenoic acid oxidase
KO	*ent*-kaurene oxidase
KS	*ent*-kaurene synthase
LOF	loss-of-function
MIGS	miRNA-induced gene silencing
miRNA miR137	MicroRNA
mRNA	messenger RNA
NaOAc	Natriumacetat
PAC	Paclobutrazol
PacBio	Pacific Biosciences

PCR	polymerase chain reaction
Pcz	Propiconazole
P-locus	encodes for pollen size within the *S*-locus
PVP	Polyvinylpyrrolidon
qPCR	quantitative real-time PCR
RNA	ribonucleic acid
Rnase	Ribonuclease
RNA-seq	RNA-sequencing
RT	room temperature 20–23 °C
S-allele	dominant *S*-locus allele
s-allele	recessive *S*-locus allele
SI	self-incompatibility
S-locus	Self-incompatibility-locus
SNP	single nucleotide polyphormism
Spec	*Spectabilis*
Taq	*Thermus aquaticus*
T-DNA	transfer-DNA
Tig	Contig
TNE	Tris NaCl EDTA
w/v	weight per volume
Ws-2	Wassilewskija-2

List of Figures

List of Tables

1. Cell size control during a plant life

The life cycle of higher plants is based on recurring phases of growth and development which can easily be divided into the following: seed germination, juvenile growth, the formation of organs to grow into a mature plant, the transition to flowering, pollination and embryogenesis and finally the production of mature seeds, which represents at the same time the start of the next generation[24]. All of these stages occur with a constancy of cell division, cell expansion as well as cell differentiation (Fig.1). Organ growth always starts with cell division which is mitosis, involving the duplication of a complete set of the genetic material. Thereafter, the final cell size and shape are based on selective transcription and translation of this genetic material via cell expansion and differentiation[25].

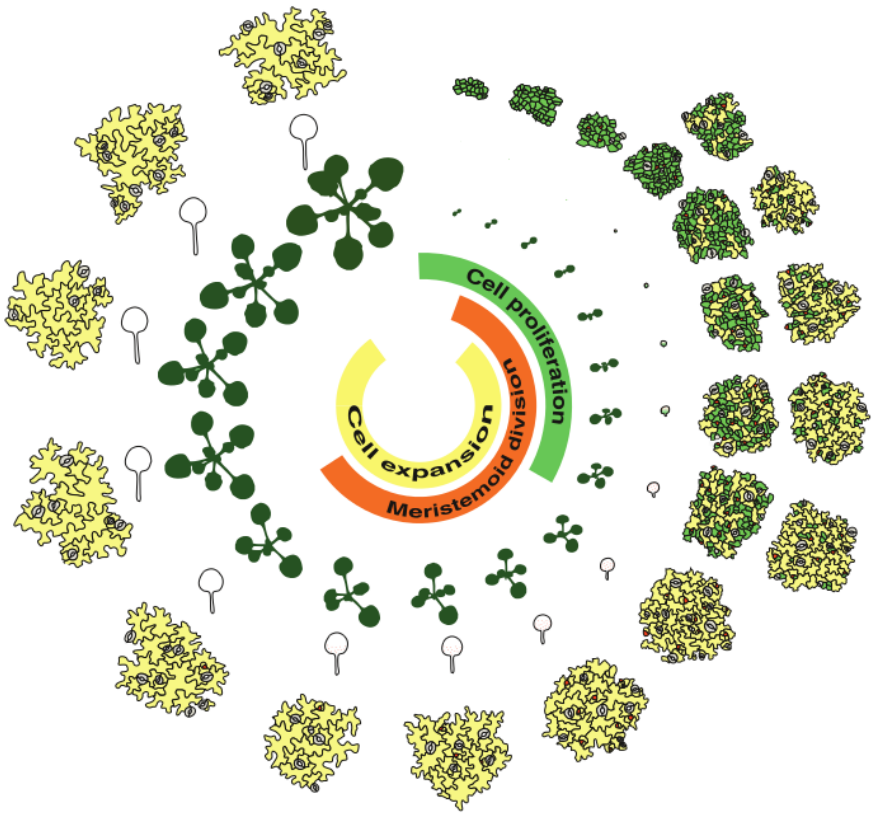


Fig. 1 Developmental stages of *Arabidopsis thaliana* at the rosette, leaf and cellular level (modified after [1])

The developmental timeline starts in the centre of the circle with a cell proliferation phase (in green) and a cell expansion phase (in yellow). Note that leaves exhibit an additional meristemoid division phase (in red) necessary for later stomata cell development. Double cellular drawings represent differences at the tip (outer) and base (inner) of the leaf.

Research often focuses on cell division processes, as these define the (overall) starting point of organ growth[1]. However, in the life cycle of a higher plant, it is the cell expansion that allows the beginning of a new generation, starting with the development of the embryo sac within the seed. Previous studies showed that seed coat growth initiated after fertilization is rather driven by cell expansion than cell division, as the cell number of the integument cells stays unchanged[26,27]. In parallel, the central vacuole rapidly expands until the early heart stage of the embryo is reached[28]. The central vacuole is representing the base for the later endosperm development, that is responsible for the nutrition and therefore, the survival of the developing embryo[29] (Fig. 2). Hence, it has to be emphasized that without cell expansion it would not be possible for a new plant to emerge.

This dissertation deals with two projects, each of them investigating two different topics related to cell expansion. Cell expansion is a directed process, occurring either as cell elongation along the longitudinal axis or as cell enlargement of the complete cell along the transverse axis[30]. The two projects described in this thesis mirror these processes, as the first project is examining a mutant exhibiting overall cell enlargement and the second project is analysing two naturally occurring floral morphs differing (amongst others) in style length and anther heights due to differences in longitudinal cell elongation. Since both projects were based on different plant species, *Arabidopsis thaliana* (*Arabidopsis*) and *Amsinckia spectabilis* (*Boraginaceae*) respectively, it was decided to divide this dissertation into two parts. First, a general introduction of cell size regulation in plants will be given.

1.1 Regulation of cell size in plants

In plants, organ growth is regulated via overlapping and strongly interconnected phases: an initiation phase, a general cell division phase, a transition phase and a cell expansion phase. Many genetic regulators are known to control these developmental processes in a very strict spatial and temporal pattern, often affecting several components simultaneously[1]. In the following, the focus will be set on regulators predominantly responsible for cell expansion because these factors are particularly important regarding the two research projects described in this thesis.

Final cell size is influenced by the cell expansion rate and the cell expansion duration[31], which represent different targets of genetic regulation pathways (Fig. 2).

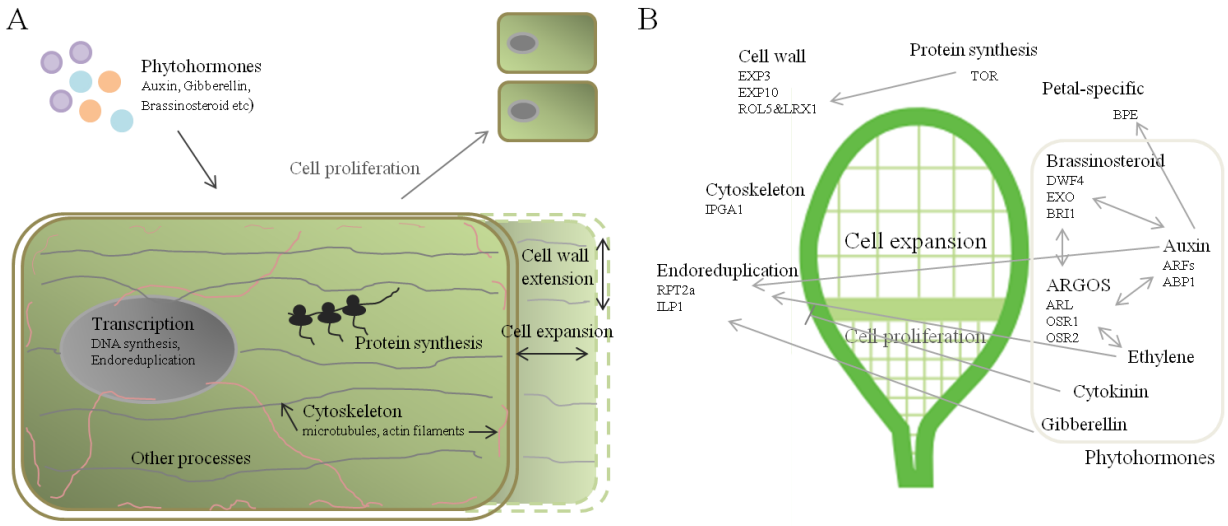


Fig. 2 Regulation of cell expansion in *Arabidopsis thaliana*

(A) Modified scheme[32] of cellular target pathways and structures effecting cell expansion **(B)** Illustration of main growth regulators of cell size (modified after [33]). Arrowheads and bars indicate positive and negative effects between different pathways, respectively. Left-right arrowheads represent interactions between pathways.

1.1.1 Cell size regulation via the cell wall

Cell expansion is based on turgor-driven size increase after cell wall loosening and de novo synthesis of cell wall components[34]. The cell wall of a growing cell stretches irreversibly depending on the enlargement of the cell volume. Since this process can be interrupted and continued, in less than a minute, the process is supposed to be dynamically regulated. The molecular process of cell expansion seems to be based on wall pH changes that affect the activity of EXPANSINs (EXP), cell-wall associated proteins that are known to control the cell wall extensibility[35]. Overexpression of EXP10 under the control of its own promoter leads to the formation of elongated petioles and larger leaves, the latter due to enlarged cells[36]. A similar phenotype can be found in EXP3-overexpressing plants exhibiting larger leaves. Here, the underlying mechanism is still unknown[37].

1.1.2 Impact of cytoskeleton on cell size

The cytoskeleton plays a central role in the plant cell growth, since it is responsible for the positioning of the necessary cell wall materials. Cellulose microfibrils are the main structural components of the cell wall. Their transport is linked to the arrangement of cortical microtubules in expanding cells[38]. One negative regulator for anisotropic petal growth is INCREASED PETAL GROWTH ANISOTROPY1 (IPGA1), a microtubule-associated protein. It co-localizes with and directly binds to cortical microtubules playing a negative role in the microtubule ordering. Thus, *ipga1-1* loss-of-function mutants exhibit longer but narrower petals and an increase of anisotropic cell expansion of the petal epidermis in the late flower development[39].

1.1.3 Effect of protein biosynthesis on cell size

A decrease of protein biosynthesis leads to smaller cells, as was shown for RNA interference (RNAi) lines of TARGET OF RAPAMYCIN (TOR), a serine/threonine kinase acting as a major growth regulator in all eukaryotes[40]. TOR is binding to rRNA gene promoters to activate their expression and therefore determine the number of translationally active ribosomes. Thus, a reduction in soluble protein content due to a decreased number of polysomes could explain the cell size reduction in RNAi lines[41]. Since TOR is regulating cell growth by inducing different molecular pathways, a further explanation, especially for the root cell size increase in TOR-overexpressing lines[42], can be found in its effect on cell wall structures via REPRESSOR OF LRX1 (ROL5) and LRR-EXTENSIN1 (LRX1)[43].

1.1.4 Endoreduplication as further cause for increased cell size

Strongly discussed in recent studies is the impact of genome size and therefore nucleus size in plant cell sizes[44]. Although new evidence was found to contradict this theory[45], there is a general trend that cells with greater DNA content are larger. Robinson et al. proved this effect by showing enlarged petal cells and organ sizes of tetraploid and octaploid *Arabidopsis* plants compared to their diploid wild-types[46]. In this regard, another interesting phenomenon was found in *Arabidopsis thaliana* because its leaf cell sizes are shown to be associated with endoreduplication, a cell cycle variant, allowing several rounds of genome duplications without subsequent mitosis[47,48].Thus, *Arabidopsis* ploidy levels range between 2C stomata cells (C represents the amount of unreplicated DNA in haploid cells) and 32C trichome cells within the same leaf[49]. Even more extreme cases of this linkage between cell size and increased

endoreduplication rounds were observed in other plant species, such as a ploidy level of 8192C in the largest *Phaseolus coccineus* suspensor cells[50]. Cell cycles usually follow a strict sequence of Gap phase 1 (G1-phase), DNA synthesis phase (S-phase), Gap phase 2 (G2-phase) and finally the mitotic phase (M-phase). However, endocycling cells skip the complete mitotic phase to prevent chromosome segregation and cell division[2] (Fig. 3). The exit from the M-phase is controlled by the inactivation of mitosis-promoting factors, complexes of CYCLIN-DEPENDANT KINASEs (CDKs) and mitotic CYCLINs (CYCs), CDK/CYCA or CDK/CYCB, respectively[51]. Thus, CDKB1 associated with A2-type CYCs is repressed in endoreplicating cells. Since *cdkb1* mutants exhibit only mild defects and are still able to undergo endoreduplication, the major regulator of endoreduplication remains still unknown[52]. SIAMESE RELATED (SMR) as well as the ANAPHASE-PROMOTING COMPLEX/CYCLOSOME (APC/C) act as CDK inhibitors and are therefore responsible for the proteolysis of CYCs by the 26S proteasome[2].

Pointing towards a mechanistic link between endoreduplication and cell size is REGULATORY PARTICLE AAA-ATPASE 2a (RPT2a), a subunit of the 26S proteasome. If downregulated, leaves exhibit enlarged cells with increased ploidy levels[53]. Furthermore, an upregulation of the transcriptional repressor INCREASED LEVEL OF POLYPLOIDY1 (ILP1) results in enlarged cells of hypocotyls and leaves as well as increased ploidy levels compared to wild-type plants. The mutant phenotype can be explained by ILP1 repressing CYCA expression[54].

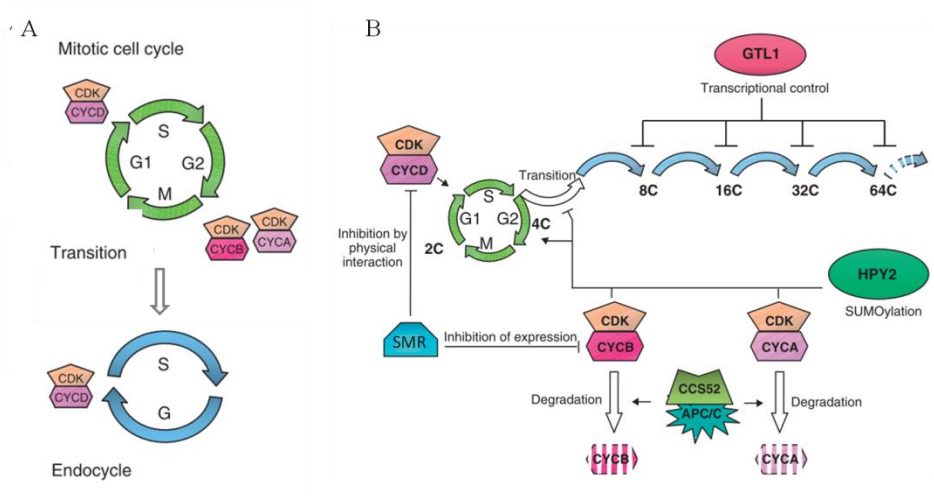


Fig. 3 Developmental control of endocycles and cell growth in plants (modified after [2])

(A) The mitotic cell cycle is marked by a repetitive pattern of G1, S, G2 and M-phases. Endocycling cells undergo several rounds of S phase by skipping the M phase, resulting in a doubled amount of DNA content per endocycle. CYC/CDK complexes are responsible for the regulation of the mitotic cycle progression at the transition from G1 to S and G2 to M phases. Reduced levels of CYCA/CDK and CYCB/CDK are necessary to enter the endocycle. **(B)** Regulation of endocycle entry and exit. SMR decreases CYCB/CDK expression which - in concert with CCS52/APC/C complex-mediated proteolysis to the degradation of the CYC/CDK complexes by the 26S proteasome. Additional repressors of either the entry or the progression of the endocycle are the SUMO E3 ligase HPY2 and the factor GTL1, respectively.

1.1.5 Phytohormones as cell size regulators

Phytohormones are known to act as signaling transmitters via affecting the expression of their responsive genes, which are often responsible for growth regulation[55]. Accordingly, auxin, cytokinins as well as ethylene were shown to determine cell size along the longitudinal as well as the transversal axis[34]. Many cell size regulations are clearly linked to one specific phytohormone pathway, since most factors directly act on the synthesis or signaling of single phytohormones. BRs (essential hormones for plant growth and development) for example, are bound by the membrane localized receptors BRASSINOSTEROID-INSENSITIVE 1 (BRI1) and BRI1-ASSOCIATED RECEPTOR KINASE 1 (BAK1) after enzymatic synthesis. The main regulatory point of the BR signaling transduction is the dephosphorylation of BRASSINAZOLE-RESISTANT 1 (BZR1) and BRI1-EMS-SUPPRESSOR 1 (BES1) protein complex (BZR1/BES1) by BRI1 SUPPRESSOR 1 (BSU1) phosphatase, leading to gene expression of BR-responsive genes and amongst others to cell elongation[56,57] (Fig. 4). Changes in cell size compared to wild-types can easily be caused by mutated proteins of BR synthesis or signaling. Therefore, enhanced expression of BRI1 receptor results in elongated leaf petioles[58], the same

was observed by ectopic expression of a BR biosynthesis enzyme DWARF4 (DWF4)[59], both leading to higher BR levels in the cells. Further, an upregulation of BR- responsive genes, e.g. EXORDIUM (EXO) causes larger leaf cells and promoted root growth[60].

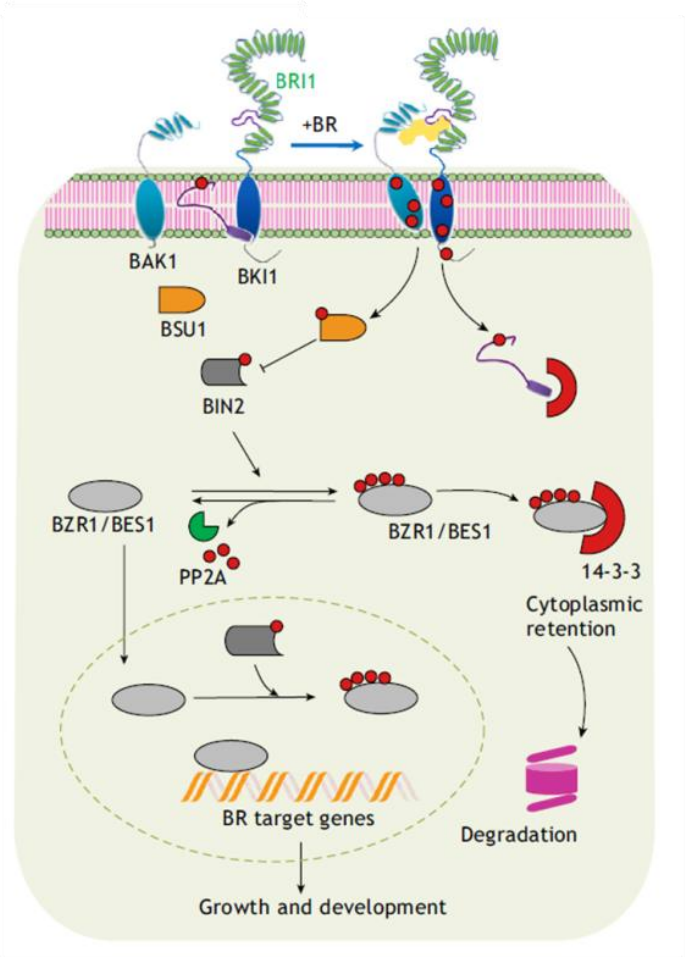


Fig. 4 Schematic representation of Brassinosteroid signalling pathway (modified after [3])

In the absence of BR, BIN2 phosphorylates BZR1/BES1 proteins. Further binding to 14-3-3 proteins leads to cytoplasmic inactivation and eventually to degradation. Upon binding of BR by BRI1, a BAK1/BRI1 heterodimer is formed, activating an intracellular phosphorylation cascade that results in the dephosphorylation and therefore activation of BZR1/BES1 proteins allowing the BR response to happen.

An auxin-linked example for a cell size regulator, acting in only one pathway, is AUXIN BINDING PROTEIN 1 (ABP1). Ectopic expression results in enlarged leaf cell sizes and a receptor-like increased auxin-binding capacity was shown to be the base of ABP1 function[61].

However, the presence of cross talks between the single phytohormone pathways complicates a strict functional splitting of known regulators regarding their genetic mechanisms. One representative example is AUXIN RELATED FACTOR 2 (ARF2), a transcription factor that

mediates gene expression in response to auxin. *Arf2* mutants exhibit thickened stems, enlarged seeds, embryos, leaves and sepals, that are caused by increased cell expansion[62]. Although a clear link of ARF2 to the auxin pathway is not known, the auxin-response element (TGTCTC) was found to be enriched in the promoters of BR-responsive genes[63]. Following this, it has been described that the BR-regulated-BRASSINOSTEROID-INSENSITIVE 2 (BIN2) kinase can phosphorylate ARF2 resulting in a loss of DNA-binding and repression activities. Consistent with this, treatment with brassinazole (BRZ), a BR synthesis inhibitor, is resetting most of the differential expressed genes back to wild-type level in *arf2* mutants[64].

Further examples for the cross talk between the phytohormone pathways are proteins of the AUXIN REGULATED GENE INVOLVED IN ORGAN SIZE (ARGOS) gene family: Members of this family are ARGOS, ARGOS-LIKE (ARL) and ORGAN SIZE RELATED (OSR) 1 and 2. Recent studies identified ARGOS as founding member of this family inducible by auxin. Transgenic plants expressing sense and antisense ARGOS cDNA exhibit enlarged aerial organs based on a prolongation of cell proliferation via an elevated expression of AINTEGUMENTA (ANT) and CYCD3;1[65]. ARGOS overexpressing plants were additionally shown to increase cell expansion[66]. Ectopic expression of OSR1 and 2[67,68] as well as ARL[69] resulted in enlarged organs, caused by an accelerated rate of cell expansion[68]. All three proteins act redundantly and are closely related homologues, since they share a conserved OSR domain, which is sufficient to explain the promoted organ growth[67,68]. Although the exact functional mechanisms are still unknown, a clear influence of phytohormones on these three candidates has been shown. Thus, ARL is upregulated by BR and might act downstream of BRI1 since ectopic ARL expression partially rescues cell expansion in the *bri1-119* mutant[69]. OSR2 as well is induced by BR[67] , whereas OSR1 expression is induced by ethylene but repressed by ABA and BR[68]. Due to highest expression levels of ARL in flowers, recent studies discuss the inhibition of ethylene signaling as the major function of ARL. Increased ethylene signaling was shown to cause premature flower ripening in advance of fertilization. ARL might act in preventing this[70].

In some cases, phytohormones can also cause organ-specific cell expansion. For example, a petal-specific isoform of the basic helix-loop-helix (bHLH) transcription factor BIG PETAL (BPEp) was shown to interact with AUXIN RESPONSE FACTOR8 (ARF8) limiting cell expansion in petals. This interaction is most likely influenced by local auxin levels[71].

Recent studies found evidence that phytohormones can also affect cell size via endoreduplication. The decreased ploidy level in *ga1-11* mutants grown in the dark, for instance, could be rescued by exogenous GA4+7 to a wild-type like ploidy level. Ethylene, on the other side, led to an additional round of endoreduplication on light- and dark grown seedlings[72]. Auxin biosynthesis, transport and signaling loss-of-function mutants undergo additional endocycles up to 256C, indicating that also auxin plays an important role regarding the switch between mitotic and endocycle. Cytokinin, on the contrary, seems to act antagonistically to auxin in *Arabidopsis* roots by suppressing the entrance into the endocycle, as cytokinin biosynthesis and signaling loss-of-function mutants exhibit a delay in this transition[73]. Although the detailed mechanisms of the impact of phytohormones on the endocycle are still unknown, possible check points could be the transcriptional regulation of SMR and APC by hormones additionally to other developmental regulators[44].

Regarding the first project of this thesis, a more detailed insight about the effect of gibberellins on cell size will be given in the following.

2. Organ-size increase caused by an EMS-mutation in the *INCURVATA11* gene

2.1 Introduction

2.1.1 Gibberellin pathway and its impact on cell size

2.1.1.1 Gibberellin biosynthesis

Gibberellins (GA) are a class of tetracyclic diterpenoid carboxylic acids responsible for the regulation of cell division and cell expansion during plant growth and development acting throughout the entire life cycle of a higher plant, starting with seed germination, hypocotyl- and stem elongation, leaf expansion, flowering and fruit development[74]. Thus, GA-deficiency results in dwarf phenotypes of the entire plant, whereas increased GA content leads to a typical GA-overdose phenotype, including enlarged organs due to more and larger cells and earlier flowering[75,76] Until today 136 different GAs have been described in bacteria, fungi and plants, although only few of them, such as GA_1, GA_3, GA_4 and GA_7, are bioactive[4,77]. In *Arabidopsis*, the major bioactive form is GA_4, since overall GA_4 levels are highly increased compared to GA_1[77].

The biosynthesis of GAs requires a set of enzymes, including terpene cyclases (TPS), membrane-associated cytochrome P450 monooxygenases as well as soluble 2-oxoglutarate-dependent dioxygenases (2OGDs). It is localized in plastids, the endomembrane system and the cytosol, thus occurring in three main steps (Fig. 5). The initiation of GA biosynthesis is the two-step-conversion of geranylgeranyl-diphosphate (GGDP) to *ent*-kaurene catalyzed by two TPSs, *ent*-copalyl diphosphate (*ent*-CDP) synthase (CPS) and *ent*-kaurene synthase (KS), occurring in the proplastids[78]. Intriguingly in plants, both synthases are encoded by highly conserved single copy genes[75] and are not under feedback control of the GA signaling pathway, differentiating them from later GA-biosynthesis genes. Overexpression of *CPS* and *KS* rescued the dwarf phenotypes of the corresponding loss-of-function mutants but did not lead to a GA-overdose phenotype if transformed into wild-type plants. Therefore, accumulation of early GA-biosynthesis intermediates, like CPS and KS, do not cause higher expression of later GA-biosynthesis genes, representing the ability of plants to ensure GA homeostasis already at these early time points of GA-pathway. These very early intermediates allow the first steps of

controlling the GA homeostasis via the regulation of their expression. *KS* (*GA2*), for example, is expressed constitutively with a peak in flower organs, whereas *CPS* (*GA1*) only at a low level, most likely maintaining the ability to start GA synthesis at any required time point[79].

The endoplasmatic reticulum presents the location for six additional oxidative steps leading to the formation of the GA precursor GA_{12} by *ent*-kaurene oxidase (KO) and *ent*-kaurenoic acid oxidase (KAO). *At*KO is a multifunctional cytochrome P450 (CYP701A) encoded by the single copy and highly conserved *GA3* gene, expressed most highly in flowers[80] and not feedback controlled by GA-signaling similar to *GA1* and *GA2*[81]. *At*KAO1 and *At*KAO2, on the other hand, are members of the CYP88A subfamily, and therefore the first non single-copy genes in the GA-biosynthesis. The two *Arabidopsis* homologues function redundantly. Only double mutants exhibit the typical GA-deficient dwarf phenotype, whereas both single mutants possess a wild-type like phenotype[82]. Additionally, they exhibit their expression peak during seed germination and in young developing plant organs, which is typical for most of the early GA biosynthesis intermediates[83]. The last step of GA synthesis, localized in the cytosol, starts with the hydroxylation of GA_{12} into GA_{53} by a GA 13-oxidase (GA13ox), giving rise to the two-branched pathway to finally form three bioactive GAs. GA_{12} represents the starting point for the non-13-hydroxylation branch leading to bioactive GA_3 und GA_4, whereas GA_{53} starts the 13-hydroxlyation pathway resulting in the formation of bioactive GA_1[4]. Interestingly, the encoding gene for GA13ox has not been identified yet. Due to the lack of *Arabidopsis* GA13ox mutants, overexpression of CYP714B1 and CYP714B2, two rice GA13ox enzymes, was analyzed in *Arabidopsis*, resulting in a dwarf phenotype due to increased levels of GA_1 and decreased levels of GA_4; supporting the assumption that GA_4 is more active in promoting plant growth. Of interest is also the rice double mutant *cyp714b1 cyp714b2*, displaying a wild-type-like phenotype, that exhibits increased expression of GA_4 and of all cytosolic GA_4 intermediates[84]. Thus, a lack of GA_1 seems to shift the GA synthesis towards increased production of GA_4.

In both cytosolic pathway branches, several oxidation steps of various GA intermediates are catalyzed by 2-oxoglutarate/Fe(II)-dependant dioxygenases (2OGDs): GA 20-oxidases (GA20ox or *GA5)* followed by 3β-hydroxylation by GA 3-oxidases (GA3ox or *GA4*) as final step [4]. Therefore, the overall concentration of bioactive GAs is determined by the activity of GA20ox and GA3ox 2OGDs. *Arabidopsis* possesses five GA20ox paralogues, *At*GA20ox1 to *At*GA20ox5, exhibiting partially overlapping expression patterns. These five enzymes are

responsible for different developmental functions, although *At*GA20ox1, *At*GA20ox2 and *At*GA20ox3 seem to be of higher importance than *At*GA20ox4 and *At*GA20ox5[76,85]. *Arabidopsis ga20ox1* mutants possess as main phenotype a reduction in total stem height, whereas overexpression of GA20ox1, GA20ox2 and GA20ox3 causes a typical GA overdose phenotype including reduced seed dormancy, early flowering and elongated hypocotyls and internodes, enlarged leaves due to an increase in cell size, early flowering, longer anther filaments, more seeds per silique and elongated siliques[81,86]. GA20ox1 mainly influences cell elongating pathways, whereas GA20ox2 rather acts in flowering time and in determining silique length[74]. Another study analysed GA20ox antisense lines exhibiting darker and epinastic cotyledons, which could be rescued via exogenous GA_3 application[76].

Arabidopsis possesses four genes encoding for GA3ox enzymes[81] In contrast to the wild-type-like phenotype of *atga3ox2* mutant, *atga3ox1* exhibits a semidwarf phenotype. The double mutant shows a smaller leaf diameter, a shortened plant height and a more severe seed germination defect compared to the *atga3ox1* single mutant, which can be explained by partially redundant functions. The fact that, *At*GA3ox3 and *At*GA3ox4 are not upregulated in the *atga3ox1 atga3ox2* double mutant and also that they are specifically expressed in reproductive developmental stages (compared to *At*GA3ox1 and *At*GA3ox2 being expressed in vegetative stages), argues for distinct roles of all four *At*GA3ox enzymes[87].

The relative importance of the non-13-hydroxylated pathway versus the 13-hydroxylated pathway seems to be species-dependant. Although the GID1 receptor possesses higher binding affinity towards GA_4[88], it is not always the predominant bioactive form as it is the case in *Arabidopsis*. Here, GA_1 was only found in high concentrations in seeds and in siliques[89]. In rice, the 13-hydroxlated pathway seems to predominate as high levels of GA_1 can be found in vegetative tissues, whereas GA_4 exhibits a peak in anthers[90]. Obviously, shifts between the 13-hydroxylated and 13-nonhydroxylated pathway can happen in *Arabidopsis*, since feeding of *ga1-1* mutants with radioactive labeled GA_4 resulted amongst others in the production of GA_1[91]. Overall, the impact of both pathways and therefore the 13-oxidase remains unknown.

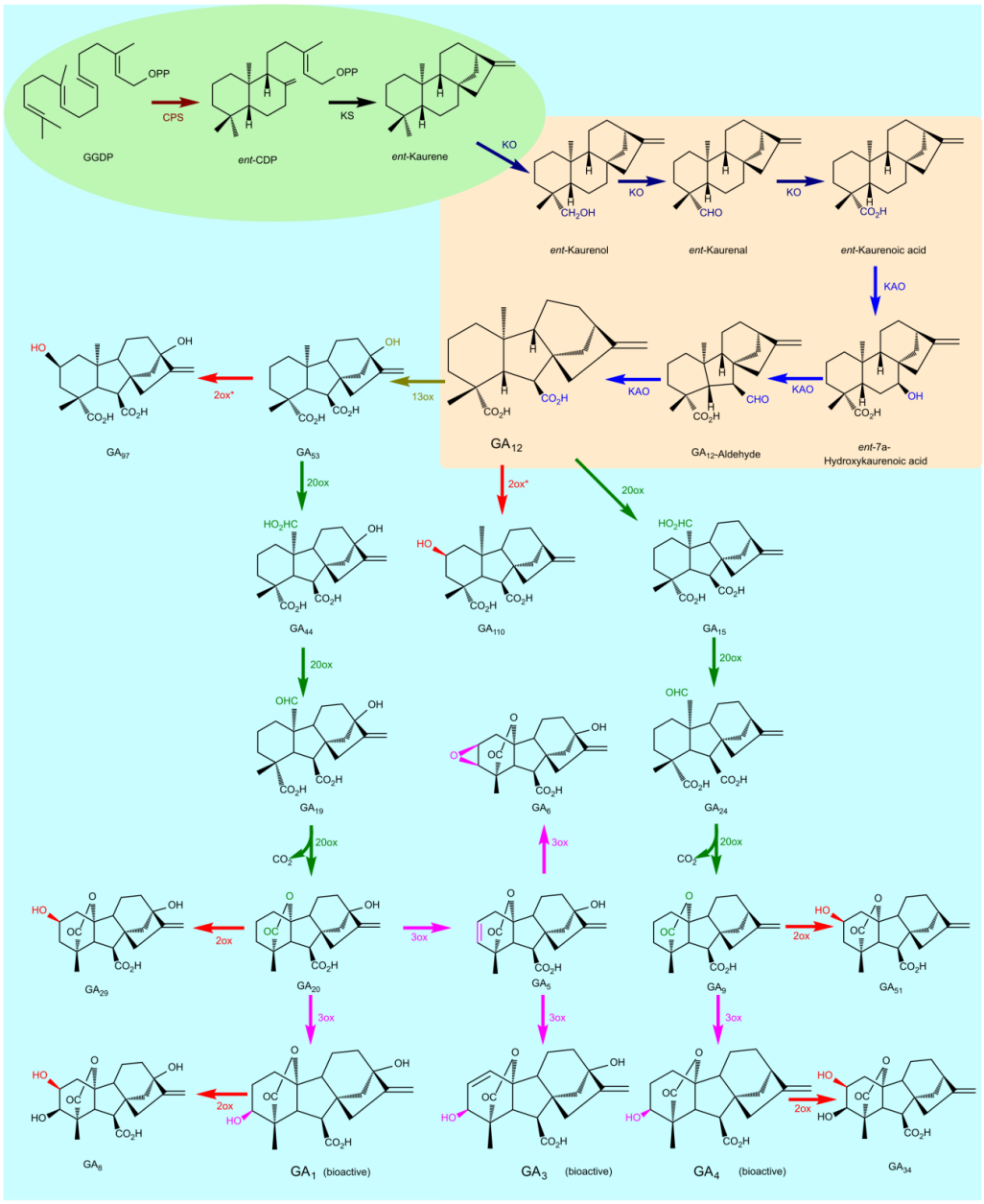


Fig. 5 Overview of gibberellins biosynthesis and deactivation on higher plants (modified after [4])

Three main steps of the GA biosynthesis are represented according to their subcellular compartments: the plastids, the endomembrane system and the cytoplasm. Enzyme groups and their modifications are highlighted in colour.

2.1.1.2 Gibberellin transport and signaling

Regarding the following steps of GA biosynthesis, the question arises, whether a de novo cell by cell GA production is necessary in every developmental stage and organ or if GA rather acts as a mobile signal. Application of endogenous GA_4 to one leaf of the GA-deficient *ga1-13* mutant

induces flowering[92] and therefore argues for a mobile GA form. In 2015, the NRT/PTR1 (NFP) family was identified to be responsible for transporting phytohormones such as auxin, ABA, JA and GA[93], with *At*NFP3.1 being confirmed to be unique for the GA transport in root endodermis cells localized to the plasma membrane[94]. Since *glucosinolate transporter 1* (*gtr1)* mutants exhibit impaired filament elongation and anther dehiscence that can be rescued by exogenous GA application, GTR1 was identified as a GA transporter[95].Thus, evidence for GA being transportable as a mobile signal has been found.

The key components of GA signaling are the nuclear-localized DELLA proteins, belonging to the plant-specific GRAS gene family of putative transcription factors. The *Arabidopsis* genome contains five DELLA genes *GIBBERELLIN-INSENSITIVE (GAI), REPRESSOR OF GA1-3 (RGA), RGA-LIKE (RGL) 1*,*2* and *3*, all of them possessing two conserved N-terminal domains, DELLA and VHYNP, which are responsible for the interaction between DELLA and the GA-receptor GID1 proteins[96–98]. At first, the key role of DELLA proteins in the GA signaling was demonstrated by the semi-dominant and GA-insensitive *gai-1* mutant, caused by a 51- bp in frame deletion within the DELLA motif leading to a GAI protein lacking 17 amino acids. Loss-of-function DELLA single mutants *rga* and *gai-t6* and the double mutant *rga-24 gai-t6* were able to partially and fully, respectively, rescue the GA-deficient dwarf *ga1-3* mutant phenotype, regarding amongst others leaf expansion, stem height and trichome initiation.[99]. Besides an increased cell proliferation, the quadruple DELLA-mutant *gai-t6 rga-t2 rgl1-1 rgl2-1* exhibits a prolonged duration of cell expansion, additionally arguing for the impact of GA signaling on cell size.

DELLA proteins interact with the DNA-binding domains of transcription factors or act as transcriptional co-repressors or co-activators and therefore inhibit downstream gene expression[100].

Synthesized bioactive GA is perceived by its soluble receptor protein GIBBERELLIN INSENSITIVE DWARF 1 (GID1). Three genes *AtGID1a, AtGID1b* and *AtGID1c* can be found in *Arabidopsis* and are expressed in all tissues and possess overlapping but also distinct roles, since single mutants do not show any phenotype. *Atgid1a atgid1b* and *atgid1a atgid1c* double mutants exhibit growth defects and developmental defects such as reduced stem elongation and lower male fertility, whereas the triple mutant is only able to germinate after removal of the seed

coat. Exogenous GA treatment does not rescue the severe dwarf phenotype that only allows the growth of a few millimeters in height within one month. Thus, the *atgid1a atgid1b atgid1c* triple mutant is GA-insensitive[101]. Crystal structure studies proved GID1 to undergo a conformational change allowing the formation of a GA-GID1-DELLA protein complex due to binding of bioactive GA, whereas no direct interaction between GA and DELLA was found[102]. Following this, the association to the E3 ubiquitin-ligase Skp1-Cullin-F-box (SCF) complex via the F-box protein SLEEPY1 (SLY1) allows the degradation of DELLA proteins via 26S proteasome and thus releases previously bound transcription factors to bind and activate promoter regions of GA responsive genes[102] (Fig. 6).

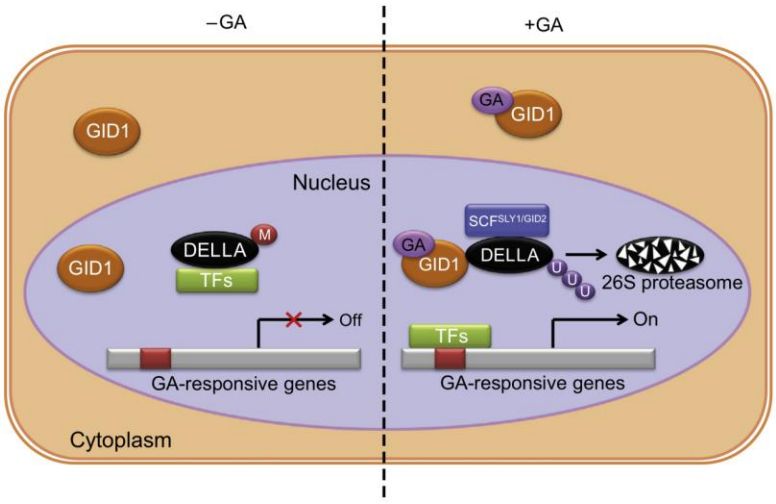


Fig. 6 The regulatory model for DELLA-mediated GA signalling (modified after [4])

DELLA proteins act as central repressors for GA signaling. In the absence of GA, DELLAs restrain growth via binding transcription factors into inactive protein complexes. The binding of GA by its receptor GID1 allows the interaction between GID1 and DELLA. The GA-GID1-DELLA complex formation initiates the binding to the F-box protein component of SCF/SLY (in *Arabidopsis*) and SCF/GID2 (in rice) and induces DELLA degradation via the 26S proteasome pathway due to polyubiquitination. The release of the previously bound transcription factors follows and allows the active expression of GA-responsive genes and thus is promoting growth. M: posttranslationally modified DELLA, U: ubiquitinated DELLA.

2.1.1.3 Inactivation and feedback regulation of GA pathway

A very strict and precise temporal and spatial regulation of the bioactive GA content is of highest importance for plants, as a rapid reaction to changing environments is essential. Thus, a complex system of GA deactivation and feedback regulation can be found for GA synthesis (Fig. 5). 2β-hydroxylations via GA 2-oxidases (GA2ox), another 2OGD, is the main mechanism for inactivating most of the GA intermediates, allowing a decrease of GA concentration at any step

of the biosynthesis[103]. The *Arabidopsis* genome possesses seven *At*GA2ox genes, whereby *At*GA2ox1 to *At*GA2ox6 were identified as C19GA2ox genes, responsible for hydroxylation of GA_1 and GA_4 as well as their precursors. *At*GA2ox7 and *At*GA20x8 act as inactivators of C20GA2ox and thus hydroxylate GA_{12} and GA_{53}. Therefore, C19GA2ox enzymes regulate the level of bioactive GA, whereas C20GA20x enzymes seem to inhibit earlier stages in the GA biosynthesis since ectopic overexpression leads to dwarf phenotypes due to decreased GA levels[74,104]. A recently discovered GA-inactivation mechanism is catalyzed by two cytochrome P450 monooxygenases belonging to the CYP714 family, *At*CYP714A1 and *At*CYP714A2. CYP714A1 overexpressing plants exhibit an extreme GA-deficient dwarfed phenotype found to be based on a role in inactivating non-13-hydroxy GAs, amongst others GA_{12} and maybe GA_4. CYP714A2 on the other hand uses KAOs as substrate for C13-hydroxylation, causing the production of weakly bioactive GA_1, since ectopic overexpression of CYP714A2 plants exhibit increased GA_1 levels compared to wild-type. Interestingly, the *cyp714A1 cyp714A2* double mutant has a strongly increased biomass due to increased cell size in the 30-50% enlarged cotyledons and rosettes, increased stem height, larger petals, bigger seeds as well as earlier flowering compared to wild-type[105]. Since CYP714A2 is additionally oxidating GA_{12}, both CYP714 enzymes compete for non-13-hydroxy substrates. Thus, they might be responsible for determining the ratio of GA_1 and GA_4[106], similar to CYP714D1 and CYP714Bs in rice[84].

Feedback regulation is an additional mechanism to maintain GA homeostasis in plants[81]. The main targets are the 2OGDs, whereas early GA biosynthesis genes, such as *CPS*, are rather not used for this mechanism[99]. For instance negative feedback regulation was shown for *At*GA20ox enzymes. Therefore, *atga20ox1*, *atga20ox2* and *atga20ox3* single mutants were found to increase expression of at least one of the other two family genes, to maintain the needed bioactive GA_4 content[74]. A similar feedback regulation was described for *At*GA3ox genes and a feedforward regulation shown for *At*GA2ox genes (Fig. 5), the first leading to downregulation, the latter to upregulation following an exogenous GA application[107]. These feedback loops are not only established within one single gene family, since *At*GA3ox genes have been shown to be upregulated in *atga20ox2* and *atga20ox1 atga20ox2* mutants, too[74]. Further, CYP714A1 and CYP714A2 overexpressing plants exhibit a strong upregulation of *AtGA20ox1* and *AtGA3ox1*. Since the bioactive GA_4 level is still decreased in these plants, this feedback regulation is still not sufficient to reach the wild-type status[106]. Since feedback regulation is based on a recognition

point to allow for a well regulated system, scientists went on to search for the responsible factor. Several studies have pointed at DELLA proteins being the switch that researchers were looking for. Loss-of-function mutants that are lacking important signaling components such as GID1[101] and GID2/SLY[108] were found to exhibit an increased expression of *AtGA3ox* and *AtGA20ox* as well as elevated levels of bioactive GA[96]. The main argument for DELLA proteins, conveying the feedback to GA biosynthesis and GA receptor genes[109], was given by several studies, describing that plants with non-degradable and therefore over-stabile DELLAs exhibit strongly increased GA levels[110]. A further study with loss-of-function DELLA mutants displaying decreased bioactive GA content[111], supports this theory. The complete mechanism of the DELLA mediated feedback regulation is still not resolved, although several components have already been identified[112,113].

2.1.1.4 Examples for DELLA regulated targets

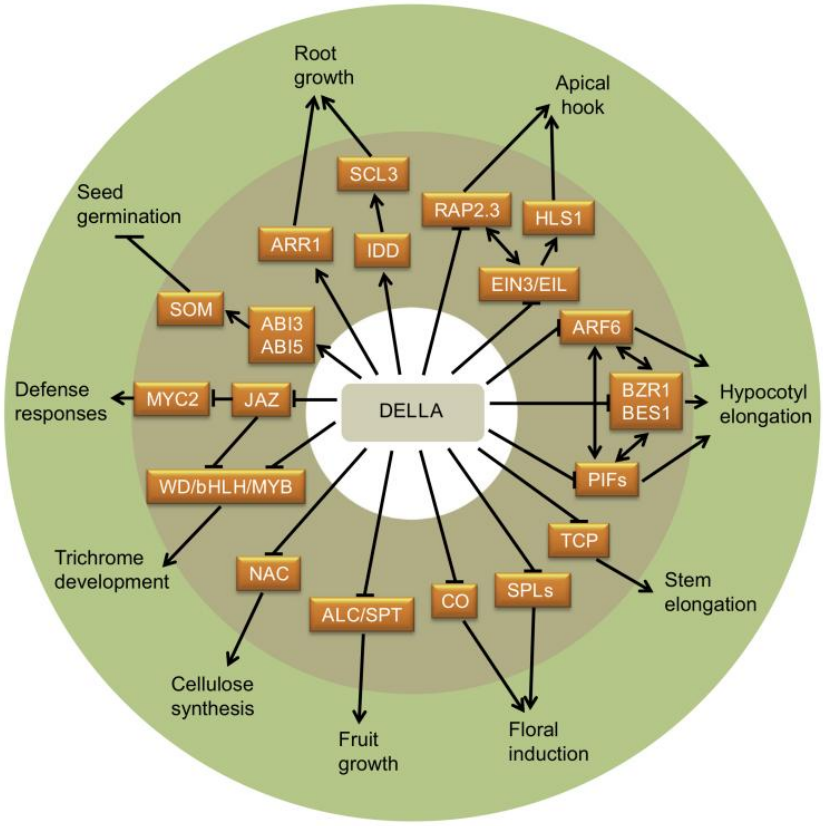


Fig. 7 DELLA proteins act as key regulators of plant growth and development (modified after [4])

GA signal transduction is mediated via interaction of DELLA proteins with multiple target proteins affecting a huge amount of different regulatory pathways.

As described above, DELLA proteins are proven to function not redundantly but rather separately yet still in an overlapping manner. They are responsible for regulating plant growth and development in a very complex manner influencing, for example seed development, leaf expansion, stem elongation, trichome initiation, flowering time and apical dominance[98]. Therefore, the main task of the DELLA protein is the integration of hormonal and environmental signals to adjust corresponding GA levels (Fig. 7)[103]. In the following, several examples for DELLA regulation mechanisms are given.

One major impact of GA on plant development is the transition from the vegetative phase to flowering in *Arabidopsis*. Here, RGA interacts directly with microRNA156 (miRNA156)-targeted SQUAMOSA PROMOTER BINDING-LIKE 3 and 9 (SPL3 and SPL9). These two transcription factors activate the expression of two MADS box transcription factors, *APETALA1* (*AP1*) and *FRUITFUL* (*FUL*), as well as miR172. This microRNA targets the mRNA of AP2-like transcription factors, such as SCHNARCHZAPFEN (SNZ) and SCHLAFMUTZE (SMZ), which are negative regulators of *FLOWERING LOCUS T* (*FT*) expression and promote flowering when downregulated. Thus, the transition to flowering is repressed by the RGA-SPL3-interaction via inactivating miR172 in leaves. Further, MADS box genes are inactivated at the shoot apices under short day conditions[114].

An interplay of the BR, Auxin and GA pathway has been identified to regulate hypocotyl cell elongation. DELLA proteins interact with the auxin response factor ARF6, forming a BZR1-ARF6-PIF4-DELLA module. This results in RGA-dependant blocking of protein-DNA and protein-protein-interactions of ARF6 and the bHLH transcription factor PHYTOCHROME INTERACTING FACTOR 4 (PIF4), that, in concert, causes shorter hypocotyls[115]. The PIF homolog SPATULA (SPT) is responsible for seed dormancy, fruit growth and the slowdown of cotyledon expansion. Since *spt-11* and *spt-12* mutants exhibit strongly widened cotyledon cells, SPT acts as growth repressor regulating mainly the same target genes as DELLA proteins RGA and GAI. Due to the fact that SPT is additionally inhibited by these DELLA proteins, a molecular circuit has arisen, balancing the final cotyledon cell size to prevent extreme overgrowth[116]. SPT together with bHLH proteins ALCATRAZ (ALC) and INDEHISCENT (IND) are also DELLA-inhibited targets of a further GA-dependent regulation regarding the valve margin development and fruit opening[117]. The interplay of two MADS-box proteins

SHATTERPROOF 1 and 2 (SHP1/2) acting upstream and in parallel with IND causes the cell specification of the valve margin layer. *Shp1* and *shp2* mutants were shown to rescue the low fertility of 35S::GA2ox pollen and are therefore proven to also influence pollen tube growth in a GA-dependent manner[118].

Further, an effect of DELLA proteins on the microtubule organization was identified via the interaction with the prefoldin complex (PFD), a cochaperone, which is required for correct tubulin folding. A GA-dependent DELLA degradation releases PFD and allows the formation of α/β-tubulin heterodimers in the cytoplasm, whereas the absence of bioactive GA leads to blocked PFD complex that stays captured in the nucleus resulting in severely affected microtubule organization[119]. Other regulatory pathways show combined control of GA-dependency as well as the cytoskeleton in regard of a strongly spatial and temporal regulation of cell expansion, which will be explained in the following chapter.

2.1.1.5 Trichomes as special case of cell size regulation via GA

Arabidopsis trichomes have been an advantageous study model for the genetic analysis of the strongly spatial regulation of cell expansion, since they form an unique cell shape[120] (Fig. 8A). Wild-type trichome development occurs through two growth phases. The first one is based on four endoreduplication cycles until the final ploidy level of 32C is reached. In parallel to each endocycle, the cell grows out from the leaf epidermal surface and a branching event occurs. The second phase is characterized by strong cell elongation via vacuolization. The exact temporal pattern of the four endoreduplications was reported controversially in different publications[5,7,120–122] and is amongst others regulated by different phytohormones. The cytoskeleton is an additional factor that varies between these two growth phases. Thus, studies described the dependency on tubulins during the first phase, changing into a dependency on actin during the second growth phase. Therefore, disruption of the microtubules via tubulin inhibitors led to inhibited outgrowth and mostly unbranched trichomes[123]. In contrast, the treatment with actin inhibitors resulted in distorted malformed trichomes exhibiting an abnormal organization of the actin cytoskeleton[124].

Tetraploid plants possess trichomes with additional branches compared to their diploid wild-type plants[122], which argues for an effect of endoreduplication on trichome morphogenesis. Glabrous

leaves with smaller and less branched trichomes were found in *glabra1* (GL1) and *glabra3* (GL3) mutants, two bHLH transcription factors, exhibiting a lower DNA content compared to wild-type. Together with the GL3 homolog ENHANCER OF GL3 (EGL3), these two factors cause the formation of a protein complex, required for trichome initiation. TRIPTYCHON (TRY), a MYB-type transcription factor, can bind to the N-terminus of GL3 and EGL3 and is therefore inactivating this protein complex[125]. Thus, *try* mutants displaying increased DNA content have larger trichomes with more branches[126]. However, phytohormones, especially GA and JA, activate trichome initiation via the described GL1/GL3/EGL3 complex (Fig. 8B). As recent studies described, this activation pathway is DELLA-dependent, since *gai* mutants barely exhibit any trichomes, except for a few particularly two-branched ones. Loss-of-function mutants of SPINDLY (SPY), which was identified to negatively regulate GA response via increasing the activity of RGA in wild-type plants[127], possess more and particularly four-branched trichomes. Recent studies identified GLABROUS INFLORESCENCE STEMS (GIS), a C2H2 transcription factor, due to loss-of-function (LOF) mutants that clearly exhibit a negative effect on trichome initiation. Genetic interaction studies and mRNA expression profiles concluded that GIS is acting downstream of GAI. Additionally, g*is spy* double mutants resemble *gis* single mutants, placing GIS in an epistatic role to SPY[6].

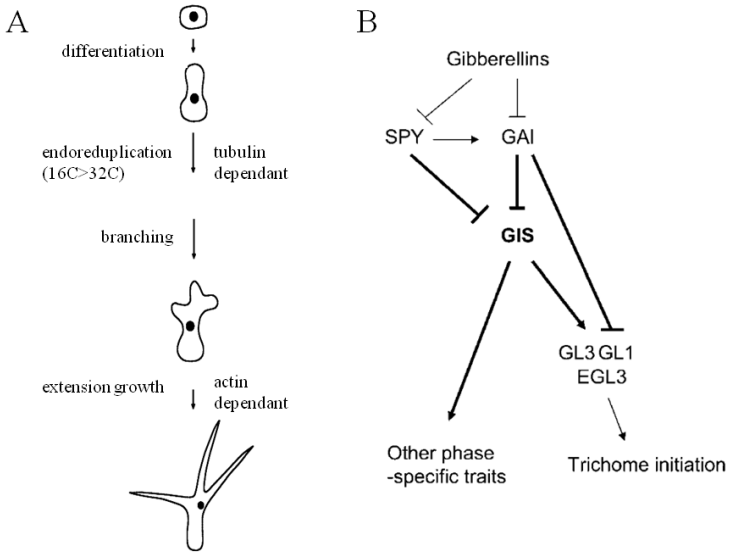


Fig. 8 Trichome morphogenesis in Arabidopsis thaliana (modified after [5–7])

(A) Drawings of the growth phases of wild-type trichomes, including different morphogenetic events and differently mediated via the cytoskeleton. **(B)** Overview of the regulation of trichome initiation by GIS

2.1.1.6 Impact of GA biosynthesis on flower development

As described in chapter 2.1.1.4, GA signaling displays a strong effect on flowering via SPL and FT regulation in a photoperiod-dependant way[114]. GA itself is responsible for the upregulation of FT under long-day conditions (LD) and therefore rescues the delayed flowering in the *ga1-3* mutant caused by FT downregulation, if exogenously applied. Under short-day conditions (SD), FT expression is less induced by GA treatment[128]. Further, GA treatment induces flowering of the non-flowering *ft* mutant under LD, pointing towards a GA-dependant but FT-independent pathway, which might be regulated via LEAFY (LFY) and SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1). Thus, GA increases LFY expression levels via SOC1, two factors that promote flowering, but can also be activated independently by GA[114,129] (Fig. 9A).

Besides its role in flower promotion, LFY induces early expression of interacting MADS box transcription factors according to the ABCE model, such as the B-class (stamen and petals) and C-class (carpel and stamen) genes *APETALA* (*AP3*) and *AGAMOUS* (*AG*) and is therefore responsible for normal floral organ development[130,131](Fig. 9B). In contrast to the vegetative stage, LFY expression is not GA-dependant in the flower meristem (FM) of *ga1-3* mutant, whereas AP3 and AG expression levels were shown to be upregulated under exogenous GA application[132]. The importance of GA for flower development was shown in expression studies of early floral tissues detecting GA3ox1 during sepal initiation (stage 3 of FM) and stays at least until stamen filament differentiation (stage 7)[110]. GA3ox1 expression is additionally used by AG to build a positive feedback loop, since maintenance of AG expression is required to ensure correct floral organ and particularly stamen development[133,134]. This GA maximum in stamen compared to the remaining floral organs, together with a clear correlation between stamen and petal size in *ga20ox ga20ox2* mutants, points towards stamens as the source for bioactive GA for petal size[8,74]. In general, flower development under SD seems to be more sensitive to GA_4 than to GA_1 as exogenous application to wild-type plants indicated[135], fitting to LFY being more responsive to GA_4[92]. However, a more highly concentrated GA_1 treatment leads to the same strong phenotypes regarding flowering as was detected with lower concentrated GA_4 applications[135], indicating that GA_1 can overtake GA_4 responses.

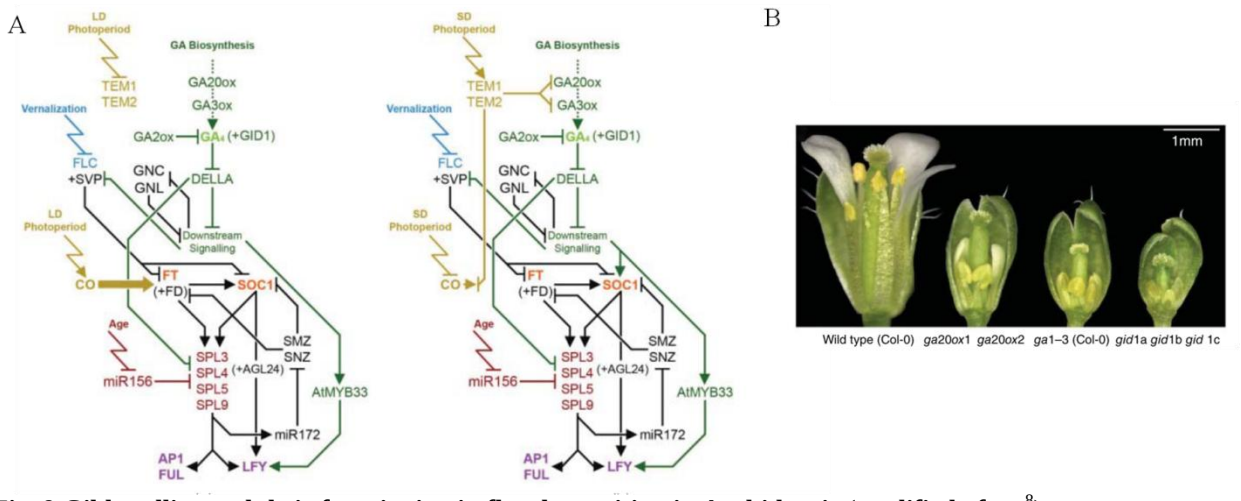


Fig. 9 Gibberellins and their functioning in floral transition in Arabidopsis (modified after [8])

(A) Representation of GA signalling (green) network linking photoperiod (yellow), age (red) and vernalisation pathway (blue) based on the expression levels of *FT* and *SOC1*, both promoting flowering by activating the expression of floral identity genes LFY, AP1 and FUL. Under LD conditions (left), FT is upregulated by CO-dependant regulation. Under SD-conditions (right), flowering is induced via GA-dependant upregulation of SOC1 and LFY, whereas FT is not upregulated. Additionally, GA regulates flowering via DELLA-mediated inhibition of SPL protein activity and negative regulation of environmentally-sensitive flowering inhibitors such as SVP. Arrowheads and bars indicate positive and negative regulation of downstream targets, respectively. **(B)** Comparison of floral phenotypes of selected GA biosynthesis and signalling mutants to wild-type. Note that growth of all floral organs is reduced, with petals and stamens showing greatest sensitivity. Anthers of *ga1-3* and *gid1a gid1b gid1c* mutants exhibit premature developmental arrest.

2.1.2 2-oxoglutarate /Fe(II)-dependant dioxygenases and their function in plant metabolism

The second largest enzyme superfamily of the plant genome is represented by 2-oxoglutarate/Fe(II)-dependant dioxygenases (2OGDs)[10]. These enzymes are responsible for a wide range of oxidative reactions, including hydroxylations, desaturations, demethylenation, ring closure, ring cleavage, epimerization, rearrangement, halogenations and demethylation. Not only in plants, 2OGDs are of highest importance as oxidizing catalysts[136], since this superfamily can equally be found in microorganisms, fungi and mammals[9,137]. 2OGDs are soluble non-heme containing proteins, localized in the cytosol. Their main function is catalyzing the oxidation of a substrate (R-H) that is accompanied by the decarboxylation of 2OG and thus leads to the formation of succinate and carbon dioxide (R-H+ 2OGD + O_2 = R-OH + succinate + CO_2). For this enzymatic catalysis, molecular oxygen (as co-substrate) as well as ferrous iron Fe(II) (as cofactor) are required[138] (Fig. 10B).

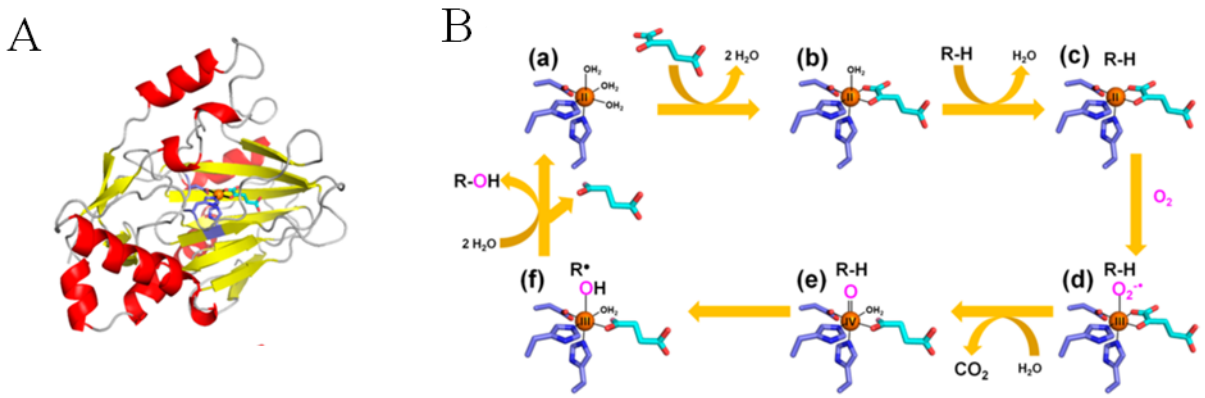


Fig. 10 Structure and mechanism of a representative 2OG-dependant oxygenase (modified after[9])

(A) Graphical structure TauD (PDB ID: 1GY9) with red α helices, yellow β strands, grey unstructured regions, Fe (orange sphere), metal ligands (sticks with blue carbons) and 2OG (sticks with cyan carbons). The major grouping of β strands forms a double-stranded β helix core. **(B)** Simplified hydroxylation mechanism

The overall sequence similarities of 2OGDs are rather low, but all share a carboxyl-terminal conserved region, the 2OG-FeII_Oxy motif [139], consisting of the following sequence motif, His-Xaa-Asp/Glu- (Xaa)n-His, that is responsible for the formation of the catalytic triad, a double stranded β-helix core, that binds Fe(II)[140] (Fig. 10A). Kawai et al. designed a phylogenetic tree of all predicted 2OGDs in 6 plant species, including *Arabidopsis thaliana* with more than 130 identified 2OGDs, representing 0.5% of the total genome[10]. Based on their protein sequence similarity, they were classified into three classes: DOXA, DOXB and DOXC. Interestingly, this phylogenetic classification corresponds to functional differences. DOXA consists of plant homologs of the *Escherichia coli* AlkB[141], the representative member of DNA-repair via oxidative demethylation of alkylated nucleic acids and histones. One *Arabidopsis* representative is ALKBH2, exhibiting in vitro repair activities and complementation abilities of repair deficiencies of the *E. coli alkB* mutant[142].

DOXB is a plant-specific class, including 2OGDs acting in cell wall protein synthesis via proline 4-hydroxylation. Cell wall proteins, such as extensins, proline-rich proteins and arabinogalactan proteins, exhibit proline residues that are posttranslationally modified by prolyl 4-hydroxylases, leading to the formation of glycoproteins. As an example, P4H5 was shown to be responsible for root hair elongation in *Arabidopsis* via peptidyl-proline hydroxylation on extensins to ensure the correct cell wall self-assembly[143].

The most varied 2OGD class is DOXC, playing important roles in specialized metabolism of many phytochemicals, such as phytohormones and flavonoids. The importance of 2OGDs in the GA pathway has already been explained in chapter 2.1.1. 2OGDs were also identified to act in the pathways of ethylene, auxin, salicylic acid and alkaloid biosynthesis and metabolism[136]. DOX class-specific main catalytic mechanisms are shown in Figure 11.

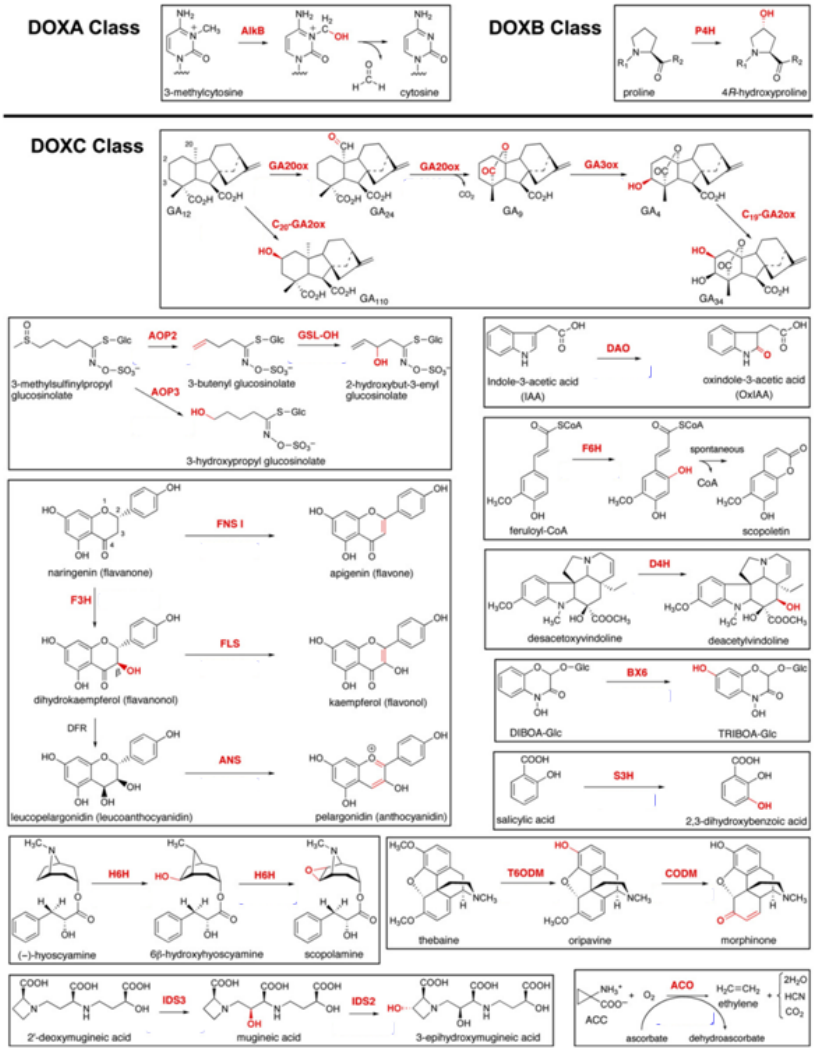


Fig. 11 Experimentally defined functions for plant 2OGDs (modified after [10])

2.1.2.1 INCURVATA11 as a special case of 2OGDs

INCURVATA11 (ICU11) was shown to play a special role within the 2OGD family. Although the catalytic triad as well as a prolyl 4-hydroxylase domain (xaa-Pro-Gly) is conserved in the ICU11 protein, Kawai et al. were not able to place it in any of his presented DOX classes of 2OGDs. Until 2018, nothing was known about the function of ICU11. Finally, Mateo-Bonmatí et al. published the first information about it, when they identified a mutant exhibiting curved leaves and early flowering. The causal locus was mapped to At1g22950 and was called INCURVATA11 due to the leaf phenotype of the mutant (Fig. 13A). A C→T insertion in exon 1 led to a premature Stop codon, and was found to be a loss-of-function mutation, accordingly. In Ws-2 background, a second mutant with a T-DNA insertion in intron 2, exhibited a similar phenotype, confirming that At1g22950 was the causal gene. Interestingly, ICU11 appeared to be very highly conserved in metazoa including humans with its paralogue OGFOD2 (Fig. 12).

ICU11 is the founding member of a small gene family that was named CUPULIFORMIS (CP) (Fig. 13B). All members of the CP family (CP2 (At3g18210), CP3 (At5g43660), CP4 (At1g48740), and CP5 (At1g48700)) possess the P4Hc domain, whereas none of them were integrated into the 2OGDs analysis by Kawai[10].

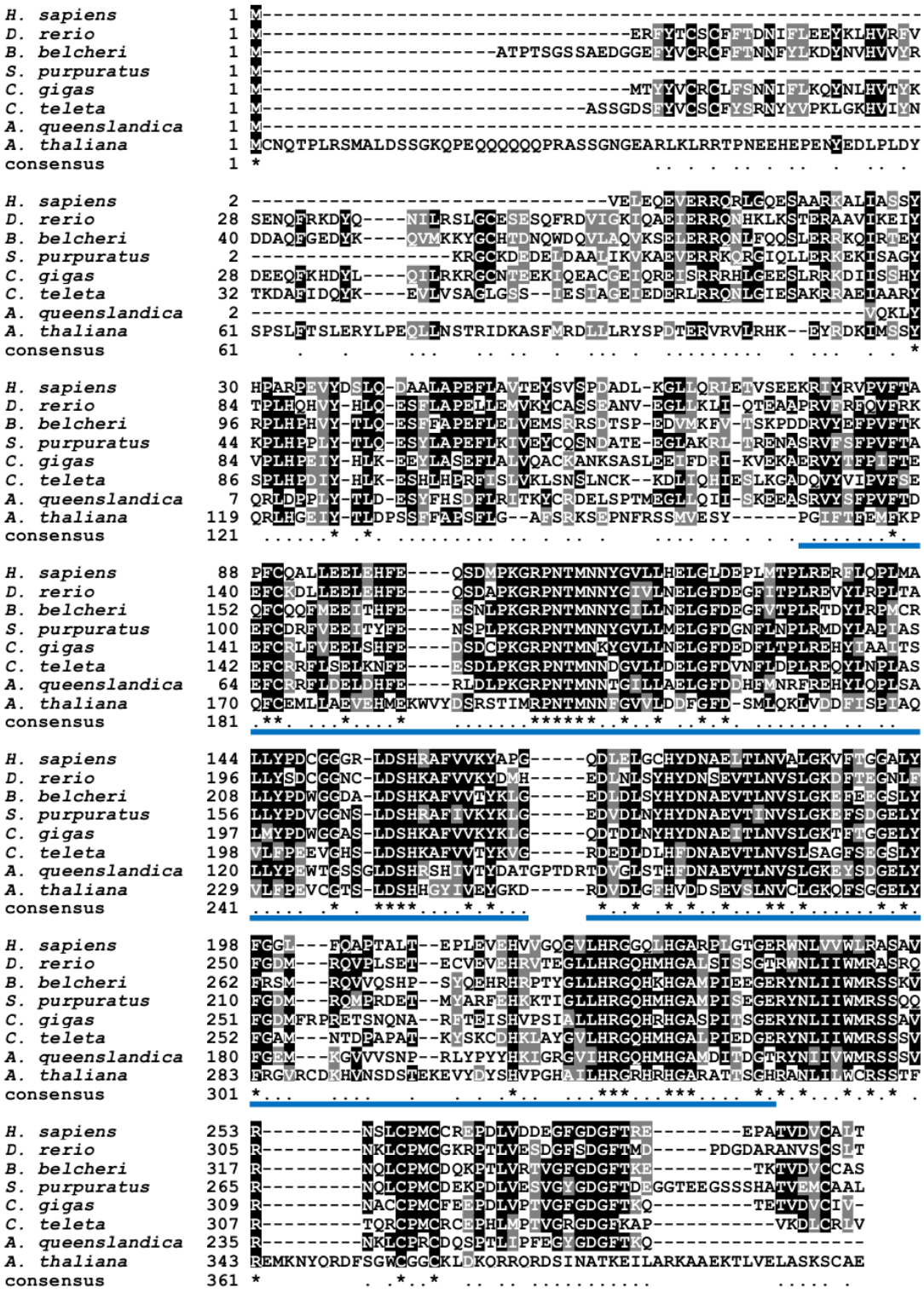


Fig. 12 Alignment of the amino acid sequences of ICU11 orthologs in Metazoa (modified after [11])

Identical and similar residues are shaded in black and gray, respectively. Numbers indicate residue positions. A blue bar under the consensus line indicates the putative prolyl 4-hydroxylase domain of ICU11. Protein sequences were retrieved from NCBI.

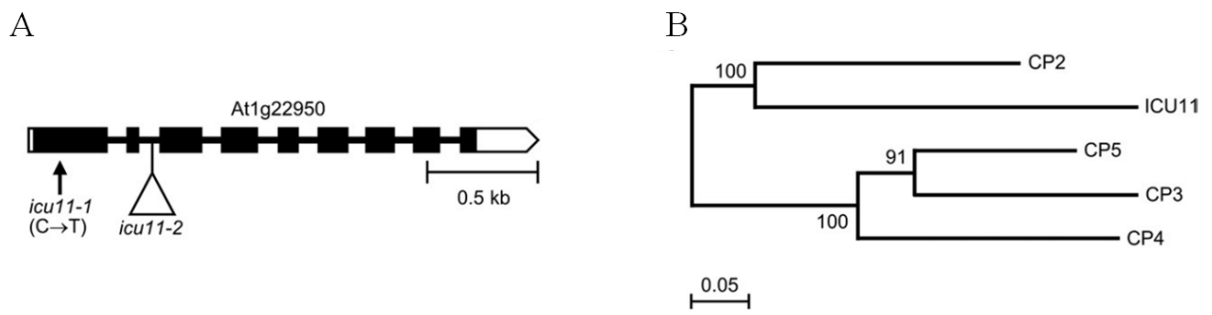


Fig. 13 Genetic structure and phylogenetic analysis of ICU11 and CP protein family (modified after [11])

(A) *ICU11* gene structure including positions of *icu11* mutations. Boxes and lines display exons and introns, respectively. Black boxes represent translated regions. White boxes untranslated regions. Triangle indicates T-DNA insertion and the vertical arrow the SNP for *icu11-2* and *icu11-1*, respectively. **(B)** Phylogenetic analysis of CP family proteins. Scale bar represents 5% amino acid sequence changes

Although CP2 is the closest relative of ICU11 both do not act in a redundant manner, since *cp2* single mutant exhibit a wild-type-like phenotype, whereas *icu11-1* showed epinastic cotyledons and hyponastic first and second true leaf pairs, due to decreased cell size in the palisade mesophyll cells. It also exhibited premature flowering and decreased fertilized ovules compared to its wild-type S96. However, the double mutant *icu11 cp2* skips the vegetative growth and starts flowering immediately after germination, similar to the phenotypes of the loss-of-function mutants of Polycomb-group genes (PcG) EMBRYONIC FLOWER 1 and 2 (EMF1/2)[11,144].

PcG genes are major regulators of the chromatin state via gene repression and are therefore responsible for the establishment and maintenance of cell identity [130]. PcG proteins act in two main types of complexes POLYCOMB-REPRESSIVE COMPLEX 1 (PRC) 1 and 2 in metazoans, and in plants. They are particularly responsible for shutting off gene expression of important developmental regulators to prevent their presence at the wrong stage or tissue during the entire life cycle. The catalytic core of *Arabidopsis thaliana* PRC1 consists of LIKE HETEROCHROMATIN PROTEIN1/TERMINAL FLOWER2 (LHP1/TFL2), EMBRYONIC FLOWER1 (EMF1) and the RING domain proteins. AtPCR2 is formed by CURLY LEAF (CLF), SWINGER (SWN) and MEDEA (MEA). Different biochemical reactions maintain strictly controlled dynamics regarding the repressed state of the chromatin to ensure the correct expression patterns allowing all different phase transitions from gametophyte to sporophyte, embryo to vegetative and vegetative to reproductive state. Here, PRC1 and PRC2 act together but also independently and they can recruit each other if required. PRC2 proteins trimethylate

lysine 27 of histone 3 (H3K27me3), whereas PRC1 recognizes this epigenetic mark and is responsible for maintenance of this repressive state. Additionally, PRC1 was shown to monoubiquitilate histone H2A (H2AK119ub/H2AK121ub) via its ligase activity[145].

Additionally to the *emf1/2*-like phenotype of *icu11 cp2* double mutant, further double mutants of *icu11* and components of the epigenetic silencing machinery exhibit similar or even more severe phenotypes[11]. Further, the RNA-Seq of *icu11-1* exhibited hundreds of misexpressed genes, including *AP1*, *FUL*, *SNZ*, *FT*, *SPT*, *SPL3*, *ALC* and *SHP1*&*2*. Furthermore, some of them possessed a decrease in H3K27me3 levels in their promoter regions. Thus, ICU11 was supposed to be the founding member of a gene family acting via an unknown mechanism as epigenetic repressor[11].

Although the Mateo-Bonmatí publication seemed to enlighten the molecular function of ICU11, it was decided to pursue this project, for a deeper analysis of the phenotypic differences induced by the varying mutant *icu11* alleles and to follow up on the results contradictory to the published findings.

2.1.3 Preliminary work

The starting point of this project was an EMS screen done on the line *paps1-4*, a mutant of the *Arabidopsis* POLY(A)-POLYMERASE 1 (PAPS1) which the Lenhard lab worked on. The background *paps1-4* line showed a slight increase in petal size. In order to find POLY(A)-POLYMERASE 1 (PAPS1)-interacting genes, the EMS screen aimed to find mutants exhibiting an enhanced phenotype. The most promising double mutant that was found possessed strongly enlarged petals. Further analysis uncovered that the single *enhancer of paps1 (eop1)* mutant showed larger petals on its own, leading to the assumption that the phenotype of the double mutant could be explained by an additive phenotype instead of a genetic interaction between PAPS1 and the causal gene of *eop1*. At this point, *eop1* started to be a project on its own. The following chapters describe the characteristics of the *eop1* mutant and different approaches followed up on to reveal the molecular basis of the *eop1* phenotype.

2.2. Materials and Methods

2.2.1 Biological materials and growth conditions

The *Arabidopsis thaliana* (L.) Heynh. ecotype Columbia-0 (Col-0) was used in all experiments. The *eop1* (*icu11-3*) mutation was identified in an EMS-mutagenesis screen in the mutant background of *paps1-4*. To remove the *paps1-4* mutation and enable comparison between *eop1* single mutants and Col-0 the double mutant was backcrossed three times to Col-0. Seeds of the *icu11-2* (FLAG_402G04) and *icu11-4* (FLAG_207D09) T-DNA insertion lines as well as the Wassilewskija-2 (Ws-2) line were provided by the Versailles Arabidopsis Stock Center[147]. Unless otherwise stated, all lines studied in this work were homozygous for the indicated mutations. T-DNA insertion sites were confirmed using the combination of a gene-specific (Suppl. 6.2) and the LB4 primer for PCR- based genotyping.
The tetraploid Col-0 line, used as FACS control, was provided by Duarte D. Figueiredo.

Seeds were surface-sterilized with sodium hypochloride based on Iii[148] with a modified incubation time of not more than 5 min and a washing step with 70% EtOH: The dried seeds were either put on half-strength Murashige-Skoog (MS) medium (Duchefa Biochemie, Haarlem, Netherlands) or directly on soil. A vernalization for at least 2 days at 4°C was carried out before the seeds were put to normal long day growth conditions, such as cycles of 16 h light and 22°C followed by cycles of 8 h at 16°C, a humidity of 70% as well as a light level of 150 µmol $m^{-2}s^{-1}$.

Two electro competent bacterial strains were used for this thesis: *Escherichia coli* XL1-blue (Agilent, Santa Clara, California, USA) and *Agrobacterium tumefaciens* GV3101 strain (Gold Biotechnology, St Louis, Missouri, USA). Both were grown in Luria Broth (LB medium) under 37°C and 28°C, respectively. Overnight cultures were shaken at 220 rpm using an INFORS HT Ecotron incubator (Bottmingen, Switzerland).

2.2.2 Phenotyping

Organ size measurements

Dissected organs, such as 5^{th} and 6^{th} leaves close to senescence, as well as petals, styles, stamen and sepals from the 10^{th} -12^{th} fully opened flowers of the main stem were flattened and scanned

at a resolution of 3200 dpi (HP ScanJet 4370). Organ area analysis was carried out using ImageJ (http://rsbweb.nih.gov/ij/). Images of seed size and seeds per siliques (close to senescence) were taken using the microscope Olympus SZX12 (Shinjuku, Tokio, Japan) and the Olympus BX51 microscope using an AxioCam ICc3 camera (Zeiss).

Nail polish imprints

Imprints of adaxial pavement cells of cotyledons 14 days after germination were carried out according to a modified nail polish based protocol[149]. Cotyledons were carefully laid on smeared dental impression material (President Plus Light Body (4626), Colténe) until the material was completely solidified. After peeling off the cotyledon, the cell impression was filled with a thin layer of clear nail polish. The dried nail polish layer was observed and pictured under a light microscope, Olympus BX51. The digitalized pictures were photo merged using Adobe Photoshop (https://www.adobe.com/de/products/photoshop.html) and cell sizes and numbers were analyzed using ImageJ.

Low-melt agarose gel imprints

Gel imprints of adaxial sides of petals were carried out in an adapted protocol from Horiguchi[150]. A drop of 2% low-melt agarose containing 0.01% bromophenol blue was pre-warmed at 50°C and placed on a pre-warmed glass slide. The plant material was immediately put on the smeared droplet. As soon as the gel solidified, the petal was carefully removed. After the complete drying of the agarose layer for 10 min, the imprint could be pictured under a light microscope, Olympus BX51. In the following, the digitalized pictures were photo merged using Adobe Photoshop (https://www.adobe.com/de/products/photoshop.html) and analyzed using ImageJ.

Flow cytometry

Flow cytometry samples were analyzed at the Max-Planck-Institute for Molecular Plant Physiology (Potsdam) according to the established Galbraith's protocol[151]. Three biological replicates of fresh rosette leaves per genotype were harvested, wrapped in wet tissue and chopped while immersed in 1 ml homogenization buffer (supplemented with 10 mM dithiothreitol (DTT), 50 mg/ml RNase and 50 mg/ml propidium iodide (PI)). The suspension was thoroughly mixed, filtered through a 20 mm filter (CellTrics Partec, Germany) and kept in dark for 10 min. The ploidy level of the nuclei was calculated based on their DNA content using a

FACS Aria II flow cytometer (BD Bioscience, California, USA). 3 biological replicates of tetraploid and diploid Col-0 individuals were used as control lines, acting as references regarding their PI fluorescence. PI fluorescence was excited at 488 nm using a blue laser and recorded using a 585 nm/42 band-pass filter as well as a 556 nm long-pass dichroic mirror. On average 2800 nuclei (2013–3400) were recorded per plant to reach an unequivocal signal of its ploidy level.

Confocal laser microscopy

To analyze cellular expression pattern of ICU11, confocal laser scanning microscopy images were obtained and processed using the operator software ZEISS Zen lite for the ZEISS LSM 880 confocal microscope (ZEISS, Oberkochen, Germany). GFP was excited at 488 nm with an argon ion laser and its emission analyzed between 499 and 544 nm. To differentiate between nuclear and nonnuclear localization of ICU11-GFP, root tips as well as lateral roots of 10 days old seedlings were stained with a 10 µg ml^{-1} solution of 4',6'-Diamidin-2-phenylindol (DAPI) (Carl Roth, Karlsruhe, Germany) and excited for 405-466 nm with a violet laser.

2.2.3 Chemical treatments

Oryzalin

Seeds of *eop1* and Col-0 genotypes were germinated and grown for 10 d on oryzalin (Sigma-Aldrich, St. Louis, Missouri, USA) containing 0.5x MS-media (0.1 µM, 0.2 µM, 0.4 µM, 0.6 µM, 0.8 µM, 1 µM and 1.2 µM). Plates were scanned at a resolution of 1200 dpi (HP ScanJet 4370). Root length measurements were carried out using ImageJ.

Propiconazole

Seeds of both genotypes were sown out on 0.5x- MS media, half of each supplemented with 1 µM propiconazole (Pcz) (dissolved in 100% EtOH) (Sigma-Aldrich, St. Louis, Missouri, USA). 3 h light treatment was followed by 5 days of etiolated growth in the dark but otherwise normal growth conditions. Plates were scanned at a resolution of 1200 dpi (HP ScanJet 4370). Hypocotyl lengths were measured using ImageJ.

Paclobutrazol

Seeds of both genotypes were grown on media containing different concentrations of Paclobutrazol (dissolved in ddH_2O) (Duchefa Biochemie, Haarlem, Netherlands):0, 0.1 µM, 0.2 µM, 0.3 µM, 0.4 µM, 0.5 µM, 1 µM, 2.5 µM, 5 µM, 10 µM, 20 µM, 25 µM and their germination rate was calculated[152].

2.2.4 Genetic analysis

DNA-extraction

Besides few modifications, DNA was extracted from fresh leaf material according to the previously described protocol[153]. The plant material was ground with a mill together with 380 µl TNE extraction buffer (200 mM Tris, pH 8.0, 250 mM NaCl, 25 mM EDTA), for 4 min. This step was followed by a quick spin and the adding of 20 µl 10% Sodium Dodecyl Sulfate (SDS). The mixed samples were centrifuged for 5 min at 3000xg at RT. Afterwards, 150 µl of 5 M potassium Acetate (pH 5.2) were added and the final steps of the protocol were followed.

Genotyping

All primers used for genotyping all *icu11* alleles used in this study, are presented in Supplement 6.2. For marker design, CAPS[154] and dCAPS[155] were used that are based on natural or induced restriction-endonuclease-sensitive polymorphisms. PCRs were accomplished according to manufacturer's instructions of MyTaq DNA Polymerase (Bioline by Meridian Bioscience (Cincinnati, Ohio, USA)) protocol. Necessary restriction digests were carried out according to manufacturer's instructions of New England Biolabs Inc. (NEB, Ipswich, Massachusetts, USA) using marker specific enzymes.

RNA extraction and splicing analysis

Three biological replicates of total RNA of 14 days old seedlings of both genotypes was extracted with a TRIzol-based protocol[156]. The harvested material was immediately frozen in liquid nitrogen and homogenized manually. 800 µl TRIzol (Life Technologies, Carlsbad, California, USA) were directly added and each sample was inverted until the powder was completely suspended. 5 min incubation at RT and further 5 min on ice were followed by another inverting step. Afterwards, 200 µl chloroform was pipetted to the solution, followed by vortexing and a 5 min on ice incubation. All samples were then centrifuged for 10 min at

13,000 rpm at 4°C and the around 600 µl of the aqueous phase were transferred into a new microcentrifuge tube. 700 µl of isopropanol were added and the solution was inverted to precipitate the RNA for 20 min at RT. The RNA was pelleted via a 10 min centrifugation at 13,000 rpm at 4°C. After removing the supernatant, the pellets were washed once with 75% EtOH, that was followed by a 5min spin at 13,000 rpm at 4°C. In the following the pellets were dried completely and dissolved in 30 µl H_2O. Finally, the dissolved RNA was incubated for 10 min at 60°C improving solubility and removing the EtOH leftovers.

The genomic DNA of each RNA sample was removed by using TURBO^{TM} DNase (Invitrogen by Thermo Fisher Scientific, Waltham, Massachusetts) according to the manufacturer's instructions. Reverse transcription followed using oligo(dT) and the Superscript III Reverse Transcriptase (Invitrogen). cDNAs of exon7-exon8 region of wild-type and mutant *ICU11* transcripts were amplified using LBO77 & LBO79.

Gene expression

To quantify expression levels of *ICU11* (used oligos LBO108 & LBO110, with exception of the *icu11-2* line: LOB305 & LBO306 (published oligos[11]) and *GA20ox1* and *GA20ox2* in wild-type lines as well as in all used *icu11* lines, total RNA of three biological and three technical replicates each was extracted and the genomic DNase digested samples were reverse transcribed (described in chapter 2.4). Quantitative-RT-PCR was performed using the LightCycler® 480 (Roche, Basel, Switzerland) according to manufacturer's instructions of the SensiMix SYBR Low-ROX kit (Bioline). Used primers are listed in Supplement 6.2.

Molecular cloning and plant transformation

Overexpression (p35S::*ICU11::YFP)* and rescue (p*ICU11*::g*ICU11*) constructs as well as the transgene lacking the catalytic function (*pICU11::*g*ICU11* HDH > SAS) were amplified in PCR reactions from genomic DNA of Col-0 wild-type seedlings. The genomic region 917 bp upstream of the transcription starting site was defined as promoter of *ICU11*. Nucleotide exchanges were introduced via mismatch primers following a published site-directed mutagenesis protocol[157]. All used primers and vectors are included in Supplement 6.2-6.3. The p35S::GFP::term was provided from Adrien Sicard as a derivate of pBluescript II KS (StrataGene, pBlueMLAPUCAP), that was together with pJET from the CloneJET PCR CLONING Kit (Thermo Fisher Scientific, Waltham, Massachusetts, USA) taken for all

constructs for subcloning using the CloneAmp™ HiFi PCR Premix (Takara, Saint-German-en-Laye, France). For Sanger sequencing and all cloning steps, the constructs were either gel extracted or cleaned after PCR using the NucleoSpin Gel and PCR Clean-up kit (MACHEREY-NAGEL). Sequencing was performed at LGC Genomics, Berlin, Germany).

pGPTV-HPT[158] served as final plant transformation vector so that the validated fragments were transformed into its AscI site via electroporation using MicroPulser Electroporator #1652100 (BioRad, Hercules, California, USA).

MIGS constructs were amplified according to the publication of de Felippes[159] using the previously described p35S::GFP::term as backbone.

Target sites for gRNAs of CRISPR/Cas9 constructs were found using CHOPCHOP (http://chopchop.cbu.uib.no/). pCBC-DT1T2 (addgene, https://www.addgene.org/) was used for PCR-amplification of all possible combinations of four different gRNAs, followed by Golden Gate cloning of all fragments into the final expression vector pBEEH104E (addgene).

All final vectors were transformed into *Agrobacterium tumefaciens* GV3101 strain (Gold Biotechnology, St Louis, Missouri, USA) and floral dipped into the corresponding *A. thaliana* lines according to the published protocol from Clough et al.[160].

Statistical analysis

Statistical analyses were performed using RStudio[161] or Microsoft Excel 7. The multiple comparisons of phenotypic means were calculated by using Tukey's HSD post hoc test using the agricolae package implemented in R. Two-sample comparisons were carried out using a two tailed Student's t-test, if populations were normally distributed; if not Mann-Whitney-U-test was performed. Significances of populations of a sample size below or equal to five were calculated using Kolmogorov-Smirnow-test. Experimental measures and resulting statistical tests were independent of each other and thus did not require adjustment for multiple testing. The null hypothesis was rejected at $p \leq 0.05$. Sample distributions are presented by box plots using ggplot2 package implemented in RStudio. Here, the middle line represents the median, the upper and lower box border correspond to the 75th and 25th percentile, respectively. Whiskers stretch until the maximum and the minimum values comprised with 1.5 interquartile ranges. Outliars are demonstrated by single dots outside the whiskers. Grey dots surrounding the boxplots represents sample size and sample values.

2.3 Results

2.3.1 Eop1 exhibited larger petals due to increased cell size

The EMS-mutant *eop1* exhibited strongly enlarged petals, with an average increase of 27% (p = 1,9 x 10^{-14}) of petal size compared to Columbia (Col-0) (Fig. 14). Surprisingly, also the heterozygous plants showed an enlargement of this trait. Although the difference in petal size differed only by 9% in comparison to Col-0, it was already highly significant with a p-value of 1.8 x10^{-6}. Hence, the *eop1* mutation acts as a semi-dominant mutation.

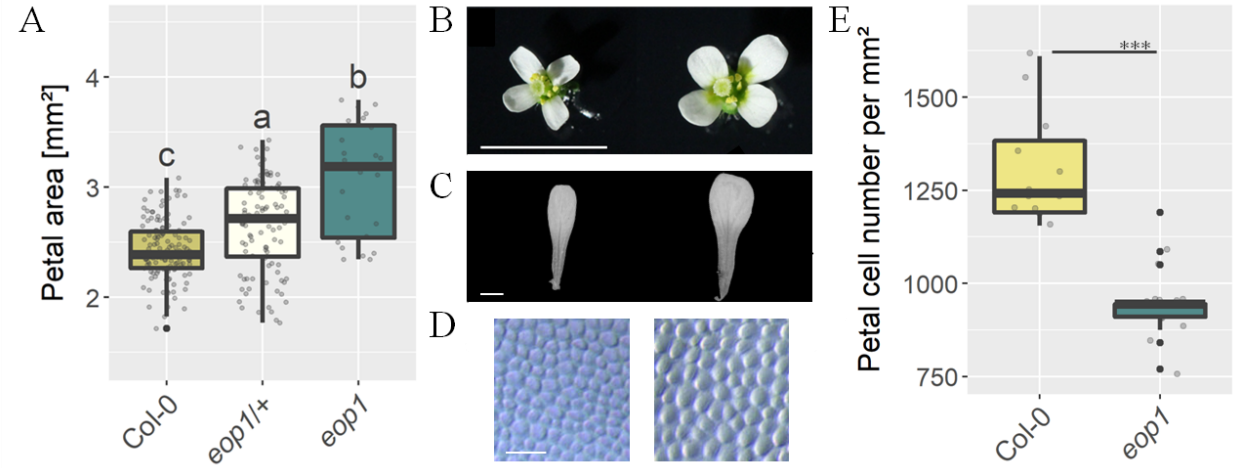


Fig. 14 Comparison of *eop1* mutant petals against wild-type

(A) Quantification of petal size of wild-type, as well as heterozygous and homozygous *eop1* lines; Letters indicate significant difference as determined by a Tukey's HSD test. **(B-D)** Images of wild-type (left) and *eop1* flowers (right), petals and adaxial petal imprints; scale bars: 5mm (A), 1mm (B), 50 µm (C). **(E)** Quantification of petal cell number per mm^2

Referring to the cellular basis of the petal size increase in *eop1*, petal imprints were prepared and analyzed. Figure 14D-E shows that *eop1* mutants clearly exhibit a significantly decreased cell number within the same area compared to Col-0 (p = 1.5 x 10-8, Mann-Whitney-U test), meaning that the *eop1* mutation promoted an increase in cell size of about 27%. Accordingly, the petal size increase of 27% (Fig. 14) can completely be explained by an increase in cell size in the petals of mutant plants.

2.3.2 *Eop1* showed additional phenotypes

2.3.2.1 Organ size increases

During the work with the *eop1* lines additional organ size phenotypes have been noticed and analyzed aiming to find the pathway which EOP1 is acting in.

One of the most prominent phenotypes of *eop1* is the size difference in cotyledons. Thus, wild-type and *eop1* adaxial epidermal cells of cotyledons were imprinted in nail polish, allowing an overview regarding cell size differences. *Eop1* mutant plants exhibited an enlargement of 31% of total cotyledon size compared to Col-0 ($p = 3.9 \times 10^{-6}$, Mann-Whitney-U test). Since the cell number per same area was significantly decreased in *eop1* mutants (Fig. 15A-D), it was concluded that the observed phenotype is based on a difference in cotyledon cell size. This cell size enlargement was demonstrated to be responsible for 29% ($p = 7.6 \times 10^{-6}$, Mann Whitney-U test) of the total cotyledon increase, supporting the assumption that the *eop1* mutation exclusively led to an increase in cell size and rather not cell number. Next, floral organ sizes were measured. Carpel and sepal areas exhibited a significant difference between *eop1* and Col-0 (Fig. 15F-H), whereas no significant difference was found by comparing the area of the stamen ($p = 0.95$). The increase of carpel size is approximately 36% ($p = 2.5 \times 10^{-7}$) and the increase is 12% for sepals ($p = 9{,}4 \times 10^{-7}$, Mann-Whitney-U test) compared to Col-0.

Additionally, also the stem diameter, the leaf area of fully mature leaves 5 and 6, as well as the seed size exhibited significant size increases in *eop1* mutants compared to Col-0 with p-values of 6.6×10^{-16}, 5×10^{-4}(Mann-Whitney-U test), and 2×10^{-16} (t-test) respectively (Fig. 15E). Moreover, the stem heights were measured periodically until plants finished flowering (Fig. 15K-D). Over the whole stem flowering period, *eop1* mutants exhibited a significantly elongated stem (lowest $p = 0.00783$, Mann-Whitney-U test), which correspond to earlier flowering phenotype.

To sum up, the *eop1* mutation leads not only to enlarged petals but also to an increase in cotyledon area (due to larger cotyledon cells), carpel and sepal area, as well as stem diameter and height. Therefore, the effect of the *eop1* mutation was found to be more general and not petal-specific as originally assumed.

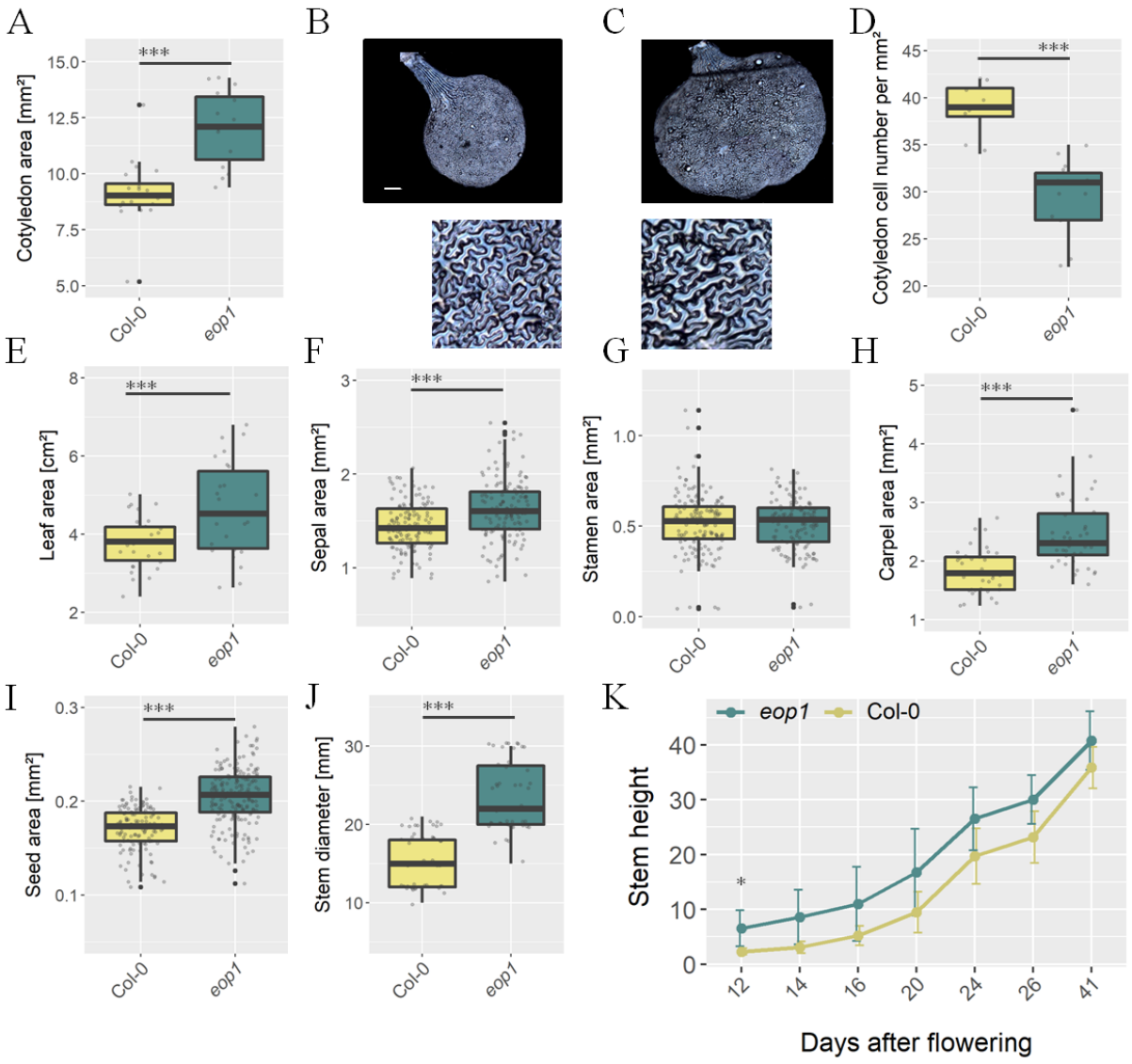


Fig. 15 Additional growth phenotypes of *eop1* mutant

(A-D) Comparison of cotyledon size between Col-0 (left) and *eop1* (right) **(B)** Imprint of adaxial wild-type cotyledon **(C)** and of *eop1*; scale bar: 50 µm, cut-outs represent 1mm^2 **(E-J)** Quantification of further growth phenotypes of *eop1* in comparison to Col-0 **(K)** Stem growth curves of both genotypes; note the non equidistanced scale of the x-axis. Asterisks represent significance ($p > 0.001$) determined by t-test or Mann-Whitney-U-test as indicated in the text.

2.3.2.2 Branching and silique phenotypes

Further analysis of all phentypes of the *eop1* mutant led to the detection of significant differences regarding the growth of the shoot branches. The *eop1* mutants possessed significantly fewer rosette branches ($p = 1.9 \times 10^{-8}$), but more axillary branches ($p = 2 \times 10^{-4}$, Mann-Whitney-U test) compared to Col-0 (Fig. 16A-B).

No significant differences were detected with regard to seed number per silique (p = 0.72) and number of siliques (p = 0.13, Mann- Whitney-U test) (Fig. 16C-D), implying no differences in seed production and in the amount of produced inflorescences.

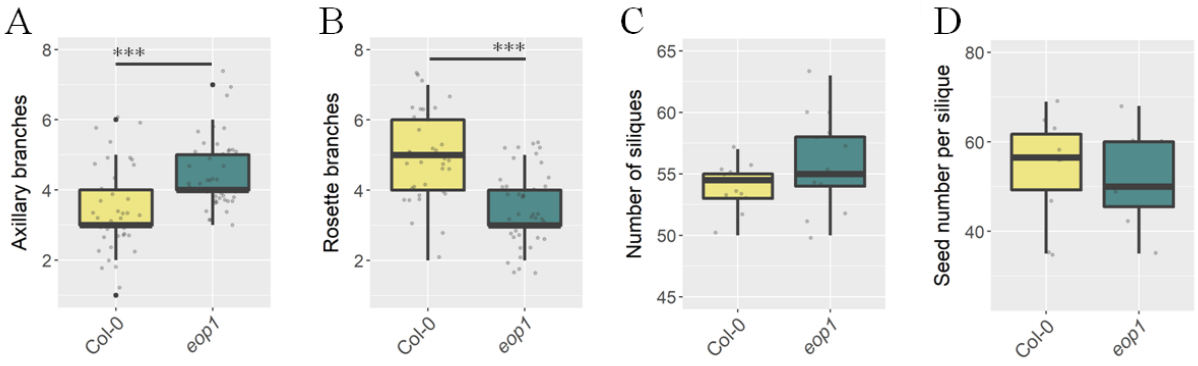


Fig. 16 Branching and siliques phenotypes of *eop1* mutant

(A-D) Quantification of further phenotypes of *eop1* in comparison to Col-0; Asterisks represent significance (p > 0.001) determined by t-test or Mann-Whitney-U-test as indicated in the text.

2.3.2.3 Trichome growth exhibits additional branching events

In addition to the size phenotypes, a trichome phenotype of the *eop1* mutants was indicated by their "hairy" appearance. Thus, a closer look was taken at the trichome number, length or branching. Microscopy of the first true leaf pair of 20 d old seedlings revealed that *eop1* trichomes had undergone additional branching events, explaining the "hairy" appearance (Fig. 17). Therefore, 100% of the analyzed Col-0 trichomes exhibited the expected three-branched pattern, compared to only 35% of the analyzed *eop1* trichomes exhibiting this appearance. Most of the *eop1* trichomes possessed four (42%) and five (23%) branches, respectively. Based on this, trichome branching is influenced by *eop1* mutation.

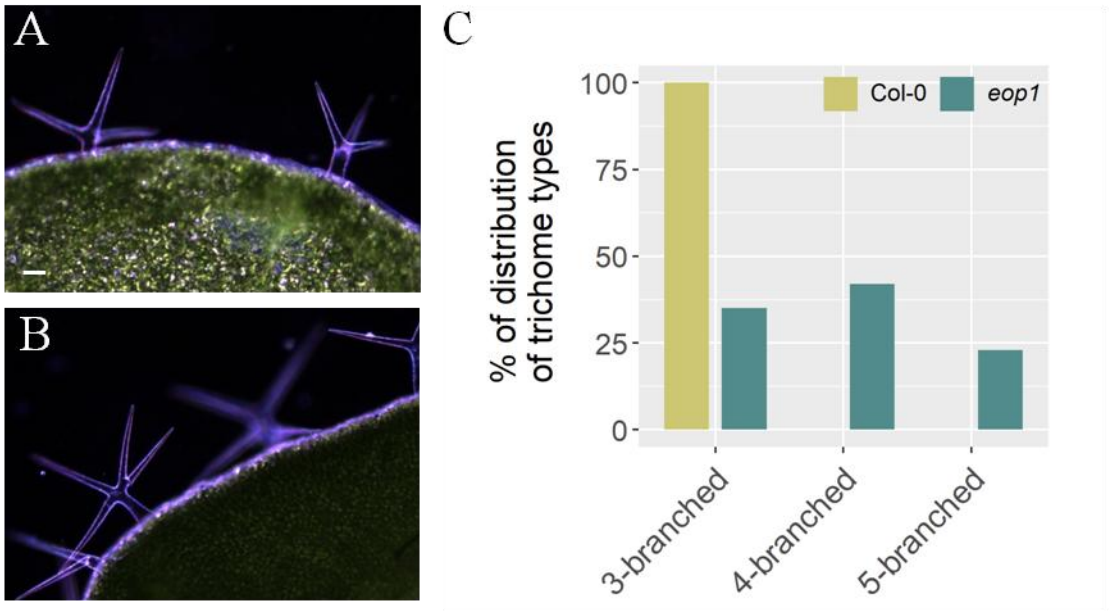


Fig. 17 Trichome branching phenotype of *eop1*

(A-B) Binocular microscope images of adaxial sides of the first true leaf pair of wild-type (A) and *eop1* (B) **(C)** Quantification of trichome branched types of Col-0 and *eop1*; scale bar: 100 µm

2.3.2.4 Ploidy level measurements

Since trichome development, especially each branching event, is initiated by an individual endoreduplication cycle[5], it was assumed that the ploidy level might be changed in *eop1* mutants compared to Col-0. Thus, flow cytometry of propidium iodide-stained nuclei of three biological *eop1* replicates was performed. Diploid (2n) and tetraploid (4n) Col-0 plants were taken as controls. For this analysis, three independent biological replicates of fully matured first true leaf pairs of each genotype were analyzed. The diploid lines of the *Arabidopsis* wild-type leaves showed a typical pattern of cells according to their different ploidy levels[49], whereas a shift towards more nuclei with higher ploidy levels was found for the tetraploid samples. Overall, the *eop1* mutants exhibited a ploidy level pattern more similar to the diploid Col-0 line than to the tetraploid line (Fig. 18). Although few significant differences in the percentage of 2C, 8C and 16C cells compared to the 2n controls ($p = 0{,}0308$, $p = 0{,}001$, $p = 0{,}0495$, respectively, Kolmogorov-Smirnov test) were observed, it was decided to not follow up on these measurements, as the results were not convincing enough to conclude that ploidy changes and therefore endoreduplication, might truly be affected by the *eop1* mutation. Facing the similar amount of 32C nuclei and the fact that no increase (or even occurrence) of 64C or even higher

ploidic cells could be observed, the trichome branching phenotype of the *eop1* mutants is most likely not caused by additional endoreduplication cycles.

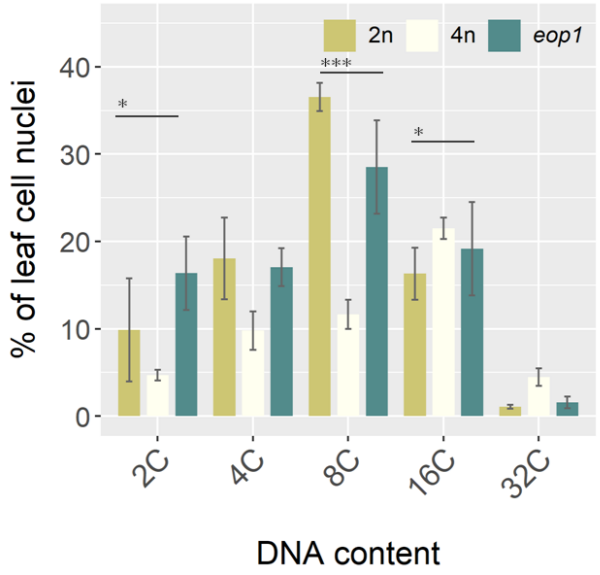


Fig. 18 Comparison of the distribution of ploidy levels of *eop1*, diploid and tetraploid wild-type leaf cells

Asterisks represent significance (*** = p > 0.001; * = p > 0.05) determined by t-test or Mann-Whitney-U-test as indicated in the text. Note that significances are only shown for diploid Col-0 and *eop1* mutant.

2.3.2.5 Oryzalin treatment

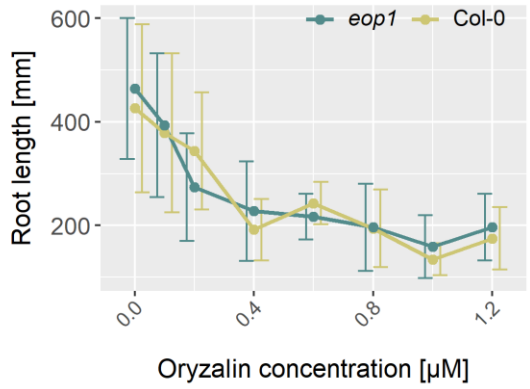


Fig. 19 Comparison of root growth inhibition by oryzalin between both genotypes

Trichome morphogenesis was reported to be influenced by the cytoskeleton via differences in microtubule and actin filament formation in *A. thaliana* (as described in chapter 2.1.1.5)[123]. Therefore, *eop1* and Col-0 seeds were germinated and grown for 10 d on oryzalin containing

media (0.1 µM, 0.2 µM, 0.4 µM, 0.6 µM, 0.8 µM, 1 µM and 1.2 µM). This dinitroaniline herbicide depolymerizes microtubules and prevents the polymerization of new ones by binding to alpha tubulin[162] during all stages of the mitotic cycles[163]. Accordingly, root growth is strongly impaired in plants treated with oryzalin (Fig. 19). Comparing the behaviour of Col-0 and *eop1* seedlings to this treatment, no significant ($p > 0.05$, Mann-Whitney-U test) difference was detected. Therefore, *eop1* mutation does not cause obvious changes in the microtubule formation. Since branching is particularly dependent on tubulin[121], it is rather unlikely that the *eop1* mutation is influencing the actin cytoskeleton, leading to the assumption that the cytoskeleton is not strongly affected by this mutation.

2.3.2.6 Effect of phytohormones inhibitors on *eop1* mutation

Having excluded influences of the *eop1* mutation on cytoskeleton and on endoreduplication, the effect of the *eop1* mutation on organ and cell size could still be caused by increased concentrations of phytohormones[55]. Cell elongation is particularly depending on BR signaling and GA-inactivated DELLA transcription factors as well as auxin levels[164]. Thus, a brassinosteroid and a gibberellin inhibitor were used to investigate their impact on the rescue of *eop1* mutants.

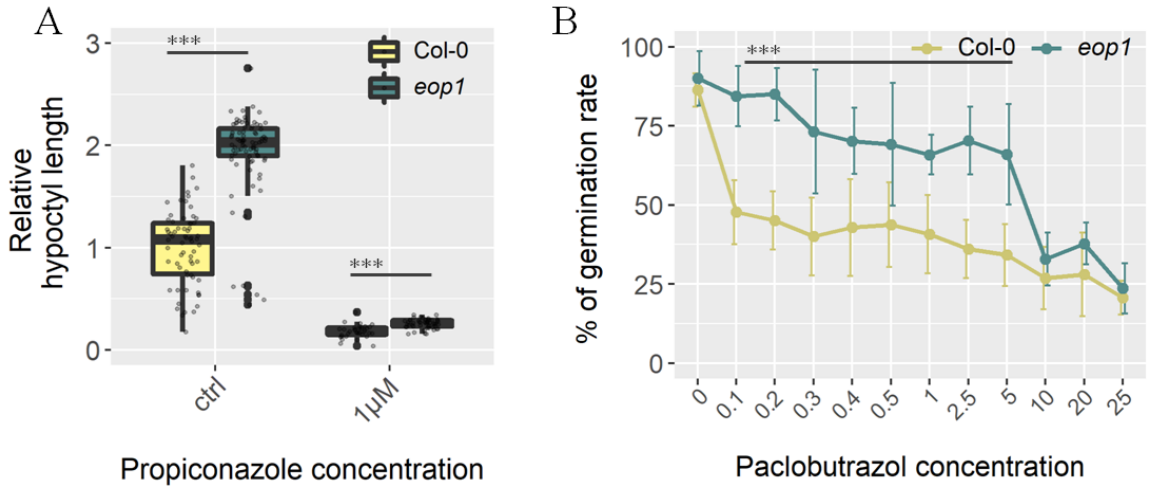


Fig. 20 Comparative treatment with the phytohormone inhibitors propiconazole and paclobutrazol of *eop1* and wild-type

(A) Propiconazole treatment; Relative hypocotyl length was calculated as ratio to Col-0 control plants. **(B)** Paclobutrazol treatment; Note the not equidistanced scale of the x-axis. Asterisks represent significance (*** = $p > 0.001$) determined by t-test or Mann-Whitney-U-test as indicated in the text.

Propiconazole (Pcz) has been reported to impair hypocotyl length via the inhibition of BR biosynthesis[165]. A similar effect of Paclobutrazol (PAC) on GA pathway was discovered[166].

In case the increased cell size of *eop1* is due to an overactive BR or GA response, a higher insensitivity to Pcz and PAC, respectively, would be expected.

Pcz treatment resulted in Col-0 hypocotyls that were decreased in length for about 82% whereas *eop1* mutants were even more shortened exhibiting only 13% of their untreated length (Fig. 20A). Analysis showed that although in both cases there is a significant difference in the sensitivity to Pcz ($p = 4.01 \times 10^{-5}$, Mann-Whitney-U test), *eop1* seedlings seemed to be more sensitive than Col-0 seedlings. Thus, a repetition including smaller steps in Pcz concentrations would allow for a clearer impression, it can be assumed that a change in BR response is most likely not the causal reason for elongated stem cells in *eop1* mutants.

A different approach was used for testing the effect of Paclobutrazol on e*op1* mutants. Since GA is responsible for seed germination, the germination rate of treated and untreated mutants and wild-type plants was compared. This experiment (Fig. 20B) revealed a significant difference in the germination ability in the presence of PAC between both genotypes ($p = 4.1 \times 10^{-10}$ for 0.1 µM PAC, Mann-Whitney-U test). The data show that *eop1* mutants are less sensitive to PAC than Col-0 as a significantly higher amount of seedlings was able to germinate up to a PAC concentration of 5 µM, leading to the conclusion that the *eop1* mutation positively stimulates the GA pathway in *A. thaliana.*

2.3.3 Identification of INCURVATA11 as gene affected by the *eop1* mutation

The mapping of the gene causing the *eop1* mutant phenotype started with a previous analysis of Illumina sequencing data from two different F2-pools of plants (Fig. S2). One pool of plants exhibited enlarged petals and represented the EMS-double mutant phenotype *eop1/paps1-4* (see chapter 2.1.3). The second pool consisted of plants with petals showing a wild-type phenotype. Analysis of the sequencing results revealed that the causal mutation is tightly linked to two markers flanking an interval of 545 kb. The fine mapping of this genomic region was conducted under the assumption that two thirds of the phenotypic wild-type pool consisted of plants that were heterozygous for the *eop1* mutation. According to this, a ratio of 2:1 for the causal SNP was expected in this pool, whereas 100% of the double mutant pool (exhibiting larger petals)

should contain the EMS-mutation at this position. The examination led to nine candidate SNPs that were used for marker design. 80 double mutants, arising from about 400 individuals of the F2 mapping population, were PCR-genotyped at these nine SNPs. 18 recombinant plants were found that allowed for the detection of the causal SNP: a C to T point mutation was identified at position 8125559 of chromosome 1 in the At1g22950 gene. This gene codes for INCURVATA11 (ICU11), which belongs to the 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily proteins.

2.3.4 Different alleles of *icu11*

To further prove that *ICU11* is the causal gene for the *eop1* phenotype, *T-DNA* insertion lines, mostly leading to the disruption of the gene, were analyzed with the aim to identify additional *icu11* alleles showing similar phenotypes as *eop1*. Therefore, 19 *T-DNA* insertion lines were ordered and PCR-genotyped. In contrast to 17 non-segregating lines, two FLAG lines (FLAG_207D09 and FLAG_402G04) in Ws-2 (Wassilewskija) wild-type background, were found to segregate and to exhibit enlarged petal sizes of approximately 15% and 9-26%, respectively (Fig. 21). This finding supported the idea that *eop1* mutation affects the *ICU11* gene. The *T-DNA* insertion site of FLAG_207D09 is 68 bp upstream of the transcription starting site of the gene at position 8127239, and thus part of the 5'UTR, whereas the *T-DNA* of FLAG_402G04 is inserted at position 8126671, located in the second intron of At1g22950 (Fig. 21D). A loss-of-function allele in S96 wild-type background with a NMU (N-nitroso-N-methylurea mutagenesis) mutated transition from C to T in the first exon, and the FLAG_402G04 line had already been described in literature[11]. These alleles are called *icu11-1* and *icu11-2*, respectively. Therefore, the *eop1* mutant allele was renamed to *icu11-3* and the FLAG_207D09 line was named *icu11-4*.

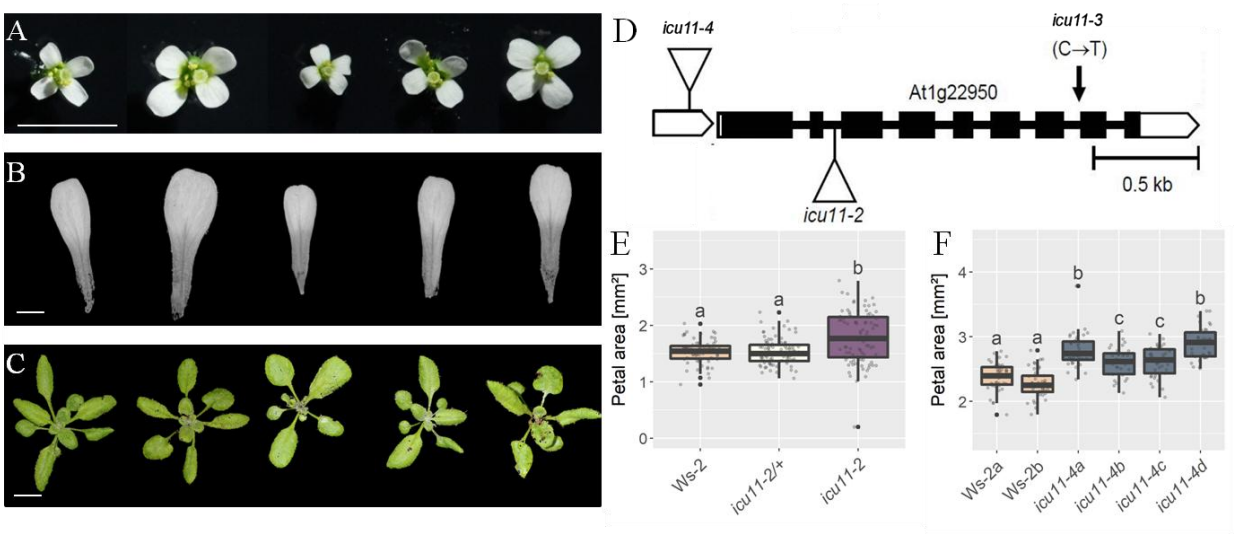


Fig. 21 Comparative overview of *icu11-2* and *icu11-4*

(A-B) Flower and petal sizes of Col-0, *icu11-3*, Ws-2, *icu11-2*, *icu11-4* (from left to right); scale bar: 5mm (A), 1mm (B) **(C)** 3-weeks-old rosettes (order according to A-B; scale bar = 1cm **(D)** Genetic structure[11] of ICU11 including the positions of mutations; Boxes and lines display exons and introns, respectively. Black boxes represent translated regions. White boxes represent untranslated regions. Triangle indicates T-DNA insertion for *icu11-2* and *icu11-4* and the vertical arrow the SNP for *icu11-3*. **(E)** Quantification of petal area of heterozygous and homozygous *icu11-2* in comparison to Ws-2 **(F)** Quantification of four independent *icu11-4* lines compared to two independent Ws-2 lines. (E-F) The letters indicate significant differences as determined by a Tukey's HSD test.

2.3.4.1 Pre- and posttranscriptional changes caused by mutations of the *icu11* alleles

Due to the different localizations and types of the three mutations within the *ICU11* gene, all three alleles were screened regarding their pre- and posttranscriptional changes on *ICU11* mRNA and protein levels. *T-DNA* insertions mostly disrupt the gene[147]. Hence, a decrease in relative mRNA expression levels was expected for *icu11-2* and *icu11-4* (Fig. 22A). Surprisingly, both mutant lines showed a higher relative expression of *ICU11* compared to Ws-2 wild-type. For *icu11-4*, the difference was significant with a p of 0.0132 (Kolmogorov-Smirnov test). *T-DNA* insertions upstream of the transcription starting site can lead to an overexpression of the upstream gene[167], since *T-DNA* genes are under CaMV 35s promoter control[168,147]. Thus, an induction of a higher expression of the genes located downstream might explain this finding. Interestingly, an increased expression of *ICU11* has also been found in the *icu11-2* line, although here the *T-DNA* is inserted in the second intron, leading to a premature stop codon. However, both *T-DNA* lines unexpectedly showed an increase of *ICU11* mRNA level.

In *icu11-3* no significant difference regarding the relative *ICU11* mRNA expression was detected compared to Col-0 wild-type. Thus, not the mRNA level itself, but another effect of the mutation is causing the mutant phenotype. Further investigations to the C to T point mutation of *icu11-3* revealed that the transition occurs at the last nucleotide of intron 7. As the first two and the last two nucleotides of introns are responsible for the correct splicing of the mRNA[169], the *ICU11* cDNA of the *icu11-3* line was analysed regarding size differences compared to Col-0 (Fig. 22B). Two DNA bands appeared, that differed in size compared to the Col-0 control band. The two cDNA fragments were therefore further analyzed by sequencing. This analysis revealed that the larger band arise from a non-spliced intron 7, whereas the smaller band is lacking intron 7 and 44 nucleotides of exon 8 (Fig. 22C, cDNA and predicted protein sequences in Fig. S1-S2). That means that the *icu11-3* mutation causes changes in the splicing pattern as two different splicing forms of *ICU11* exist in this mutant line. Both splicing forms lead to a premature stop codon. In summary, two of the three mutant *icu11* alleles cause an overexpression of the *ICU11* transcript, whereas the third one exhibits modified mRNA splicing which indicates that most likely also the resulting mutant proteins will also show modifications. Unexpectedly, both, transcriptional and posttranscriptional changes, of the *ICU11* gene resulted in a similar phenotype regarding petal size. If *icu11-2*, *icu11-3* and *icu11-4* are true loss-of-function mutants or rather gain-of-function mutants remained unclear at this point.

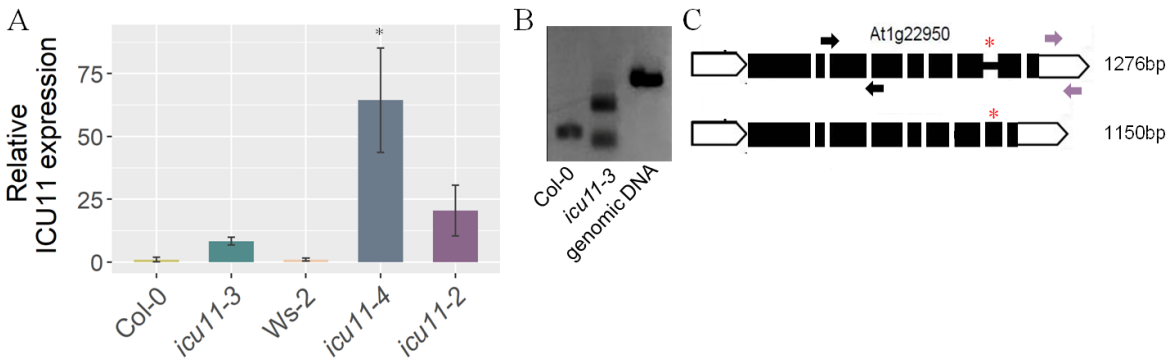


Fig. 22 Pre- and posttranscriptional changes caused by *icu11* mutations

(A) Comparison of relative expression of all three *icu11* mutations and their respective wild-types. Asterisks represent significance (* = $p > 0.05$) determined by t-test or Mann-Whitney-U-test as indicated in the text. **(B)** Agarose gel electrophoresis of RT-PCR produced with specific primers to amplify the full-length mRNA of the *ICU11* gene from wild-type and *icu11-3* **(C)** Schematic representation of both splicing versions of *icu11-3* (modified after [11]). Red asterisks indicate positions of premature stop codons. Arrows indicate qPCR primer positions for *icu11-3* & *icu11-4* (black) and *icu11-2* (purple). Note that *icu11-2* possesses T-DNA insertion within black amplified region.

2.3.4.2 Allelic relationships

The unexpected results of the relative *ICU11* mRNA expression of all three *icu11* alleles questioned the earlier assumptions that the analyzed lines are loss-of-function mutants. To gain more insights, a genetic complementation experiment (Fig. 23A) was carried out. First, homozygous *icu11-3* and *icu11-4* plants were crossed to heterozygous *icu11-2/+* plants, respectively, since *icu11-2* was published as a loss-of-function mutant according to previous studies[11]. Afterwards, petal sizes of the F1 population were analyzed. If the parental line was a crossing between a homozygous gain-of-function mutant (GOF) and a heterozygous loss-of function (LOF) mutant, the measured F1 petal sizes should be separable into two groups. The first group would be expected to contain slightly enlarged petal sizes compared to wild-type petals due to the combination of a wild-type copy from the *icu11-2/+* line and one mutant GOF copy of the *icu11-3* or *icu11-4* line, respectively, representing a heterozygous GOF situation. The second group would be expected to exhibit similar petal sizes as the wild-type control petals, as the copy of the gain-of-function allele should complement the loss-of-function copy of the *icu11-2*/+ line. Small deviances could be expected due to possible dose-dependant effects of the two different *icu11* mutations.

The analysis revealed that all individuals of the F1 populations showed an increase in petal size (Fig. 23D) and also in overall organ size (Fig. 23B-C) compared to Col-0 and Ws-2 backgrounds. The combination of two different mutant alleles led to a significant different increase in petal size of about 23% (*icu11-3/icu11-2*) and 27% (*icu11-4/icu11-2*) compared to the respective wild-type line, similar to the phenotype of the parental homozygous mutant lines in this experiment.

Taken together, no genetic complementation could be detected in any of the tested mutant allele combinations, indicating that all three *icu11* mutations act in a similar way regarding the resulting ICU11 protein and the linked pathways.

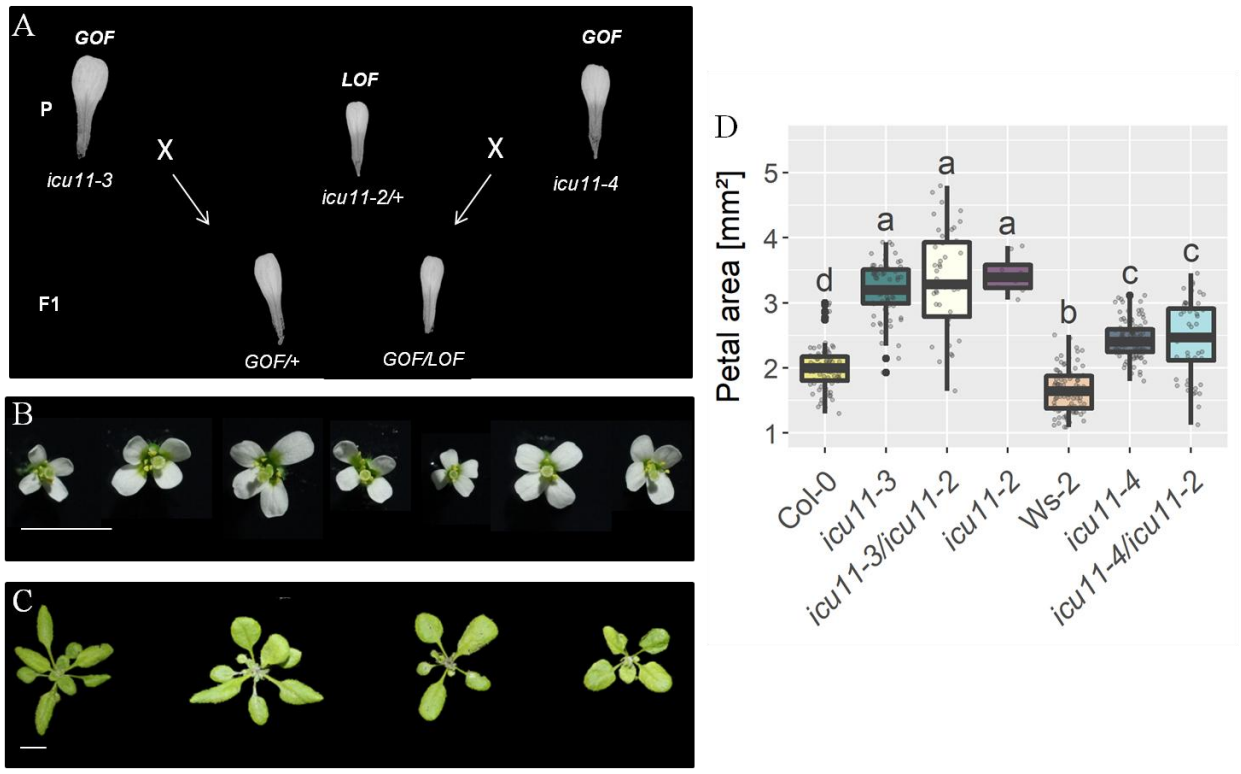


Fig. 23 Overview of complementation assay

(A) Theoretical model of complementation assay **(B)** Comparison of flower sizes of all genotypes (Col-0, *icu11-3*, *icu11-3/icu11-2*, *icu11-2*,Ws-2, *icu11-4*, *icu11-4/icu11-2* (from left to right))and **(C)** 3-weeks-old rosettes (Col-0, *icu11-3/icu11-2*, Ws-2, *icu11-4/icu11-2*). Scale bar: 5mm (A) and 1cm (B). **(D)** Quantification of petal areas of all genotypes; The letters indicate significant differences as determined by a Tukey's HSD test.

qPCR comparison to *icu11-1*

The published loss-of-function mutant *icu11-1* showed a significant increase in the relative expression of several MADS-box transcription factors as well as FLOWERING LOCUS T (FT)[11]. To gain further insights into the behaviour of the three *icu11* alleles, *icu11-4* was taken to analyze the relative expression of the same downstream genes. If *icu11-4* represented a true overexpressing line, a decrease in the gene expression of the MADS-box transcription factors would be expected, if the expression patterns depend on *ICU11* mRNA expression (Fig. 24). Strikingly, the expression of all tested MADS-box transcription factors was upregulated in *icu11-4* compared to Ws-2 background. The relative *APETALA 3* (*AP3*) and *SEPALLATA 3* (*SEP3*) expressions were also tested in the *icu11-3* line, similarly exhibiting a strong upregulation in the mutant. The *FT* expression, on the other hand, was clearly downregulated in the tested *icu11-4* line compared to Ws-2 background. *Flowering Locus T,* which promotes the

transition from vegetative growth to flowering, is expressed in a circadian rhythm. Thus, its expression peak is present in the morning and strongly repressed in the afternoon[170]. Therefore, differences of *FT* expression could be explained by different harvesting times and by a reinforced regulation in the tested *icu11* mutants compared to their wild-type background and are not necessarily depending on the single mutant *icu11* allele. Since three *icu11* alleles behave in a similar expression pattern regarding the analyzed downstream genes, it seems very likely that any changes of the *ICU11* mRNA and protein levels lead to the same phenotype observed for all *icu11* alleles.

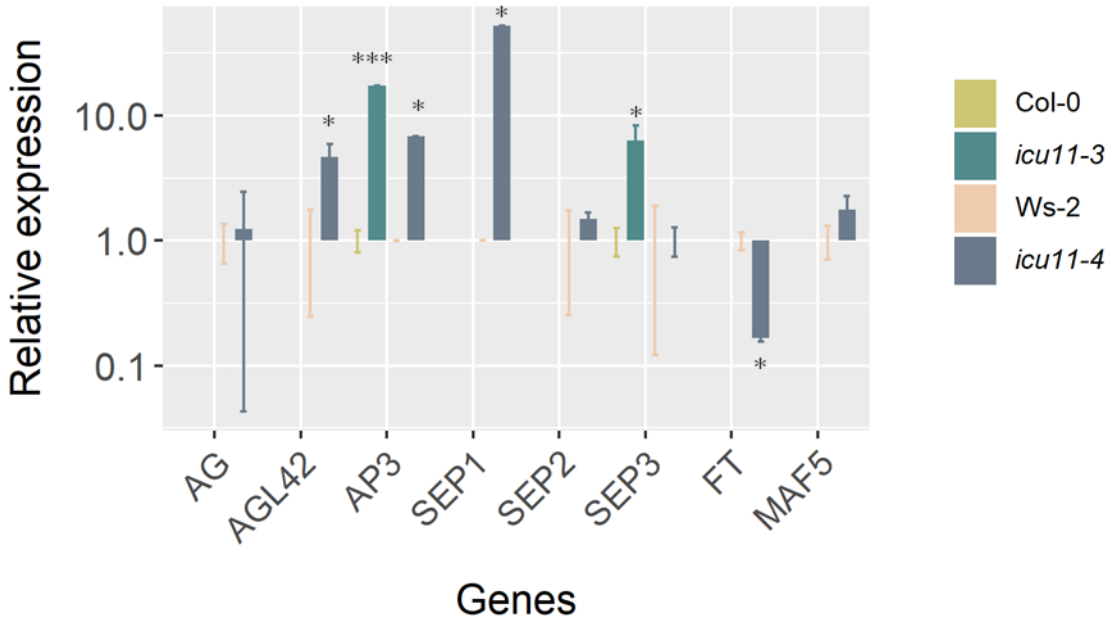


Fig. 24 Relative expression of MADS-box transcription factors of four genotypes

Expression levels are calculated as ratio to their respective wild-type levels. Note the logarithmic y-axis. Asterisks represent significance (* = p > 0.05, *** = p > 0.001) determined by t-test or Mann-Whitney-U-test as indicated in the text.

2.3.4.3 Knockdown and knockout lines of ICU11

Doubting the loss-of-function behaviour of the three *icu11* alleles, another approach to gain a knockdown or knockoff line was aimed for, using two different techniques. First, a Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR associated protein 9 (CAS) (CRISPR/Cas9) system[171] carried out with three different combinations of two single

guide RNAs, that revealed only one mutant plant out of 450 PCR-genotyped individuals in T1 and T2. Unfortunately, the resulting mutation was a complete deletion of intron 2 leading to a wild-type phenotype, proving the functioning of the system but lacking a helpful outcome.

The second approach was to design miRNA-induced gene silencing (MIGS)[159] lines. This method is based on an *Arabidopsis thaliana*-specific 22-nucleotide miRNA (miR137) that causes gene silencing via the activation of expression of trans-acting small interfering RNAs (tasiRNAs). Recent studies showed that the fusion of fragments of genes downstream of a miR137 target site led to a significant knockdown of a single gene. Thus, two different miR137-constructs were designed, consisting either of 496 bp of the N-terminal domain of the *ICU11* cDNA or the following 496 bp of the C-terminal domain and transformed into Col-0 plants. The first construct never led to T1 plants, and it was decided to keep going only with the second line.

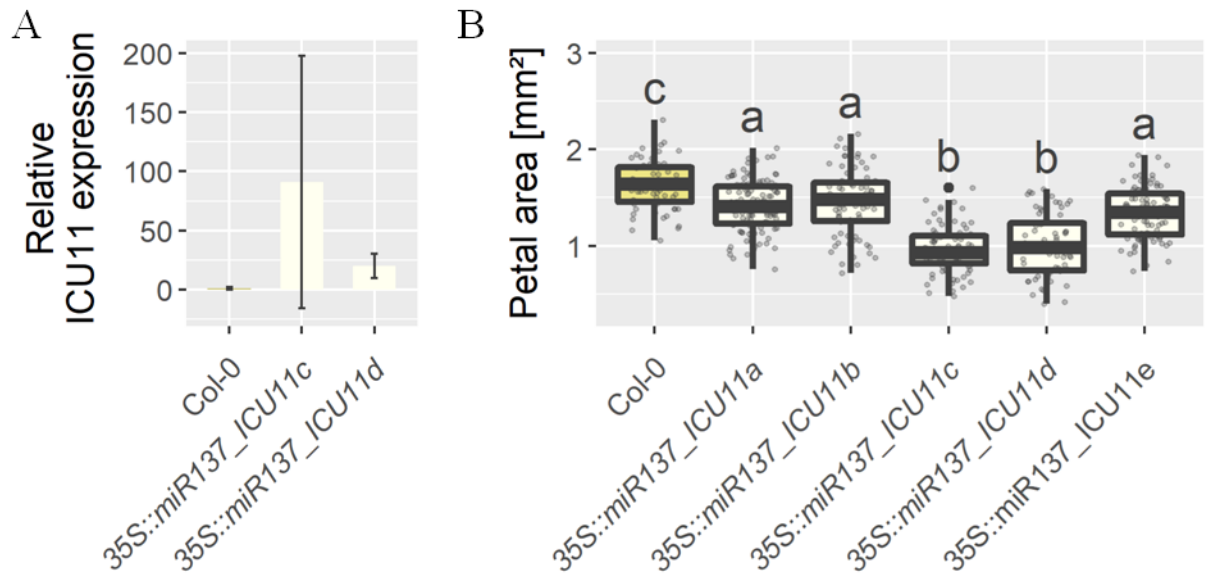


Fig. 25 Overview of 35S::miR137:ICU11 lines

(A) Relative ICU11 expression levels of 35S::miR137_ICU11 lines compared to wild-type **(B)** Quantification of petal sizes of five independent 35S::miR137_ICU11 T2 transformants compared to wild-type; The letters indicate significant differences as determined by a Tukey's HSD test.

To measure the impact of the gene silencing, petal areas of five independent T2 MIGS lines were analyzed and found to be smaller compared to Col-0 background ($p < 8.7 \times 10^{-5}$), which was promising regarding the knockdown of ICU11 (Fig. 25B). Unfortunately, the relative expressions

levels of *ICU11* mRNA of the two lines, that exhibited the smallest petal sizes were not decreased ($p = 0.23$ and $p = 0.27$, respectively, Kolmogorov-Smirnov test) compared to Col-0 (Fig. 25A). According to this, no trustworthy knockdown or knockoff line of ICU11 could be gained.

2.3.4.4 Transgenic lines

Aiming at a better understanding regarding the genetic regulation of *ICU11*, different transgenes were introduced in both wild-type backgrounds, Col-0 and Ws-2, as well as *icu11-3* and *icu11-4* and at least 10 plants of a minimum of three independent lines of their corresponding T2 generations, unless additionally explained, were analyzed regarding petal size.

Overexpression of *ICU11*

The first analyzed transgene was the genomic full length *ICU11* under control of a 35S::promoter (35S::g*ICU11*). Measuring the petal sizes of at least seven individuals of each T1 line, the effect of overexpressed *ICU11* was examined (Fig. 26). Focusing on the T1 lines in Col-0 and Ws-2 background, highly significantly enlarged petals (21% and 50%, respectively) were measured ($p = 3.2 \times 10^{-8}$ and $p < 2 \times 10^{-16}$, respectively). Similar results were found for the T1 transformants in *icu11-3* background with an additional petal size increase of 14% and $p = 7.3 \times 10^{-06}$). Thus, overexpressing *ICU11* resulted in a stronger effect on petal size than the *icu11-3* mutation. Comparing 35S::g*ICU11* (Col-0) to the same transgene in *icu11-3* produces a petal size difference of only 4%, indicating on the other hand, that the *icu11-3* mutation is slightly adding up the petal size compared to the overexpression. However, a higher sample size of T1 individuals might rule out this difference.

Surprisingly, no significant increase of petal size could be found for the 35S::g*ICU11* (*icu11-4*) T1 individuals ($p = 0.62589$), as it exhibited a petal size increase of only 1% compared to the untransformed *icu11-4* line, that showed by itself an increase of 51% compared to the Ws-2 wild-type background. Regarding this result, the *icu11-4* mutation obviously led to an identical effect as the overexpression of *ICU11*. Finally, it was concluded that all three *icu11* mutants possess an overexpression phenotype, with *icu11-4* displaying the strongest phenotype regarding petal size.

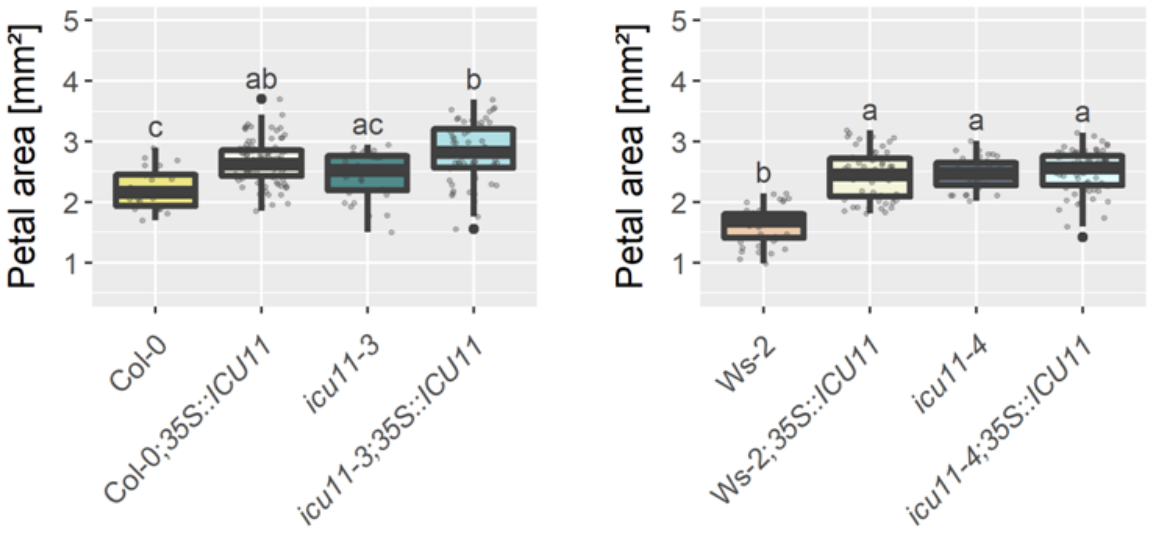


Fig. 28 Comparative representation of T2 ICU11 overexpression lines against their respective wild-types

The letters indicate significant differences as determined by a Tukey's HSD test.

Rescue of *icu11*-3 and *icu11-4*

To confirm that *ICU11* is the responsible gene for the *icu11-3* and *icu11-4* phenotypes, a rescue construct was designed and transformed into these mutants as well as into their respective wild-type lines. The rescue construct consisted of 917 bp upstream of the transcription starting site of the *ICU11* gene, including the assumed promoter region, and was followed by the full-length genomic *ICU11* (g*ICU11*). Petal areas of averaged T2 lines are shown in Figure 27. Focusing on the *icu11-4* mutant, a clear nearly complete rescue was detected ($p = 0.0045$), as the transgenic *icu11-4* petals are only 0.5% larger than the Ws-2 petals. Thus, the rescue construct led to a 29% ($p = 9.76 \times 10^{-40}$) decrease of petal area compared to the original increase of a 41% enlarged *icu11-4* petal sizes in comparison to their untransformed wild-type.

A different result was detected for the *icu11-3* transformants. Although the transformed T2 plants exhibited a decrease in petal size of 38% ($p = 2.49 \times 10^{-43}$) compared to the 26% increase in the untransformed mutant lines, a very similar decrease in petal area of the transformed Col-0 plants (32%, $p = 7.05 \times 10^{-20}$, Mann-Whitney-U test)) was discovered. This strong negative impact on the transformed Col-0 plants can either be explained by feeding damage of thrips or by stress caused by hygromycin selection. However, the facts that the Ws-2 line, the *icu11-4* control and T2 lines were grown in parallel under same conditions, and that the other transgenic lines (in Col-0 background) did not show such a strong decrease in petal size, contradict the previous

explanation. Additionally, the influence of a transgene is often dependant on the location, resulting in stronger or weaker effects depending on the genetic position and also the expression rate of the transgenic insertion. Analyzing a higher number of independent T2 lines could enlighten this vague result. Nevertheless, a slight rescue of around 6%, can be assumed also for the *icu11-3* T2 lines.

With regard to the overexpression phenotypes and the mutation types of the mutant *icu11* alleles, the detected rescue effect is unexpected. One or two additional *ICU11* wild-type copies from the rescue construct should neither change the mutant phenotype, nor rescue it, under the condition that the background mutation represents an overexpression of *ICU11*. However, the rescue analysis rather argues for loss-of-function *icu11* mutants and provides a further proof that *ICU11* is the causal gene for the observed phenotype.

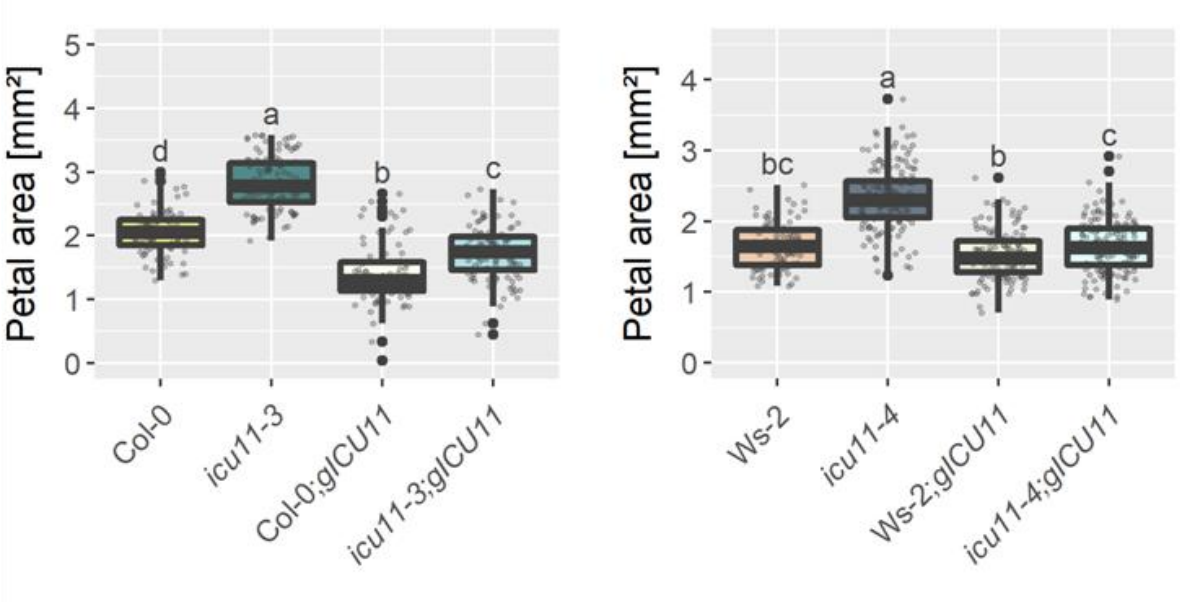


Fig. 29 Comparative representation of T2 ICU11 rescue lines against their respective wild-types

The letters indicate significant differences as determined by a Tukey's HSD test.

Transgenic *ICU11* construct lacking catalytic function

As described in chapter 2.1.2 the functioning of 2-OG-Fe(II)-dependant oxygenases is mainly determined by the 2OG-FeII_Oxy motif[139], consisting of the sequence motif His-Xaa-Asp/Glu-(Xaa)n-His, that is responsible for the formation of the catalytic triad, a double stranded β-helix core, that binds Fe(II)[140]. To test if the observed phenotype of the *icu11* mutants is dependant on

this motif, a site-directed mutagenesis was performed. Therefore, the responsible amino acids were exchanged by other amino acids with different chemical properties, to prevent the fulfilling of their original function. The positively charged and basic histidines (H) were changed into polar but uncharged serine (S), whereas the acidic and negatively charged aspartic acid (D) was mutagenized into non-polar and hydrophobic alanin (A).The transgene g*ICU11* HDH > SAS was transformed into *icu11-3* and *icu11-4* and their respective wild-types under the control of its genomic promotor. A rescue effect of the mutant lines would argue for ICU11 acting independently of the 2OG-FeII_Oxy motif or rather indicate that the observed phenotype is not linked to the 2OGD catalytic function. The results of the measured petal areas of independent T2 lines are shown in Figure 28. In general, the transgene and/or the selection led to a decrease in petal size of about 14% and 11% for Col-0 and Ws-2 wild-types ($p = 0.000933$ and $p = 0.0001524$). The transgenic *icu11-3* and *icu11-4* lines showed a decrease in petal size of 19% and 16%, similar to that of the respective wild-type T2 lines. Especially in comparison to the very clear rescue of the *icu11-4* T2 lines, a true rescue effect of the transgene lacking the 2OGD function is rather unlikely. Thus, the phenotypes of at least these two *icu11* alleles seem to be dependent on a functional His-Xaa-Asp-(Xaa)n-His triad and therefore the 2-OGD catalytic function of ICU11.

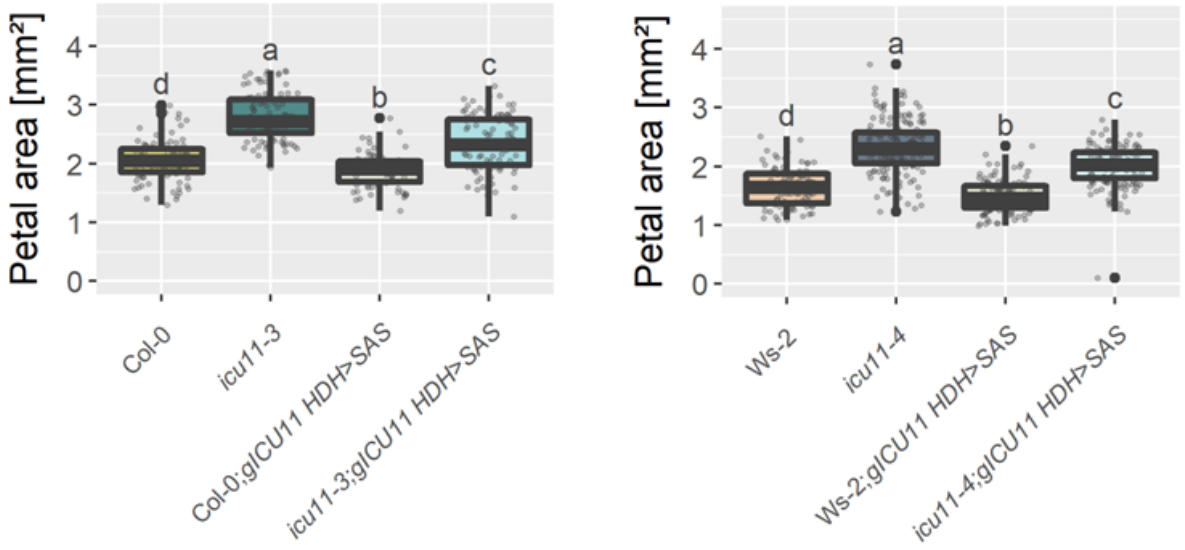


Fig. 28 Comparative representation of T2 ICU11 lines lacking its catalytic 2OGD function against their respective wild-types

The letters indicate significant differences as determined by a Tukey's HSD test.

2.3.5 Localization of 35s:ICU11

As basal characteristic of a protein, the localization within the cell can give an additional insight in its function. Therefore, the overexpression construct (35S::g*ICU11*) was cloned to YFP upstream of the stop codon of the genomic *ICU11*. Root apices of nine-days-old Col-0 T2 seedlings were taken to check the localization of ICU11 with the help of a confocal microscope. The YFP signal could clearly be detected in the cytoplasm (Fig. 29), surrounding the vacuole (Fig. 29H). Additionally, an overlap with the DAPI-stained nuclei could not been detected (Fig. 29D-F). Although a slight nuclear localization could not be excluded without further analysis under better microscopic resolution, it can be concluded that ICU11 exhibits a clear cytoplasmic localization under standard LD conditions. Regarding, the influence of the *icu11-3* mutation on the GA-pathway described in chapter 2.3.2.6, the detected cytoplasmic localization provides a further argument for the ICU11 protein acting in GA biosynthesis. To restrict the localization to the first, second or third part of the GA-synthesis, occurring in the plastid, the endoplasmic reticulum or the cytosol[4], it would be necessary to conduct further analyses with specific dyes for these different cell organelles and compartments.

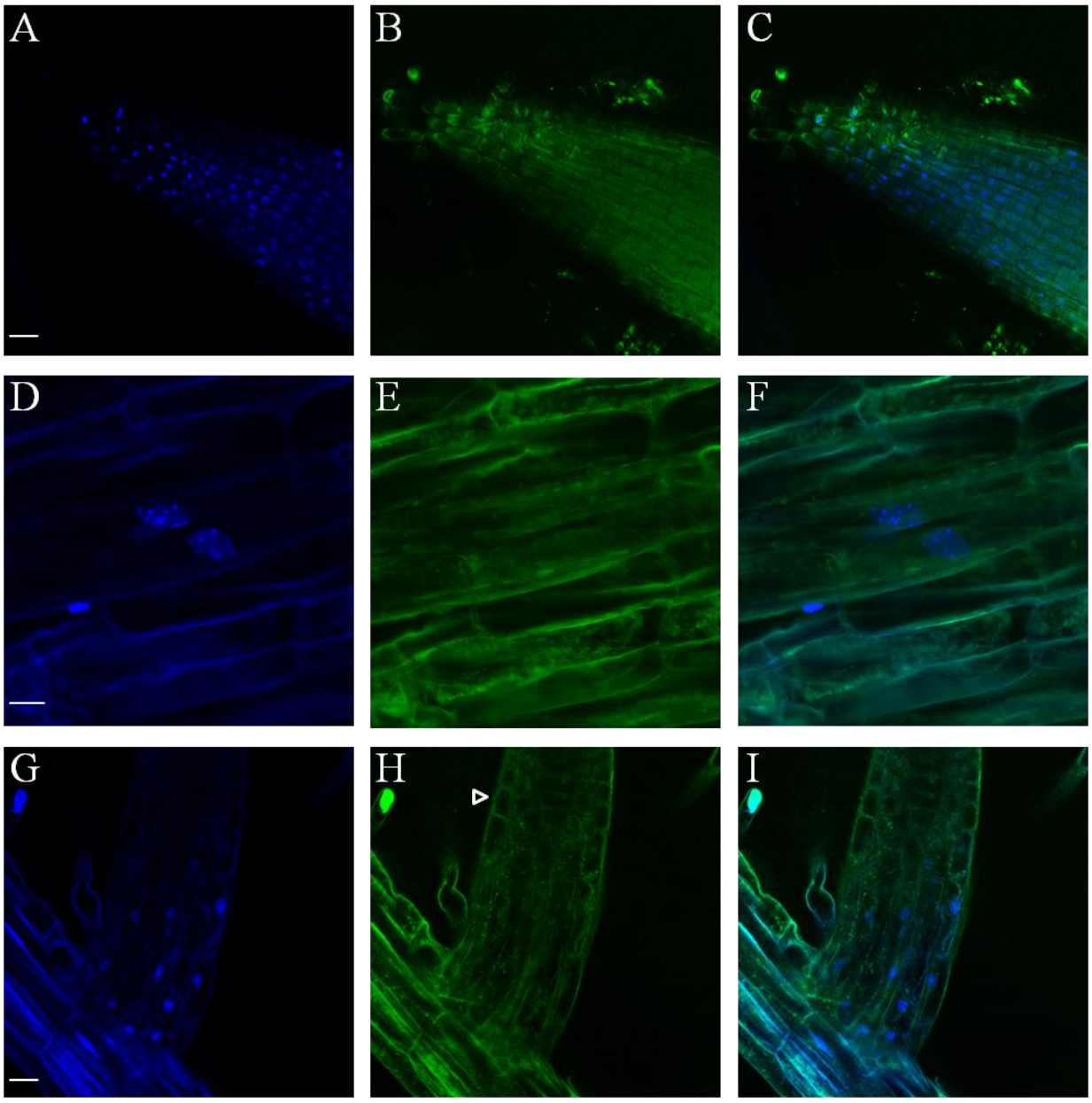


Fig. 29 Subcellular localization of ICU11. Confocal laser scanning micrographs of 35S::*ICU11* in Col-0

Different parts of the root are shown **(A-C)** Apex, **(D-F)** elongation zone **(G-I)** Lateral root outgrowth. Note the clear cytoplasmic YFP-signal surrounding the vacuole (white arrow). Scale bar: 20 µm

2.3.6 Expression levels of key genes of the GA biosynthesis and signaling pathway

As detected in chapter 2.3.2.6, ICU11 seems to have an influence on the GA pathway, especially on GA synthesis, implied by the localization analysis described in chapter 2.3.5. To support this theory, a qPCR was performed on the major bioactive GA-synthesis genes GA20ox1 and

GA20ox2[76] of the three *icu11* lines and on two T2 MIGS lines (Fig. 30). The expression pattern of GA20ox1 did not show a significant difference in any of the analyzed lines. Focusing on GA20ox2, significant changes for the *icu11-3* (p = 0.000129123, Kolmogorov-Smirnov test) and the *icu11-2* (p = 3.41788 x 10-5, Kolmogorov-Smirnov test) lines were detected compared to their wild-types. Therefore, ICU11 was shown to influence the GA synthesis via increasing GA20ox2 expression at least in the *icu11-2* and *icu11-3* mutants. Since it was not possible to clarify the type of the mutations of the *icu11* alleles, it could not be determined, if ICU11 increases or decreases GA20ox2 expression under wild-type conditions.

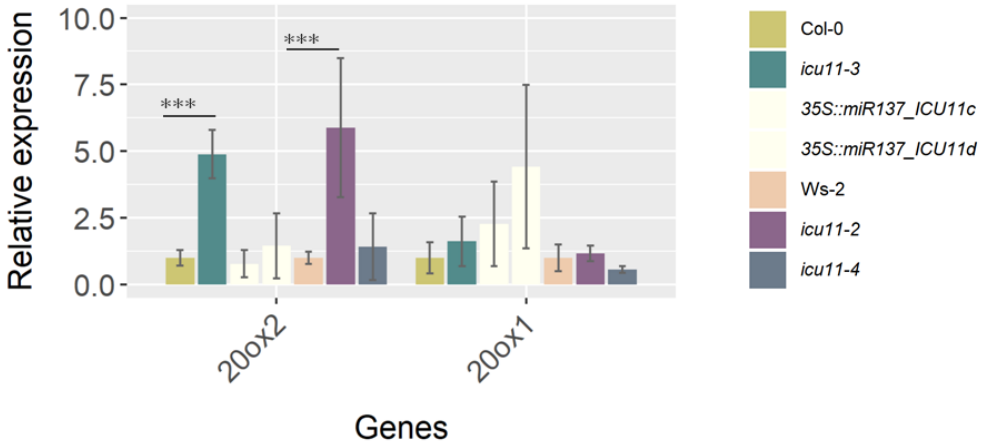


Fig. 30 Comparative representation of GA-synthesis gene expression of all genotypes

Asterisks represent significance (*** = p > 0.001) determined by t-test or Mann-Whitney-U-test as indicated in the text.

2.4 Discussion

2.4.1 Proposing a model of the relationship of *icu11* alleles

In recent years, mutations that lead to an increase in plant biomass gained more and more interest especially regarding crop yield improvement[172–174]. Thus, the *eop1* mutant found via an EMS-screen, which exhibited not only a 26% increase in petal size but also overall increased organ sizes based on enlarged cell sizes, deserves attention also from this point of view. The analysis of the *eop1* mutant was aimed at detecting the causal gene as well as the molecular mechanism underlying the observed phenotype. Sequencing and fine mapping resulted in a C to T transition of the *ICU11* gene. Additionally, two *icu11* T-DNA insertion lines were found to exhibit a similar phenotype, thus confirming *ICU11* as causal gene for the increased biomass phenotype of the *eop1* mutant. With the progression of the project, several unexpected and also contradictory results turned out to complicate the understanding of how ICU11 and its three *icu11* mutant alleles (*icu11-2, icu11-3* and *icu11-4*) act in *Arabidopsis thaliana*. Having analyzed the underlying transcriptional changes of *ICU11* caused by the three mutations, an increased mRNA expression level for the *icu11-2* and the *icu11-4* lines was detected. This contradicted the published knock down shown for the *icu11-2* mutant[11], especially since the amplified *ICU11* region corresponded exactly to the published one due to the usage of the published oligos. However, different growth conditions and technical methods might explain this conflicting result. The *icu11-3* mutation was shown to change the splicing pattern of *ICU11* and is therefore most likely leading to truncated protein versions due to premature STOP codons. Summarizing this analysis, both an increase of *ICU11* mRNA level and modified proteins lead to a similar phenotype, a result, that is rather unexpected. Furthermore, the complementation assay revealed that all three mutant *icu11* lines exhibit a similar behaviour regarding their type of mutation (LOF or GOF), since F1 crossing lines carrying one copy of two different *icu11* mutated genes show the e phenotype as homozygous mutants. An additional puzzling result was the transgenic overexpression of *ICU11* (*35S::ICU11*) in *icu11-3*, *icu11-4* and their respective wild-types, which phenotypically resembled the untransformed *icu11* mutants with strong increases in petal size, representing the overall increased biomass. The overexpression analysis on its own would argue for the *icu11* mutants to be gain-of-function (GOF) mutants, supporting the measured

increase in mRNA *ICU11* levels. The fact that the *icu11-4* mutant with highest *ICU11* expression and strongest petal size increase with 51% did not show significantly increased petal sizes with an additional overexpression transgene, specifically encourages the theory that the *icu11* mutants act as GOF mutants, with *icu11-4* representing an overexpressing mutation by itself. Confusingly, the *icu11-4* line turned out to be rescued by a homozygous wild-type *ICU11* transgene under the genomic promoter. This was unexpected since an overexpressing mutant should not be influenced by an additional wild-type gene copy. This rescue effect questions the GOF mutant theory and argues rather for LOF mutants. The T2 MIGS lines gave rise to even more contradictory results, exhibiting decreased petal sizes, but increased *ICU11* mRNA expression, which did not help to solve the confusions. Clearly the fact that *icu11-4* can be rescued proposes the strongest argument for the *icu11* alleles to be LOF mutants, although the increased mRNA levels are contradictory. It needs to be considered, that a high mRNA level does not necessarily result in an increased protein level. Since any analyses of the presence of mutated proteins, for example SDS-PAGE followed by western blot with specific anti-ICU11 antibodies in comparison to the wild-type ICU11, have not been examined, no assumption can be made regarding the absence or size of the mutated proteins. An autogenous regulation, in which proteins directly regulate their own transcription, as it is known e.g. for RNA-binding proteins[175] and others[176,177], or a positive feedback loop used as regulation effect caused by an inactive or not functioning protein[100,178], might be a plausible explanation for the detected increase of mutant *icu11* mRNA level. However the mutant *icu11* proteins might actually not be able to fulfill their function. Regarding the *icu11* mutants, nonfunctional or absent proteins might occur by truncated protein versions or mRNA silencing due to T-DNA dependant splicing changes or degradation effects[179]. Still, the underlying mechanism of the resemblance of the overexpressing and the LOF phenotype remains unclear.

The following model is based on the assumption that an overexpression of *ICU11* inhibits the function of the protein, and thus causes the same effect like a LOF protein. This model is derived from the idea that ICU11 uses oligomerization as regulation mechanism between active and inactive state. Since the overexpression leads to a LOF phenotype, ICU11 is most likely more active in lower, and inactive in higher oligomerization states. Similar to other well examined proteins[180,181] an ICU11 homodimer might be inactive because of monomeric parts are wedged into the active site of the bound monomer. An activation and therefore dissociation of the ICU11

monomers could occur via a favoured binding to a ligand or by signaling events that induce an open conformation into the dimer[182]. An overexpression possibly leads to an even higher oligomerization degree, hindering the ligand binding or conformational opening and is thus amplifying the inactive ICU11 state. According to this model, a truncated or misspliced *icu11* version would not be able to bind to its interacting partner and thus not fulfill its wild-type function.

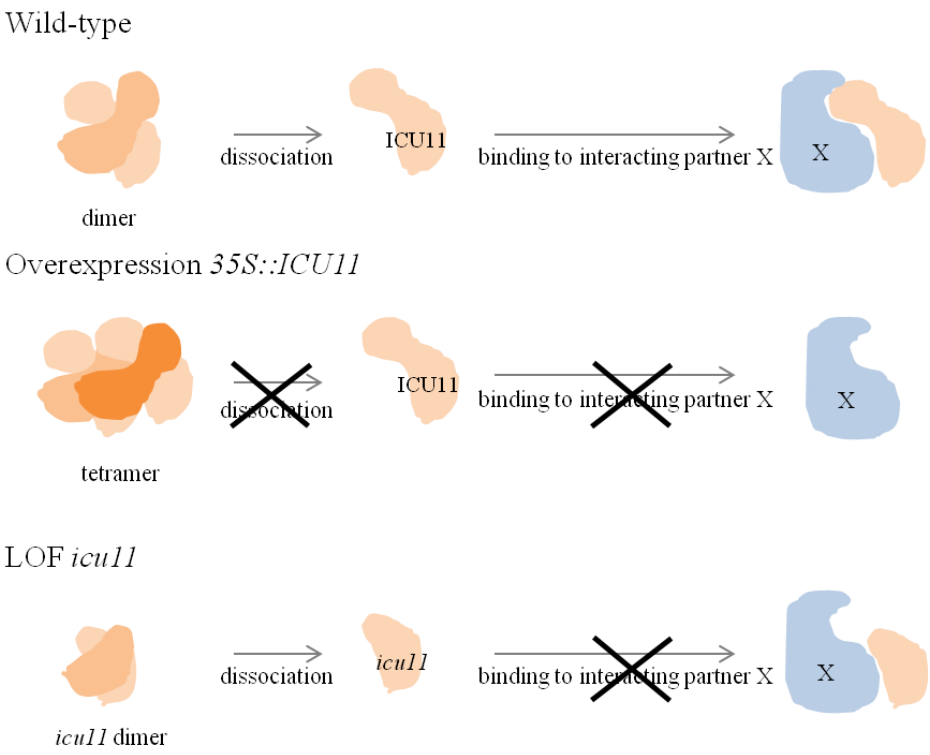


Fig. 31 Proposed model for functional effects of different changes on ICU11 protein based on oligomerization-mediated active and inactive states

Note that LOF icu11 protein is represented smaller due to the lack of the required domain that allows interaction with the binding partner.

Of course, the proposition of this model raises the question, if 2OGDs are known to dimerize. Crystallographic analyses found 2OGD structures mainly consisting of a double-stranded β-helix (DSBH) core fold. α-helices localized at the N-terminus, as well as the C-terminus of these DSBH were shown to allow dimerization. Specifically ribosomal oxygenases, such as the human OGFOD1 and JMJD6, as well as the *Arabidopsis* JMJ24 were found to form dimers[183,184].

To check, if ICU11 indeed occurs in an oligomeric state, it would be necessary to examine a native SDS-PAGE, followed by a western blot with specific anti-ICU11 antibodies. Protein interaction in form of dimerization could be detected either via an *in vivo* protein assay, such as a Yeast-two-hybrid assay or Bimolecular fluorescence complementation (BiFC) assay, or via an *in vitro* protein assay, such as pull-down or by immunoprecipitation. If different oligomerization states of ICU11 could be shown, further analyses regarding enzyme activity could be performed. Therefore, the knowledge of the molecular function and thus one or more substrates of ICU11 would be necessary to prove the assumed model or to give deeper knowledge about the regulatory mechanism of ICU11 function.

2.4.2 ICU11 influences the GA pathway in *Arabidopsis thaliana*

This study indicated an influence of ICU11 on the GA pathway in *Arabidopsis thaliana*, At least *icu11-3* showed a significantly higher germination rate to the GA-inhibitor PAC in comparison to Col-0. Additionally, a 5- to 6-fold increase of one main GA synthesis gene Ga20ox2, known to play a major role in floral development, was detected[85]. Focusing on the phenotype of all three *icu11* alleles, similarities to a GA-overdose phenotype were observed, starting with early flowering and elongated hypocotyls and internodes[74,103]. These GA-overdose phenotypes were found to be caused by increased GA20ox1, GA20ox2 or GA20ox3 expression, fitting to the previously described results of this study. A closer look to the regulation of GA-biosynthesis reveals an additional accordance regarding the CYP714A1 and CYP714A2 proteins, as were shown to catalyze a recently discovered GA-inactivation mechanism[106]. Although these enzymes have different functions – such as inactivating non-13-hydroxy GAs for CYP714A1 and the production of GA_1 via C13-hydroxylation for CYP714A2, both participate in the interplay of determining the ratio of GA_1 and GA_4 due to their competition for non-13 hydroxy substrates. The phenotype of the *cyp714a1 cyp714A2* double mutant was described with an overall increase of biomass about 30-50%, demonstrated by enlarged cotyledons and rosettes, increased stem height and larger petals, bigger seeds and earlier flowering compared to their wild-type[84,105]. This list perfectly overlaps with the phenotype of the *icu11* mutants. A possible role of ICU11 in regulating the balance of GA_1 and GA_4 could also provide another explanation for the similar *icu11* phenotype independent of overexpressing or LOF mutations. Thus, overexpression of either the non-13-hydroxy or 13-hydroxy GA pathway would cause an increase in overall GA

content and therefore could lead to a GA-overdose phenotype. A somewhat reduced non-13-hydroxy or the 13-hydroxy GA-synthesis pathway, on the other hand, might be responsible for a shift towards an increase of the other pathway to ensure the balancing of a fixed ratio between GA_1 and GA_4, as it is known for rice[84], causing as well an increase in overall bioactive GA content. Previously published data showed a 6.29-fold increase of CYP714A1 in the *icu11-1* mutant[11], supporting the impression that ICU11 plays a role in the regulating mechanism these two cytochrome P450 monooxygenases are responsible for. In general, the cytosolic localization of ICU11-YFP allows the assumption that ICU11 might influence the third step of GA-biosynthesis (localized in the cytosol) in a direct manner. The published RNA-Seq data of the Mateo-Bonmatí lab, present typical DELLA-regulated genes, such as AP1, FUL, SNZ, FT, SPT, SPL3, ALC, SHP1&2 amongst the most differentially expressed genes[11], reducing flowering time, elongating hypocotyl cells and increasing cotyledon cell size according to their expression pattern in *icu11-1*. This too is supporting an impact of ICU11 on the GA pathway. To test if GA content is significantly changed in the *icu11* mutants compared to their wild-types, a Liquid chromatography/Mass spectrometry (LC/MS) approach might be worth to examine. Moreover, an analysis of the given ratios of GA_1 and GA_4 amount would help to exclude or to confirm the assumptions made so far.

Still, it has to be kept in mind that the observed influence of ICU11 on the GA-pathway might only be a side effect of its true molecular function, since one major argument is conflicting with this theory. ICU11 is highly conserved in metazoa up to humans[11], which is not to be expected for a phytohormone-linked gene and is accordingly not the case for other GA-pathway genes[75,81]. A further contradictory result might be the localization study, because ICU11 was published to be located in nuclei[11]. A possible explanation for these different observations is that protein localizations can vary depending on present growth conditions. Furthermore, a nucleic localization could not be completely excluded in this study and a weak nucleic localization could still be sufficient to fulfil a function in the nucleus. According to the model proposed in chapter 2.4.1, an overexpression of ICU11 might lead to a non-functional protein and therefore an accumulation in the cytosol. A cytosolic expression is additionally contradicting to the molecular function as epigenetic repressor involving histone modification, as assumed by the Micol Lab.

However, independent of the complete discovery of the molecular function of ICU11, mutations in the *ICU11* gene might be of great value for research interested in increased crop biomass, since it might lead to similar phenotypes in other species. Therefore, ICU11 could provide a new regulatory point for organ growth of crops, addressing the increasing demand for nutrition of a growing world population.

2.4.3 Recent study reveal ICU11 function

After the lab work for this project was finished, a further study of ICU11 was published[185], showing that ICU11 is physically associated with the Polycomb Repressive Complex 2. It was proven via immunoprecipitation that ICU11 interacts with EMBRYONIC FLOWER 1 (EMF1), LIKE HETEROCHROMATIN PROTEIN1 (LHP1), and TELOMERE_REPEAT_BINDING FACTORS (TRBs). Moreover, ICU11 was shown to act as a regulator for the epigenetic switching of FLC chromatin between the active H3K36me3 state and the silenced H3K27me3 state fundamental for the cold-induced PcG-mediated silencing during the vernalization process.

Thus, the observed link to the GA-pathway raises the question, if ICU11 might also affect the chromatin state of GA-biosynthesis genes, such as GA20ox2. Pleiotropic functions have been observed also for 2OGDs acting as human demethylases, such as JMJD6[183], which is involved in the development of cancer and thus related to growth via directly interacting with RNA[186]. Recently, the regulation of plant growth was shown to be directed by a complex interplay of a flexible chromatin structure, phytohormone signaling and gene expression reprogramming[187]. Thus, ICU11 might be an additional factor acting within this crosstalk.

3. Chapter transition to Amsinckia project

The first project of this thesis developed into a challenging project. The lack of coherent explanations for the observed results, which often contradicted previously published results, were a recurrent theme during the complete working time. According to Thomas Henry Huxley, it was decided to start a second project "*Amsinckia*" in hope for a better outcome.

4. *Amsinckia*, a peculiar case of self-compatible heterostyly and its repeated transition to selfing

4.1 Introduction

As a consequence of their sessile lifestyle, plants are immobile and lack the mobility of virtually all animals[188]. Thus, an active search for a suitable partner is out of range for plants, which lead to the development of different mechanisms to improve reproductive success. One of these mechanisms is the development of hermaphroditism, which unites both sexes within the same flower[189] (Fig. 32B) and thus allows the possibility for selfing. Besides the advantages of self-fertilization regarding the absence of mating partners, selfing can also be favoured in environments without pollinators and can therefore lead to improved colonization ability[190]. On the contrary, self-compatibility has been proposed as evolutionary "dead-end", causing a doubled or quadrupled species extinction rate[190]. Thus, selfing in hermaphroditic flowers is often found to be counter-regulated by self-incompatibility (SI). SI, found in approximately 40% of flowering plant species of at least 100 different families[191], ensures outbreeding capacity and therefore genetic diversity. This argument is supported by studies that demonstrated inbred offspring being less fit than outbred offspring[192]. Thus, the resulting inbreeding depression is one of the major selective forces influencing plant mating.

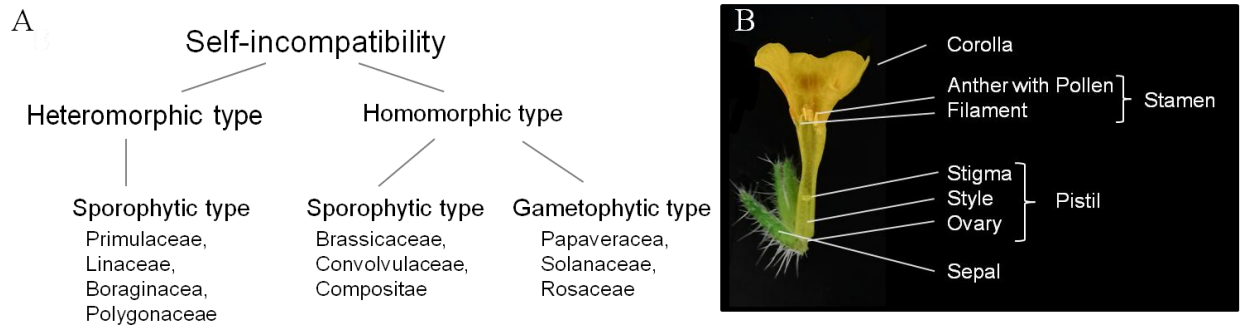


Fig. 32 Overview of the self-incompatibility types and flower anatomy of a heteromorphic species

(A) Classification of self-incompatibility types and typical plant families that show respective self-incompatibility types **(B)** Flower anatomy of *Amsinckia spectabilis* (*Boraginaceae*); note the separation of the sexual organs

Within plant kingdom, different mechanisms of SI can be found (Fig. 32A). Lewis classified two main types of SI, the heteromorphic and the homomorphic type[193]. Whereas the heteromorphic SI is characterized by two (distyly) or three (tristyly) different flower morphologies (morphs) within one species - mostly allowing successful pollination only between different morphs -, no obvious difference between the flowers of homomorphic systems can be detected. In the following, homomorphic and heteromorphic SI will be separately explained in more detail.

4.1.2 Homomorphic self-incompatibility

Homomorphic SI is widely distributed in approximately 50% of all angiosperms[17].To ensure SI in homomorphic species, two ways of genetic control of the pollen phenotype determine male and female compatibility. Thus, the homomorphic SI is further grouped into gametophytic (haploid generation) and sporophytic (diploid generation) SI. Gametophytic SI (GSI) is represented by the genotype of the pollen itself, whereas the genotype of the pollen parental plant (e.g. anther tapetum (nutritive cell layer of anthers)) determines the sporophytic SI (SSI). Regarding the genetic control of homomorphic SI, the very first published scientists, analyzing the phenomenon, Lewis and Bateman[193,194], developed the theory of at least two tightly linked genes, expressing the male and female determinants per individual and exhibiting a range of different alleles within the same species, the so called *S*-locus. Thanks to new techniques allowing molecular insights, recent studies confirmed this theory by showing that the *S*-locus of many species is represented by a single polymorphic locus. Although the uncovered *S*-loci of different species revealed different SI-mechanisms, represented by different kinds of included *S*-genes, one common pattern could be detected in all of them. Thus, each *S*-locus consists of at least two strongly linked transcriptional units, consisting of the female and the male *S*-determinant. Within one species, variants of these *S*-determinant units, the so called *S*-haplotypes, are present. The self/nonself-recognition is based on protein-protein interactions between the male and female determinant of the same *S*-haplotype. Thus, a pollen-pistil combination of both carrying the same *S*-determinants (self) will lead to rejection, whereas a combination derived from different (non-self) *S*-haplotypes will allow fertilization[195]. The pollen self-recognition of GSI species causes the arrest of the self-pollen tube growth. In contrast, in SSI species pollen hydration is inhibited (Fig. 33). In SSI (diploid-dependant), each pollen grain has two different *S*-haplotypes, that are expressed in the tapetal cells, which would be rejected by

pistils carrying either one or the other *S*-haplotype, in case of codominance between the present *S*-haplotypes. Thus, at least four different *S*-haplotypes must be present in a population to allow the finding of a mating partner. According to this, a given population with four codominant *S*-haplotypes S_1, S_2, S_3 and S_4, would only allow one compatible match between S_1S_2 and S_3S_4, whereas S_1S_2 pollen would be rejected by S_1S_3 as well as S_1S_4 pistils, representing a strong mating restriction and thus leading to mate limitation. Allelic interactions of all present *S*-haplotypes, are of high importance within populations. Therefore, a dominance hierarchy, such as the dominance of S_2 over S_1, which would allow pollen with S_1S_2 tapetal cells to fertilize S_1S_3 pistils, is counter acting this effect in SSI species, as it was shown for *Brassica*[196]. Further studies detected additional relationships between different *S*-haplotypes, like codominance and dominance hierarchies that differ between pistils and pollen, and the fact that dominance relationships are nonlinear[197].

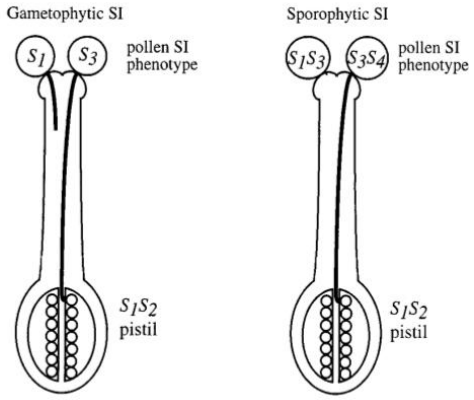


Fig. 33 Illustration of the homomorphic gametophytic and sporophytic self-incompatibility types (modified after [12])

The pollen SI phenotype of the gametophytic type is determined by the *S*-haplotype of the haploid pollen genome so that each pollen grain carries either the paternal or the maternal *S*-allele. For the sporophytic type, the pollen SI phenotype is determined by the *S*-genotype of its diploid parent, so that each pollen grain carries the determinants of two *S*-alleles. In both types, matching of the *S*-haplotypes of pollen and pistil tissue result in rejection of the pollen. As a special case for the sporophytic type, dominance (S1 is dominant over S3) behaviour of different *S*-alleles can lead to reinforced or weakened rejection probability.

Many *Brassicaceae* species served as classical examples of SSI, such as *B. rapa* (30 identified *S*-haplotypes[198]), *B. napus*, *B. oleracea* (50 identified *S*-haplotypes[199]) and *Arabidopsis lyrata*, which allowed the uncovering of the molecular mechanism of SI-response in the *Brassica*

family. Here, the *S*-locus consists of three genes, *S-LOCUS RECEPTOR KINASE* (*SRK*), *S-LOCUS PROTEN 11* (*SP11*) or *S-LOCUS CYSTEINE RICH* (*SCR*) and *S-LOCUS GLYCOPROTEIN* (*SLG*), with SRK as female determinant and SP11/SCR as male determinant. SRK is highly expressed in the plasma membrane of the stigma papilla cells, whereas SP11 is particularly found in the anther tapetum, accumulating in the pollen coat during pollen maturation. During pollination, self-SP11 penetrates the papilla cell wall and binds SRK, followed by induction of autophosphorylation of SRK, which in parallel acts as a trigger for a signalling cascade leading to the rejection of self-pollen via a not completely revealed mechanism (Fig. 34). Nonself-SP11 is not able to bind to the present SRK, thus the pollen germination is not inhibited and fertilization may follow. SLG was shown to act as a supporter of the SI-response in some *S*-haplotypes[200–202].

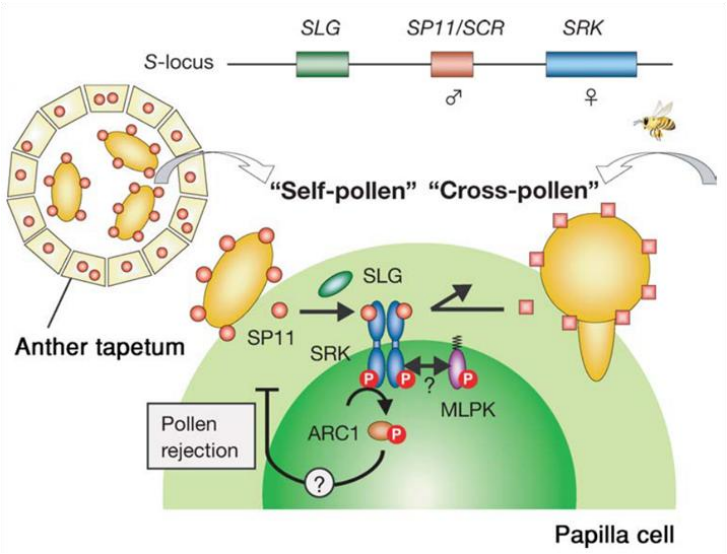


Fig. 34 Self-recognition SI system in *Brassicaceae* (modified after [13])

The *S*-locus encodes female and male *S*-determinants, SRK and SP11/SCR, respectively. Self-pollination is followed by the self-specific SP11 binding to its respective SRK, resulting in the stabilization of active SRK dimeric form positioned in the plasma membrane, that leads to SI-response mediated self-pollen rejection.

Further molecular characterization of SI occurred in two other families, *Papaveraceae* and *Solanaceae*, both exhibiting gametophytic SI. The SI-response in *Papaveraceae* consists of two genes, PAPAVER RHOEAS STYLE S (PrsS) and PAPAVER RHOEAS POLLEN S (PrpS), both co-evolving polymorphic proteins[203]. Here, only the molecular characteristics of the female determinant is well analysed. PrsS encodes a secreted stigma protein and is expressed by papilla

cells of the stigma. Although the exact molecular mechanism of self-rejection is still unclear, it was shown by application of recombinant PrsS to self-pollen, that this mechanisms includes a Ca2+ influx into outgrown tubes, immediately after germination. Ca2+ acts as trigger for a signal transduction finally leading to the depolymerization of actin and microtubules and amongst others to DNA fragmentation, both leading to growth inhibition and apoptosis[204]. The *Papaveraceae* male determinant PrpS was assumed to act as pollen surface receptor directly interacting with PrsS[205] (Fig. 35).

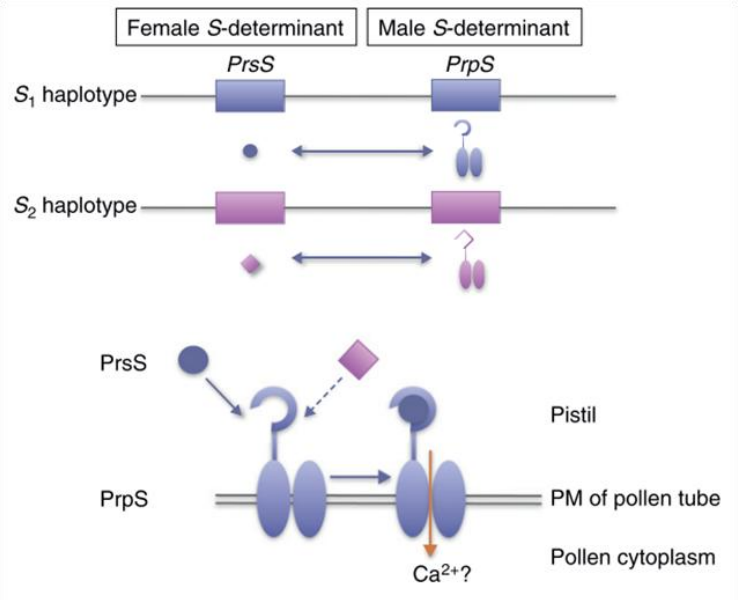


Fig. 35 Self-recognition SI system in *Papaveraceae* (modified after [14])

The *S*-locus encodes female and male S-determinants, PrsS and PrpS, respectively. Self-pollination is followed by the self-specific PrsS binding to its respective PrpS on the pollen plasma membrane, resulting in a Ca^{2+} influx in the pollen tube that leads to SI-response-mediated programmed cell death.

The SI-mechanism in *Solanaceae*, *Rosaceae* and *Scrophulariaceae* is maintaned by two *S*-locus genes *S*-RNase and *S*-LOCUS F-BOX BROTHER GENES (SLF/SFB), the first representing the female determinant and the latter the male determinant. The fact that here the SI-response is based on nonself-recognition, but not self-recognition, displays the main difference to the SI-systems of *Brassicacea*e and *Papaveraceae*[206]. *S*-RNases are highly expressed in the extracellular matrix of the style and enter each growing pollen tube. Nonself-SFLs binds to *S*-RNases and causes its detoxification via F-box protein activity of SLF, that acts as substrate-recognition subunit of the E3 ubiquitin ligase SCF (Skp1-Cullin1-F-box) complex. In self-pollen

tubes, the SLFs are not able to bind and therefore inactivate S-RNase, so that its cytotoxic ribonuclease activity degradates the pollen RNA and therefore inhibits further pollen tube growth[203](Fig. 36).

To ensure the function of this nonself-recognition system, only one present *SLF* gene per *S*-locus is not enough. Thus, recent studies found proof for the previously assumed collaborative non-self-recognition model, where several *SLF* genes are linked to one *S*-RNase within one *S*-haplotype. Therefore, 16-20 *SLF* genes per *S*-locus were found in self-incompatible *Petunia* species[207,208], that are enabled to interact with one or more *S*-RNases of other *S*-haplotypes.

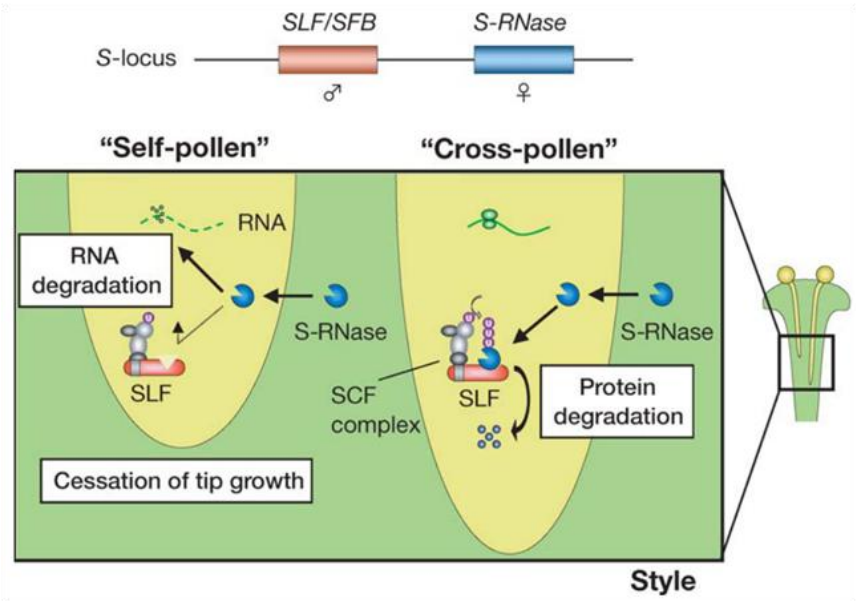


Fig. 36 Nonself-recognition SI system in *Solanaceae* (modified after [13])

The *S*-locus encodes a single female and several male *S*-determinants, *S*-RNase and SLF/SFBs, respectively. In the case of self-pollination, an interaction between self-SLFs and the respective *S*-RNase is absent. The active *S*-RNase, degrades the pollen RNA, leading to its growth inhibition. Non-self SLFs degrade the *S*-RNase, so that pollen RNA stays untouched and pollen tube growth can proceed.

Based on the nonself-recognition SI system several possible scenarios for the evolutionary break down of SI, such as the acquisition of a new SLF that detoxifies the self S-RNase[206], were developed. Moreover, it was demonstrated that in many gametophytic SI species the development of self-compatible (SC) species, either induced or as a natural variant, occured via polyploidization[15]. In tetraploids, the *S*-RNase based SI allows the increase of different SLFs and therefore the recogniton of all present *S*-RNases via competitive interaction, thus leading to non-

inhibited pollen tube growth. Here, the genetic background is based on the unreduced pollen of diploids, containing two expressed *S*-alleles. Especially heterozygous pollen (e.g. S_1S_2) possess SLFs that detoxify both *S*-RNases of the S_1 and S_2-alleles of the maternal genotype ($S_1S_1S_2S_2$ in any possible combination) and therefore represents a selfcompatible cross[209](Fig. 37).

A closer look to the *Solanaceae* family suggests, that polyploidization always occurs simultaneously with the development of SC, since most analyzed polyploid species are self-compatible[209]. *Rosaceae* on the other hand, provide examples which show only weak SC induced by polyploidization, as it was found in apple[210], but also species, such as sour cherry (*Prunus cerasus* L.), that contradict this theory, while possessing SI despite of genome duplication. Lewis explained this observation by a dominance relationship or the abscence of interaction of the two pollen *S*-haplotypes, resulting in a similar behaviour as haploid pollen[211]. Additionally, recent studies detected further mutations of the *S*-locus that allowed SC in *Rosaceae*[212,213].

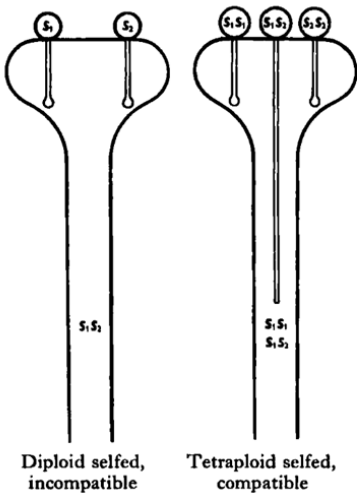


Fig. 37 Self-incompatibility behaviour of diploid and tetraploid plants (modified after [15])

Diploid pollen (containing two *S*-alleles) of tetraploid plants increases the amount of degradable *S*-RNases due to the possible increase of different encoded SLFs. Thus, heterozygous pollen allows selfing via the detoxification of both present *S*-RNases in the style.

4.1.3 Heteromorphic self-incompatibility

Next to SI-mechanisms that exclusively happen "inside" of the flower organs, evolution found an alternative way to promote SI via affecting the overall flower morphology (herkogamy) and their intramorph compatibility. Both together provide the basal ideas of heteromorphic SI. The herkogamous sexual polymorphisms of flowers within one species that differ reciprocally in their relative positions of pistils and stamens already attracted Charles Darwin. As a result, his book "The different forms on flowers on plants of the same species" (1877&1884), represents until today the base for all research that has been done in this field[15,194,214,215]. Besides enantiostyly (i.e. mirror-imaged plants with left-styled and right-styled flowers within one individual or within a population), heterostyly is the most analyzed type of reciprocal herkogamy. Whereas the homomorphic SI allows the presence of several dozens of different *S*-haplotypes within one population, the stylar polymorphism in heteromorphic SI allows only two or three different polymorphism within one population. Thus, populations of heterostylous species consist of either two (distyly) or three (tristyly) floral morphs, that differ reciprocally in their position of the sexual organs, meaning that one morph possesses a short style but high-positioned anthers (thrum flowers or S (short-stylous)-morphs), whereas the other morph exhibits the reciprocal pattern with a long style but low-positioned anthers (pin flowers or L (long-stylous)-morphs)[193]. Tristyly shows a similar pattern, where an additional mid-morph (M-morph) is included. Further phenotypes are often found to be linked to heterostyly, which are morph-dependant differences of pollen size, exine structure as well as stigmatic papillae length. All together describe the so called heterostylous syndrome (Fig. 38A)[216].

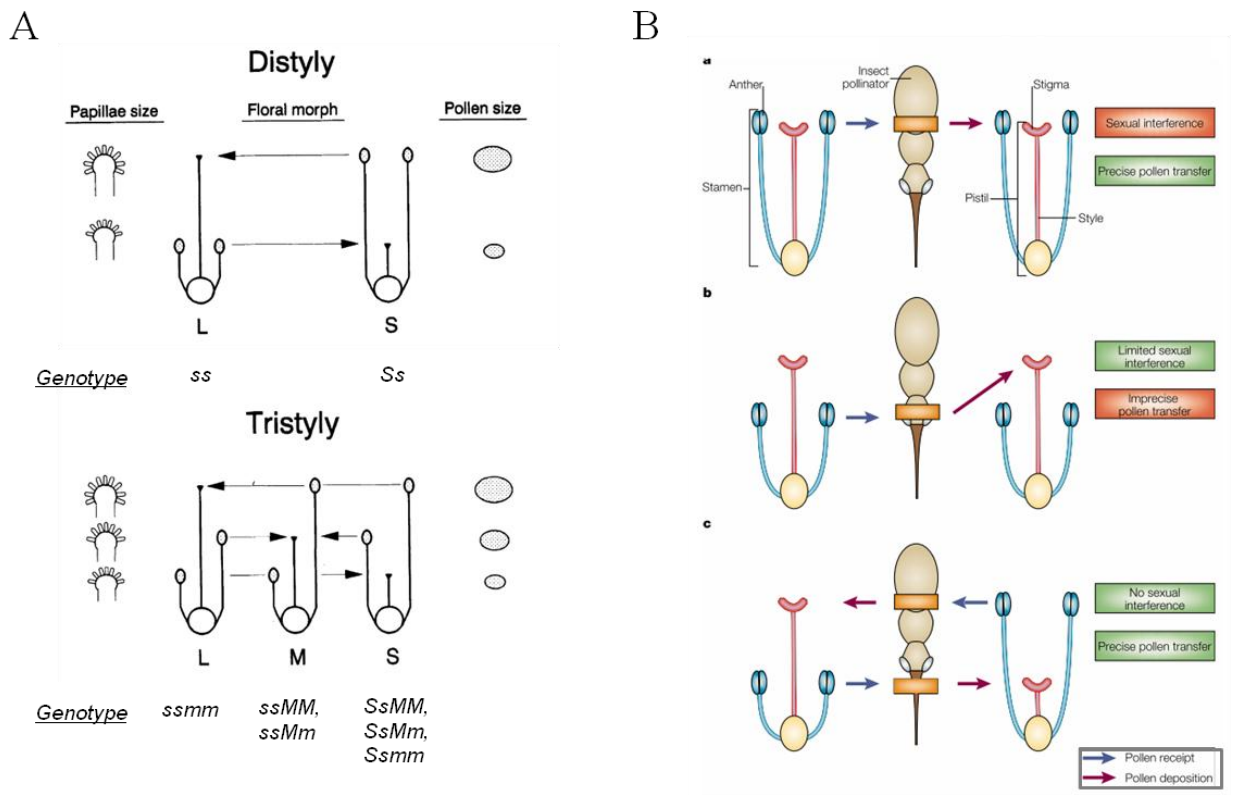


Fig. 38 Overview of basal theories of heterostyly (modified after [16–18])

(A) Heterostylous syndrome. Compatible pollination is indicated by arrows. All other pistil-pollen combinations are assumed to yield in less or no seed set. Genotypes of the floral morphs with the most common patterns of inheritance are indicated below. **(B)** Illustration of basal principle of "Cross promotion theory" explained by pollen transfer of different floral designs via pollinators. Thus, separated positioning of sexual organs allows pollen deposition on different parts of the pollinator's body and transferring it preferentially to similar positioned styles. Costs are shown in red boxes, benefits in green. (a) equivalent height of anthers and stigmas (b) one floral morph with separated sexual organs (herkogamy) (c) population with two floral morphs with reciprocally positioned sexual organs (distyly). Note that (c) provides the best mechanism regarding low chances for sexual interference and most precise pollen transfer.

Heterostyly has evolved independently in at least 28 animal-pollinated angiosperm families (convergent evolution) and was assumed to be responsible for the promotion of cross-pollination and outbreeding[217]. The dependency on animal pollination simultaneously is the reason for the previously mentioned limited maximum of three stylar polymorphisms, as it will be explained in the following.

Darwin assumed that the reciprocal positioning of the sexual organs promotes legitimate (nonself) pollination directed via animal-pollinators (Fig. 38B). Specifically, pollinators transfer

the pollen from one floral morph to the other way more precisely than within the same morph. Thus, low-positioned pollen from L-morphs is efficiently transferred to the S-morph style via the head of the pollinator and reciprocally via the abdomen of the pollinator. According to this, the probability of loading illegitimate self- or intramorph-pollen on one stigma is strongly reduced. Therefore, the dependency on the small size of the pollinator body and the limited number of clearly separated body sites explains the low number of floral morphs in heterostyly. Accordingly, Darwin's "cross promotion hypothesis" says that due to heterostyly, intermorph pollination is enhanced compared to intramorph pollination, indicating that heterostyly acts as an "outbreeding mechanism"[18]. Taking into consideration that most heterostylous species possess additional SI via intramorph incompatibility, which also leads to at least strongly reduced self-fertilization and guarantees outcrossing, as well as reduces mating with one half (distyly) or one third (tristyly) of the individuals within a population, Darwin's hypothesis seems rather incomplete. More precise investigation regarding the dependency of fitness of the offspring from male or female functioning resulted in a clearer explanation. Reciprocal herkogamy increases the male component of fitness via reducing male gamete wastage on incompatible stigmas, whereas intramorph incompatibility prevents unnecessary costs on the maternal side by protecting of self-fertilization[18]. Thus, precise pollen transfer without the costs for sexual interference and self-pollination is the main achievement of heterostyly compared to monomorphic animal-pollinated species [216].

4.1.3.1 Genetic models of the heteromorphic self-incompatibility locus

More than a century ago, Bateson and Gregory showed, via inter- and intramorph crossings of the heterostylous *Primula sinensis*, that the genetics of heterostyly follows Mendel's laws on the basis of a single locus, called *S*[218]. Thus, they demonstrated that L- intramorph crosses always lead to a 100% L-morph F1 generation, whereas S-intramorph crosses resulted in a 3:1 S-morph:L-morph ratio, as would be expected for a heterozygous situation at the *S*-locus (*S*/*s*) in the S-morphs and as well for a homozygous but recessive situation for the L-morphs (*s*/*s*) (Fig. 38A). Intermorph crossings, on the other hand, combined with SI in the analyzed species, resulted in a 1:1 ratio of S-morphs and L-morphs, leading to the conclusion that homozygous dominant *P. sinensis* plants (*S*/*S*) do never exist. Apart from few dominance reversed exceptions[219], the results of Bateson's study gave rise for more than 100 years of research, first in hope for uncovering the genetic architecture of the *S*-locus, and second for resolving the basis of

the convergent evolution of heterostyly, its link to SI and if same or different mechanisms, structures, as well as included genes can be found in the 28 heterostylous families.

The first insights into the genetic architecture of the *S*-locus was given by Alfred Ernst[220], who found also homostylous individuals in large populations of *Primula hortensis* and *Primula viscosa,* possessing either high anther position and long styles or low anther positions and short styles. Further some individuals exhibit pollen sizes that were unlinked to the anther height, meaning small sized pollen were found on high-positioned anthers and the other way around. However, self-incompatibility was always linked to pollen size and style length. These observations allowed him to build up a first model for the heterostylous S-locus in *Primula* consisting of at least three tightly linked genes, named *G*, *A* and *P* locus. Here, the *G*-locus encodes for style length, the *A*-locus for anther height and the *P*-locus for pollen size. Thus, a dominant *G*-locus leads to a short style, and the dominant *A* locus to high positioned anthers. Therefore, the S-morph possesses a *GPA/gpa* genotype, whereas the L-morph displays a *gpa/gpa* genotype. This model fulfils the characteristics of a supergene, meaning a chromosomal cluster of tightly linked loci that are responsible for the phenotypic expression of different behavioural or developmental characteristics, here the formation of the two morphs[221]. Regarding the occurrence of homostyles (*gPA* or *Gpa*) two different theories were developed. Whereas Ernst supposed mutations within the *S*-locus genes to be causal, Dowrick argued for rare recombination events at the *S*-locus as basal mechanism[222]. Over the time, additional genes (up to seven in total) were discussed to be included in the *S*-locus, determining male and female self-incompatibility due to the assumption that a pleiotropic effect – such as combining style length and female incompatibility encoded by a single gene – seemed rather unlikely considering physiology[222]. Finally, Lewis and Jones supported the supergene model for the heterostylous *S*-locus proposed from Alfred Ernst, with few add-ons. Accordingly, they assumed a diallelic *S*-locus, maintained via strongly suppressed recombination and thus most likely with a centromeric localization. Further, homostyles and other unusual phenotypes were proposed to be originated by rare recombination events[223].Up to this, the confirmation that the supergene model also holds for the *S*-locus of heterostylous species unrelated to *Primula,* was still missing. Referring to two studies from the beginning of the 21st century, it was proven that the supergene model also fits to other species, such as buckwheat and *Turnera subulata.* In buckwheat (*Fagopyrum*) also homostylous (*Sh*) variants occurred in natural populations. Additionally, similar dominance

relationships of the different *S*-alleles as in *Primula* (*S* > *Sh* > *s*) could be shown via interspecific crosses of a distylous forms with a homostylous form, suggesting that the long-homostylous allele (*gPA*) was derived from a recombination event at the *S*-locus[224]. Further for *Turnera subulata*, x-ray treatment of *S*-pollen, inducing deletions at the *S*-locus, led to few long-homostylous F1 offspring after crossing to L-morphs. Thus, also in *Turnera subulata* at least two different loci for anther height and style length are parts of the *S*-locus supergene[225].

Besides these confirmed *S*-locus characteristics in different heterostylous species, the disassortative mating of S- and L-morphs causes the occurrence of the dominant *S* allele only in a heterozygous state and thus ensures a 1:1 morph ratio in heterostylous populations. This leads to the prediction of additional genetic *S*-locus properties, such as high sequence diversity caused by a greater effect of genetic drift as well as an accumulation of transposons and repetitive sequences[18,19].

Regarding the evolution of the heterostyly supergene, two different models have been developed, the first by Charlesworth and Charlesworth[226] and the second by Lloyd and Webb[223]. The Charlesworth model is based on a self-compatible homostylous ancestor with high levels of inbreeding depression that developed at first a new mutation, leading to a pollen type that is self-incompatible and shortly after a second female mutation causing compatibility with the new pollen type. This diallelic SI system is thus the first part to evolve on the way to the heterostylous SI, which would be followed by further mutations leading to the reciprocal herkogamy of the two morphs. Here, it was supposed that a dominant mutation led to the short-styled phenotype, followed by a recessive mutation, deciding for high anther positioning. These morphological differences would enhance compatible, thus intermorph, animal pollination and therefore selecting for the presence of both morphs in the populations at equal ratio, under the condition that both style and anther mutations were tightly linked to the SI locus.

The Lloyd and Webb model, on the other hand, argues from the starting point that most families, in which heterostylous species occur, possess approach herkogamy, meaning a longer style than anther height. Thus, this is assumed to be the base for the evolution of reciprocal herkogamy which was introduced by a dominant mutation leading to a short style and causing style polymorphism within the population. Also here, a preference for intermorph compared to intramorph pollination would support a second mutation causing higher anther position

combined with short styles, leading also to an anther–height polymorphism. Finally, diallelic SI could have evolved via advantageous adaptations of pollen regarding their growth through different style types. Additionally, a linkage between the morphological genes and the SI genes was assumed as basal mechanism, but also not morph-linked genes that might arise anywhere in the genome, could lead to advantageous morph-specific pollen behaviour, and could therefore be favoured by selection[18,19].

The main argument against the Charlesworth model is the unlikely ancestor state with high levels of inbreeding depression due to SC and the evolution of heterostylous SI to escape via outbreeding. Hence, as many homostylous forms (mostly showing high selfing efficiency) are derived from heterostylous and non-herkogamous forms, they strongly contradict this argument. Furthermore, there is a wide range of SI types that can be found in heterostylous species, such as cryptic self-incompatibility[227] and differences in the strength of SI, as well as no SI at all[17,18]. That is another fact that rather supports the Lloyd and Webb model.

4.1.3.2 Evolution of homostyly

In several heterostylous species, the breakdown of heterostyly and, if present, SI, occurred via the transition to fully selfing homostylous forms, such as *Amsinckia*[228], *Eichhornia*[229], and *Turnera*[230], three genera, where a distylous ancestor was proven via phylogenetic studies. According to the floral morph, the homostylous variant is derived from short-homostylous and long-homostylous variants, with equal anther and stigma height. In some families, as in *Primula*, these transitions are associated with polyploidization, which is not an indispensable characteristic of homostyly. Others possess different or additional phenotypes, such as smaller flowers promoting selfing efficiency. But also autonomous selfing itself is not inevitably linked to homostyly, as it was shown for two natural occurring homostylous species of *Turnera*, that either exhibit large flowers and retain herkogamy[231] and residual diallelic SI[232]. Taken these facts together, it is likely that evolution found different ways to develop the transition to homostyly. In general, selfing homostylous populations develop within biogeographical patterns, such as spreading to regions, where reliable pollinators cannot be found, thus leading to high selective pressure on floral morph variants, that assure the reproductive success the most[18,233].

4.1.3.3 Unravelling the genetic architecture of the heteromorphic *S*-locus

Particularly, the last decade brought a great deal of new findings regarding the genetic architecture of the heterostylous *S*-locus of different species. Since similar patterns were found, the basal characteristics will be explained for *Primula* in the following (Fig. 39). First studies found morph-specifically expressed genes allowing the development of two *S*-linked but not *S*-gene-markers[234], that set the starting point for unravelling the genetic architecture of the *S*-locus in *Primula*. In 2016, Huu[235] identified *CYP734A50* to be the causal gene for the *G*-locus by showing its exclusive expression in S-styles. Additionally he confirmed the absence of this gene in long homostyles or strongly reduced expression, caused by a mutation leading to an amino acid exchange and a deletion, respectively, as well as presence in all S-morphs of several heterostylous *Primula* species. Further, virus-induced-gene-silencing (VIGS) of *CYP734A50* in *P. forbesii* resulted in long homostyles. Additionally, the molecular mechanism of how CYP734A50 inhibits style growth could be discovered. Based on homologous CYP450 genes, found in other plant species, such as tomato and *A. thaliana*[236,237], all of them possessing BR inactivating functions, BR levels of S- and L-morph styles where measured. Results exhibit strongly increased BR levels in L-morph styles, indicating that the short style morph is caused by low BR levels. Further experiments detected an L-morph-specific style cell size increase that was attributed to the BR concentration, since exogenous treatment of S-morphs with BR allowed rescue of the short S-morph styles, leading to long homostylous variants. Regarding self-incompatibility, the presence of naturally occurring long homostylous *P. vulgaris* variants, which show strongly reduced CYP734A50 expression and loss of SI, argues for the fact that a single gene indeed controlls style length and female incompatibility. In the same year, a second study published assemblies of the dominant *S*- and the recessive *s*-allele sequence based on BAC contigs[238,239]. Here, the basal characteristic regarding the genetic architecture of the *S*-locus in *Primula* was uncovered, which is the hemizygosity of the *S*-locus. Thus, this region comprises an insertion of around 280 kb only present on the dominant *S*-allele, which prevents any recombination by crossing-over within this region. In *Primula*, this region was found to contain five predicted genes: *GLOBOSA 2* (*GLO2*), a *CYTOCHROME P450* (*CYP734A50*), a *KELCH-REPEAT F-BOX PROTEIN* (*KFB*), a *PUMILIO-LIKE RNA-BINDING PROTEIN* (*PUM*) as well as a gene encoding for a protein with a highly conserved C-terminal domain (*CCM*)[240]. Compared to the L-morph possessing one copy of a *CYCLIN-LIKE F-BOX GENE* (*CFB*), the

inserted region of the dominant *S*-chromosome is flanked by one copy of *CFB* on either side. Further, *GLO2* was assumed to be the causal genes for the *A*-locus in *Primula*, due to the observation of short homostyles carrying a transposons insertion in the *GLO2* gene. On a later study from Huu et al, confirmed *GLO2* to encode the *A*-locus, demonstrated by VIGS of *GLO2* in *P. forbesii*. Here, silencing resulted in low anther positioning and thus leading to short homostylous flowers derived from S-morph individuals[241]. Further, no difference in pollen size or male incompatibility of the VIGS plants compared to untreated plants could be detected. Additionally, a difference in cell size underlies the morph-specific anther phenotype, due to the fact that cell elongation of the corolla is promoted beneath the inserted position of the anthers compared to the cells above. However, the molecular mechanism that causes cell elongation remains unclear until now.

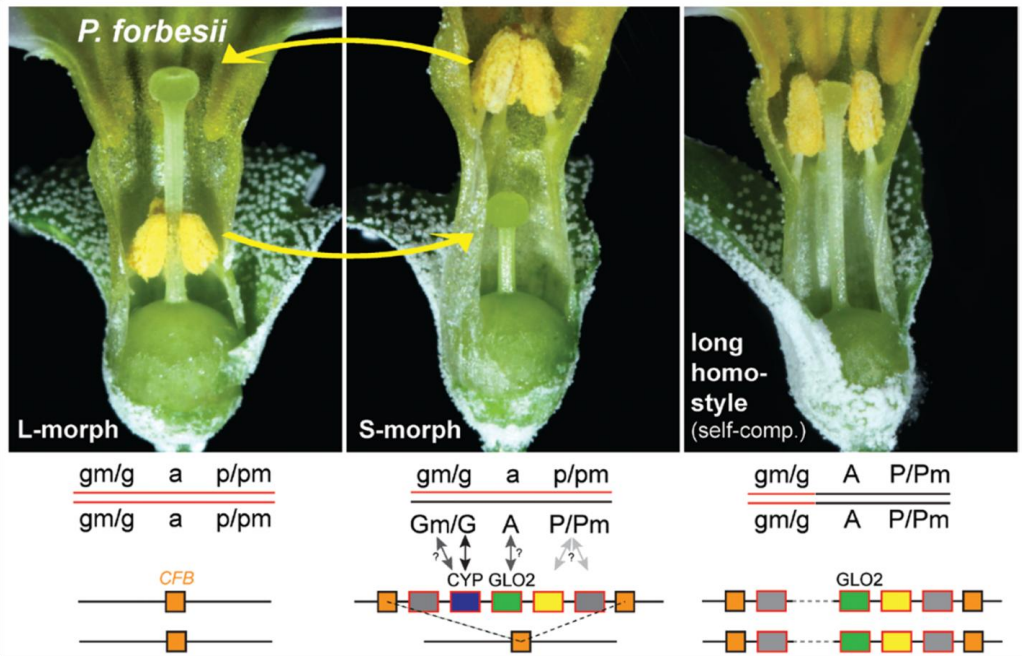


Fig. 39 Overview of molecular and genetic basis of heterostyly in *Primula* (modified after [19])

Different anther-pistil positioning of L-morph, S-morph and long homostyle (from left to right) of *Primula forbesii* are shown. Arrows indicate compatible crosses. Genetic architecture of the diallelic *S*-locus is represented below the pictures according to the Ernst model[220], including the loci for style length (*G*), female incompatibility (*Gm*), anther height (*A*), pollen size (*P*) and male incompatibility (*Pm*). Capital letters indicate dominant alleles over small lettered recessive ones. The five *Primula S*-locus genes (coloured boxes) are included further below. Double-headed arrows indicate links of genes to specific loci, different gray shades refer to different strength of proof regarding the link.

The discovery of the hemizygosity of the *S*-locus simultaneously implies the base for the evolution of homostylous forms. Since a hemizygous chromosomal region is excluded from

homologous recombination, it is unlikely that homostyly developed via rare recombination events but rather via mutations of the *G*- or the *A*-locus[19,235,238], as originally proposed by Ernst[220].

Additionally, genetic mapping and phylogenetic analysis of the *GLO2* study allowed the suggestion that *CYP734A50*, and *GLO2*, most likely have derived from a step-wise duplication, rather than a segmental duplication, since the ancestral *CYP734A* and *GLO* paralogues (*CYP734A51* and *GLO1*) were shown to be unlinked. Thus, a segmental duplication would be based on a structural rearrangement linking both loci and unlinking them afterwards. Phylogenetic analysis further supported the step-wise duplication theory by implying *CYP734A* to be the first locus that duplicated, followed by the recent duplication of *GLO1*. *GLO2* (and possibly also *CYP734A50*) underwent a neofunctionalization most likely on protein level, since *GLO1* acts as B-class homeotic gene being responsible for the identity of petal and stamen. These findings support both, the Charlesworth model and the Lloyd and Webb model regarding the order of the origin of the *S*-locus genes, proposing the *G*-locus polymorphism to be the first that arose followed by the *A*-locus.

Matching patterns of the characteristics of the *Primula S*-locus were found in several heterostylous species, such as *Linum grandiflorum*, where a thrum-style specific gene (TSS1) was detected and a hemizygous *S*-locus was assumed[242]. In three *Fagopyrum* species *S*-LOCUS EARLY FLOWERING 3 (S-ELF3) was found to be exclusively expressed in short styles and only present in S-morph DNA, assuming also here a hemizygous structure of the *S*-locus. Disruption of this gene in two self-compatible species led to long homostyles and further argues for S-ELF3 to be responsible for the *G*-locus[243]. As a special case, short homostyly in *Fagopyrum esculentum* was found to be caused by a non-*S*-locus mutation[244]. In *Turnera subulata* three hemizygous genes, *Ts*SPH1, *Ts*YUC6 and *Ts*BAHD, were detected, with the latter only expressed in short styles, *Ts*YUC6 in anthers and *Ts*SPH1 in filaments. Additionally, a long homostylous mutant exhibited a deleted *Ts*BAHD gene and a short homostylous mutant possessed no *Ts*SPH1 expression, indicating that *Ts*BAHD represents the *G*-locus and *Ts*SPH1 the *A*-locus in *Turnera*. Further, homologous genes of *Ts*BAHD were found to inhibit BR in *Arabidopsis (At*BAHD), similar to the *Arabidopsis* CYP734A50 closest homolog PHYB ACTIVATION-TAGGED SUPRESSOR1 (*At*BAS1). Transgenic expression of *Ts*BAHD in

A. thaliana led to BR-deficient phenotypes with differences in severity in a dose-dependant manner. Additionally, BAS1 expression was shown to be downregulated in the mutant lines, whereas DWF4 and CONSTITUTIVE PHOTOMORPHGENESIS AND DWARFISM (CPD), encoding for two BR biosynthesis-involved enzymes, were upregulated[245]. All together, these results revealed convergent evolution in a biochemical mechanism though acting on different genes between *Turnera subulata* and *Primula*, both leading to inhibited style length growth in the S-morph flowers. Thus, *Ts*BAHD is causing decreased style cell elongation and thus a short-style, whereas *Ts*YUC6 was shown to play a role in auxin synthesis. Therefore, an incompatibility mechanism determined by non-matching BR and auxin concentrations between pollen and style was assumed, leading to compatible crosses of low BR of L-morph styles and low auxin concentrations of pollen from high positioned anthers of S-morphs and reciprocally. The structural resolution via BAC clones of the *Turnera S*-locus revealed two inverted regions, one consisting of three genes and the other of 14 genes, surrounding the hemizygous *S*-locus genes. These inverted regions are intensifying the suppression of recombination via crossing over[246].

4.1.3.4 *Amsinckia*, a genus of the *Boraginaceae*, exhibiting heterostyly without SI

The angiosperm family *Boraginaceae* consists of around 1600 species, divided among at least 110 genera. Phylogenetic analysis based on morphological traits, three chloroplast DNA regions and one nuclear ribosomal DNA region, revealed uncommon characteristics of this family. First, within the *Boraginaceae* the heterostylous breeding system developed at least 12 times[247]. Second, different types regarding stylar conditions and flower morphs can be found, since it includes genera, for example *Lithodora*[248], with both distyly and stigma-height dimorphism (either approach or reverse herkogamy), although several theoretical models support distyly as ancestral state for the *Boraginaceae*[247–249]. And third, different combinations between these style-types and all forms of self-incompatibility[233] up to self-compatibility, for example in *Pulmonaria*[250], are existing within this family. However, the high number of independent origins of heterostyly argues for a possible removal of the linkage between SI and the responsible locus for style length, as was supported by a study in *Anchusa*[251,252]. Thus, *Boraginaceae* represent a great model for studying the evolution of heterostyly and its non-obligatory association to SI.

Here, one *Boraginaceae* genus, *Amsinckia*, revealed a fourth special characteristic, that is the repeated breakdown of heterostyly to homostyly (Fig. 40). *Amsinckia* is a genus of annual plants primarily settled in California, but is also naturally occurring north to British Columbia, east to Utah, south to Baja California and occasionally in southern America. It can also be found in Europe, South Africa, central and North America, Alaska, the Yukon as well as in Australia[253].The genus consists of approximately 20 species or small groups of close relatives, that differ particularly in their rate of self-fertility and associated flower morphology[228,254]. Five taxa are distylous and predominantly outcrossing or exhibit mixed mating systems, while the remaining taxa are homostylous and mostly selfing. Thus, the higher the stigma-anther-separation, the higher is the outcrossing rate[20] (Fig. 40C). Phylogenetic analysis of the variation of restriction sites in chloroplast DNA revealed that the ancestral states of the genus are distyly and outcrossing, since the self-fertilizing taxa are of recent origin like the heterostylous ones[228,247].

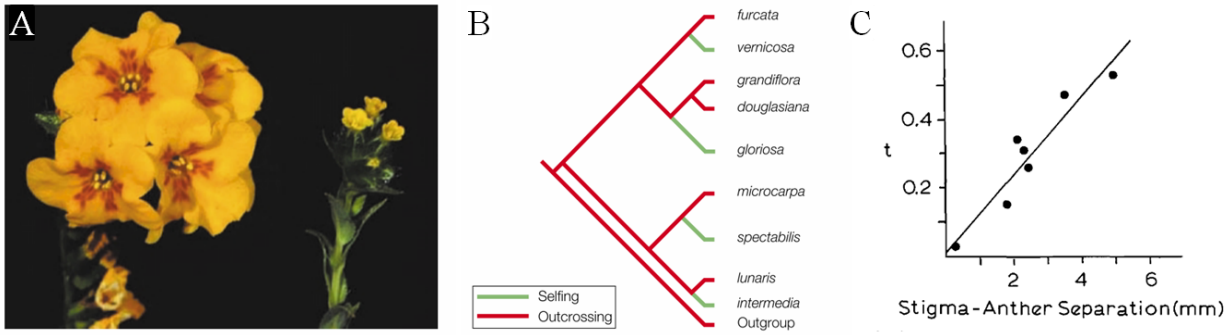


Fig. 40 *Amsinckia* genus as origin for multiple transitions to selfing(modified after [16,20])

(A) Flower size difference of outcrosser *Amsinckia furcata* and its small flowered, predominantly selfing sister taxa *Amsinckia vernicosa*. **(B)** Phylogenetic reconstructions of mating systems based on restriction site variation in the chloroplast genome in the genus *Amsinckia*. Results assume that the ancestral state is distyly. **(C)** Correlation of outcrossing rate of populations (t) with averaged stigma-anther distances in flowers within the populations.

Besides the differences in mating systems, also morphological traits clearly differ between the hetero- and the homostylous taxa. Thus, distyly is represented by two large flower morphs that possess reciprocal differences in style and stamen length. L-morphs are described as long-styled with low-positioned anthers at the base of the corolla, whereas S-morphs have a short style close by the nectary and anthers that are high-positioned[255]. Additionally, also pollen size differences between the two morphs have been detected, with S-morphs producing larger pollen grains.[254].

Heterostyly in *Amsinckia* underlies the classic one locus (*S*-locus) genetic model, with the S-morph being dominant (*Ss*) to the L-morph (*ss*)[255], consistent with the *Primula S*-locus model. The homostylous taxa, on the other hand, are often accompanied by various degrees of flower size reduction[228]. Unexpectedly, in *Amsinckia*, distyly is not coupled with an obvious sporophytic SI, although previous studies, discussed the presence of a cryptic SI. Thus, compared to intramorph pollen, intermorph pollen was shown to preferentially fertilize in *Amsinckia grandiflora*[256] and *Amsinckia douglasiana*[257]. Three years later, the reduced seed production in intramorph crosses of *A. grandiflora* was explained as a result of inbreeding depression rather than competitive pollen tube growth[258]. Surprisingly, natural populations of distylous *Amsinckia* represent approximately a 1:1 morph ratio, leading to the assumption that homozygous dominant (*SS*) plants are absent[227,259,260]. This raises the question, how self-fertilization can be prevented, if SI is not present. A recent study suggests several external factors as putative causes. One reason could be the pollinators equipped with a short proboscis only allowing pollen transfer between the long-level sex organs, as it was shown for the bee *Apis florae*, the pollinator of *Luculia pinceana* (*Rubiaceace*)[215]. Without SI, there might be different ways to prevent self-fertilization in distyly species. The absence of illegitimate progeny could be a result of inbreeding depression due to higher rates of deleterious (recessive) mutations at the seed germination phase that are expressed in homozygotes. However, a recent analysis of the outcrosser *A. douglasiana* compared to the selfer *A. gloriosa* demonstrated no significant difference in the accumulation of deleterious mutations[261]. This argues against inbreeding depression being the reason for the absence of homozygous dominant individuals in natural *Amsinckia* populations. A pollinator study with *A. grandiflora* found *Anthophora edwardsii* to be the main pollinator, preferentially transferring pollen from S-morph to L-morph, but L-morph stigmas received more total pollen and also proportionally more illegitimate pollen. Still the pollinator behaviour was shown to not influence the changing morph ratios within three seasons[260], arguing against the exclusive responsibility of pollinators of preferring intermorph crossings.

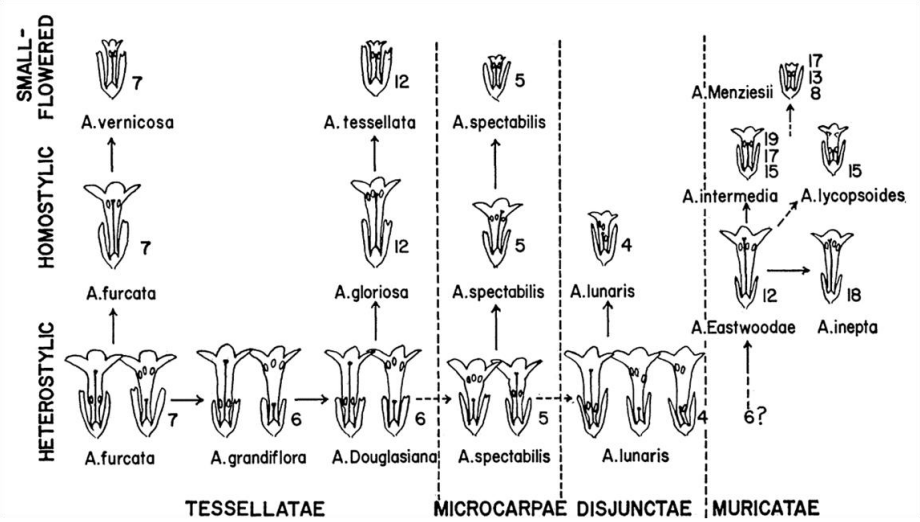


Fig. 41 Overview of *Amsinckia* species and their relationships (modified after [21])

Numbers indicate respective haploid chromosome numbers.

The *Amsinckia* genus was divided in four taxonomic sections based on cytogenetics and morphological traits[21,255]: *Tessellatae*, *Microcarpae*, *Disjunctae* and *Muricatae* (Fig. 41*)*. The latter one represents the only section without heterostyly and additionally differs due to an irregular series of high chromosome numbers. In the interest of this project, the three remaining sections will be focused on. The *Tessellatae* are supposed to contain the basic member of the *Amsinckia* genus *A. furcata*. This fact is based on its chromosome number (n = 7) and the assumption that there might have been an evolutionary series of reduction in chromosome number, leading to n = 4 in the *A. lunaris* (*Disjunctae*) via the series of *A. grandiflora* (n = 6), *A. douglasiana* (n = 6) and the large flowered distylous *A. spectabilis* (n = 5). According to cytogenetics, the first branch representing the breakdown of heterostyly, is based on distylous *A. furcata* and leads to small flowered and homostylous *A. vernicosa*. The left group of the *Tessellatae* all produce smooth nutlets. *A. douglasiana* represents the founder species for the second selfing branch via homostylous and large-flowered *A. gloriosa* to small-flowered *A. tessellata*. Unexpectedly, this evolutionary transition was accompanied by the duplication of the genome, leading to tetraploid *A. gloriosa* and *A. tessellata*, which additionally argues for heterostyly to be the ancestral state. Polyploidization might have caused this transition to homostyly. However, it does not necessarily explain the breakdown of heterostyly, since the development to fully selfing and small flowering species occurred at least twice without genome

duplication. A further characteristic of the *Tessellatae* is the increase in stigmatic papillae length of L-morphs compared to S-morphs, which is an additional distinct feature compared to the *Microcarpae*, that exhibit the reverse pattern[254]. The differentiation between *A. lunaris* and the *Muricatae* section occurred on the base of cytogenetics and scent. Thus, *A. lunaris* is characterised by a sweet fragrance, distinguishable from the sourish scent typical for the *Muricatae*. Apart from this, also the *Disjunctae* reveal a third breakdown of heterostyly.

The *Microcarpae* consist of three closely related forms that are still able to intercross and to form hybrids. A further special characteristic of the *A. spectabilis* forms is that they naturally occur not only in strictly heterostylous populations, found in Nipomo and Santa Maria, and in strictly homostylous populations, but also in mixed hetero- and homostylous populations, for example close to Montana de Oro. Homostylous populations consist of a range of different flower sizes, particularly found in north and south California, while the central part is marked by large-flowered heterostylous forms. Only the two most northern populations, close to Zmudowksi, are isolated small flowering homostylous forms[20,22] (Fig. 42).

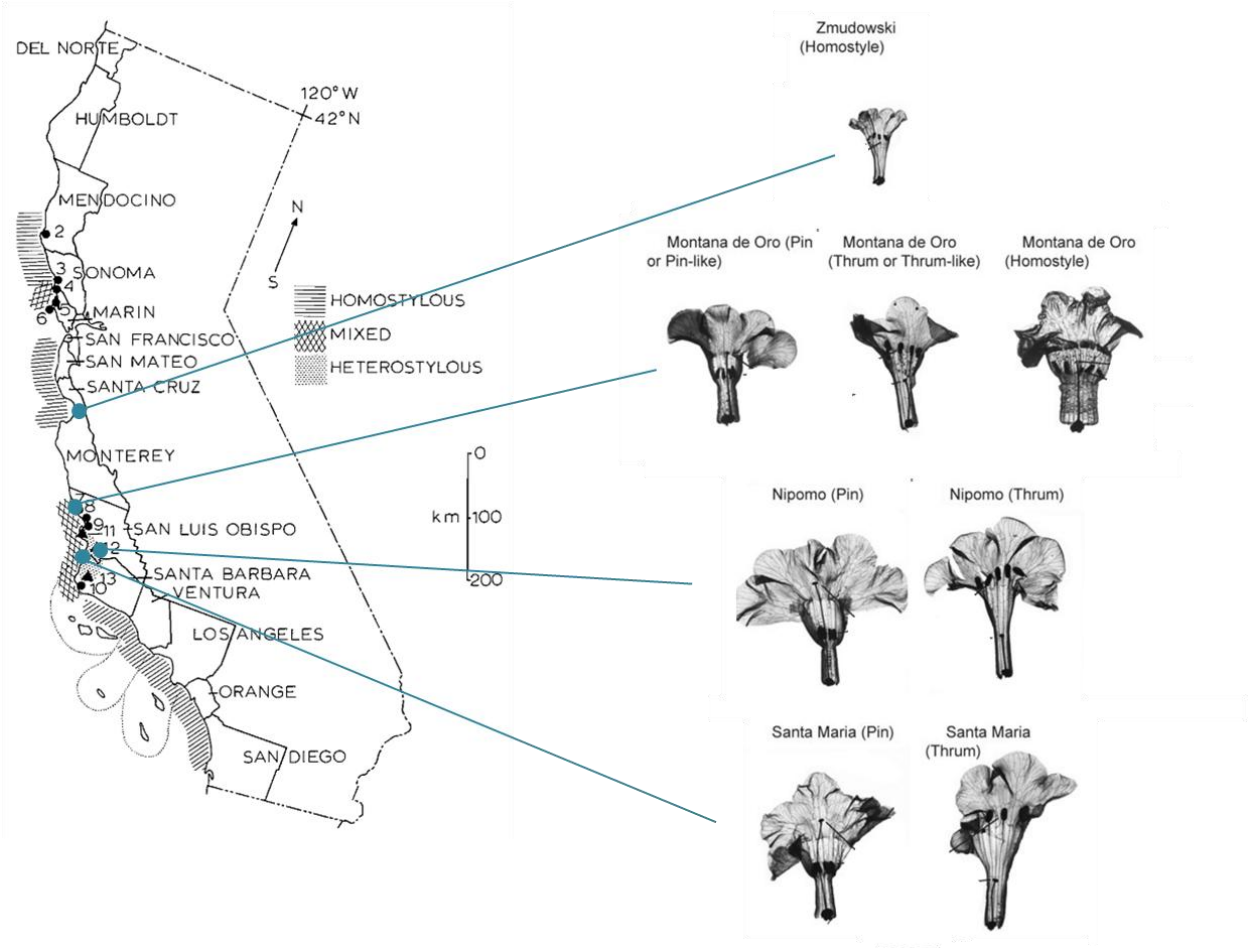


Fig. 42 Geographical range of heterostylous, homostylous and mixed populations of *Amsinckia spectabilis* in California (modified after [20,22])

Various collected forms are shown on the right side and linked to their localizations. Nipomo (91003) and Santa Maria (91004) are distylous, consisting of pins and thrums. Montana de Oro (95002) is a mixed population, consisting of morphs that vary from pin or pin-like through homostylous to thrum or thrum-like. The Zmudowski State Beach (91011) population is small-flowered long homostylous.

4.2 Aim of the project

The special and uncommon characteristics of the *Amsinckia* genus generated a great deal of interest in the Lenhard lab, forming the basis for this study. Especially, the revelation of the genetic basis of the *S*-locus in the heterostylous and the small-flowered fully selfing *Amsinckia spectabilis* form was of primary importance, in hope for a better understanding of the evolutionary background regarding heterostyly and its breakdown in species that do not possess an obvious self-incompatibility that is genetically associated to the heteromorphic phenotype.

More detailed, the aim of this project was to answer the following questions. 1) Which and how many phenotypic differences can be found in the S- and the L-morph plants that would indicate an approximate quantity of loci that are genetically determined by the *Amsinckia S*-locus? 2) Is it possible to reveal the genetic architecture of the *S*-locus in S- and L-morphs and thus, to detect the mechanism that keeps up the suppressed recombination?

3) Can candidate genes of the *S*-locus be identified? According to recent studies, organ growth related mechanisms or genes would be expected, since contrasted growth rates of stamen and pistil were proven to be causal for the development of the two morphs[262].

4) How is the maintenance of the 1:1 morph ratio assured in naturally occurring populations?

5) Is it possible to gain information regarding the genetic architecture of the homostylous *S*-locus? Can similarities to the L-or the S-morph S-locus be found, and provide hints regarding the ancestral morph the long homostylous form was derived from? Is the transition to homostyly in *Amsinckia spectabilis* caused by an *S*-locus linked event, such as mutation or recombination, or could a non-*S*-locus modifier gene be responsible, possibly explaining the almost seamless transitions regarding the differences in floral organ positioning within the mixed populations?

4.3 Materials and Methods

4.3.1 Biological material and growth conditions

Four different *Amsinckia spectabilis* lines were used for this study and were kind gifts from M. Johnston and D. Schoen. Their geographical origins are described in detail in chapter 4.1.3.4. Two different populations of the heterostylous *Amsinckia spectabilis* form were taken. The S to L cross of two parental plants of the 91004 population from Santa Maria displayed the base for the 150 F1 individuals that were taken for the genome and transcriptome analysis. The PacBio Sciences sequencing was based on two F3 plants, that were derived from repeated selfing, a homozygous recessive (L-morph) and a homozygous dominant S-morph, both developed form a S-morph selfed individual, originated from the Nipomo population (91003). The small homostylous individual that was sequenced via genome Illumina sequencing belonged to the 91011 population, collected near Zmudowski[22]. The large-flowered homostylous Montana de Oro population (95002) was only grown for photographing.

Seeds were surface-sterilized, by rinsing them very shortly with 100% EtOH, followed by the transfer to pots, covered with 1-2 cm of soils and stratisfied for 3-7 days at 4°C. Afterwards they were put into a shaded place for 2-3 days at 22°C and transferred to two different growth conditions, in the following. 91004 and 91011 were grown under short day conditions (8 h light/16 h dark), whereas the Nipomo population was grown under long day conditions (16 h light and 8 h dark). All populations were grown at 22°C over day and 16°C at night, at humidity of 70%, and under a light level of 150 µmol $m^{-2}s^{-1}$. Usually, germination could be observed after 2-3 days, allowing the seedlings to be shifted to full light.

4.3.2 Morphological measurements

Phenotyping

To analyze flower phenotypes from all lines, pictures of the first three most opened flowers of the main shoot inflorescence were taken by a camera (Canon EOS 60D (Ota, Tokio, Japan) after removal of the vertical half of the corolla tube. Style length and anther height were digitally analyzed using ImageJ (http://rsbweb.nih.gov/ij/).

Pollen size and number were determined with the help of the counting chamber of a Neubauer hemocytometer. The analyzed pollen was harvested from the last two unopened flowers. An overnight incubation at 60 degree in an open 1.5ml Eppendorf tube allowed for the anthers to open. 30 µl of 5% Tween-20 were added, followed by a 2 min vortex step that led to the release of the pollen. Afterwards, the samples were sonicated (Bandelin, SONOREX, Type RC31) for 20 min. 10 µl of the sample solution were filled into one counting chamber and pictures were taken using the light microscope, Olympus BX51. Pollen size measurement and counting was carried out using ImageJ.

Statistical analysis

Statistical analyses were conducted in RStudio[161] or Microsoft Excel 7. The multiple comparisons of phenotypic means were calculated by using Tukey's HSD post hoc test using the agricolae package implemented in R. For two-sample comparisons, a two-tailed Student's t-test assuming unequal variances was used for normally distributed data, whereas the Mann-Whitney-U-test was applied for not normally distributed samples. Experimental measures and resulting statistical tests were independent of each other and thus did not require adjustment for multiple testing. The null hypothesis was rejected at $p < 0.05$. Sample distributions are presented by box plots using ggplot2 package implemented in RStudio. Here, the middle line represents the median, the upper and lower box border correspond to the 75^{th} and 25^{th} percentile, respectively. Whiskers stretch until the maximum and the minimum values comprised with 1.5 interquartile ranges. Outlayers are demonstrated by single dots outside the whiskers. Grey dots surrounding the boxplots represent individual sample values.

Low-melt agarose gel imprints

Gel imprints of style and corolla tissue were carried out in an adapted protocol from Horiguchi[150]. A drop of 2% low-melt agarose containing 0.01% bromophenol blue was pre-warmed at 50°C and placed on a pre-warmed glass slide. The plant material was immediately put on the smeared droplet. As soon as the gel solidified, the petal was carefully removed. After the complete drying of the agarose layer for 10 min, the imprint could be pictured under a light microscope, Olympus BX51. In the following, the digitalized pictures were photo merged using Adobe Photoshop (https://www.adobe.com/de/products/photoshop.html) and analyzed using ImageJ.

Aniline blue staining

Aniline blue staining[263] was used to detect pollen germination and pollen tube growth through the style. Flowers close to opening were emasculated and left to grow for 12 h to achieve pistil maturation. This step was followed by hand pollination with anthers. 12 h after pollination, the pistils were dissected to remove the ovary walls and fixed in a 9:1 (ethanol:acetic acid) for more than 2 h. Afterwards the samples were washed with 70% EtOH for ~30 min followed by 1 N NaOH overnight treatment. Finally, the aniline blue staining (0.1% [w/v] aniline blue pH 8.9, 0.1M K3PO4) was carried out for another 6 h for observation under ultraviolet illumination (SteREO Lumar V12, (Zeiss, Oberkochen, Germany)). In the following, the digitalized pictures were analyzed using ImageJ.

4.3.3 Genome and Transcriptome analysis

DNA-extraction for genotyping

The harvested *Amsinckia* plant material was frozen and ground in liquid nitrogen. The fine powder was suspended in 750 µl of 65°C-prewarmed CTAB buffer (100 mM Tris (pH 8.0), 2% [w/v] CTAB, 30 mM EDTA, 2 M NaCl, 2% PVP) and mixed thoroughly. Afterwards, the mixture was incubated at 65°C for 30 min and vortexed every 10 min. After that, 630 µl of Chloroform:isoamyl alcohol (C:I) 24:1 were added. The mixture was shaken until both emerged phases were fully emulsified, followed by centrifugation at 13,000 rpm for 5 min at 4°C. Next, 630 µl of the aqueous phase was transferred into a new 1.5 ml microcentrifuge tube. This extraction step was repeated, starting again with adding 630 µl of C:I to the transferred solution. After a second centrifugation, about 500 µl of the aqueous phase was collected and precipitated with 50 µl of 3 M NaOAc (pH5.2) and 500 µl of isopropanol for 5min at RT. In the following, the solution was centrifuged for 10min at 13,000 rpm at RT. The pellet was washed once with 70% EtOH and dissolved in 50 µl TE supplemented with RNase. Finally, the samples were incubated at 37°C for 5 min to allow the remaining EtOH to evaporate[264].

Preparation of the genomic library and bioinformatics analysis

The DNA was taken for genotyping and the *S*-locus marker tests. Individual S-, L-morph (Santa Maria population) and homostylous (Zmudowski population) DNA was used for shearing for an insert size of 400 bp using Covaris® S220 Ultrasonicator (Covaris, Woburn, Massachusetts,

USA), forming the base for the NEBNext Ultra II FS DNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA) that was carried out according to the manufacturer's instructions using Sera-Mag Speedbeads™ (Thermo Fisher Scientific, Waltham, Massachusetts, USA) for each clean-up step. DNA size distribution and quality were tested with an Agilent High Sensitivity D1000 ScreenTape System (Agilent) and the DNA concentrations were measured using Qubit® 2.0 Fluorometer. Pools of different libraries were combined based on equimolar ratios and paired-end sequenced using the Illumina NextSeq 500 platform generating approximately 80-200 million 150 bp read pairs for each library. Reads were demultiplexed by Illumina indices and assembled to obtain a reference using SPAdes[265] and Megahit[266]. In the following, reads were mapped to the assembled S-morph genome using Bowtie 2[267]. To compare L- and S-morph genomes, log2 fold changes of the morph-specific read counts were calculated after a manual addition of 5 reads, to account for stochastically appearing false mapping reads and to enable calculation of the ratio in case of 0 values.

Transcriptomic analysis

Style and corolla tissue (including attached anthers) of four oldest flower buds, that remained after removing the open flowers and buds that would open within the next 24 h per inflorescence were manually dissected of the 150 individuals of the Santa Maria population and sampled. Three replicates per morph and organ were collected, each consisting of 25 same-morph individuals. Total RNA of these 12 samples was extracted using a Trizol based protocol[156] as described in chapter 2.2.

The RNA quality was checked on an Agilent-2100-Bioanalyzer (Fa. Agilent Technolgies GmbH, Böblingen). Then, the genomic DNA was digested using TURBO™ DNase (Invitrogen by Thermo Fisher Scientific, Waltham, Massachusetts, USA) according to the manufacturer's instructions. 500 ng of each DNase digested RNA was taken for library preparation using the TruSeq RNA Library Prep Kit v2 (Illumina, San Diego, California). Fragmentation aimed at a median insert size of around 195 bp. Each cleaning step was carried out using Sera-Mag Speedbeads™ (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and the final library quality and concentration measurements were done using Qubit® 2.0 Fluorometer and an Agilent High Sensitivity D1000 ScreenTape System (Agilent). All 12 libraries were pooled

based on equimolar ratios and paired-end sequenced using the Illumina NextSeq 500 platform generating approximately 80-200 million 150 bp read pairs for each library.

Bioinformatics analysis of the transcriptome

Reads were demultiplexed by the used Illumina indices and mapped to the assembled S-morph genome using Bowtie 2[267]. Each contig was divided into overlapping pieces via sliding window with a window size of 500 bp and a step size of 250 bp. Finally, all reads for each window were counted using BEDTools[268] forming the base to calculate the fold changes of the 4 times enriched S-morph genes compared to L-morph. For this calculation 10 counts were added manually to each present count to account for stochastically appearing false mapping reads and to enable calculation of the ratio in case of 0 values. Marker regions were blasted with BLASTn to find close relatives.

Genotyping for morph-specific segregation of the designed markers

All candidate markers were blasted against the S-morph reference genome to find regions with high similarities. Sequence differences between these regions, such as SNPs, deletions or insertions, allowed for the design of markers that should discriminate between these regions. For marker design CAPS[154] and dCAPS[155] were used that are based on natural or induced restriction-endonuclease-sensitive polymorphisms. PCR and digests were carried out according to chapter 2.2.

Preparation for PacBioSciences sequencing

For single-molecule real-time sequencing (SMRT) (Pacific BioSciences)[269] the extraction of high molecular weight DNA was carried out using the NucleoBond® HMW DNA (Macherey-Nagel Düren, Germany) starting with a maximum of 500 mg individual plant material consisting of very young tissues, such as flower buds and very young side branches. The size distribution was measured using the Agilent High Sensitivity D1000 ScreenTape System (Agilent) and was determined to have a median size of 65 kb. Samples were brought to the lab of Axel Himmelbach located at the IPK (Gatersleben, Germany) for library preparation and sequencing. Adaptor ligations and the preparation of the DNA-polymerase-primer complexes, were performed according to the manufacturer's instructions of the SMRTbell Express Template Preparation Kit 2.0 (Pacific Biosciences, Menlo Park, California, USA). Finally, three SMRT

cells were used to sequence the dominant homozygous S-morph library and one SMRT cell for the L-morph on the PacBio Sequel System on continuous long read mode to reach the maximum of read length.

Bioinformatics analysis of PacBio sequencing

Assembling the genome of the 3 runs of the homozygous S-morph individual based on the PacBio sequences (roughly 37 x coverage for the S- and estimated 9 x coverage for the L-morph individual) using canu[270] resulted in an assembly with an NG50 of 7 Mb (9 contigs), considering an estimated genome size of 300 Mb. The RNA-Seq samples were mapped against the S-morph PacBio genome reference using STAR-2.7.3a, To calculate the homozygosity score over the genomes of L- and S-morphs, variant calling was carried out using samtools/bcftools (http://samtools.github.io/bcftools/call-m.pdf)[271].

4.3.4 Brassinosteroid treatment

For the Brassinosteroid treatment approximately 16 still unopened flower buds per inflorescence were once injected with solutions containing 0.01 mM, 0.02 mM or 0.1 mM homobrassinolide (dissolved in EtOH absolute), or a mock solution containing the corresponding maximal EtOH concentration to the 0.1 mM homobrassinolide. Once fully opened, flowers were bisected, photographed and organ sizes were determined using ImageJ.

4.4 Results

4.4.1 Lines of *Amsinckia spectabilis*

The following study is based on the work with three different *A. spectabilis* lines (Fig. 43), including two populations of a heterostylous form (collected in Nipomo and Santa Maria) with large flowers that exhibited typical orange coloured freckles close to the corolla mouth, which might function as ultraviolet recognition pattern for pollinators. Here, the L-morph is defined by a long style and low-positioned anthers, whereas the S-morph is marked by a short style and high-positioned anthers. The S-morph *S*-locus is assumed to be heterozygous (*Ss*), since progeny of a selfed S-morph approximately show a 1:1 L:S-morph ratio. Meanwhile the L-morph must be homozygous recessive (*ss*), since only L-morph progeny can be found after selfing[227,259]. The second large-flowered line (collected near Montana de Oro) was lacking these orange spots and possessed a partially selfing efficiency and a long homostylous phenotype. The third line (collected near Zmudowski State Beach) was also long homostylous that was associated with small flowers and the capacity of complete autonomous selfing[20,22,228,253,272].

Fig. 43 Comparison of inflorescences of *Amsinckia spectabilis* variations (modified after[21])

Note different flower sizes as well as different intensities of orange spots around the corolla opening (mouth). Scale bar: 5mm

4.4.2 Phenotypes of *Amsinckia spectabilis* L-morph and S-morph

To identify the phenotypes determined by the *S*-locus, the two different morphs of the Santa Maria population were analysed regarding their phenotypes (Fig. 44). As expected, both show a reciprocal positioning of the sex organs.

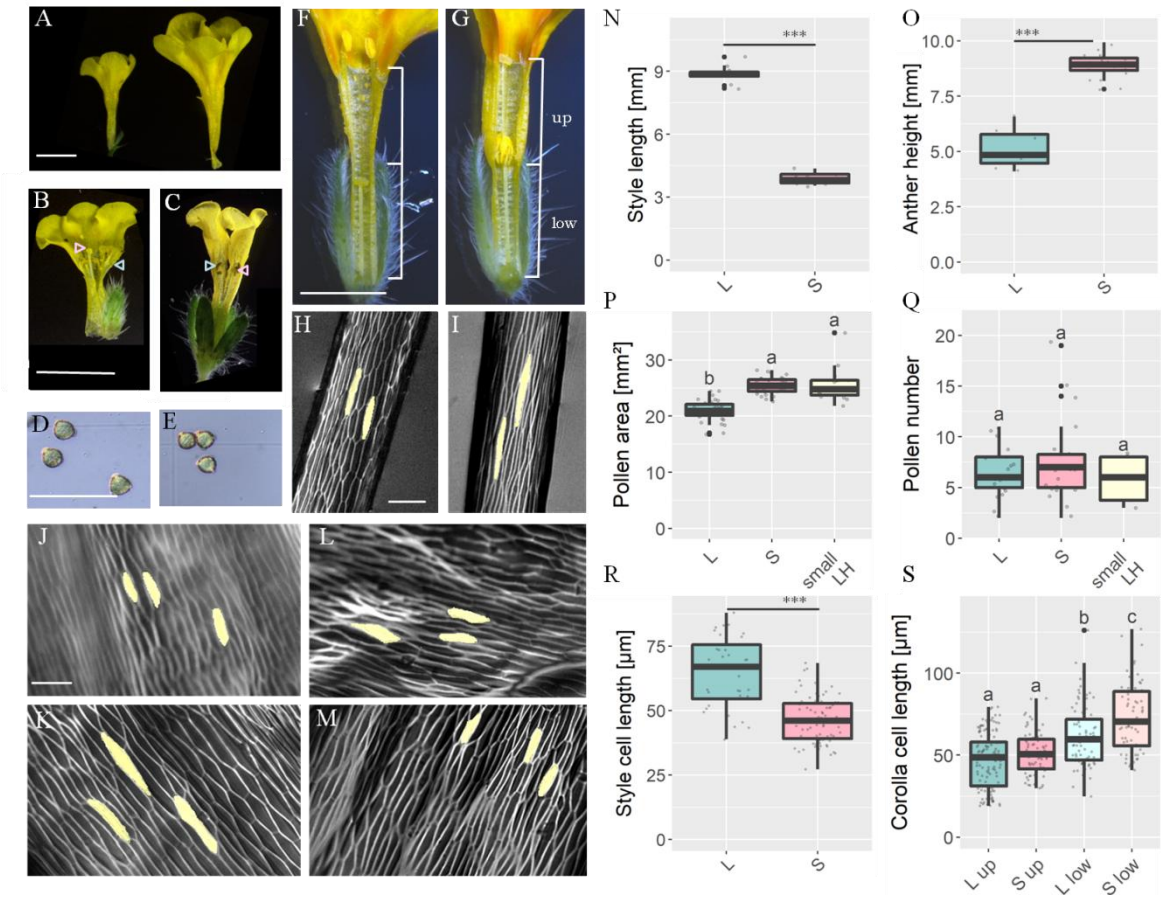


Fig. 44 Phenotypes of L- and S-morphs of *Amsinckia spectabilis*

(A) Flower size comparison of the small-flowered selfing variation to the heterostylous variation. Scale bar: 5 mm. **(B-C)** Dissected flower to show sex organ positioning (pink arrow: stigma; blue arrow: anthers) of a small-flowered long homostylous individual from the Zmudowski population (b) and large-flowered long homostylous individual from the Montana de Oro. Scale bar: 5 mm. **(D-E)** Pollen size of S-morph (D) and L-morph (E). Scale bar: 25 µm **(F-G)** Dissected flowers from the heterostylous Santa Maria population, note the reciprocally positioned sex organs, with short style and high anther height (S-morph (F)) and the long style and low anther height (L-morph (G)) Illustration explain the subdivision into the upper and lower corolla part according to anther attachment position. Scale bar: 5 mm **(H-I)** Agarose gel imprints of S-morph style (H) and L-morph style (I). Single cells are coloured. **(J-K)** Agarose gel imprints of upper (J) and lower (K) corolla tissue of a S-morph **(L-M)** Agarose gel imprints of upper (L) and lower (M) corolla tissue (as illustrated in F-G) of a L-morph. (J-M) Single cells are coloured, Scale bar: 50 µm **(N-Q)** Quantification of morph-specific phenotypes. L: L-morph, S: S-morph, small LH: small-flowered long-homostyle. **(R-S)** Quantification of cell length of style and corolla tissue; Asterisks represent significance (*** = $p > 0.001$) determined by t-test or Mann-Whitney-U-test as indicated in the text. The letters indicate significant differences as determined by a Tukey's HSD test.

Starting with style length, a significant difference ($p = 3.67 \times 10^{-15}$, student's t-test) can be found, with an average style length of 8.9 mm for the L-morphs compared to 3.9 mm on average in S-morphs, displaying a 56% style length decrease found in S-morphs (Fig. 44N). Agarose gel imprints were used as follow up to detect if the style length difference is rather due to a difference in cell number or in cell elongation. Analysis identified cell size as one cause of the observed organ size difference ($p = 9.31 \times 10^{-10}$, Mann-Whitney-U test) with an averaged decrease of style cell length in S-morphs of 28% compared to L-morph cell lengths (Fig. 44R). The remaining 28% of style length difference might additionally be explained by a higher cell number in L-morph styles that could not be investigated with the used imprint technique. Also, individual experimental differences could cause the missing 28%, since style length and style cell length measurements were done on different individuals.

A recent study[254], found a significant pollen size difference between S-and L-morph, which could be confirmed in this study ($p = 1.17 \times 10^{-16}$, Mann-Whitney U-test). Here, S-morph pollen grains were 21% larger compared to L-morph pollen. A significant difference in pollen number per flower could not be detected. A similar pattern could be detected for the small-flowered long-homostylous selfing population, indicating a rather S-morph-like genetic situation at the *P*-locus (Fig. 44P-Q).

Anther height displayed a significant increase (p = 1.48 x 10-5, Mann-Whitney-U-test) of 58% in S-morphs compared to L-morphs, as was expected by the definition of the heterostyly syndrome (Fig. 44O).

To discover the cause of anther height differences agarose gel imprints of the upper and lower cells of the inside of the corolla tubes (divided according to the anther attachment at the corolla) were conducted (Fig. 44 F-G). While the corolla cells of the upper areas did not show a significant size difference between both morphs, the lower tube cells differed significantly between L- and S-morph (p = 1.4 x 10-5, Mann-Whitney U test), with an S-morph specific increase of corolla cell length of about 20% (Fig. 44S) indicating that the high anther position is at least partially caused by a promoted cell elongation, corresponding to the style length in L-morphs. The remaining 38% of anther height difference, which cannot be explained by cell elongation, might be caused by an additional increase of cell number. In summary, these results support the assumption that the dominant *A. spectabilis S*-allele carries the information for style length decrease, promoted corolla cell elongation, which is leading to a high anther position, and

an increase of pollen size. Thus, fitting to the proposed *S*-locus model for *Primula* by Alfred Ernst, the *S*-locus of *A. spectabilis* seems to possess at least a *G*-locus, an *A*-locus and a *P*-locus.

Besides determining organ size or positioning, the *G*- locus in *Primula* transmit female self-incompatibility[19,235]. Moreover, there are contradictory assumptions regarding the presence or absence of self-incompatibility in *Amsinckia*[227,256,257]. Thus, it was necessary to examine the degree of SI in *A. spectabilis*. Therefore, reciprocal intermorph crosses and intramorph crosses were carried out via hand pollination and the fully matured seeds were counted (Fig. 45B). Since one *Boraginaceae* characteristic is the presence of only four ovules per flower[227,272], four seeds per flower were expected to be the maximum seed set. Similar rates of successful seed maturation of about 70% were detected for both intermorph crosses S (♂) to L (♀) and L (♂) to S (♀), as well as for the S to S intramorph cross, indicating that for the S-morphs no obvious intramorph SI could be observed. Regarding the intra-L-morph cross, however, a significant decrease in matured seeds was detected with only around 43% of the maximum possible seed set.

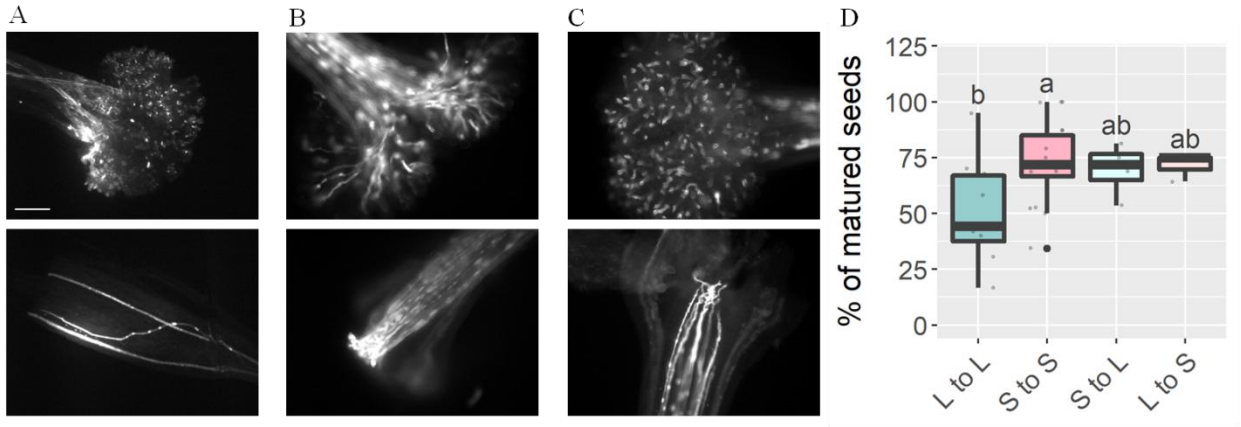


Fig. 45 Selfing efficiency of heterostylous *Amsinckia spectabilis*

(A-C) Aniline blue stained stigma (upper images) and style (lower images) of L to L crosses (A), S to S (B) and S (♂) to L (♀) (C). Scale bar 50 µm **(D)** Quantification of seed set of all possible inter- and intramorph crossings ((♂) to (♀)). The letters indicate significant differences as determined by a Tukey's HSD test. L:L-morph, S:S-morph

These findings were supported by aniline blue stained styles 12 hours after pollination (Fig. 45A-C). Here, the outgrowth of L-pollen tubes in the L-style was clearly reduced compared to the intermorph crossings and to the intramorph S-cross. These results would rather have been expected for the intramorph S-cross, since a prezygotic rejection of *S*-pollen tube inside a short (*Ss*) style could be an explanation for the absence of *SS* plants in the *Amsinckia* populations.

However, according to this analysis the *S*-locus in *A. spectabilis* does not inherit a strong SI-mechanism.

4.4.3 Development of homozygous dominant (*SS*) plants

Further proof for the absence of strong SI in *A. spectabilis* was the occurence of dominant homozygous (*SS*) plants during this study that were gained via two selfing rounds of a S-morph individual. Already the development of mature and germinable seeds argues against a present strong SI, supporting the previous finding. Thus, eight F2 plants were grown and morph phenotyped, resulting, in a ratio of 7:1, shifting towards more S-morphs. The result deviated from the 2:1 S:L morph ratio that would have been expected for absent (either due to inhibited seed maturation or a decreased germination fitness level and therefore dying) *SS* individuals. The observed accumulation of S-morphs led to the assumption, that dominant homozygous plants might already be present within these F2 plants. To investigate this theory, all eight plants underwent a third selfing round, so that from each F2 S-morph twenty individual F3 plants were grown and phenotyped, expecting 100% of the progeny to be S-morph in case the parental F2 plant was already dominant homozygous. In parallel, all eight F2 plants were taken for comparative phenotyping regarding their floral morphologies (Fig. 46). Here, style length, anther height and the resulting style-anther-distance were shown to divide the eight individuals in three size-related groups, with four plants, exhibiting an even stronger S-morph phenotype as the remaining three S-morph plants ($p = 6.72 \times 10^{-6}$), meaning an even shorter style (66% compared to 57%) and higher positioned anthers (100% compared to 74%) and therefore an increased style-anther-distance. The pollen area did not differ significantly between the S-morph individuals. Since a homozygous dominant *S*-locus would be expected to transmit an increased total expression of the responsible *S*-genes due to the doubled copy number, an intensified S-morph phenotype would be expected, as it was found for four F2 individuals. The outcome of the F3 individuals resulted in the confirmation of these four F2 parents to be homozygous for the dominant *S*-allele, since their progeny consisted of S-morph individuals without exception.

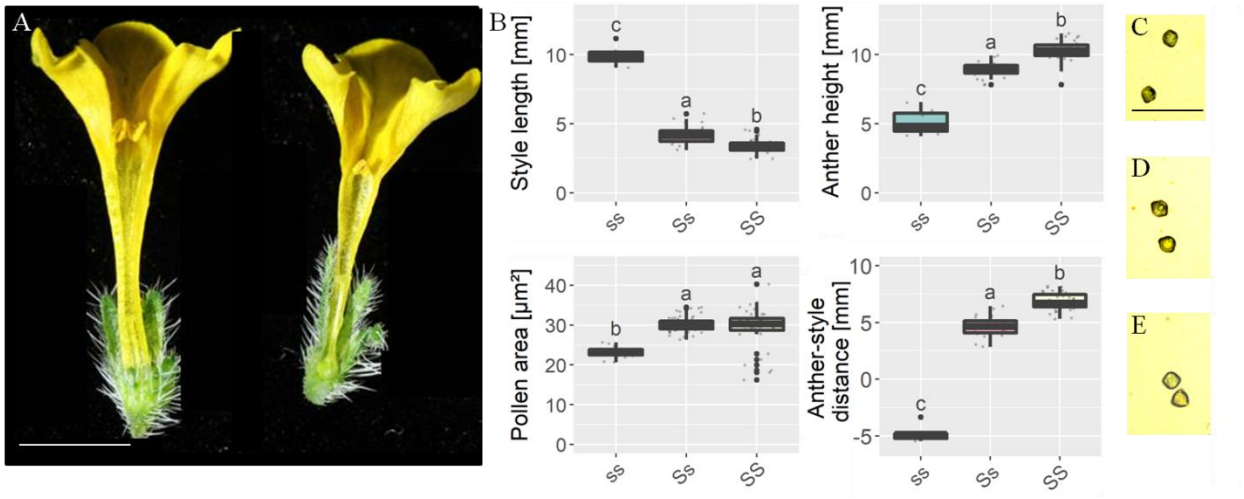


Fig. 46 Overview of phenotypes of dominant homozygous (*SS*) plants

(A) Dissected flowers of a homozygous dominant (*SS*) individual (left) and a heterozygous S-morph (right). Scale bar: 5mm **(B)** Quantification of flower phenotypes of *SS* plants compared to homozygous recessive L-morphs (*ss*) and heterozygous S-morphs (*Ss*). **(C-E)** Comparison of pollen size of L-morph (C), heterozygous S-morph (D) and homozygous dominant S-morph **(E);** Scale bar: 25 µm; The letters indicate significant differences as determined by a Tukey's HSD test

4.4.4 BR levels are not causal for style length differences in *Amsinckia spectabilis*

Brassinosteroids play a role in determining the style length in other heterostylous species, as it was shown for CYP734A50 in *Primula*[235] and for *Ts*BAHD in *Turnera subulata*[273]. Both factors act as BR inactivators that cause decreased style growth and are therefore responsible for the short style in S-morph plants. According to this, it was tested, whether different BR concentrations in the *A. spectabilis* flowers lead to changes in style length. Thus, all closed flower buds of S-morph and L-morph plants were treated with different BR concentrations. One week later, the style lengths of the treated flowers were measured (Fig. 47).

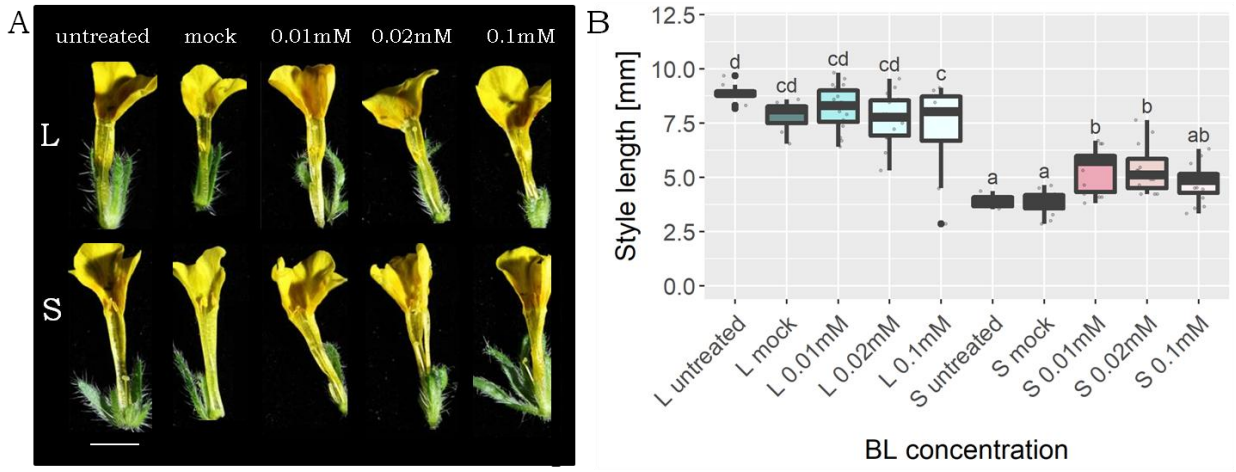


Fig. 47 Reactions of flower buds of S- and L-morphs of *Amsinckia spectabilis* to brassinolide treatment

(A) Dissected flowers of L-morph (top) and S-morph (low) illustrate style size differences during brassinolide (BL) treatment; scale bar: 5mm **(B)** Quantification of style length of S- and L-morphs at different BL concentrations, The letters indicate significant differences as determined by a Tukey's HSD test

The treatment led to a general decrease in style length as it was observed for about 11% for the mock treated L-morphs compared to the untreated ones ($p = 1.4 \times 10^{-13}$). This style shortening pattern was also found for the BR-treated L-morphs which all possessed shorter styles compared to the untreated control L-morph plants. Nevertheless, the lowest BR concentration of 0.01mM BR led to slightly (but not significantly) longer styles, higher BR concentrations resulted in even less different style length compared to the mock treatment. Regarding the S-morph treatment, the treatment itself did not led to a significant difference in style length, which might be due to the general short S-morph style length. As soon as BR was added, a significant increase in style length of the S-morphs was observed ($p \leq 0.00175$). The lowest BR treatment with 0.01mM allowed 40% elongated styles compared to the mock treatment. Higher BR concentrations could not cause a further increase, but rather a decrease, which is expected for BR concentrations above an optimum acting in a growth inhibitory way[274]. Thus, an increase in flower BR concentration is not able to completely rescue the style length phenotype of S-morphs. Thus, it cannot be concluded that a gene encoding for a BR-pathway inhibitor is exclusively representing the *G*-locus in *A. spectabilis.*

4.4.5 Genome-wide identification of the *S*-locus in *Amsinckia spectabilis*

4.4.5.1 Comparative transcriptomics of L- and S-morph style and corolla tissue

To gain first information about the genetic architecture of the *S*-locus in *Amsinckia spectabilis* a differential expression analysis of L-morph and S-morph style and corolla material was performed. Three biological replicates, each consisting of the four oldest completely closed flower buds of one inflorescence from 25 same-morph individuals were collected. After removal of the sepals and the ovary tissue, the remaining flowers were divided into two samples of styles and corolla tubes, including the stamen. Thus, 12 samples in total with six samples per morph were taken for RNA-Seq analysis. In parallel, one S-morph, one L-morph and a small-flowered long homostylous (LH) morph individual were taken for genomic Illumina sequencing. The fact that the heterozygous S-morph *S*-locus exclusively contains the genetic information for both morphs (compared to the recessive L-morph *S*-locus), determined the S-morph reads to be taken as base for the genome assembly. Regarding the genomic sequencing, two different assemblers SPAdes (estimated genome size: 195 Mb) and Megahit (estimated genome size: 248 Mb) were used and compared, leading to the decision to proceed with the Megahit assembly due to a higher average length of assembled contigs and more assembled nucleotides (Fig. 48A). In the following, L- and S-morph transcripts, as well as the L- and S-morph genomes were mapped again the S-morph assembly. This mapping was screened via window sliding for morph-specific expression differences that would hint towards the approximate quantity of *S*-locus (linked) genes and possible candidate markers of the *S*-locus. Thus, expression of the *S*-locus specific genes should only be present in the S-morph samples, whereas significantly fewer (plus mapping or assembly-error based) reads should be found in L-morph samples. Additionally, a tissue-specific filter was included, selecting for markers with significantly higher expression in S-morph styles compared to S-morph corolla, aiming to find *G*-locus candidates. A quite high amount of regions exhibiting strong expression differences were detected (Fig. 48B). This result either argues against a small responsible genomic *S*-locus region, as it was found in *Primula*[235] and *Turnera*[273], containing only few responsible genes, or it is simply a result of the highly fragmented genome assembly. Thus, the causal *S*-locus region possibly got lost due to a lack of coverage beyond the threshold. Further, a shift of the distribution within the overall expression differences towards positive L/S log2 fold changes indicates more regions with higher expression

in L-morph samples compared to S-morph samples that either might be due to higher expression in L-morphs or lower expression in S-morphs or a combination of both. Technical errors, such as unequal and therefore morph-dependant polyadenylation (only polyadenylated mRNA was purified for library preparation) or morph-specific differences of read quality resulting in mapping differences) might be causal as well as biological reasons, which cannot be concluded for sure at this point.

Differentially expressed regions were filtered based on a four-fold reduction of L over S (L/S ratio) and a two-fold enrichment of S- style over S-corolla, 219 candidate markers were found (Fig. 48C). An additional filter was based on the genomic level, with preferably exclusive read counts in the S-genome, resulting in 40 top candidates. Due to the overall fragmented assembly it was decided, to screen through all 219 candidates, ensuring to not miss an interesting marker. These candidate markers were manually selected regarding sufficient assembled reads per marker and repeated occurrence of the same marker (due to the sliding window method). The candidate markers were also screened for too many genomic L-morph reads, finally resulting in 22 possibly *S*-locus linked markers (L/S transcript ratios are shown in Tab. 1).

All identified markers were blasted against the S-morph Illumina reference assembly, resulting in two quite similar regions within the S-genome for nearly all 22 markers possessing at least 93% similarity to the candidate marker (Tab. 2). These findings were the base for two further analyses. Firstly, CAPS or dCAPS markers were possible to design for only 11 of the 22 markers, due to insufficient length of the identified region. The markers were designed to amplify both regions and differentiated them via morph-specific restriction site polymorphisms. This resulted in the confirmation of a morph-specific segregation for 7 of the tested 11 markers for 48 samples (25 L-morphs and 23 S-morphs) (Fig. 49 & Fig. S3). Inconsistencies in the genome reference resulting in incorrectly designed primers could explain the lack of confirmation of more assumed candidate markers or sequencing, assembly and mapping errors might also cause false positive candidates. A confirmation of the RNA-seq data via newly designed primers based on structurally annotated marker genes of the assembly will help to gain a better understanding.

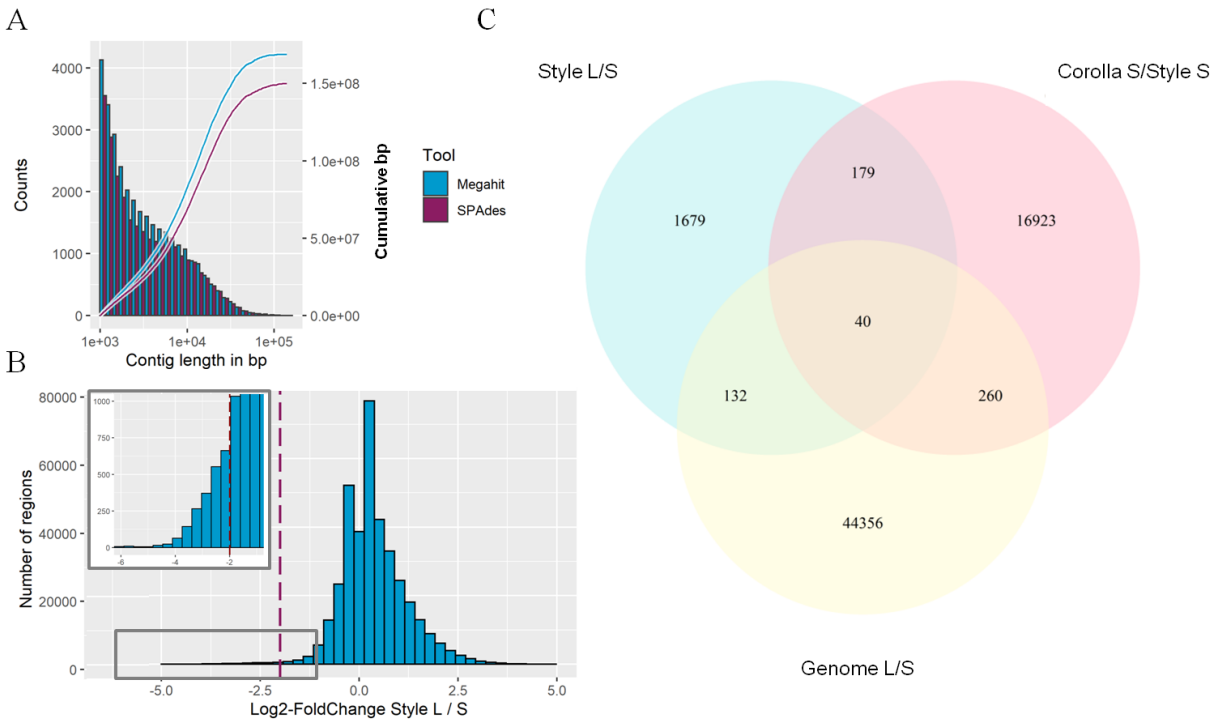


Fig. 48 Overview of DNA- and RNA-Seq analysis and selection of candidate markers

(A) Comparative analysis of assembly tools Megahit and SPAdes regarding distribution of assembled contig lengths (bars and left y-axis) and assembled amount of nucleotides in relation to contig length (lines and right y-axis). **(B)** Distribution of log2 fold changes of L- and S-morph transcripts mapped to S-morph genome. Regions are based on 500 bp sliding windows. Note the high peaks around the middle axis and many very low counts of regions with highest expression differences. Line marks first cut-off for candidate markers. Zoom (grey box) shows continuous distribution of transcripts with most specific S-morph expression. **(C)** Venn diagram of selection process for candidate markers.

Marker	log2 fold change L/S transcripts	Putative gene	PCR-confirmed
1	-3,3784	beta-1,3-galactosyltransferase 7-like	
2	-6,8804	40s ribosomal protein S11	
3	-4,7847	copper transporter 2-like	X
4	-3,8227	No hit	
5	-4,8788	Gibberellin regulated protein,snakin-2-like	X
6	-4,0410	No hit	
7	-3,8463	dystroglycan-like	X
8	-4,7465	tubulin beta-2 chain-like isoform	
9	-3,3882	Phosphatidylinositol N-acetylglucosaminyltransferase subunit P	
10	-4,3971	40S ribosomal protein S16-like	
11	-4,9708	ATP-citrate synthase alpha chain protein 1-like	X
12	-5,1422	probable indole-3-pyruvate monooxygenase YUCCA10	
13	-5,1430	translation machinery-associated protein 22 isoform X1	
14	-5,1571	Zinc finger protein CONSTANS-LIKE 15	
15	-3,2410	annexin D2-like	
16	-5,8192	tubulin beta-1 chain	X
17	-5,8562	protein BOBBER 1	
18	-2,8969	palmitoyl-protein thioesterase 1-like	
19	-7,7623	arabinogalactan peptide 23	X
20	-5,4804	40S ribosomal protein S13	X
21	-4,2690	30S ribosomal protein S6B	
22	-3,1441	acetyl-CoA-benzylalcohol acetyltransferase-like	

Tab. 1 List of candidate markers

Representation of log2 fold changes of L/S transcript ratio. Note that ratios were calculated as morph-specific (not tissue-specific) averaged transcript read counts.

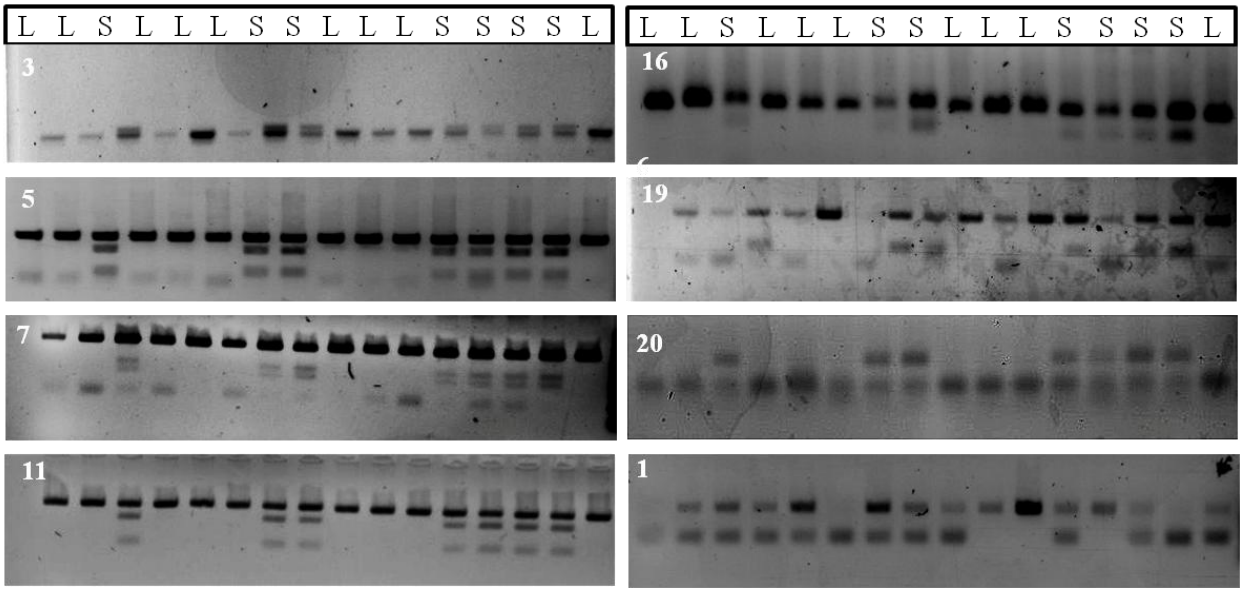
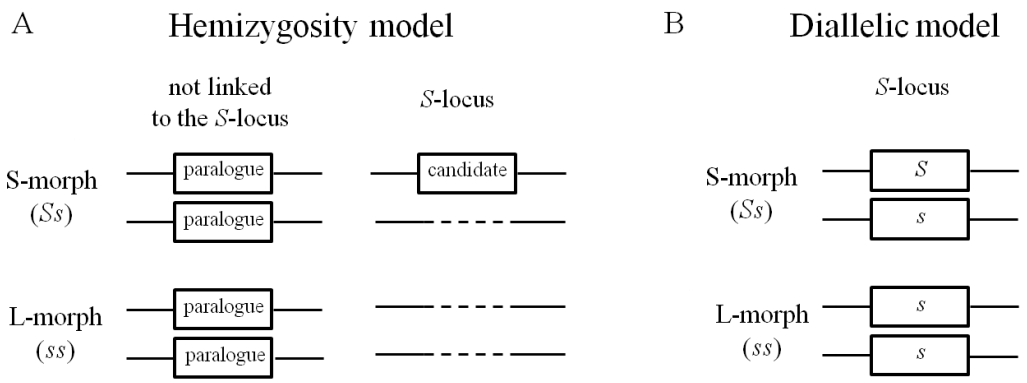


Fig. 49 Morph-specific segregation on genomic level of candidate markers

White numbers refer to marker numbers of Tab.1. Marker 1 represents an example for non-morph-specific segregation as comparison. Note that S-morph samples possess a digest pattern that cannot be found in L-morph samples. DNA bands that can be found in each sample refer to the amplification of the second similar region or allele that is present in both morphs.

Secondly, the genetic base of the relation between the candidate marker and the identified second highly similar region was investigated (identities shown in Tab. 2). Thus, two hypotheses were developed that might explain the observed pattern of (Fig. 50). The first was developed according to the *Primula* model, that assumed a gene duplication to be causal for the development of the *S*-locus[19]. Thus, the two identified regions were supposed to be originated from one *S*-locus specific gene (the candidate), and from the precursor gene located somewhere not linked to the *S*-locus in the genome (the paralogue), respectively. The second hypothesis is based on the assumption that the two hits arose from two different alleles, the dominant *S*- and the recessive *s*-allele. Proving one of these theories would also argue for or against a hemizygous region that contain the candidate markers. Accordingly, the genome based read count S:L ratio of these two hits in both morphs were calculated (Suppl. 7.1). The expected read count ratio for the paralogue theory approximately would be twice as high for the paralogue compared to the *S*-locus candidate gene in the S-morph, due to a genomic two-copy situation of the paralogue and the hemizygous *S*-locus architecture of the candidate gene. For the L-morph paralogue a similar read ratio to S-morph would be expected, whereas no reads for the candidate gene should be

found here. Regarding the second theory, both found S-morph hits should appear in similar ratios and adding both should result in the L-morph allele read count ratio, due to a heterozygous diallelic *S*-locus situation in the S-morphs and a homozygous situation in the L-morphs. With one exception, the read ratios of all candidate regions were found to fulfil the conditions of the second theory with L:S ratios of the second BLAST hit between 0.412-0.83 (Tab. 2) and were thus further arguing against a hemizygous but for a diallelic *S*-locus architecture in *Amsinckia spectabilis*.



Expected genetic read count ratios S/L

Genetic read counts	S-morph	L-morph	S/L
candidate	1	0	NaN
paralogue	2	2	1

Genetic read counts	S-morph	L-morph	S/L
S-allele	1	0	NaN
s-allele	1	2	0.5

C

>*S*-allele_Marker_1
TTCTTCTTCATAAAATAAATAAATCCCTCTTCTTCCACC
TTATAGGAACACCAGCTGGCGGCTGTTAACGAGCATT
CGGATGTTTTAAGTTCAAATGAATGCAGTCGAAAGAG
CCTCTTCACCTTCTCCACACCTCTCATGAACCCACTTA

>*s*-allele_Marker_1
TTCTTCTTCATAAAATAAATAAATCCCTCTTCTTCCACC
TTATAGGAACACCAGCTGACGGCTGTTAACGAGCATT
CGGATGTTTTAAGTTCAAATGAATGCAGTCGAAAGAG
CCTCTTCACCTTCTCCACACCTCTCGTGAACCCACTTA

Genetic read counts	S-morph	L-morph	S/L
S-allele	49	18	2.72
s-allele	46	92	0.5

Fig. 50 Models explaining the presence of two highly similar sequences of candidate markers within the S- and L-morph genomes

(A) Model based on a hemizygous *S*-locus architecture **(B)** Model based on diallelic *S*-locus architecture. Comparison of the expected genomic read count ratios of both models are presented in the tables below. Note that S/L ratios of paralogue and *s*-allele represent the main difference (blue squares). NaN: not a number. **(C)** Marker 1 (see also Tab. 1) as example for the diallelic model. Pink and blue letters in sequences indicate polymorphisms of the two detected highly similar alleles.

Marker	Identity of BLAST hits on assembly (%)	Contigs in genome assembly	Genomic read counts		Ratio S/L
***S*-allele** ***s*-allele**			***S***	***L***	
1	100	k141_103992	49	18	2,722
	98	k141_156335	46	92	**0,500**
2	100	k141_140146	41	5	8,200
	99	k141_51496	59	89	**0,663**
3	100	k141_196847	38	6	7,000
	95	k141_196908	42	85	**0,494**
4	100	k141_205251	25	6	4,167
		No 2nd hit			
5	100	k141_215832	42	5	10,200
	98	k141_357057	42	102	**0,412**
6	100	k141_250207	51	45	1,133
		k141_280729	26	67	**0,388**
7	100	k141_254140	38	5	7,600
	95	k141_254080	78	94	**0,830**
8	100	k141_258212	12	9	1,333
	97	k141_78230	70	61	**1,148**
9	100	k141_260889	45	10	4,500
	93	k141_180029	59	82	**0,720**
10	100	k141_289497	49	9	5,444
	97	K141_91290	40	89	**0,449**
11	100	k141_309157	48	5	9,600
	97	k141_309108	28	66	**0,424**
12	100	k141_31321	19	5	3,800
	99	k141_291779	65	91	**0,714**
13	100	k141_314453	34	5	6,800
	99	k141_86659	35	45	**0,778**
14	100	k141_333738	32	5	6,400
	99	k141_71305	73	103	**0,709**
15	100	k141_338387	32	17	1,882
	94	k141_364857	80	107	**0,748**
16	100	k141_349598	19	5	3,800
	99	k141_164051	40	49	**0,816**
17	100	k141_374145	29	5	5,800
	96	k141_372327	34	90	**0,378**

18	100	k141_42659	16	5	3,200
	97	k141_156700	78	103	**0,757**
19	100	k141_71391	17	5	3,400
	98	k141_44704	51	72	**0,708**
20	100	k141_73500	26	5	5,200
	99	k141_98871	54	80	**0,675**
21	100	k141_76744	39	5	7,800
	97	k141_76755	34	88	**0,386**
22	100	k141_841	56	39	1,436
	99	k141_380247	26	62	**0,419**

Tab. 2 Representation of genomic read counts ratio for all 22 candidate markers indicating a diallelic *S*-locus-model

Bold ratio numbers indicate L/S ratio of the *s*-allele (described in Fig. 50), pointing towards diallelic markers.

Still, the presence of a hemizygous region cannot be excluded at this point, due to the highly fragmented Illumina genome assembly that might have caused the loss of the possibly small hemizygous region.

4.4.5.2 Pacific BioSciences single-molecule real-time sequencing

To improve the genome assembly, single-molecule real-time sequencing (SMRT), developed by Pacific BioSciences (PacBio), was performed. PacBio sequencing produces longer reads and thus contigs compared to Illumina sequencing (Fig. 48A & 51A) that allows sequencing through repetitive regions and therefore a strongly improved structural understanding of the analysed genome[275].

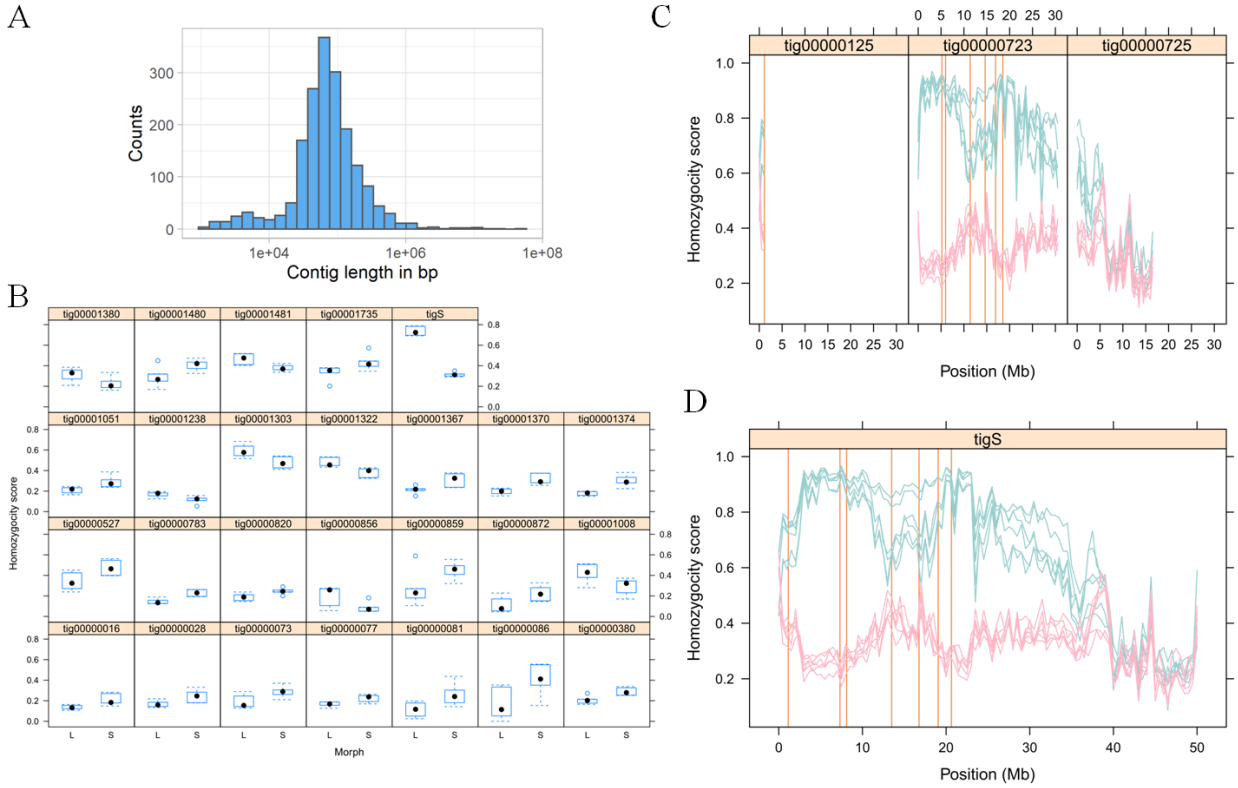


Fig. 51 Overview of PacBio Sequencing results

(A) Distribution of contig length based on PacBio sequencing. **(B)** Overview of all contigs exhibiting a significantly different homozygosity score between L- and S-morph. **(C)** Illustration of the homozygosity score of the three largest contigs (tig00000125, tig00000723, tig00000725) with the highest significant morph-specific differences. **(D)** tigS is formed based on combining the three contigs from (C) (plus four smaller contigs) most likely including the *S*-locus. Orange lines indicate 7 PCR-confirmed markers; pink: S-morph samples; bluegreen: L-morph samples

To ensure that the supposed hemizygous region will not get lost again within the assembly due to low coverage within the heterozygous *S*-locus of the S-morph, a dominant homozygous S-morph (*SS*) individual was taken for PacBio sequencing together with a L-morph individual, both originated from the same selfed heterozygous S-morph ancestor. The PacBio assembly led to an estimated genome size of approximately 300 Mb. Based on the assumption that the *S*-locus

region would be marked by a significantly higher homozygosity in L-morphs *(ss*) compared to S-morphs (*Ss*), the following analysis was performed. Here, the transcriptome data of L- and S-morphs were mapped against the PacBio *SS* assembly and SNP variances (variant calling) between the two morphs were calculated forming the basis for generating the homozygosity score. Thus, contigs with highest significant differences of the homozygosity score between S- and L-morphs were assumed to point towards a region possessing the expected characteristics of the *S*-locus. Contigs that showed a significant homozygosity difference between both morphs are shown in Figure 51B (all in Fig. S4-5). Here, the tigS contig (top right Fig. 51B) represents a combination (final refinements of the transitions between each contig are still missing) of the following 7 contigs in the corresponding order: tig00000125 (1.5 Mb), tig00000126, tig00000127, tig00000128, tig00000723 (30 Mb), tig00000724 and tig00000725 (17 Mb), shown in Figure 51D. The three largest contigs are represented in Figure 51C and exhibited the highest significance differences. Thus, tigS includes the region that differs the most between the two morphs regarding homozygosity.

In parallel, the positions of the 22 previously proposed markers, were located within the assembly. All 22 candidate markers were found to be located in two of the seven contigs (Suppl. 7.4) that were combined to tigS, (orange lines represents seven PCR-confirmed markers, Fig 52C-D), strongly supporting the assumption that tigS contains the *S*-locus and further that all 22 marker are linked to this region. Therefore, they were assumed as true positive *S*-locus genes or are at least linked to it. The complete tigS possesses a size of 50 Mb, whereby the presumable *S*-locus is located between 0-23 Mb. This region most likely marks the boundaries of the region that is exposed to suppressed recombination, due to the fact that the difference in homozygosity clearly decreases downstream of this position. Additionally, the PCR-confirmed markers are located over this 23Mb region. Within this region, only very small regions with hemizygous character (exclusively present in S-morph but absent in L-morph) could be defined. One of them includes a duplication of a putative TCP14 class homolog (position within tigS: 8440818-8441840 and 8411318-8412340, Suppl. 7.5). A PCR-based confirmation of this possible hemizygous region would be needed to allow a final conclusion regarding the molecular nature if this insertion. However, no hemizygous region consisting of more than two genes, was detected. This result rather argues against a hemizygous region as exclusive reason for the suppressed recombination of the *S*-locus, since at least three different morph-specific phenotypes (*G*-, *A*- and

P-locus) are expected to be genetically determined by it. Therefore, the genetic architecture of the *Amsinckia spectabilis S*-locus differs from that revealed for other heterostylous species such as *Turnera* and *Primula*, providing a so far unknown genetic base of the *S*-locus organization.

Marker	Position on tigS contig	Similarities of surrounding regions
5	1127707	Unclear
19	7313157	L-like
16	8111108	L-like
20	13491224	Unclear
3	16766643	Unclear
11	19051984	L-like
7	20638077	L-like

Tab. 3 Basal similarities of long-homostylous genome to the S- or the L-morph genome within the surrounding regions of the 7 confirmed markers

Basal similarities were determined by manual investigation of the surrounding regions of the marker positions. Markers are sorted according to their positions on tigS.

Additionally, the genome data of the small-flowered long homostyle individual, together with the S-morph and L-morph genomic Illumina reads, were mapped against the S-morph PacBio assembly. To gain more insight into the basal genetic mechanism of the transition from the large-flowered heterostyly to the small-flowered fully selfing *Amsinckia spectabilis* form, the long homostylous genome reads were compared to the genomic L-and S-morph Illumina reads based on overall similarities (manually screened in a 1 kb window surrounding the 7 PCR-confirmed marker positions) (Tab. 3). Since all seven markers are distributed over the assumed *S*-locus, a change in similarities towards L- or S-morph within the markers (assumed chromosomal breaking point) would argue for a recombination event that led to the long homostylous version. On the other hand, a similarity to one morph over the complete region rather supports one or more mutations to be responsible for the new phenotype. Surprisingly, the long homostylous was found to resemble the L-morph in most of the seven markers, raising the question, if the long homostylous form has emerged from the L-morph. These findings rather argue for a mutation to be causal for the development of the new floral morph, than for a recombination event.

4.5 Discussion

Amsinckia represents a special case of heteromorphic SI, due to its heterostylous flowers, unusually combined with no obvious SI and the repeated breakdown of this heterostyly to homostyly and finally the transition to full selfing[256–258]. This unexpected pattern of SI led to many unanswered questions. The results of this PhD project provide first and very interesting answers that allowed a better understanding of the genetic base of the *Amsinckia spectabilis S*-locus. As a starting point, a precise analysis of the specific phenotypes of the two floral morphs in the heterostylous *Amsinckia spectabilis* was carried out. In general, all previously published phenotypic differences[254,255], such as style length (56%) and anther height differences (58%) and morph-specific pollen size differences (21%), could be confirmed. Additionally, it was demonstrated that 28% and 20% of the observed differences in style length and anther height, respectively, can be explained by different cell lengths within the style and the corolla tissue. To be able to determine, whether a further increase in cell number is added to the cell size increase and is therefore responsible for the remaining 28% and 26% of style organ size and anther height differences, further investigations need to be done. Here, confocal microscopy of the complete style and corolla cells could help to answer this question.

Furthermore, inter- and intramorph crosses formed the experimental design for the analysis regarding the presence of SI in *Amsinckia spectabilis*. S to L, L to S and S to S crosses exhibited a similar degree of mating efficiency with approximately 70%. Unexpectedly, the L to L cross possessed a significantly lower mating efficiency, proven by a decreased seed set and by reduced pollen tube growth through the style found in aniline blue stained flowers. These results were rather unexpected, since a SI-mechanism, if present, would be particularly expected to happen in intramorph S-crosses to prevent the production of mature *SS* seeds. Nevertheless, also a reduced L to L reproductive success might somehow enhance preferential intermorph crossings over intramorph crossing. Additionally, the development of dominant homozygous *SS*-plants after two rounds of selfing presents a further argument against a (strong) SI in *Amsinckia spectabilis*. Overall, these results strongly imply that the *Amsinckia spectabilis S*-locus encodes at least three traits, most likely genetically determined by three different loci, the *G*-, the *A*- and the *P*-locus, whereas no obvious hint for SI could be observed. Still, it remains unclear, how the 1:1 morph ratio in naturally occurring populations is maintained. The absence of homozygous dominant

(*SS*) plants might be a possible explanation. Since no prezygotic and prematuration rejections of *SS* seeds could be detected, a fitness decrease, such as a delayed or inhibited germination might be plausible. Corresponding examples could be found in other species, exhibiting sexual dimorphism, such as *Rumex nivalis*[276] or *Spinacia oleraceae*[277] with heavier male seeds and earlier germination than female seeds and *Silene latifolia*[278] with sex-dependent differences in dormancy and survivorship. Furthermore, a slower or decreased total stem height leading to a pre-flowering developmental arrest, caused by a lack of light due to overgrowth of neighbouring plants, could cause this result. Moreover, lower attractiveness for pollinators, of the *SS* plants compared to *ss* and *Ss* plants, might be an additional explanation. Regarding this, field studies with cameras filming pollinator visits of *Bombyliid* flies, butterflies and bees (in the genera *Anthophora*, *Osmia Bombus* and *Synhalonia),* which are all known to pollinate *Amsinckia*[279], could provide a better understanding. Here, pollination between and within the different *Amsinckia spectabilis* forms could be compared, possibly allowing an assessment, whether pollinator body size and behaviour could exclusively explain the equal L:S ratios in natural populations. Another molecular explanation might be a dose-dependent fitness-reducing effect of two copies of the dominant *S*-allele compared to one copy in the heterozygous S-morph. However, an interaction of the listed and additional effects might be the missing answer regarding the balancing of the 1:1 morph ratio.

Further analysis regarding the genetic architecture of the *S*-locus was based on RNA-Seq of L- and S-morph flower buds divided into corolla and style tissue. In parallel, individual genomic Illumina sequencing of L- and S- (*Ss* and *SS*) morphs and of a long homostylous small-flowered individual, was carried out. Based on the RNA-Seq, 22 markers were found that represented a significant differential expression between L- and S-morph. The seven markers could be confirmed to segregate in a morph-specific manner based on PCR-analysis. For all of them predominantly diallelic behaviour could be calculated via read counting and thus there is no indication for one of them to be a hemizygous marker in the *Amsinckia spectabilis S*-locus. The subsequent PacBio assembly allowed the confirmations of all 22 markers to be *S*-locus or *S*-locus linked, since all of them were found to be located on the two contigs that possessed the most significantly different homozygosity score between the two morphs, as it would be expected for a region with suppressed recombination compared to the rest of the genome. Five smaller contigs were combined with these two and formed a region of 50 Mb, very likely including the

Amsinckia spectabilis S-locus with an approximated size of 23 Mb. No hemizygous region could be detected within this area consisting of more than two predicted genes. Still, a duplicated putative TCP14 class I gene was found to be inserted in the S-morph genome that might be a plausible candidate for the *A*-locus. Besides causing further phenotypes, this gene was shown to increase cell proliferation in floral tissues in *Arabidopsis thaliana*[280], where its role in stamen filament elongation was demonstrated[281]. Whether the *Amsinckia spec.* putative TCP14 homologue possesses a similar function and if this function is responsible for a S-morph-floral phenotype cannot be concluded without further experiments. Here, the virus-induced-gene silencing (VIGS)[282,283], which was developed for *Amsinckia spectabilis* in the Lenhard lab during a master project (unpublished data), would provide a fast and efficient way to test if degraded TCP14-mRNA would change the S-morph flower phenotype into a short homostylous form. Since no indications for hemizygous encoded *G*- and *P*-loci, or three different hemizygous loci at all, were found, another mechanism must explain the suppressed recombination of the *S*-locus that allows the stabilization of the two floral morphs in *Amsinckia spectabilis* populations. On the experimental level, a supporting argument for the absent or at least strongly decreased recombination of the *S*-locus was the lack of recombinants within all PCR-marker tested 48 individuals. Still, a higher sample size would be necessary to estimate the level of recombination frequency. According to previous studies on other heterostylous species, such as *Turnera*[246], inverted chromosomal areas can also explain the absent recombination within the *S*-locus. Here, the hemizygous *S*-locus is surrounded by two inverted segments, including 14 and three genes, respectively. Until today, for all revealed plant *S*-loci a hemizygous region was found to be causal, thus no example with an inverted region exclusively responsible for the suppressed recombination is known in the plant kingdom. In animals, on the other hand, an inverted supergene (> 100 Mb) was found to cause the determination of two morphs, including head colour as well as reproductive behaviour, of the white-throated sparrow *Zonotrichia albicollis*. Interestingly, the very rarely occurring homozygous dominant sparrow offspring exhibits a strong size reduction in males, whereas females scarcely experience parental care, both representing a strong fitness decrease[284]. Whether an inverted region is causal for the suppressed recombination in *Amsinckia spectabilis S*-locus cannot be concluded at this point. Further analysis, especially of very long sequencing reads would be necessary to detect possible inversion sites. Therefore, a repetition of the L-morph PacBio sequencing needs to be carried

out, since the coverage here was too low to serve as comparison to the homozygous S-morph PacBio assembly. In case that still no responsible morph-specific presence/absence behaviour of genes (hemizygosity) will be detected for the *Amsinckia spectabilis S*-locus, the question for the molecular mechanism underlying the morph-specific differences arises, especially with such a high similarity between the dominant *S*- and recessive *s*-alleles. Therefore, literature provides different but also overlapping results regarding supergenes that are exposed to strongly suppressed recombination. Regarding the previously described white-throated sparrows, a decreased expression of the inverted genes of the dominant allele was detected in heterozygous birds. A more detailed understanding of these expression differences is still missing. Thus tissue-specific effects or a very recent degradation of the dominant allele were assumed to be causal[284]. However, these findings show strong similarities to neo-sex chromosomes, where functional degradation is accompanied by lower average gene expression of the dominant chromosome. For instance, the neo-Y chromosome in *Drosophila miranda* exhibit a significantly lower expression of 80% of the genes compared to the neo-X chromosome regardless of the level of amino acid substitutions or the degree of functionality of the translated proteins. Here, the accumulation of deleterious mutations was found to be causal, leading to reduced fitness, due to the inefficiency of natural selection acting on a non-recombining chromosome[285]. In the plant kingdom, similar observations for sex chromosomes were made in *Rumex hastatulus.* Again, selection inefficiency due to suppressed recombination was detected to be causal for reduced gene expression of the dominant Y-allele and an accumulation of amino acid substitutions. Additionally, 28% of the genes of the ancestral X-chromosome were hemizygous, indicating gene loss on and therefore degeneration of the Y-chromosome[286]. However, two theories might be proposed regarding the underlying molecular mechanism of the *s*- and *S*-allele-specific differences of the *Amsinckia spectabilis S*-locus. First, similarly to previously described examples, differential expression might be responsible. Second, proteins on the *S*-allele possibly possess new or improved functions, which might explain the morph-specific differences (as it was found for GLO2 in *Primula*[241]), similar to dominant negative mutations. For a detailed understanding, further analyses are needed. Based on an improved PacBio sequencing with repeated circular read-through of each DNA-fragment resulting in a strongly reduced sequencing error rate (compared to the available PacBio assembly based on continuous long read mode), a structural annotation of tigS genes, including possible promoter regions, 5’ and 3’ UTR and introns would be necessary

for a comparative study regarding causal amino acid substitutions, deletions, insertions, frameshifts of both alleles. Regarding the first theory, a new mapping of the RNA-Seq data against this improved PacBio assembly should allow for a statement regarding expression differences of the dominant and recessive alleles within tigS. Further investigations, especially functional analyses of candidate genes, are required to determine, whether detected differences are exclusively responsible for the formation of the two floral morphs.

Regarding the evolution of the small-flowered long homostylous *Amsinckia spectabilis* form, a analysis detected higher similarities to the L-morph as to the S-morph genome for most of the seven markers within the tigS region. Thus, it might be possible that the long homostylous form derived from the L-morph, pointing towards a mutated *A*-locus that resulted in high positioned anthers. Overall, this comparison could only be carried out in a very superficial manner, due to the high sequencing error rate in the PacBio assembly and the different original populations the individuals were taken from for Illumina genome analysis and PacBio sequencing. For a more detailed investigation, sequence comparison based on variant calling would help to get a more precise picture of the genomic architecture of the long-homostylous form. A promising approach, might be an additional next generation sequencing of the intermediate long-flowered long homostylous version. Further, more detailed analyses are needed to reveal, if long homostylous forms developed by mutation or by rare recombination events.

The unusual occurrence of repeated transition toward selfing in *Amsinckia* still remains unclear. An ecological explanation might be the fact that *Amsinckia* are colonizers of unstable or pioneer habitats, a characteristic that would be simplified by homostyly and selfing, and thus a trait nature would select for[21,255,287]. Further disadvantages of heterostyly are the reduced amount of matching mating partners within one population, a weak heteromorphic SI and the dependency on pollinators, that might cause a relaxation of the frequency-dependant selection that maintains the two floral morphs within one population and therefore the promotion towards homostyly and selfing[233].

Regarding the evolutionary base of the *S*-loci in the *Boraginaceae* family, sequencing analyses of other species would be required. For example, as closely related species and for an intra-genus comparison *Amsinckia douglasiana* might be a good candidate. Published RNA-Seq data of

phylogenetically more distanced species, such as *Lithospermum multiflorum* or *Oreocarya*[288], could help to reveal, whether the *Boraginaceae S*-loci are originated from one ancestor, or if a similar genetic event happened repeatedly. According to this, two possible theories were proposed. First, the ancestral state consists of both polymorphic alleles and remained only in few of the *Boraginaceae* genera, or second, one of the alleles entered the genera via hybridization with a closely related species that was followed by introgression, as it is known for the white-throated sparrows[284].

Thus, many questions regarding the heterostyly and its repeated transition to homostyly in *Amsinckia spectabilis* still remain to be answered and will hopefully be resolved in the near future.

5. References

1. Gonzalez N, Vanhaeren H, Inzé D. Leaf size control: Complex coordination of cell division and expansion. *Trends Plant Sci*. 2012;17(6):332-340. doi:10.1016/j.tplants.2012.02.003
2. Breuer C, Ishida T, Sugimoto K. Developmental control of endocycles and cell growth in plants. *Curr Opin Plant Biol*. 2010;13(6):654-660. doi:10.1016/j.pbi.2010.10.006
3. Planas-Riverola A, Gupta A, Betegoń-Putze I, Bosch N, Ibaneş M, Cano-Delgado AI. Brassinosteroid signaling in plant development and adaptation to stress. *Dev*. 2019;146(5). doi:10.1242/dev.151894
4. Gao X, Zhang Y, He Z, Fu X. *Gibberellins*.; 2017. doi:10.1016/B978-0-12-811562-6.00004-9
5. Hülskamp M. Cell morphogenesis: How plants split hairs. *Curr Biol*. 2000;10:R308-R310. doi:10.1016/S0960-9822(00)00437-1
6. Gan Y, Kumimoto R, Liu C, Ratcliffe O, Yu H, Broun P. GLABROUS INFLORESCENCE STEMS modulates the regulation by gibberellins of epidermal differentiation and shoot maturation in Arabidopsis. *Plant Cell*. 2006;18(6):1383-1395. doi:10.1105/tpc.106.041533
7. Schwab B, Folkers U, Ilgenfritz H, Hülskamp M. Trichome morphogenesis in Arabidopsis. *Philos Trans R Soc B Biol Sci*. 2000;355(1399):879-883. doi:10.1098/rstb.2000.0623
8. Plackett ARG, Wilson ZA. Gibberellins and Plant Reproduction. In: *Annual Plant Reviews*. Vol 49. ; 2016:323-358. doi:10.1002/9781119312994.apr0540
9. Caitlyn Q Herr and Robert P. Hausinger. Amazing Diversity in Biochemical Roles of Fe(II)/2-Oxoglutarate Oxygenases. *Trends Biochem Sci*. 2018;43(7):517-532. doi:10.1016/j.physbeh.2017.03.040
10. Kawai Y, Ono E, Mizutani M. Evolution and diversity of the 2-oxoglutarate-dependent dioxygenase superfamily in plants. *Plant J*. 2014;78:328-343. doi:10.1111/tpj.12479
11. Mateo-Bonmatí E, Esteve-Bruna D, Juan-Vicente L, et al. INCURVATA11 and CUPULIFORMIS2 are redundant genes that encode epigenetic machinery components in arabidopsis. *Plant Cell*. 2018;30(7):1596-1616. doi:10.1105/tpc.18.00300
12. Mccubbin AG, Kao T. Molecular Recognition and Response in Pollen and Pistil Interactions. *Annu Rev Cell Dev Biol*. 2000;16:333-364. doi:10.1146/annurev.cellbio.16.1.333
13. Takayama S, Isogai A. Self-Incompatibility in Plants. *Annu Rev Plant Biol*. 2005;56:467-489. doi:10.1146/annurev.arplant.56.032604.144249
14. Iwano M, Takayama S. Self / non-self discrimination in angiosperm self-incompatibility. *Curr Opin Plant Biol*. 2012;15(1):78-83. doi:10.1016/j.pbi.2011.09.003
15. Lewis BYD, Innes J. Incompatibility in flowering plants. 1949:472-496.

16. Barrett SCH. The evolution of plant sexual diversity. *Nat Rev Genet*. 2002;3(4):274-284. doi:10.1038/nrg776
17. Barrett SCH, Cruzan MB. Incompatibility in heterostylous plants. In: *Genetic Control of Self-Incompatibility and Reproductive Development in Flowering Plants*. ; 1994:189-219. doi:10.1007/978-94-017-1669-7_10
18. Barrett SCH, Shore JS. *New Insights on Heterostyly : Comparative Biology, Ecology and Genetics*.; 2008.
19. Kappel C, Huu CN, Lenhard M. A short story gets longer: Recent insights into the molecular basis of heterostyly. *J Exp Bot*. 2017;68(21-22):5719-5730. doi:10.1093/jxb/erx387
20. Ganders FR. Breeding systems and genetic variation in Amsinckia spectabilis (Boraginaceae). *Can J Bot*. 1985;63:533-538.
21. Ray PM, Chisaki HF. Studies on Amsinckia . II . Relationships Among the Primitive Species. *Am J Bot*. 1957;44(6):537-544.
22. Bartkowska MP, Johnston MO. QUANTITATIVE GENETIC VARIATION IN POPULATIONS OF AMSINCKIA SPECTABILIS THAT DIFFER IN RATE OF SELF-FERTILIZATION. *Evolution (N Y)*. 2009;63-5(1957):1103-1117. doi:10.1111/j.1558-5646.2008.00607.x
23. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*. 2003;31(13):3784-3788. doi:10.1093/nar/gkg563
24. Goldberg RB. Plants: Novel developmental processes. *Science*. 1988;240:1460-1467. doi:10.1126/science.3287622
25. Meyerowitz EM. Genetic control of cell division patterns in developing plants. *Cell*. 1997;88(3):299-308. doi:10.1016/S0092-8674(00)81868-1
26. Figueiredo DD, Batista RA, Roszak PJ, Hennig L, Köhler C. Auxin production in the endosperm drives seed coat development in Arabidopsis. *Elife*. 2016;5:1-23. doi:10.7554/eLife.20542
27. Fiume E, Coen O, Xu W, Lepiniec L, Magnani E. Growth of the Arabidopsis sub-epidermal integument cell layers might require an endosperm signal. *Plant Signal Behav*. 2017;12(8):8-10. doi:10.1080/15592324.2017.1339000
28. Beauzamy L, Fourquin C, Dubrulle N, Boursiac Y, Boudaoud A, Ingram G. Endosperm turgor pressure decreases during early Arabidopsis seed development. *Dev*. 2016;143(18):3295-3299. doi:10.1242/dev.137190
29. Brown RC, Lemmon BE, Nguyen H, Olsen OA. Development of endosperm in Arabidopsis thaliana. *Sex Plant Reprod*. 1999;12(1):32-42. doi:10.1007/s004970050169
30. Bereterbide A, Hernould M, Castera S, Mouras A. Inhibition of cell proliferation, cell expansion

and differentiation by the Arabidopsis SUPERMAN gene in transgenic tobacco plants. *Planta*. 2001;214(1):22-29. doi:10.1007/s004250100584

31. Horiguchi G, Tsukaya H. Organ size regulation in plants: Insights from compensation. *Front Plant Sci*. 2011;2:1-6. doi:10.3389/fpls.2011.00024
32. Gonzalez N, Beemster GT, Inzé D. David and Goliath: what can the tiny weed Arabidopsis teach us to improve biomass production in crops? *Curr Opin Plant Biol*. 2009;12(2):157-164. doi:10.1016/j.pbi.2008.11.003
33. Czesnick H, Lenhard M. Size Control in Plants — Lessons from Leaves. *Cold Spring Harb Perspect Biol*. 2015;7:1-16.
34. Cosgrove DJ. Growth of the plant cell wall. *Nat Rev Mol Cell Biol*. 2005;6(11):850-861. doi:10.1038/nrm1746
35. Cosgrove DJ. Diffuse growth of plant cell walls. *Plant Physiol*. 2018;176(1):16-27. doi:10.1104/pp.17.01541
36. Cho HT, Cosgrove DJ. Altered expression of expansin modulates leaf growth and pedicel abscission in Arabidopsis thaliana. *Proc Natl Acad Sci U S A*. 2000;97(17):9783-9788. doi:10.1073/pnas.160276997
37. Kwon YR, Lee HJ, Kim KH, Hong SW, Lee SJ, Lee H. Ectopic expression of Expansin3 or Expansinβ1 causes enhanced hormone and salt stress sensitivity in Arabidopsis. *Biotechnol Lett*. 2008;30(7):1281-1288. doi:10.1007/s10529-008-9678-5
38. Smith LG, Oppenheimer DG. Spatial Control of Cell Expansion By the Plant Cytoskeleton. *Annu Rev Cell Dev Biol*. 2005;21(1):271-295. doi:10.1146/annurev.cellbio.21.122303.114901
39. Yang Y, Chen B, Dang X, et al. Arabidopsis IPGA1 is a microtubule-associated protein essential for cell expansion during petal morphogenesis. *J Exp Bot*. 2019;70(19):5231-5243. doi:10.1093/jxb/erz284
40. Deprost D, Yao L, Sormani R, et al. The Arabidopsis TOR kinase links plant growth, yield, stress resistance and mRNA translation. *EMBO Rep*. 2007;8(9):864-870. doi:10.1038/sj.embor.7401043
41. Krizek BA. Making bigger plants: key regulators of final organ size. *Curr Opin Plant Biol*. 2009;12(1):17-22. doi:10.1016/j.pbi.2008.09.006
42. Ren M, Qiu S, Venglat P, et al. Target of Rapamycin Regulates Development and Ribosomal RNA Expression through Kinase Domain in Arabidopsis. *Plant Physiol*. 2011;155:1367-1382. doi:10.1104/pp.110.169045
43. Leiber RM, John F, Verhertbruggen Y, Diet A, Knox JP, Ringli C. The TOR pathway modulates the structure of cell walls in Arabidopsis. *Plant Cell*. 2010;22(6):1898-1908. doi:10.1105/tpc.109.073007

44. Lang L, Schnittger A. Endoreplication — a means to an end in cell growth and stress response. *Curr Opin Plant Biol*. 2020;54:85-92. doi:10.1016/j.pbi.2020.02.006
45. Wang H, Dittmer TA, Richards EJ. Arabidopsis CROWDED NUCLEI (CRWN) proteins are required for nuclear size control and heterochromatin organization. *BMC Plant Biol*. 2013;13(1):1-13. doi:10.1186/1471-2229-13-200
46. Robinson DO, Coate JE, Singh A, et al. Ploidy and size at multiple scales in the arabidopsis sepal. *Plant Cell*. 2018;30(10):2308-2329. doi:10.1105/tpc.18.00344
47. Breuninger H, Lenhard M. Control of tissue and organ growth in plants. *Curr Top Dev Biol*. 2010;91(C):185-220. doi:10.1016/S0070-2153(10)91007-7
48. Sugimoto-Shirasu K, Roberts K. "Big it up": Endoreduplication and cell-size control in plants. *Curr Opin Plant Biol*. 2003;6(6):544-553. doi:10.1016/j.pbi.2003.09.009
49. Melaragno JE, Mehrotra B, Coleman AW, Coleman W. Relationship between Endopolyploidy and Cell Size in Epidermal Tissue of Arabidopsis. *Plant Cell*. 2013;5(11):1661-1668.
50. Brady T. Feulgen Cytophotometric determination of the DNA Content of the Embryo Proper and Suspensor Cells of Phaseolus coccineus. *Cell Differ*. 1973;75:65-75.
51. Inzé D, De Veylder L. Cell Cycle Regulation in Plant Development. *Annu Rev Genet*. 2006;40(1):77-105. doi:10.1146/annurev.genet.40.110405.090431
52. Nowack MK, Harashima H, Dissmeyer N, et al. Genetic Framework of Cyclin-Dependent Kinase Function in Arabidopsis. *Dev Cell*. 2012;22(5):1030-1040. doi:10.1016/j.devcel.2012.02.015
53. Sonoda Y, Sako K, Maki Y, et al. Regulation of leaf organ size by the Arabidopsis RPT2a 19S proteasome subunit. *Plant J*. 2009;60(1):68-78. doi:10.1111/j.1365-313X.2009.03932.x
54. Yoshizumi T, Tsumoto Y, Takiguchi T, et al. Increased level of polyploidy1, a conserved repressor of CYCLINA2 transcription, controls endoreduplication in Arabidopsis. *Plant Cell*. 2006;18(10):2452-2468. doi:10.1105/tpc.106.043869
55. Vanstraelen M, Benková E. Hormonal Interactions in the Regulation of Plant Development. *Annu Rev Cell Dev Biol*. 2012;28(1):463-487. doi:10.1146/annurev-cellbio-101011-155741
56. Gampala SS, Kim TW, He JX, et al. An Essential Role for 14-3-3 Proteins in Brassinosteroid Signal Transduction in Arabidopsis. *Dev Cell*. 2007;13(2):177-189. doi:10.1016/j.devcel.2007.06.009
57. De Rybel B, Audenaert D, Vert G, et al. Chemical Inhibition of a Subset of Arabidopsis thaliana GSK3-like Kinases Activates Brassinosteroid Signaling. *Chem Biol*. 2009;16(6):594-604. doi:10.1016/j.chembiol.2009.04.008
58. Wang ZY, Seto H, Fujioka S, Yoshida S, Chory J. BRI1 is a critical component of a plasma-membrane receptor for plant steroids. *Nature*. 2001;410(6826):380-383. doi:10.1038/35066597

59. Choe S, Fujioka S, Noguchi T, Takatsuto S, Yoshida S, Feldmann KA. Overexpression of DWARF4 in the brassinosteroid biosynthetic pathway results in increased vegetative growth and seed yield in Arabidopsis. *Plant J*. 2001;26(6):573-582. doi:10.1046/j.1365-313X.2001.01055.x
60. Coll-Garcia D, Mazuch J, Altmann T, Müssig C. EXORDIUM regulates brassinosteroid-responsive genes. *FEBS Lett*. 2004;563(1-3):82-86. doi:10.1016/S0014-5793(04)00255-8
61. Jones AM. Auxin-Dependent Cell Expansion Mediated by Overexpressed Auxin-Binding Protein 1 *Science*. 1998;282(5391):1114-1117. doi:10.1126/science.282.5391.1114
62. Schruff MC, Spielman M, Tiwari S, Adams S, Fenby N, Scott RJ. The AUXIN RESPONSE FACTOR 2 gene of Arabidopsis links auxin signalling, cell division, and the size of seeds and other organs. *Development*. 2006;133(2):251-261. doi:10.1242/dev.02194
63. Halliday KJ. Plant hormones: The interplay of brassinosteroids and auxin. *Curr Biol*. 2004;14(23):1008-1010. doi:10.1016/j.cub.2004.11.025
64. Vert G, Walcher CL, Chory J, Nemhauser JL. Integration of auxin and brassinosteroid pathways by Auxin Response Factor 2. *Proc Natl Acad Sci U S A*. 2008;105(28):9829-9834. doi:10.1073/pnas.0803996105
65. Hu, Xie, Chua. The Arabidopsis Auxin-Inducible Gene. *Plant Cell*. 2003;15:1951-1961. doi:10.1105/tpc.013557.was
66. Rai MI. Regulators of Ethylene Signaling in Arabidopsis thaliana: CTR1 and ARGOS Family. 2016.
67. Qin Z, Zhang X, Zhang X, Feng G, Hu Y. The Arabidopsis ORGAN SIZE RELATED 2 is involved in regulation of cell expansion during organ growth. *BMC Plant Biol*. 2014;14(1):1-11. doi:10.1186/s12870-014-0349-5
68. Feng G, Qin Z, Yan J, Zhang X, Hu Y. Arabidopsis ORGAN SIZE RELATED1 regulates organ growth and final organ size in orchestration with ARGOS and ARL. *New Phytol*. 2011;191(3):635-646. doi:10.1111/j.1469-8137.2011.03710.x
69. Hu Y, Poh HM, Chua NH. The Arabidopsis ARGOS-LIKE gene regulates cell expansion during organ growth. *Plant J*. 2006;47(1):1-9. doi:10.1111/j.1365-313X.2006.02750.x
70. Kuluev B, Mikhaylova E, Ermoshin A, et al. The ARGOS-LIKE genes of Arabidopsis and tobacco as targets for improving plant productivity and stress tolerance. *J Plant Physiol*. 2019;242:1-11. doi:10.1016/j.jplph.2019.153033
71. Varaud E, Brioudes F, Szé J, et al. AUXIN RESPONSE FACTOR8 Regulates Arabidopsis Petal Growth by Interacting with the bHLH Transcription Factor BIGPETALp. *Plant Cell*. 2011;23(3):973-983. doi:10.1105/tpc.110.081653
72. Gendreau E, Orbovic V, Höfte H, et al. Gibberellin and ethylene control endoreduplication levels

in the Arabidopsis thaliana hypocotyl. *Springer*. 2016;209(4):513-516.

73. Ishida T, Adachi S, Yoshimura M, Shimizu K, Umeda M, Sugimoto K. Auxin modulates the transition from the mitotic cycle to the endocycle in Arabidopsis. *Development*. 2010;137(1):63-71. doi:10.1242/dev.035840
74. Rieu I, Ruiz-Rivero O, Fernandez-Garcia N, et al. The gibberellin biosynthetic genes AtGA20ox1 and AtGA20ox2 act, partially redundantly, to promote growth and development throughout the Arabidopsis life cycle. *Plant J*. 2008;53(3):488-504. doi:10.1111/j.1365-313X.2007.03356.x
75. Koornneef M, van der Veen JH. Induction and analysis of gibberellin sensitive mutants in Arabidopsis thaliana (L.) heynh. *Theor Appl Genet*. 1980;58(6):257-263. doi:10.1007/BF00265176
76. Coles JP, Phillips AL, Croker SJ, García-Lepe R, Lewis MJ, Hedden P. Modification of gibberellin production and plant development in Arabidopsis by sense and antisense expression of gibberellin 20-oxidase genes. *Plant J*. 1999;17(5):547-556. doi:10.1046/j.1365-313X.1999.00410.x
77. Talon M, Koornneef M, Zeevaart JAD. Endogenous gibberellins in Arabidopsis thaliana and possible steps blocked in the biosynthetic pathways of the semidwarf ga4 and ga5 mutants. *Proc Natl Acad Sci U S A*. 1990;87(20):7983-7987. doi:10.1073/pnas.87.20.7983
78. Aach H, Bode H, Robinson DG, Graebe JE. ent-Kaurene synthase is located in proplastids of meristematic shoot tissues. *Planta*. 1997;202(2):211-219. doi:10.1007/s004250050121
79. Fleet CM, Yamaguchi S, Hanada A, et al. Overexpression of AtCPS and AtKS in Arabidopsis confers increased ent-kaurene production but no increase in bioactive gibberellins. *Plant Physiol*. 2003;132(2):830-839. doi:10.1104/pp.103.021725
80. Helliwell CA, Sheldon CC, Olive MR, et al. Cloning of the Arabidopsis ent-kaurene oxidase gene GA3. *Proc Natl Acad Sci U S A*. 1998;95(15):9019-9024. doi:10.1073/pnas.95.15.9019
81. Hedden P, Phillips AL. *Gibberellin Metabolism: New Insights Revealed by the Genes*. Vol 5.; 2000.
82. Helliwell CA, Sullivan JA, Mould RM, Gray JC, James Peacock W, Dennis ES. A plastid envelope location of Arabidopsis ent-kaurene oxidase links the plastid and endoplasmic reticulum steps of the gibberellin biosynthesis pathway. *Plant J*. 2001;28(2):201-208. doi:10.1046/j.1365-313X.2001.01150.x
83. Regnault T, Daviere JM, Heintz D, Lange T, Achard P. The gibberellin biosynthetic genes AtKAO1 and AtKAO2 have overlapping roles throughout Arabidopsis development. *Plant J*. 2014;80(3):462-474. doi:10.1111/tpj.12648
84. Magome H, Nomura T, Hanada A, et al. CYP714B1 and CYP714B2 encode gibberellin 13-

oxidases that reduce gibberellin activity in rice. *Proc Natl Acad Sci U S A*. 2013;110(5):1947-1952. doi:10.1073/pnas.1215788110

85. Plackett ARG, Powers SJ, Fernandez-Garcia N, et al. Analysis of the developmental roles of the arabidopsis gibberellin 20-oxidases demonstrates that GA20ox1, -2, and -3 are the dominant paralogs. *Plant Cell*. 2012;24(3):941-960. doi:10.1105/tpc.111.095109
86. Li S, Liu Y, Zheng L, et al. The plant-specific G protein γ subunit AGG3 influences organ size and shape in Arabidopsis thaliana. *New Phytol*. 2012;194(3):690-703. doi:10.1111/j.1469-8137.2012.04083.x
87. Mitchum MG, Yamaguchi S, Hanada A, et al. Distinct and overlapping roles of two gibberellin 3-oxidases in Arabidopsis development. *Plant J*. 2006;45(5):804-818. doi:10.1111/j.1365-313X.2005.02642.x
88. Ueguchi-Tanaka M, Ashikari M, Nakajima M, et al. GIBBERELLIN INSENSITIVE DWARF1 encodes a soluble receptor for gibberellin. *Nature*. 2005;437(7059):693-698. doi:10.1038/nature04028
89. Varbanova M, Yamaguchi S, Yang Y, et al. Methylation of gibberellins by Arabidopsis GAMT1 and GAMT2. *Plant Cell*. 2007;19(1):32-45. doi:10.1105/tpc.106.044602
90. Kobayashi M, Yamaguchi I, Murofushi N, Ota Y, Takahashi N. Fluctuation and Localization of Endogenous Gibberellins in Rice. *Agric Biol Chem*. 1988;52(5):1189-1194. doi:10.1080/00021369.1988.10868799
91. Karssen CM, van Loon LC, Vreugdenhil D. *Progress in Plant Growth Regulation*.; 1992.
92. Eriksson S, Böhlenius H, Moritz T, Nilsson O. GA4 is the active gibberellin in the regulation of LEAFY transcription and Arabidopsis floral initiation. *Plant Cell*. 2006;18(9):2172-2181. doi:10.1105/tpc.106.042317
93. Chiba Y, Shimizu T, Miyakawa S, et al. Identification of Arabidopsis thaliana NRT1/PTR FAMILY (NPF) proteins capable of transporting plant hormones. *J Plant Res*. 2015;128(4):679-686. doi:10.1007/s10265-015-0710-2
94. Tal I, Zhang Y, Jørgensen ME, et al. The Arabidopsis NPF3 protein is a GA transporter. *Nat Commun*. 2016;7:1-11. doi:10.1038/ncomms11486
95. Saito H, Oikawa T, Hamamoto S, et al. The jasmonate-responsive GTR1 transporter is required for gibberellin-mediated stamen development in Arabidopsis. *Nat Commun*. 2015;6:1-11. doi:10.1038/ncomms7095
96. Willige BC, Ghosh S, Nill C, et al. The DELLA domain of GA INSENSITIVE mediates the interaction with the GA INSENSITIVE DWARF1A gibberellin receptor of Arabidopsis. *Plant Cell*. 2007;19(4):1209-1220. doi:10.1105/tpc.107.051441

97. Lantzouni O, Alkofer A, Falter-Braun P, Schwechheimer C. GROWTH-REGULATING FACTORS interact with DELLAs and regulate growth in cold stress. *Plant Cell*. 2020;32:1018-1034. doi:10.1105/tpc.19.00784

98. Dill A, Jung HS, Sun TP. The DELLA motif is essential for gibberellin-induced degradation of RGA. *Proc Natl Acad Sci U S A*. 2001;98(24):14162-14167. doi:10.1073/pnas.251534098

99. Silverstone AL, Chang CW, Krol E, Sun TP. Developmental regulation of the gibberellin biosynthetic gene GA 1 in Arabidopsis thaliana. *Plant J*. 1997;12(1):9-19. doi:10.1046/j.1365-313X.1997.12010009.x

100. Yoshida H, Ueguchi-Tanaka M. DELLA and SCL3 balance gibberellin feedback regulation by utilizing INDETERMINATE DOMAIN proteins as transcriptional scaffolds. *Plant Signal Behav*. 2014;9:1559-2324. doi:10.4161/psb.29726

101. Griffiths J, Murase K, Rieu I, et al. Genetic characterization and functional analysis of the GID1 gibberellin receptors in Arabidopsis. *Plant Cell*. 2006;18(12):3399-3414. doi:10.1105/tpc.106.047415

102. Hedden P. Gibberellins close the lid. *Nature*. 2008;456(7221):455-456. doi:10.1038/456455a

103. Hedden P, Thomas SG. Gibberellin biosynthesis and its regulation. *Biochem J*. 2012;444(1):11-25. doi:10.1042/BJ20120245

104. Schomburg FM, Bizzell CM, Lee DJ, Zeevaart J a D, Amasino RM. Overexpression of a novel class of gibberellin 2-oxidases decreases gibberellin levels and creates dwarf plants. *Plant Cell*. 2003;15(1):151-163. doi:10.1105/tpc.005975

105. Zhang Y, Zhang B, Yan D, et al. Two Arabidopsis cytochrome P450 monooxygenases, CYP714A1 and CYP714A2, function redundantly in plant development through gibberellin deactivation. *Plant J*. 2011;67(2):342-353. doi:10.1111/j.1365-313X.2011.04596.x

106. Nomura T, Magome H, Hanada A, et al. Functional analysis of arabidopsis CYP714A1 and CYP714A2 reveals that they are distinct gibberellin modification enzymes. *Plant Cell Physiol*. 2013;54(11):1837-1851. doi:10.1093/pcp/pct125

107. Thomas SG, Phillips AL, Hedden P. Molecular cloning and functional expression of gibberellin 2-oxidases, multifunctional enzymes involved in gibberellin deactivation. *Proc Natl Acad Sci U S A*. 1999;96(8):4698-4703. doi:10.1073/pnas.96.8.4698

108. Mcginnis KM, Thomas SG, Soule JD, et al. The Arabidopsis. *Society*. 2003;15(May):1120-1130. doi:10.1105/tpc.010827.increased

109. Zentella R, Zhang ZL, Park M, et al. Global analysis of DELLA direct targets in early gibberellin signaling in Arabidopsis. *Plant Cell*. 2007;19(10):3037-3057. doi:10.1105/tpc.107.054999

110. Hou X, Hu WW, Shen L, et al. Global identification of DELLA target genes during arabidopsis

flower development. *Plant Physiol*. 2008;147(3):1126-1142. doi:10.1104/pp.108.121301

111. King KE, Moritz T, Harberd NP. Gibberellins are not required for normal stem growth in Arabidopsis thaliana in the absence of GAI and RGA. *Genetics*. 2001;159(2):767-776.
112. Fukazawa J, Ito T, Kamiya Y, Yamaguchi S, Takahashi Y. Binding of GID1 to DELLAs promotes dissociation of GAF1 from DELLA in GA dependent manner. *Plant Signal Behav*. 2015;10(10):1559-2324. doi:10.1080/15592324.2015.1052923
113. Zhang ZL, Ogawa M, Fleet CM, et al. SCARECROW-LIKE 3 promotes gibberellin signaling by antagonizing master growth repressor DELLA in Arabidopsis. *Proc Natl Acad Sci U S A*. 2011;108(5):2160-2165. doi:10.1073/pnas.1012232108
114. Yu S, Galvão VC, Zhang YC, et al. Gibberellin regulates the Arabidopsis floral transition through miR156-targeted SQUAMOSA PROMOTER BINDING-LIKE transcription factors. *Plant Cell*. 2012;24(8):3320-3332. doi:10.1105/tpc.112.101014
115. Oh E, Zhu JY, Bai MY, Arenhart RA, Sun Y, Wang ZY. Cell elongation is regulated through a central circuit of interacting transcription factors in the Arabidopsis hypocotyl. *Elife*. 2014;2014(3):1-19. doi:10.7554/eLife.03031
116. Josse EM, Gan Y, Bou-Torrent J, et al. A DELLA in disguise: SPATULA restrains the growth of the developing arabidopsis seedling. *Plant Cell*. 2011;23(4):1337-1351. doi:10.1105/tpc.110.082594
117. Arnaud N, Girin T, Sorefan K, et al. Gibberellins control fruit patterning in Arabidopsis thaliana. *Genes Dev*. 2010;24(19):2127-2132. doi:10.1101/gad.593410
118. Kay P, Groszmann M, Ross JJ, Parish RW, Swain SM. Modifications of a conserved regulatory network involving INDEHISCENT controls multiple aspects of reproductive tissue development in Arabidopsis. *New Phytol*. 2013;197(1):73-87. doi:10.1111/j.1469-8137.2012.04373.x
119. Locascio A, Blázquez MA, Alabadí D. Dynamic regulation of cortical microtubule organization through prefoldin-DELLA interaction. *Curr Biol*. 2013;23(9):804-809. doi:10.1016/j.cub.2013.03.053
120. Hülskamp M, Schnittger A. Spatial regulation of trichome formation in Arabidopsis thaliana. *Semin Cell Dev Biol*. 1998;9(2):213-220. doi:10.1006/scdb.1997.0209
121. Schnittger A, Hülskamp M. Trichome morphogenesis: A cell-cycle perspective. *Philos Trans R Soc B Biol Sci*. 2002;357(1422):823-826. doi:10.1098/rstb.2002.1087
122. Perazza D, Herzog M, Hülskamp M, Brown S, Dorne AM, Bonneville JM. Trichome cell growth in Arabidopsis thaliana can be derepressed by mutations in at least five genes. *Genetics*. 1999;152:461-476.
123. Mathur J. Cytoskeleton in trichome morphogenesis. *Developm*. 1999;126:5559-5568.

124. Mathur J, Spielhofer P, Kost B, Chua NH. The actin cytoskeleton is required to elaborate and maintain spatial patterning during trichome cell morphogenesis in Arabidopsis thaliana. *Development*. 1999;126(24):5559-5568.
125. Wang Z, Yang Z, Li F. Updates on molecular mechanisms in the development of branched trichome in Arabidopsis and nonbranched in cotton. *Plant Biotechnol J*. 2019;17(9):1706-1722. doi:10.1111/pbi.13167
126. Hülskamp M, Miséra S, Jürgens G. Genetic dissection of trichome cell development in Arabidopsis. *Cell*. 1994;76(3):555-566. doi:10.1016/0092-8674(94)90118-X
127. Silverstone AL, Tseng TS, Swain SM, et al. Functional analysis of Spindly in gibberellin signaling in Arabidopsis. *Plant Physiol*. 2007;143(2):987-1000. doi:10.1104/pp.106.091025
128. Hisamatsu T, King RW. The nature of floral signals in Arabidopsis. II. Roles for FLOWERING LOCUS T (FT) and gibberellin. *J Exp Bot*. 2008;59(14):3821-3829. doi:10.1093/jxb/ern232
129. Moon J, Suh SS, Lee H, et al. The SOC1 MADS-box gene integrates vernalization and gibberellin signals for flowering in Arabidopsis. *Plant J*. 2003;35(5):613-623. doi:10.1046/j.1365-313X.2003.01833.x
130. Lamb RS, Hill TA, Tan QKG, Irish VF. Regulation of APETALA3 floral homeotic gene expression by meristem identity genes. *Development*. 2002;129(9):2079-2086.
131. Airoldi CA. Determination of sexual organ development. *Sex Plant Reprod*. 2010;23(1):53-62. doi:10.1007/s00497-009-0126-z
132. Yu H, Ito T, Zhao Y, Peng J, Kumar P, Meyerowitz EM. Floral homeotic genes are targets of gibberellin signaling in flower development. *Proc Natl Acad Sci U S A*. 2004;101(20):7827-7832. doi:10.1073/pnas.0402377101
133. Gómez-Mena C, de Folter S, Costa MMR, Angenent GC, Sablowski R. Transcriptional program controlled by the floral homeotic gene AGAMOUS during early organogenesis. *Development*. 2005;132(3):429-438. doi:10.1242/dev.01600
134. Ito T, Ng KH, Lim TS, Yu H, Meyerowitz EM. The homeotic protein AGAMOUS controls late stamen development by regulating a jasmonate biosynthetic gene in Arabidopsis. *Plant Cell*. 2007;19(11):3516-3529. doi:10.1105/tpc.107.055467
135. Xu YL, Gage DA, Zeevaart JAD. Gibberellins and stem growth in Arabidopsis thaliana: Effects of photoperiod on expression of the GA4 and GA5 loci. *Plant Physiol*. 1997;114(4):1471-1476. doi:10.1104/pp.114.4.1471
136. Farrow SC, Facchini PJ, Martens S, Mach E, Peters R, Hudson AO. Functional diversity of 2-oxoglutarate/Fe(II)-dependent dioxygenases in plant metabolism. *Front Plant Sci*. 2014;5(524):1-15. doi:10.3389/fpls.2014.00524

137. Islam MS, Leissing TM, Chowdhury R, Hopkinson RJ, Schofield CJ. 2-Oxoglutarate-Dependent Oxygenases. *Annu Rev Biochem*. 2018;87:585-620. doi:10.1146/annurev-biochem

138. Wilmouth RC, Turnbull JJ, Welford RWD, Clifton IJ, Prescott AG, Schofield CJ. Structure and mechanism of anthocyanidin synthase from Arabidopsis thaliana. *Structure*. 2002;10(1):93-103. doi:10.1016/S0969-2126(01)00695-5

139. Punta M, Coggill PC, Eberhardt RY, et al. The Pfam protein families database. *Nucleic Acids Res*. 2012;40(D1):290-301. doi:10.1093/nar/gkr1065

140. Hegg EL, Que L. The 2-His-1-carboxylate facial triad - An emerging structural motif in mononuclear non-heme iron(II) enzymes. *Eur J Biochem*. 1997;250(3):625-629. doi:10.1111/j.1432-1033.1997.t01-1-00625.x

141. Kataoka H, Yamamoto Y, Sekiguchi M. A new gene (alkB) of Escherichia coli that controls sensitivity to methyl methane sulfonate. *J Bacteriol*. 1983;153(3):1301-1307. doi:10.1128/jb.153.3.1301-1307.1983

142. Meza TJ, Moen MN, Vågbo CB, et al. The DNA dioxygenase ALKBH2 protects Arabidopsis thaliana against methylation damage. *Nucleic Acids Res*. 2012;40(14):6620-6631. doi:10.1093/nar/gks327

143. Velasquez SM, Ricardi MM, Poulsen CP, et al. Complex regulation of prolyl-4-hydroxylases impacts root hair expansion. *Mol Plant*. 2015;8(5):734-746. doi:10.1016/j.molp.2014.11.017

144. Sung ZR, Belachew A, Shunong B, Bertrand-Garcia R. EMF, an Arabidopsis gene required for vegetative shoot development. *Science*. 1992;258(5088):1645-1647. doi:10.1126/science.258.5088.1645

145. Xiao J, Wagner D. Polycomb repression in the regulation of growth and development in Arabidopsis. *Curr Opin Plant Biol*. 2015;23:15-24. doi:10.1016/j.pbi.2014.10.003

146. Jing Y, Guo Q, Lin R. The chromatin-remodeling factor pickle antagonizes polycomb repression of FT to promote flowering. *Plant Physiol*. 2019;181(2):656-668. doi:10.1104/pp.19.00596

147. Samson F, Brunaud V, Belzergue S, et al. FLAGdb/FST: a database of mapped flanking insertion sites (FSTs) of Arabidopsis thaliana T-DNA transformants. *Nucleic Acids Res*. 2002;30(1):94-97. doi:10.1093/nar/30.1.94

148. Iii BEL, Rivero L, Calhoun CS, Grotewold E, Brkljacic J. Standardized Method for High-throughput Sterilization of Arabidopsis Seeds. 2017;(128):1-7. doi:10.3791/56587

149. Wu S, Bingyu Z. Using Clear Nail Polish to Make Arabidopsis Epidermal Impressions for Measruing the Change of Stomatal Aperture Size in Immune Response. *Methods Mol Biol*. 2017;1578:243-248. doi:10.1007/978-1-4939-6859-6

150. Horiguchi G, Fujikura U, Ferjani A, Ishikawa N, Tsukaya H. Large-scale histological analysis of

leaf mutants using two simple leaf observation methods: Identification of novel genetic pathways governing the size and shape of leaves. *Plant J*. 2006;48(4):638-644. doi:10.1111/j.1365-313X.2006.02896.x

151. Galbraith DW, Harkins KR, M LM, Ayres NM, Sharma DP, Firoozabady E. Rapid Flow Cytometric Analysis of the Cell Cycle in Intact Plant Tissues. *Am Assoc Adv Sci*. 2016;220(4601):1049-1051.
152. Schaar WVANDER, Alonso-bBanco C, Leon-Kloosterziel KM, Jansen RC, Van Ooijen JW, Koornneef M. QTL analysis of seed dormancy in Arabidopsis using recombinant inbred lines and MQM mapping. *Heredity (Edinb)*. 1997;79:190-200.
153. Li H, Li J, Cong XH, et al. A high-throughpout, high-quality plant genomic DNA extraction protocol. *Genet Mol Res*. 2013;12(4):4526-4539.
154. Konieczny A, Ausubel FM. A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. *Plant J*. 1993;4:403-410.
155. Neff MM, Neff JD, Chory J, Pepper AE. dCAPS , a simple technique for the genetic analysis of single nucleotide polymorphisms : experimental applications in Arabidopsis thaliana genetics. *Plant J*. 1998;14:387-392.
156. Simms D, Cizdziel PE, Chomczynski P. TRIzol: A new reagent for optimal single-step isolation of RNA. *Focus (Madison)*. 1993;14(4):99-102.
157. Liu H, Naismith JH. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol*. 2008;8:91. doi:10.1186/1472-6750-8-91
158. Becker D, Kemper E, Schell J, Masterson R. New plant binary vectors with selectable markers located proximal to the left T-DNA border. *Plant Mol Biol*. 1992;20(6):1195-1197. doi:10.1007/BF00028908
159. De Felippes FF, Wang JW, Weigel D. MIGS: MiRNA-induced gene silencing. *Plant J*. 2012;70(3):541-547. doi:10.1111/j.1365-313X.2011.04896.x
160. Clough SJ, Bent AF. Floral dip: A simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J*. 1998;16(6):735-743. doi:10.1046/j.1365-313X.1998.00343.x
161. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
162. Morrissette NS, Mitra A, Sept D, Sibley LD. Dinitroanilines Bind α-Tubulin to Disrupt Microtubules. *Mol Biol Cell*. 2004;15(4):1960-1968. doi:10.1091/mbc.E03-07-0530
163. Morejohn LC, Bureau TE, Mol6-Bajer J, Bajer AS, Fosket DE. Oryzalin, a dinitroaniline herbicide, binds to plant tubulin and inhibits microtubule polymerization in vitro. *Planta*. 1987;172:252-264.

164. Bai MY, Shang JX, Oh E, et al. Brassinosteroid, gibberellin and phytochrome impinge on a common transcription module in Arabidopsis. *Nat Cell Biol*. 2012;14(8):810-817. doi:10.1038/ncb2546

165. Hartwig T, Corvalan C, Best NB, et al. Propiconazole is a specific and accessible brassinosteroid (BR) biosynthesis inhibitor for arabidopsis and maize. *PLoS One*. 2012;7(5):1-11. doi:10.1371/journal.pone.0036625

166. Sun S, Wang H, Yu H, et al. GASA14 regulates leaf expansion and abiotic stress resistance by modulating reactive oxygen species accumulation. *J Exp Bot*. 2013;64(6):1637-1647. doi:10.1093/jxb/ert021

167. Li L, Li L, Shi ZY, et al. Overexpression of ACL1 (abaxially curled leaf 1) increased bulliform cells and induced abaxial curling of leaf blades in rice. *Mol Plant*. 2010;3(5):807-817. doi:10.1093/mp/ssq022

168. KAY R, CHAN A, DALY M, MCPHERSON J. Duplication of CaMV 35S Promoter Sequences Creates a Strong Enhancer for Plant Genes. *Science*. 1987;236(4806):1299-1302. doi:10.1126/science.236.4806.1299

169. Brown JWS, Smith P, Simpson CG. Arabidopsis consensus intron sequences. *Plant Mol Biol*. 1996;32(3):531-535. doi:10.1007/BF00019105

170. Shim JS, Kubota A, Imaizumi T. Circadian clock and photoperiodic flowering in arabidopsis: CONSTANS is a Hub for Signal integration. *Plant Physiol*. 2017;173(1):5-15. doi:10.1104/pp.16.01327

171. Wang ZP, Xing HL, Dong L, et al. Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in Arabidopsis in a single generation. *Genome Biol*. 2015;16(1):1-12. doi:10.1186/s13059-015-0715-0

172. Whitworth AL, Mann NH, Larkum AWD. This article is protected by copyright. All rights 1 reserved. *Ultrasound Obs Gynecol*. 2006;50(6):776-780. doi:10.1111/1462-2920.12735

173. Sharma Koirala P, Neff MM. Improving seed size, seed weight and seedling emergence in Camelina sativa by overexpressing the Atsob3-6 gene variant. *Transgenic Res*. 2020;29(4):409-418. doi:10.1007/s11248-020-00208-9

174. Gupta A. Functional Analysis of Squamosa-Promoter Binding Protein Like 13 in Controlling Flowering Time , Plant Architecture , Grain Size , and Grain Number in Wheat. *Electron Theses Diss*. 2020;4080.

175. Müller-Mcnicoll M, Rossbach O, Hui J, Medenbach J. Auto-regulatory feedback by RNA-binding proteins. *J Mol Cell Biol*. 2019;11(10):930-939. doi:10.1093/jmcb/mjz043

176. Mitrophanov AY, Groisman EA. Positive feedback in cellular control systems. 2008;30(6):542-

555.

177. Cargnin F, Flora A, Di Lascio S, et al. PHOX2B regulates its own expression by a transcriptional auto-regulatory mechanism. *J Biol Chem*. 2005;280(45):37439-37448. doi:10.1074/jbc.M508368200

178. Patharkar OR, Walker JC. Floral organ abscission is regulated by a positive feedback loop. *Proc Natl Acad Sci*. 2015;2015:1-6. doi:10.1073/pnas.1423595112

179. El-brolosy MA, Stainier DYR. Genetic compensation : A phenomenon in search of mechanisms. *PLoS Genet*. 2017:1-17.

180. Jiang G, den Hertog J, Hunter T. Receptor-Like Protein Tyrosine Phosphatase α Homodimerizes on the Cell Surface. *Mol Cell Biol*. 2000;20(16):5917-5929. doi:10.1128/mcb.20.16.5917-5929.2000

181. Giraldol R, Fernández-Tornero C, Evans PR, Díaz-Orejas R, Romero A. A conformational switch between transcriptional repression and replication initiation in the RepA dimerization domain. *Nat Struct Biol*. 2003;10(7):565-571. doi:10.1038/nsb937

182. Marianayagam NJ, Sunde M, Matthews JM. The power of two: Protein dimerization in biology. *Trends Biochem Sci*. 2004;29(11):618-625. doi:10.1016/j.tibs.2004.09.006

183. Islam S, Leissing TM, Chowdhury R, Hopkinson RJ, Schofield CJ. 2-Oxoglutarate-Dependent Oxygenases. *Annu Rev Biochem*. 2018;87:585-620.

184. Deng S, Jang IC, Su L, Xu J, Chua NH. JMJ24 targets CHROMOMETHYLASE3 for proteasomal degradation in arabidopsis. *Genes Dev*. 2016;30(3):251-256. doi:10.1101/gad.274647.115

185. Bloomer RH, Hutchison CE, Bäurle I, et al. The Arabidopsis epigenetic regulator ICU11 as an accessory protein of polycomb repressive complex 2. *Proc Natl Acad Sci U S A*. 2020;117(28):16660-16666. doi:10.1073/pnas.1920621117

186. Kwok J, O'Shea M, Hume DA, Lengeling A. Jmjd6, a JmjC dioxygenase with many interaction partners and pleiotropic functions. *Front Genet*. 2017;8:1-19. doi:10.3389/fgene.2017.00032

187. Maury S, Sow MD, Le Gac AL, Genitoni J, Lafon-Placette C, Mozgova I. Phytohormone and chromatin crosstalk: The missing link for developmental plasticity? *Front Plant Sci*. 2019;10(April):1-6. doi:10.3389/fpls.2019.00395

188. Kitsios G, Doonan JH. Cyclin dependant protein kinases and stress responses in plants. *Plant Signal Behav*. 2011;6(2):204-209. doi:10.4161/psb.6.2.14835

189. Charlesworth D, Charlesworth B. Population genetics of partial male-sterility and the evolution of monoecy and dioecy. *Heredity (Edinb)*. 1978;41(2):137-153. doi:10.1038/hdy.1978.83

190. Wright SI, Kalisz S, Slotte T, Wright SI. Evolutionary consequences of self- fertilization in plants. *Proc R Soc*. 2013;280:1-10.

191. Boris Igic , Russell Lande and JR. K. Loss of Self - Incompatibility and Its Evolutionary Consequences. *Int J Plant Sci*. 2014;169(1):93-104. doi:10.1086/523362
192. Charlesworth D, Charlesworth B. Inbreeding depression and its evolutionary consequences. *Annu Rev Ecol Syst Vol 18*. 1987:237-268. doi:10.1146/annurev.es.18.110187.001321
193. Lewis D. Comparative Incompatibility in Angiosperms and Fungi I . INTRODUCTION. 1951:235-285.
194. Bateman AJ. Self-Incompatibility Systems in Angiosperms. *Heredity (Edinb)*. 1952;6:285-310.
195. Takayama S, Isogai A. Self-Incompatibility in Plants. *Annu Rev Plant Biol*. 2005;56:467-489. doi:10.1146/annurev.arplant.56.032604.144249
196. Shiba H, Iwano M, Entani T, et al. The Dominance of Alleles Controlling Self-Incompatibility in. *Plant Cell*. 2002;14(2):491-504. doi:10.1105/tpc.010378.by
197. Hatakeyama K, Watanabe M, Takasaki T, Ojima K, Hinata K. Dominance relationships between S-alleles in self-incompatible Brassica campestris L. *Heredity (Edinb)*. 1998;80(2):241-247. doi:10.1038/sj.hdy.6882950
198. Nou S, Watanabe M, Isogai A, Hinata K. Comparison of S-alleles and S-glycoproteins between two wild populations of Brassica campestris in Turkey and Japan. *Sex Plant Reprod*. 1993;6(2):79-86. doi:10.1007/BF00227652
199. D.J.Ockendon. The s-allele collection of Brassica oleracea. *Proc III Int Symp Brassicas*. 2000:25-30.
200. Shimosato H, Yokota N, Shiba H, et al. Characterization of the SP11/SCR high-affinity binding site involved in self/nonself recognition in Brassica self-incompatibility. *Plant Cell*. 2007;19(1):107-117. doi:10.1105/tpc.105.038869
201. Takasaki T, Hatakeyama K, Suzuki G, Watanabe M, Isogai A, Hinata K. The S receptor kinase determines self-incompatibility in Brassica stigma. *Nature*. 2000;403(6772):913-916. doi:10.1038/35002628
202. Takayama S, Shiba H, Iwano M, et al. The Pollen Determinant of Self-Incompatibility in Brassica campestris. *Proc Natl Acad Sci*. 2017;97(4):1920-1925. doi:10.1073/prias.040556397
203. Wheeler MJ, Graaf BHJ De, Hadjiosif N, et al. Identification of the pollen self-incompatibility determinant in Papaver rhoeas. *Nature*. 2009;459(7249):992-995. doi:10.1038/nature08027
204. Wilkins KA, Bosch M, Haque T, Teng N, Poulter NS, Franklin-tong VE. Self-Incompatibility-Induced Programmed Cell Death in Field Poppy Pollen Involves Dramatic Acidification of the Incompatible Pollen Tube Cytosol. *Plant Physiol*. 2015;167:766-779. doi:10.1104/pp.114.252742
205. Wheeler MJ, Vatovec S, Franklin-tong VE. The pollen S -determinant in Papaver : comparisons with known plant receptors and protein ligand partners. *J Exp Bot*. 2015;61(7):2015-2025.

doi:10.1093/jxb/erp383

206. Sota F, Kubo K, Takayama S. Non-self- and self-recognition models in plant self-incompatibility. *Nat Plants*. 2016;2:1-9. doi:10.1038/NPLANTS.2016.130
207. Li S, Sun P, Stephen J, Kao T. Identification of the self-incompatibility locus F-box protein-containing complex in Petunia inflata. *Plant Reprod*. 2014;27:31-45. doi:10.1007/s00497-013-0238-3
208. Toyoda M, Kawashima S, Ando T, Isogai A, Kao T, Takayama S. Collaborative Non-Self Recognition System in S-RNase – Based Self-Incompatibility. *Science*. 2010;341:796-800.
209. Tsukamoto T, Ando T, Watanabe H, Marchesi E, Kao T. Duplication of the S -locus F-box gene is associated with breakdown of pollen function in an S -haplotype identified in a natural population of self-incompatible Petunia axillaris. *Plant Mol Biol*. 2005;57:141-153. doi:10.1007/s11103-004-6852-6
210. Adachi Y, Komori S, Hoshikawa Y, et al. Characteristics of fruiting and pollen tube growth of apple autotetraploid cultivars showing self-compatibility. *J Japanese Soc Hortic Sci*. 2009;78(4):402-409. doi:10.2503/jjshs1.78.402
211. Lewis D. Competition and dominance of incompatibilizy alleles in diploid pollen. *John Innes Hortic Institution, London*. 1947:85-108.
212. Hauck NR, Yamane H, Tao R, Iezzoni AF. Accumulation of Nonfunctional S -Haplotypes Results in the Breakdown of Gametophytic Self-Incompatibility in Tetraploid Prunus. *Genet Soc Am.* 2006;1198:1191-1198. doi:10.1534/genetics.105.049395
213. Hauck NR, Yamane H, Tao R, Iezzoni AF. Self-compatibility and incompatibility in tetraploid sour cherry (Prunus cerasus L .). *Sex Plant Reprod*. 2002;15:39-46. doi:10.1007/s00497-002-0136-6
214. Charlesworth B, Charlesworth D. Darwin and Genetics. *Genet Soc Am*. 2009;766:757-766. doi:10.1534/genetics.109.109991
215. Zhou W, Barrett SCH, Wang H, Li D. Reciprocal herkogamy promotes disassortative mating in a distylous species with intramorph compatibility. *New Phytol*. 2015:1-10.
216. Dulberger R. Floral Polymorphisms and Their Functional Significance in the Heterostylous Syndrome. In: *Evolution and Function of Heterostyly*. ; 1992:41-84.
217. Webb CJ, Lloyd DG. The avoidance of interference between the presentation of pollen and stigmas in angiosperms II. herkogamy. *New Zeal J Bot*. 1986;24(1):163-178. doi:10.1080/0028825X.1986.10409726
218. Bateson W, Gregory RP. On the Inheritance o f Heterostylism in Primula. 1905.
219. Ornduff R. The Genetics of Hetreostyly in Hypericum aegypticum. *Heredity (Edinb)*.

1979;42:271-272.

220. Sohoute C, Ernst A. Zur Vererbung der morphologischen Heterostyliemerkmale. 1928.
221. Schwander T, Libbrecht R. Supergenes and Complex Phenotypes. *Curr Biol*. 2014;24(7):R288-R294. doi:10.1016/j.cub.2014.01.056
222. Dowrick PJ. heterostyly and Homostyly in Primula obconica. *Heredity (Edinb)*. 1956.
223. Barrett SCH. The genetics of heterostyly. In: *Evolution and Function of Heterostyly*. ; 1992.
224. Matsui K, Tetsuka T, Nishio T, Hara T. Heteromorphic incompatibility retained in self-compatible plants produced by a cross between common and wild buckwheat. *New Phytol*. 2003;159:701-708. doi:10.1046/j.1469-8137.2003.00840.x
225. Labonne JDJ, Tamari F, Shore JS. Characterization of X-ray-generated floral mutants carrying deletions at the S -locus of distylous Turnera subulata. *Heredity (Edinb)*. 2010;105(2):244. doi:10.1038/hdy.2010.83
226. Charlesworth D, Charlesworth B. A model for the evolution. *Am Nat*. 1979;114(4):467-498.
227. Weller SG, Ornduff R. CRYPTIC SELF-INCOMPATIBILITY IN AMSINCKIA GRANDIFLORA. *Evolution (N Y)*. 1977;31(1):47-51. doi:10.1111/j.1558-5646.1977.tb00980.x
228. Schoen DJ. EVOLUTIONARY HISTORY OF THE MATING SYSTEM IN AMSINCKIA (BORAGINACEAE). *Evolution (N Y)*. 1997;51(4):1090-1099.
229. Kohn JR, Graham SW, Morton B, et al. RECONSTRUCTION OF THE EVOLUTION OF REPRODUCTIVE CHARACTERS IN PONTEDERIACEAE USING PHYLOGENETIC EVIDEN ; CE FROM CHLOROPLAST DNA RESTRICTION-SITE VARIATION. *Evolution (N Y)*. 1996;50(4):1454-1469.
230. Truyens S, Arbo MM, Shore JS. Phylogenetic Relationships, Chromosome and Breeding System Evolution in Turnera (Turneracea): Inferences friom its Sequence Data. *Am J Bot*. 2005;92(10):1749-1758.
231. Belaoussoff S, Shore JS. FLORAL CORRELATES AND FITNESS CONSEQUENCES OF MATING-SYSTEM VARIATION IN TURNERA ULMIFOLIA. *Evolution (N Y)*. 1995;49(3):545-556.
232. Tamari F, Athanasiou A, Shore JS. Pollen tube growth and inhibition in distylous and homostylous Turnera and Piriqueta (Turneraceae). *Can J Bot*. 2001;591:578-591. doi:10.1139/cjb-79-5-578
233. Barrett SCH. Tansley review ‘ A most complex marriage arrangement ’: recent advances on heterostyly and unresolved questions. 2019:1051-1067. doi:10.1111/nph.16026
234. Li J, Webster M, Furuya M, Gilmartin PM. Identification and characterization of pin and thrum alleles of two genes that co-segregate with the Primula S locus. *Plant J*. 2007;51:18-31.

doi:10.1111/j.1365-313X.2007.03125.x

235. Huu CN, Kappel C, Keller B, et al. Presence versus absence of CYP734A50 underlies the style-length dimorphism in primroses. *Elife*. 2016;5:1-15. doi:10.7554/eLife.17956

236. Neff MM, Nguyen SM, Malancharuvil EJ, et al. BAS1:A gene regulating brassinosteroid levels responsiveness in Arabidopsis and. *Proc Natl Acad Sci*. 2015;96(26):15316-16323.

237. Ohnishi T, Nomura T, Watanabe B, et al. Tomato cytochrome P450 CYP734A7 functions in brassinosteroid catabolism. *Phytochemistry*. 2006;67:1895-1906. doi:10.1016/j.phytochem.2006.05.042

238. Li J, Cocker JM, Wright J, et al. Genetic architecture and evolution of the S locus supergene in Primula vulgaris. *Nat Plants*. 2016;2:1-7. doi:10.1038/nplants.2016.188

239. Burrows BA, McCubbin AG. Sequencing the genomic regions flanking S-linked PvGLO sequences confirms the presence of two GLO loci, one of which lies adjacent to the style-length determinant gene CYP734A50. *Plant Reprod*. 2017;30(1):53-67. doi:10.1007/s00497-017-0299-9

240. Li J, Cocker JM, Wright J, et al. Genetic architecture and evolution of the S locus supergene in Primula vulgaris. *Nat Plants*. 2016;2(12):1-7. doi:10.1038/nplants.2016.188

241. Huu CN, Keller B, Conti E, Kappel C, Lenhard M. Supergene evolution via stepwise duplications and neofunctionalization of a floral-organ identity gene. *Proc Natl Acad Sci U S A*. 2020;117(37):23148-23157. doi:10.1073/pnas.2006296117

242. Ushijima K, Nakano R, Bando M, et al. Isolation of the floral morph-related genes in heterostylous flax (Linum grandiflorum): the genetic polymorphism and the transcriptional and post-transcriptional regulations of the S locus. *Plant J*. 2012;69:317-331. doi:10.1111/j.1365-313X.2011.04792.x

243. Yasui Y, Mori M, Aii J, et al. S-LOCUS EARLY FLOWERING 3 Is Exclusively Present in the Genomes of Short-Styled Buckwheat Plants that Exhibit Heteromorphic Self-Incompatibility. *PLoS One*. 2012;7(2):1-9. doi:10.1371/journal.pone.0031264

244. Matsui K, Nishio T, Tetsuka T. Genes Outside the S Supergene Suppress S Functions in Buckwheat (Fagopyrum esculentum). *Ann Bot*. 2004;94:805-809. doi:10.1093/aob/mch206

245. Matzke CM, Shore JS, Neff MM, McCubbin AG. The turnera style s-locus gene tsbahd possesses brassinosteroid-inactivating activity when expressed in arabidopsis thaliana. *Plants*. 2020;9(11):1-13. doi:10.3390/plants9111566

246. Shore JS, Hamam HJ, Chafe PDJ, Labonne JDJ, Henning PM, McCubbin AG. The long and short of the S -locus in Turnera (Passifloraceae) . *New Phytol*. July 2019. doi:10.1111/nph.15970

247. Cohen JI. Cladistics of Boraginaceae : evolutionary relationships , taxonomy , and patterns of character evolution. *Cladistics*. 2014;30:139-169.

248. Ferrero V, Arroyo J, Vargas P, Thompson JD, Navarro L. Evolutionary transitions of style polymorphisms in Lithodora (Boraginaceae). *Perspect Plant Ecol Evol Syst*. 2009;11:111-125. doi:10.1016/j.ppees.2009.01.004

249. Cohen JI, Litt A, Davis JI. Comparative Flora Development in Lithospermum (Boraginaceae)and Implications for the Evolution and Development of Heterostyly. *Am J Bot*. 2012;99(5):797-805. doi:10.3732/ajb.1100329

250. Brys R, Jacquemyn H, Hermy M, Beeckman T. Pollen deposition rates and the functioning of distyly in the perennial Pulmonaria officinalis (Boraginaceae). *Plant Syst Evol*. 2008;273(1-2):1-12. doi:10.1007/s00606-008-0003-5

251. Selvi F. Floral biometrics of the Anchusa undulata L . group (Boraginaceae) from the central-eastern Mediterranean. *Bot J Linn Socitey*. 1998;128:251-270.

252. Philipp M, Schou O. An unusual heteromorphic incompatibility system. *New Phytol*. 1981;89:693-703.

253. Ganders FR. HETEROSTYLY , HOMOSTYLY , AND FECUNDITY IN AMSINCKIA SPECTABILIS (BORAGINACEAE). *Calif Bot Soc*. 1975;23(1):56-62.

254. Li P, Johnston MO. Comparative floral morphometrics of distyly and homostyly in three evolutionary lineages of Amsinckia (Boraginaceae). *Can J Bot*. 2001;79(11):1332-1348. doi:10.1139/cjb-79-11-1332

255. Ray PM, Chisaki HF. Studies on Amsinckia . I . A Synopsis of the Genus , with a Study of Heterostyly in it. *Am J Bot*. 1957;44(6):529-536.

256. Weller SG, Ornduff R. INCOMPATIBILITY IN AMSINCKIA GRANDIFLORA (BORAGINACEAE): DISTRIBUTION OF CALLOSE PLUGS AND POLLEN TUBES FOLLOWING INTER- AND INTRAMORPH CROSSES 1. *Am J Bot*. 1989;76(2):277-282. doi:10.1002/j.1537-2197.1989.tb11310.x

257. Casper BB, Sayigh LS, Lee SS. Demonstration of cryptic incompatibility in distylous amsinckia douglasiana. *Evolution (N Y)*. 1988;42(2):248-253.

258. Weller SG, Ornduff R. POLLEN TUBE GROWTH AND INBREEDING DEPRESSION IN. *Am J Bot*. 1991;78(6):801-804. doi:10.1002/j.1537-2197.1991.tb14482.x

259. Ganders FR. Mating patterns in self-compatible distylous populations of Amsinckia (Boraginaceae). *Can J Bot*. 1975;53:773-779.

260. Ornduff R. The Reproductive System of Amsinckia grandiflora, a Distylous Species. *Syst Bot*. 1976;1(1):57-66.

261. Schoen DJ. DELETERIOUS MUTATION IN RELATED SPECIES OF THE PLANT GENUS AMSINCKIA WITH CONTRASTING MATING SYSTEMS. *Evolution (N Y)*.

2005;59(11):2370-2377.

262. Li P, Johnston MO. Flower Development and the Evolution of Self-fertilization in Amsinckia : The Role of Heterochrony. *Evol Biol*. 2010;37:143-168. doi:10.1007/s11692-010-9091-6

263. Mori T, Kuroiwa H, Higashiyama T, Kuroiwa T. LETTERS GENERATIVE CELL SPECIFIC 1 is essential for angiosperm fertilization. *Nat Cell Biol*. 2006;8(1):64-71. doi:10.1038/ncb1345

264. Cullings K. Design and testing of a plant-specific PCR primer for ecological and evolutionary studies. *Mol Ecol*. 1992;1:233-240.

265. Bankevich A, Nurk S, Antipov D, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol*. 2012;19(5):455-477. doi:10.1089/cmb.2012.0021

266. Li D, Liu C, Luo R, Sadakane K, Lam T. Application Note MEGAHIT : An ultra-fast single-node solution for large and com- plex metagenomics assembly via succinct de Bruijn graph. *Bioinforma Adv Access*. 2015:3-4.

267. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357-360. doi:10.1038/nmeth.1923

268. Quinlan AR, Hall IM. BEDTools : a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841-842. doi:10.1093/bioinformatics/btq033

269. Korlach J, Bjornson KP, Chaudhuri BP, et al. *Real-Time DNA Sequencing from Single Polymerase Molecules*. Vol 472. 1st ed. Elsevier Inc.; 2010. doi:10.1016/S0076-6879(10)72001-2

270. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu : scalable and accurate long-read assembly via adaptive k -mer weighting and repeat separation. *Genome Res*. 2017;27:722-736. doi:10.1101/gr.215087.116.Freely

271. Li H. A statistical framework for SNP calling , mutation discovery , association mapping and population genetical parameter estimation from sequencing data. *Bioinformaticsformatics*. 2011;27(21):2987-2993. doi:10.1093/bioinformatics/btr509

272. Johnston MO, Schoen DJ. CORRELATED EVOLUTION OF SELF-FERTILIZATION AND INBREEDING DEPRESSION : AN EXPERIMENTAL STUDY OF NINE POPULATIONS OF AMSINCKIA (BORAGINACEAE). *Evolution (N Y)*. 1996;50(4):1478-1491.

273. Henning PM, Shore JS. Transcriptome and Network Analyses of Heterostyly in Turnera subulata Provide Mechanistic Insights : Are S -Loci a Red-Light for Pistil Elongation ? *Plants*. 2020;9(713):1-20.

274. González-García MP, Vilarrasa-Blasi J, Zhiponova M, et al. Brassinosteroids control meristem size by promoting cell cycle progression in Arabidopsis roots. *Development*. 2011;138(5):849-859. doi:10.1242/dev.057331

275. Rhoads A, Au KF. PacBio Sequencing and Its Applications. *Genomics, Proteomics Bioinforma*. 2015;13(5):278-289. doi:10.1016/j.gpb.2015.08.002

276. Stehlik I, Barrett SCH. Mechanisms governing sex-ratio variation in dioecious Rumex nivalis. *Evolution (N Y)*. 2005;59(4):814-825. doi:10.1111/j.0014-3820.2005.tb01755.x

277. Freeman DC, Wachocki BA, Stender MJ, Goldschlag DE, Michaels HJ. Seed size and sex ratio in spinach: application of the Trivers- Willard hypothesis to plants. *Ecoscience*. 1994;1(1):54-63. doi:10.1080/11956860.1994.11682228

278. Purrington CB, Schmitt J. Sexual Dimorphism of Dormancy and Survivorship in Buried Seeds of Silene Latifolia. *J Ecol*. 1995;83(5):795. doi:10.2307/2261416

279. Moldenke AR. Caliornia Pollination Ecology and Vegetation Types. *Phytologia*. 1976;34(4):305-361.

280. Kieffer M, Master V, Waites R, Davies B. TCP14 and TCP15 affect internode length and leaf shape in Arabidopsis. *Plant J*. 2011;68(1):147-158. doi:10.1111/j.1365-313X.2011.04674.x

281. Gastaldi V, Lucero LE, Ferrero L V, Ariel FD, Gonzalez DH. Class-I TCP Transcription Factors Activate the SAUR63 Gene Subfamily in Gibberellin-Dependent Stamen Filament Elongation. *Plant Physiol*. 2020;182(April):2096-2110. doi:10.1104/pp.19.01501

282. Park JA, Ahn JW, Kim YK, et al. Retinoblastoma protein regulates cell proliferation, differentiation, and endoreduplication in plants. *Plant J*. 2005;42(2):153-163. doi:10.1111/j.1365-313X.2005.02361.x

283. Fu DQ, Zhu BZ, Zhu HL, Jiang WB, Luo YB. Virus-induced gene silencing in tomato fruit. *Plant J*. 2005;43(2):299-308. doi:10.1111/j.1365-313X.2005.02441.x

284. Tuttle EM, Bergland AO, Korody ML, Warren WC, Gonser RA, Balakrishnan CN. Divergence and Functional Degradation of a Sex Report Divergence and Functional Degradation of a Sex Chromosome-like Supergene. *Curr Biol*. 2016;26(3):344-350. doi:10.1016/j.cub.2015.11.069

285. Bachtrog D. Expression Profile of a Degenerating Neo-Y Chromosome in Drosophila. *Curr Biol*. 2006;16:1694-1699. doi:10.1016/j.cub.2006.07.053

286. Hough J, Hollister JD, Wang W, Barrett SCH, Wright SI. Genetic degeneration of old and young Y chromosomes in the flowering plant Rumex hastatulus. *Proc Natl Acad Sci*. 2014;111(21):7713-7718. doi:10.1073/pnas.1319227111

287. Ray PM, Chisaki HF. Studies on Amsinckia . III . Aneuploid Diversification in the Muricatae A. *Am J Bot*. 1957;44(6):545-554.

288. Cohen JI. De novo sequencing and comparative transcriptomics of floral development of the distylous species Lithospermum multiflorum. *Front Plant Sci*. 2016;7. doi:10.3389/fpls.2016.01934

6. Supplementary ICU11 project: Figures and Tables

6.1 cDNA and predicted protein sequences of *icu11-3* splicing forms

Splicing form 1 (1276 bp)

TTACTCGGCACATGATTTTGAAGCCAGCTCAACAAGTGTCTTCTCTGCAGCTTTCCTTGCTAGTATCTCTTTAGTTG
CATTAATCGAATCACGCTGCCTTCTCTGTTTATCAAGTTTGCATCCTCCACACCAACCTGAGAAATCTCTCTGATAA
TTCTTCATCTCTCGAAAAGTTGAGctacaatgagaaagagagggcgaatatgaaatcccatgaaatGgttcatgaaa
ctttcattggctgatgaaaatatcttttacCTTCTACACCACAAAATCAAGTTGGCTCGGTGTCCAGAAGTTGTAGCT
CTAGCACCATGGCGATGACGTCCACGATGAAGAATGGCGTGACCAGGTACATGAGAGTAATCATAAACTTCCTTTTC
AGTGCTATCGGAATTCACATGCTTATCACATCTCACCCCTCGAAAATACAGCTCCCCGCCAGAGAATTGTTTACCCA
AGCAGACATTTAAACTAACTTCTGAGTCATCCACATGAAACCCAAGATCAACATCCCTATCTTTTCCATATTCAACA
ATATAGCCATGGTGAGAATCTAAAGAGGTTCCACAGACTTCCGGGAACAGAACTTGAGCTATAGGACTAATAAAGTC
ATCAACCAACTTCTGAAGCATGCTATCAAAGCCAAAATCATCAAGGACAACACCGAAATTGTTCATCGTATTAGGTC
TCATTATTGTGGATCTTGAATCATAAACCCATTTCTCCATATGTTCAACCTCTGCTAATAACATTTCACAAAACTGT
GGTTTAAACATTTCAAAAGTGAAAATCCCAGGATAAGATTCAACCATACTACTTCTGAAATTCGGCTCTGACTTCCG
ACTAAAAGCTCCAAGAAACGATGGCGCAAAGAAACTACTAGGGTCAAGAGTATAGATCTCTCCATGCAGACGTTGGT
AAGAAGACATGATTTTGTCTCTGTATTCTTTATGCCTCAGAACCCGAACTCGCTCAGTATCAGGGGAATATCGGAGA
AGAAGGTCTCTCATGAAACTAGCTTTATCGATTCGTGTGGAATTGAGAAGCTGCTCAGGTAAGTAACGCTCAAGAGA
GGTGAACAGAGAAGGACTATAGTCCAAAGGTAGATCTTCATAGTTCTCAGGCTCATGTTCTTCATTAGGGGTTCTCC
GAAGCTTCAATCTTGCCTCGCCGTTTCCAGAGGAGGCTCGTGGTTGTTGTTGTTGCTGCTGCTCAGGTTGTTTTCCG
GAAGAATCGAGAGCCATAGATCTAAGAGGAGTTTGATTGCACAT

Splicing form 2 (1150 bp)

TTACTCGGCACATGATTTTGAAGCCAGCTCAACAAGTGTCTTCTCTGCAGCTTTCCTTGCTAGTATCTCTTTAGTTG
CATTAATCGAATCACGCTGCCTTCTCTGTTTATCAAGTTTGCATCCTCCACACCAACCTTCTACACCACAAAATCAA
GTTGGCTCGGTGTCCAGAAGTTGTAGCTCTAGCACCATGGCGATGACGTCCACGATGAAGAATGGCGTGACCAGGTA
CATGAGAGTAATCATAAACTTCCTTTTCAGTGCTATCGGAATTCACATGCTTATCACATCTCACCCCTCGAAAATAC
AGCTCCCCGCCAGAGAATTGTTTACCCAAGCAGACATTTAAACTAACTTCTGAGTCATCCACATGAAACCCAAGATC
AACATCCCTATCTTTTCCATATTCAACAATATAGCCATGGTGAGAATCTAAAGAGGTTCCACAGACTTCCGGGAACA
GAACTTGAGCTATAGGACTAATAAAGTCATCAACCAACTTCTGAAGCATGCTATCAAAGCCAAAATCATCAAGGACA
ACACCGAAATTGTTCATCGTATTAGGTCTCATTATTGTGGATCTTGAATCATAAACCCATTTCTCCATATGTTCAAC
CTCTGCTAATAACATTTCACAAAACTGTGGTTTAAACATTTCAAAAGTGAAAATCCCAGGATAAGATTCAACCATAC
TACTTCTGAAATTCGGCTCTGACTTCCGACTAAAAGCTCCAAGAAACGATGGCGCAAAGAAACTACTAGGGTCAAGA
GTATAGATCTCTCCATGCAGACGTTGGTAAGAAGACATGATTTTGTCTCTGTATTCTTTATGCCTCAGAACCCGAAC
TCGCTCAGTATCAGGGGAATATCGGAGAAGAAGGTCTCTCATGAAACTAGCTTTATCGATTCGTGTGGAATTGAGAA
GCTGCTCAGGTAAGTAACGCTCAAGAGAGGTGAACAGAGAAGGACTATAGTCCAAAGGTAGATCTTCATAGTTCTCA
GGCTCATGTTCTTCATTAGGGGTTCTCCGAAGCTTCAATCTTGCCTCGCCGTTTCCAGAGGAGGCTCGTGGTTGTTG
TTGTTGCTGCTGCTCAGGTTGTTTTCCGGAAGAATCGAGAGCCATAGATCTAAGAGGAGTTTGATTGCACAT

Fig. S1. Sequences of *icu11-3* splicing versions in 5'-3' direction

Colour code: orange and black: repeated colour code for consecutive order of exons (exon1:black, exon2:orange and further), red: stop codon, blue; translation start and stop codons

Predicted protein sequence of splicing form 1

Met C N Q T P L R S Met A L D S S G K Q P E Q Q Q Q Q Q P R A S S G N G E A R
L K L R R T P N E E H E P E N Y E D L P L D Y S P S L F T S L E R Y L P E Q L
L N S T R I D K A S F Met R D L L L R Y S P D T E R V R V L R H K E Y R D K
I Met S S Y Q R L H G E I Y T L D P S S F F A P S F L G A F S R K S E P N F R
S S Met V E S Y P G I F T F E Met F K P Q F C E Met L L A E V E H Met E K W V
Y D S R S T I Met R P N T Met N N F G V V L D D F G F D S Met L Q K L V D D F
I S P I A Q V L F P E V C G T S L D S H H G Y I V E Y G K D R D V D L G F H V
D D S E V S L N V C L G K Q F S G G E L Y F R G V R C D K H V N S D S T E K E
V Y D Y S H V P G H A I L H R G R H R H G A R A T T S G H R A N L I L W C R
R **Stop** R Y F H Q P Met K V S **Stop** T I S W D F I F A L S F S L **Stop** L N F S R
D E E L S E R F L R L V W R Met Q T **Stop** **Stop** T E K A
A **Stop** F D **Stop** C N **Stop** R D T S K E S C R E D T C **Stop** A G F K I Met C R V

Predicted protein sequence of splicing form 2

Met C N Q T P L R S Met A L D S S G K Q P E Q Q Q Q Q Q P R A S S G N
G E A R L K L R R T P N E E H E P E N Y E D L P L D Y S P S L F T S L
E R Y L P E Q L L N S T R I D K A S F Met R D L L L R Y S P D T E R V
R V L R H K E Y R D K I Met S S Y Q R L H G E I Y T L D P S S F F A P
S F L G A F S R K S E P N F R S S Met V E S Y P G I F T F E Met F K P
Q F C E Met L L A E V E H Met E K W V Y D S R S T I Met R P N T Met N
N F G V V L D D F G F D S Met L Q K L V D D F I S P I A Q V L F P E V
C G T S L D S H H G Y I V E Y G K D R D V D L G F H V D D S E V S L N
V C L G K Q F S G G E L Y F R G V R C D K H V N S D S T E K E V Y D Y
S H V P G H A I L H R G R H R H G A R A T T S G H R A N L I L W C R R
L V W R Met Q T ***Stop*** ***Stop*** *T E K A A* ***Stop*** *F D* ***Stop*** *C N* ***Stop*** *R D T S K*
E S C R E D T C ***Stop*** *A G F K I Met C R V*

Fig. S2. Predicted protein sequences of *icu11-3* splicing versions using ExPASy tool[23]

Open reading frames are highlited in red, Stop indicates premature stop codons

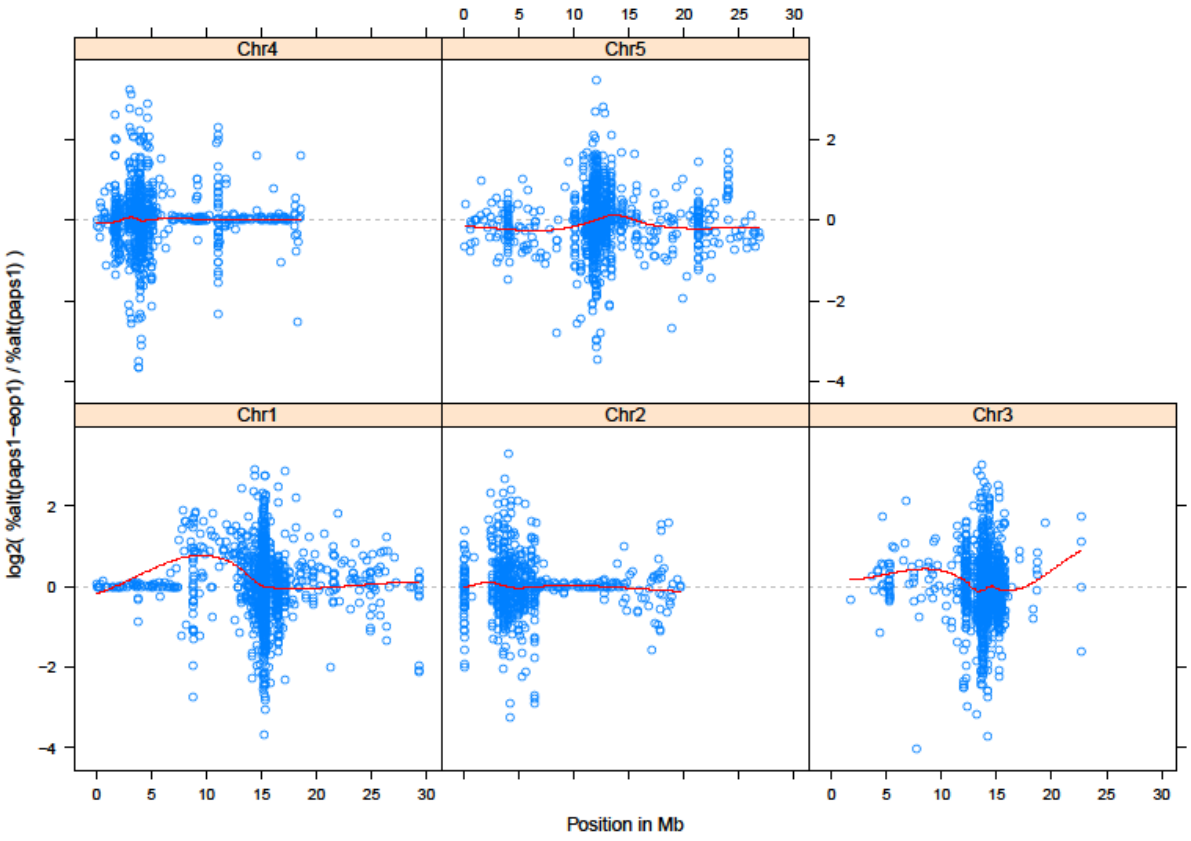


Fig. S3. Mapping of *eop1* mutation in *paps1-4/eop1* mutants compared to phenotypical wild type-like pool

Note the highest peak at the beginning of chromosome 1 around 10 Mb.

6.2 Oligonucleotide lists

ID		Amplification of	Sequence 5'-3'
LBO7	At1g22950_F	fw_Genotyping *eop1* (*icu11-3*)	GTCTTCTCTGCAGCTTTCCTTGC
LBO8	At1g22950_R	Rev_genotyping *eop1* (*icu11-3*)	ATGTACCTGGTCACGCCATTCTT
LBO77	AT1G22950qPCR_rev	Exonende 7 ICU11	GAAGTTTATGATTACTCTCATGTACC
LBO79	AT1G22950qPCR_fw	Exon 8 ICU11	GCATCCTCCACACCAACCTGAGAA
LBO87	FLAG_207D09_LP	Genotyping *icu11-4* together with LBO205	GTTGACGAGGAAGCAAAGTTG
LBO88	FLAG_207D09_RP	Genotyping *icu11-4* together with LBO205	ATCGATTCGTGTGGAATTGAG
LBO108		qPCR primer EOP_revexon exon junction	CTTCTTACCAACGTCTGCATGGAG
LBO110		qPCR primer EOP_fw2	CCCATTTCTCCATATGTTCAACC
LBO203	FLAG_402G04_LP	genotyping *icu11-2* together with LBO205	GTAACCATTGCCTCTGTGCTC
LBO204	FLAG_402G04_RP	genotyping *icu11-2* together with LBO205	AAATCATCAAGGACAACACCG
LBO205	LB4(INRA lines)	TDNA BP primer	CGTGTGCCAGGTGCCCACGGAATAGT

qPCR MADS box transcription factors

ID	Amplification of	Sequence 5'-3'
LBO206	qAG_F	CCGATCCAAGAAGAATGAGCTCTT
LBO207	qAG_R	CATTTTCAGCTATCTTTGCACGAA
LBO208	qAGL5_F	TCCGATCCAAGAAGCACGAGATGT
LBO209	qAGL5_R	TCGTTTTGCAGCTCGATTTCCCTT
LBO210	qAGL11_F	TCAATCTCCCTTTTCTGCGCGTTT
LBO211	qAGL11_R	TCAGGTCCAAGAAGCATGAGTTGC
LBO212	qAGL42_F	AGCAATCACGACTCACAAATTCAC
LBO213	qAGL42_R	AGCCTTTCTTTCTCGGACCTTTC
LBO214	qFT_F	CTGGAACAACCTTTGGCAAT
LBO215	qFT_R	AGCCACTCTCCCTCTGACAA
LBO216	qMAF5_F	TAGCGAGGTGGGGAAGCTGAAGAA
LBO217	qMAF5_R	GGCCAGAGCTATTTTCCGGTGACA
LBO218	qAP3_F	CCCTAACACCACAACGAAGGAGAT
LBO219	qAP3_R	GTTTCCTCTTGGTTTCTTGCATTC
LBO220	qSEP1_F	CTTTGGCAATGAAGCTGGATGA
LBO221	qSEP1_R	CCCTGAGACTGAGCTTGATGA
LBO222	qSEP2_F	GCACCTCCAACATGCTCAAGA
LBO223	qSEP2_R	TTCAAGTACTCTCTGTAGCTGTT
LBO224	qSEP3_F	TTAGCAGTTGAACTTAGTAGCCA
LBO225	qSEP3_R	CCAAGATCTTCTCCCAACAGAT
LBO226	qOTC_F	TGAAGGGACAAAGGTTGTGTATGTT
LBO227	qOTC_Rd	CGCAGACAAAGTGGAATGGA
LBO305	fw 3UTR ICU11 mateo group	AATATGTCAAATACCTCATCTTCT
LBO306	rev 3UTR ICU11 mateo group	TGCCTTCACTCCTCTAAGAAATT

1

Cloning strategy for ICU11 transgenes

ID	Amplification of	Sequence 5'-3'
LBO121	fw slice primer for linkervYFP into pEOP:gEOP(downstream STOP+3'UTR) out of pAs0006	Cttcaaaatcatgtgccgaggatccaggagcaggagcagg
LBO122	rev slice primer for linkervYFP into pEOP:gEOP(downstream STOP+3'UTR) out of pAS0006	Tgacatattcttgtttttgtttacttgtacagctcgtccatgcc
LBO128 New	rev primer for pEOP:gEOP into pBLUE (cutting sites SacI&KpnI)	Atatatggtaccaccaaaaggtgtttttagtttttg
LBO202	original Hbo69 pBAR/Hyg primer	GGGTCAGCACCGTTTCTGC
LBO228	fw promotor ICU11 SacI	atatatGAGCTCcgaggaagcaaagttggtgac
LBO229	rev promotor ICU11 KpnI	Atatatggtacccaggtcaaagaaagcgatac
LBO230	fw amplification CDNA ICU11 splicing1&2 SacI	atatatGAGCTCATGTGCAATCAAACTCCTCTTA
LBO231	rev amplification CDNA ICU11 splicing1&2 KpnI	atatatGGTACCTTACTCGGCACATGATTTTGAA
LBO232	rev linearization pBLUE:pEOP:gEOP for slicing with YFP (to get rid of STOP)	Ctcggcacatgattttgaa
LBO233	fw linearization pBLUE:pEOP:gEOP for slicing with YFP (to get rid of STOP)	Acaaaaacaagaatatgtcaaat
LBO234	rev amplification of pICU11:gICU11:YFP:3UTR out of pBLUE:pICU11:gICU11:YFP:3UTR for slice into pHYGMAP	Cttgagctctagaaggcgcgcgaggaagcaaagttggtga
LBO235	rev amplification of pICU11:gICU11:YFP:3UTR out of pBLUE:pICU11:gICU11:YFP:3UTR for slice into pHYGMAP	Ccattaattaatggggcgcgaccaaaaggtgtttttagttttgt
LBO237	fw linearization of cDNAICU11 Splicing out of pJET/pBLUE:cDNAICU11premSTOP for slice with YFP	Ctcggcacatgattttgaagcca
LBO236 New	fw linearization of cDNAICU11 Splicing out of pJET:cDNAICU11premSTOP for slice with YFP	Atctttctagaagatctcctacaat
LBO238	fw amplification of YFP out of pAS0006 for slicing into pBLUE:cDNAICU11premSTOP	Cttcaaaatcatgtgccgaggatccaggagcaggagcagg
LBO239	fw amplification of YFP out of pAS0006 for slicing into pBLUE:cDNAICU11premSTOP	Tggacgagctgtacaagtaaggtacccgggggatcctctag
LBO240	rev linearization of cDNAICU11 Splicing:YFP out of pBLUECDNAICU11Spli:YFP for slice to promoter	Ctcatcaaaaaaaaactaaaccct
LBO241	fw linearization of cDNAICU11 Splicing:YFP out of pBLUECDNAICU11Spli:YFP for slice to promoter	Atgtgcaatcaaactcctctta
LBO242	fw amplification of promotor ICU11 out of pBLUE:pICU11 for slicing into pBLUE:CDNASPlic:YFP	Accatggtgaattcgagctccgaggaagcaaagttggtgact
LBO243	rev amplification of promotor ICU11 out of pBLUE:pICU11 for slicing into pBLUE:CDNASPlic:YFP	Agaggagtttgattgcacatcaggtcaaagaaagcgatacgca

LBO244	fw amplification of pICU11:cDNA11SPl:YFP for slicing into pHygMAP	Attaattaatggggcgcgttactcggcacatgattttgaa
LBO245	rev amplification of pICU11:cDNA11SPl:YFP for slicing into pHygMAP	Ccattaattaatggggcgcgttacttgtacagctcgtcca
LBO246	fw site directed mutagenesis cDNA ICU11 introducing H > S and D > A substitution	Tgatcttgggttttctgtggctgactcagaag
LBO247 New	rev site directed mutagenesis cDNA ICU11 introducing H > S and D > A substitution	Agtcagccacagaaaacccaagatcaacatcc
LBO248	fw site directed mutagenesis cDNA ICU11 introducing 2ndH > S	Atcgtggacgtcatcgctctggtgctag
LBO249 New	rev site directed mutagenesis cDNA ICU11 introducing 2ndH > S	Agcaccagagcgatgacgtccacgatgaagaa
LBO250	pJET1.2forward	CGACTCACTATAGGGAGAGCGGC
LBO251	pJET1.2rev	AAGAACATCGATTTTCCATGGCAG
LBO252	YFP frame correction fw	Aatcatgtgccgagacgaattcct
LBO253	YFP frame correction rev	Gagacgaattcctgcagcccg
LBO254	rev YFP aplification out of AS006 fro slicing to pJET:cDNAicu11-3splicing2lacking7AS	Aggagatcttctagaaagatttacttgtacagctcgtcca
LBO255	fw YFP aplification out of AS006 fro slicing to pJET:cDNAicu11-3splicing2lacking7AS	Cttcaaaatcatgtgccgaggatccaggagcaggagcagg
LBO256	fw amplification pICU11 out of pJET into pJET:CDNAsplicing1&2YFP	Ctcgagtttttcagcaagatcgaggaagcaaagttggtga
LBO257	revamplification pICU11 out of pJET into pJET:CDNAsplicing1&2YFP	Agaggagtttgattgcacatcaggtcaaagaaagcgatac
LBO258	rev linearization pJET:cDNASPlicing1&2icu11-3 for slicing with pICU11 together with LBO241	Atcttgctgaaaaactcgagcc
LBO259	rev amplification of pICU11:cDNA1SplicingprematureSTOP out of pJEt into pHygMAP	Ccattaattaatggggcgcgctcggcacatgattttgaag
LBO259 New	rev amplification of pICU11:cDNA1SplicingprematureSTOP out of pJEt into pHygMAP	Attaattaatggggcgcgttactcggcacatgattttgaa
LBO260	fw slice cDNAICU11 residue changes into pHygMAP	Cttgagctctagaaggcgcgcatgtgcaatcaaactcctc
LBO261	rev slice cDNAICU11 residue changes into pHygMAP	Atccattaattaatggcgcgttactcggcacatgattttg
LBO262	fw Slicing amplification of pICU11 out of pJET:pICU11 for slicing upstream of CDNA residuec changes (ML1071)	Ctcgagtttttcagcaagatcgaggaagcaaagttggtga
LBO263	revSlicing amplification of pICU11 out of pJET:pICU11 for slicing upstream of CDNA residuec changes (ML1071)	Agaggagtttgattgcacatcaggtcaaagaaagcgatac
LBO264	fw linearization of pJET:cDNA residue 1st&2nd for slicing with pICU11	Atgtgcaatcaaactcctcttag

LBO265	rev linearization of pJET:cDNA residue 1st&2nd for slicing with pICU11	Atcttgctgaaaaactcgagcc
LBO266	linearization of pICU11:cDNA1sy&2nd for slicing to ML1071	Cttgagctctagaaggcgcgcgaggaag caaagttggtga

Crisp/Cas9 cloning

ID	Amplification of	Sequence 5'-3'
LBO281	Crisp Cas9 F1 gRNA1	ATATATGGTCTCGATTGATCGAGAGCCA TAGATCTAAGgttttagagctagaaatagcaagttaaaat
LBO282 New	Crisp Cas9 F2 gRNA3	ATTATTGGTCTCTAAACCTCTTGAGCGTTA CTTACCTGcaatctcttagtcgactctaccaata
LBO283	Crisp Cas9 R1 gRNA2	ATATATGGTCTCGATTGGGCGAGGCAAGA TTGAAGCTTgttttagagctagaaatagcaagttaaaat
LBO284	Crisp Cas9 R2 gRNA4	ATTATTGGTCTCTAAACTCTTCTCCGATATT CCCCTGAcaatctcttagtcgactctaccaata
LBO303	fw crispcas ICU11 genotyping	Ctaaaccctagaatgcgtatcgctttc
LBO304	rev crispcas ICU11 genotyping	Cagaaatataagctcgaactaacaagaat

MIGS cloning

ID	Amplification of	Sequence 5'-3'
LBO191	rev pAS95 linearization incl miR173	Ttcgcttgtagagaaaaatcactcgagcccgggtaccctgtcc
LBO192	fw cEOP1 overhang miR173	Gatttttctctacaagcgaagcgaggcaagattgaagctt
LBO193	rev cEOP1 amplification overhang pAS95	Caggtcgactctagaggatccatcgtattaggtctcattattg
LBO194	fw cEOP2 overhang miR173	Gatttttctctacaagcgaaaacaatttcggtgttgtcct

GA biosynthesis genes

ID	Amplification of	Sequence 5'-3'
LBO334	AtGA20ox1 At4g25420 fw qPCr	CTCATGAATACACGAGCC
LBO335	AtGA20ox1 At4g25420 rev qPCr	TGATACACCTTCCCAAATG
LBO336	AtGA20ox2 At5g51810 fw qPCr	ATGCTCACCGTTTGATGG
LBO337	AtGA20ox2 At5g51810 rev qPCr	CCTTCCCAAACTGCTCG

6.3 Vector list

ID	Transgene	Bacteria	Plant

		resistance	selection
AS95	35S:MCS:Term (hindIII/EcorI fragment) from ML596 in pAS77	amp	
pLB25	pAS95:35s:MIGS:cEOP1	amp	
pLB26	pAS95:35s:MIGS:cEOP2	amp	
pLB27	pHygMAP-3 (ML1071)	kan	Hyg
pLB28	pHygMAP:35s:miR137:cEOP2	kan	Hyg
pLB29	pHygMAP:35s:gEOP:YFP(not inframe)	amp	Hyg
pLB30	pJET:pICU11#16	amp	
pLB31	pJET:pICU11:gICU11#16	amp	
pLB32	pJET:pICU11gICU11YFP3'UTR #10	amp	
pLB33	pAS95:35sgEOPYFP(corr) #2	amp	
pLB34	pJET:cDNAICU11wt #7_1	amp	
pLB35	pBLUE ML939	amp	
pLB36	pJET:cDNAicu11-3SPlicing1 #15_12	amp	
pLB37	pJET:pICU11:Splicing1#17	amp	
pLB38	pJET:pICU11:Splicing2#5	amp	
pLB39	ML1071:2nd residue change # 19	kan	Hyg
pLB40	ML1071:pICU11:gICU11:YFP:3UTR # 17	kan	Hyg
pLB41	ML1071:1st residue change # 1	Kan	Hyg
pLB42	pJET:pICU11:1st residue change # 31	Amp	
pLB43	ML1071:35sgEOPYFP(corr)	Kan	Hyg
pLB44	pJET:pICU11: cDNA ICU11 1st residue change #1_77	Amp	
pLB45	pJET:pICU11: cDNA ICU11 2nd residue change #2_60	Amp	
pLB46	pJET:icu11-3 cDNA Splicing 1 #15_12 (with intron7)	Amp	
pLB47	pJET:icu11-3 cDNA Splicing2 #15_6 (-44AA)	Amp	
pLB48	ML1071:pICU11:1st residue change #52	Kan	Hyg
pLB49	ML1071:pICU11:Splicing1 icu11-3 #20	Kan	Hyg
pLB50	ML1071:pICU11:Splicing2 icu11-3 #1	Kan	Hyg
pLB51	pBEEH104E:CrispCas9gRNA1&4 #3	Kan	BASTA
pLB52	pBEEH104E:CrispCas9gRNA1&3#1	Kan	BASTA
pLB53	pBEEH104E:CrispCas9gRNA2&4 #1	Kan	BASTA
pLB54	ML1071:pICU11:2ND residue change #5	Kan	Hyg

7. Supplementary *Amsinckia* project: Figures and Tables

7.1 Read count ratios for *S*- and *s*-alleles

	blastn	Transcriptome read counts corolla			Transcriptome read counts style			Transcriptome read counts corolla			Transcriptome read counts corolla			Genomic read counts			Genomic read counts +5			Genomic ratio *S/L*
***S*-allele**	**identity**	**L1**	**L2**	**L3**	**L1**	**L2**	**L3**	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**	**S**	**H**	**L**	**S**	**H**	**L**	
***s*-allele**																				
k141_103992	100	30	49	37	63	61	56	354	238	189	1211	825	825	44	63	13	49	68	18	2,722
k141_156335	98	376	484	381	818	865	483	278	215	200	667	470	450	41	0	87	46	5	92	0,500
k141_140146	100	4	8	4	0	3	2	1730	1117	1119	2503	1551	1463	36	0	0	41	5	5	8,200
k141_51496	99	1125	1461	1484	1157	946	977	651	336	665	605	612	538	54	57	84	59	62	89	0,663
k141_196847	100	0	0	0	1	1	0	167	89	44	453	652	244	37	4	1	38	9	6	7,000
k141_196908	95	133	82	55	605	355	501	71	31	44	267	284	172	37	39	80	42	44	85	0,494
k141_205251	100	17	27	10	22	26	12	524	375	302	536	378	287	20	12	1	25	17	6	4,167
Kein 2. Hit	-																	5	5	
k141_215832	100	0	0	0	2	2	0	23	30	29	768	481	492	46	0	0	42	5	5	10,200
k141_357057	98	119	161	147	1562	1494	950	49	53	64	1029	402	529	37	49	97	42	54	102	0,412
k141_250207	100	0	0	0	0	1	0	14	14	7	29	42	25	46	32	40	51	37	45	1,133
k141_280729	96	172	355	403	692	751	438	105	167	112	478	299	373	21	15	62	26	20	67	0,388
k141_254140	100	0	0	0	0	0	0	52	38	9	281	261	162	33	0	0	38	5	5	7,600
k141_254080	95	111	151	73	727	643	450	131	81	73	596	678	420	73	48	89	78	53	94	0,830
k141_258212	100	20	46	53	20	32	38	1282	1318	846	808	1601	1306	7	0	4	12	5	9	1,333
k141_78230	97	3089	3098	3912	5935	2468	3382	2667	1614	1973	2010	3037	2957	65	55	56	70	60	61	1,148
k141_260889	100	2	9	5	2	4	3	22	62	25	20	26	36	40	0	5	45	5	10	4,500
k141_180029	93	73	69	73	72	36	37	40	42	24	17	31	16	54	70	77	59	75	82	0,720
k141_289497	100	15	89	59	17	41	21	1012	618	759	839	703	739	44	4	4	49	9	9	5,444
k141_91290	97	2124	2465	2551	1750	1250	1322	1162	607	1218	801	783	1003	35	47	84	40	52	89	0,449
k141_309157	100	15	16	17	15	10	6	798	426	423	1336	668	648	43	0	0	48	5	5	9,600

k141_309108	97	1308	1412	1449	1683	1777	1130	657	468	613	1396	673	735	23	38	61	28	43	66	0,424
k141_31321	100	1	2	5	1	0	1	555	550	278	548	273	208	14	17	0	19	22	5	3,800
k141_291779	99	1287	1752	1073	512	454	215	1269	1287	865	1194	505	522	60	24	86	65	29	91	0,714
k141_314453	100	3	10	6	3	3	5	623	431	395	624	623	424	29	1	0	34	6	5	6,800
k141_86659	99	877	1140	821	1552	786	1191	542	422	514	498	626	486	30	17	40	35	22	45	0,778
k141_333738	100	0	3	2	1	0	0	598	394	323	320	376	284	27	0	0	32	5	5	6,400
k141_71305	99	1114	738	502	1386	646	838	524	312	442	463	610	595	68	54	98	73	59	103	0,709
k141_338387	100	13	4	20	48	34	30	227	210	132	589	365	393	27	18	12	32	23	17	1,882
k141_364857	94	1199	827	1009	3025	3004	1725	1858	1315	1388	6448	3606	3807	75	48	102	80	53	107	0,748
k141_349598	100	2	7	3	2	2	6	900	1278	556	504	696	636	14	1	0	19	6	5	3,800
k141_164051	99	1428	1907	1608	2039	684	934	831	1263	883	450	721	826	35	24	44	40	29	49	0,816
k141_374145	100	4	10	6	3	1	1	1028	615	518	828	1044	831	24	5	0	29	10	5	5,800
k141_372327	96	1070	1306	1334	1866	1140	1355	821	569	669	658	854	854	29	47	85	34	52	90	0,378
k141_42659	100	19	68	41	23	19	40	346	364	258	428	308	247	11	0	0	16	5	5	3,200
k141_156700	97	424	644	601	718	336	548	286	242	256	316	260	221	73	62	98	78	67	103	0,757
k141_71391	100	1	22	2	2	4	3	1995	864	882	7114	5325	4169	12	0	0	17	5	5	3,400
k141_44704	98	156	303	251	878	899	488	122	149	101	973	374	457	46	44	67	51	49	72	0,708
k141_73500	100	10	29	4	3	10	2	1076	900	720	915	841	756	21	0	0	26	5	5	5,200
k141_98871	99	2764	3155	3317	2497	1807	1910	1439	954	1392	1320	1127	1061	49	43	75	54	48	80	0,675
k141_76744	100	2	16	7	2	5	1	389	258	267	357	261	201	34	32	0	39	37	5	7,800
k141_76755	97	664	701	646	490	413	411	410	255	390	305	214	217	29	0	83	34	5	88	0,386
k141_841	100	118	133	169	104	101	111	1580	986	1042	1465	998	906	51	19	34	56	24	39	1,436
k141_380247	99	2781	3287	3336	2601	2143	2068	1571	1053	1635	1501	1107	1033	21	37	57	26	42	62	0,419

Tab. S1 Read counts and *S/s* ratios of 22 *S*- and their corresponding *s*-alleles for individual transcriptome and genome samples

7.2 Oligonucleotide list *Amsinckia* marker

ID	Candidate	Sequence 5'-3'
LBO309	fw candidate G gene	GTGCATTACAAATATTCTAAGGACACTTT
LBO310	rev candidate G ene	AGTCCTGCTCCACTTCCAACTC
LBO311	fw Beta-1,3-galactosyltransferase	CTTCTTCATAAAATAAATAAATCCC
LBO312	rev Beta-1,3-galactosyltransferase	GTGGAGAAGGTGAAGAGGCTCTTTCG
LBO313	fw arabinogalactan	AAAATACTGATCAAATAAAAAA
LBO314	rev arabinogalactan	GGGTGCTTTGGTTGGAGCTT
LBO315	fw atp synthase	GACACAAGAAAGATGAATTACTTGG
LBO316	rev atp synthase	GTAAGAAGATCAGAGAGTATGA
LBO317	fw ribosomal protein	TAAACATCTATTACAAAACCTCC
LBO318	rev ribosomal protein	AAGAAGGCATTCACTGGA
LBO319	fw ribosomal protein2	CACCTGAAATCAGCTCTCCG
LBO320	rev ribosomal protein2	CAGCTTGGGGGAGTTCGCTT
LBO321	fw tubulin 1 chain	CTCCCAGTTGGGTCAATCCCAT
LBO322	rev tubulin1 chain	AAACAAACCCATTTTTTACCCCCC
LBO323	fw dystroglycan	AGAAACGATGAAACTGCTCAA
LBO324	rev dystroglycan	AAGATCAATCTAGTTTTAGTC
LBO325	fw atp citrate synthase	AATAGAAAAACAAACATCAACTAGTGG
LBO326	rev atp citrate synthase	ACTGGTCTTGCAAAAATGAGGAAGC
LBO327	fw no hit rosa	AACATTATTGAAATGTTGAAGGAGCATGT
LBO328	rev no hitrosa	TCATCTCATATCTATTTAACGCGGGAAT
LBO329	fw annexin	AGATACAAATCACCATTGTGACAAGCT
LBO330	rev annexin	TTCCACTTGACCGTGCCGTTGCTAAG

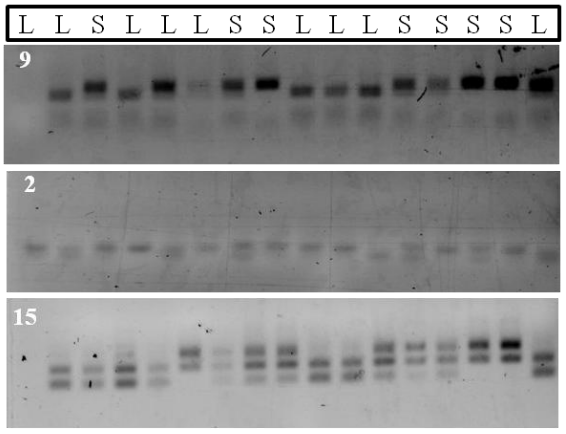


Fig. S4 dCAPS markers with non-morph-specific segregation

White numbers indicate marker numbers of Tab.2

7.3 Homozygosity scores for all PacBio Contigs

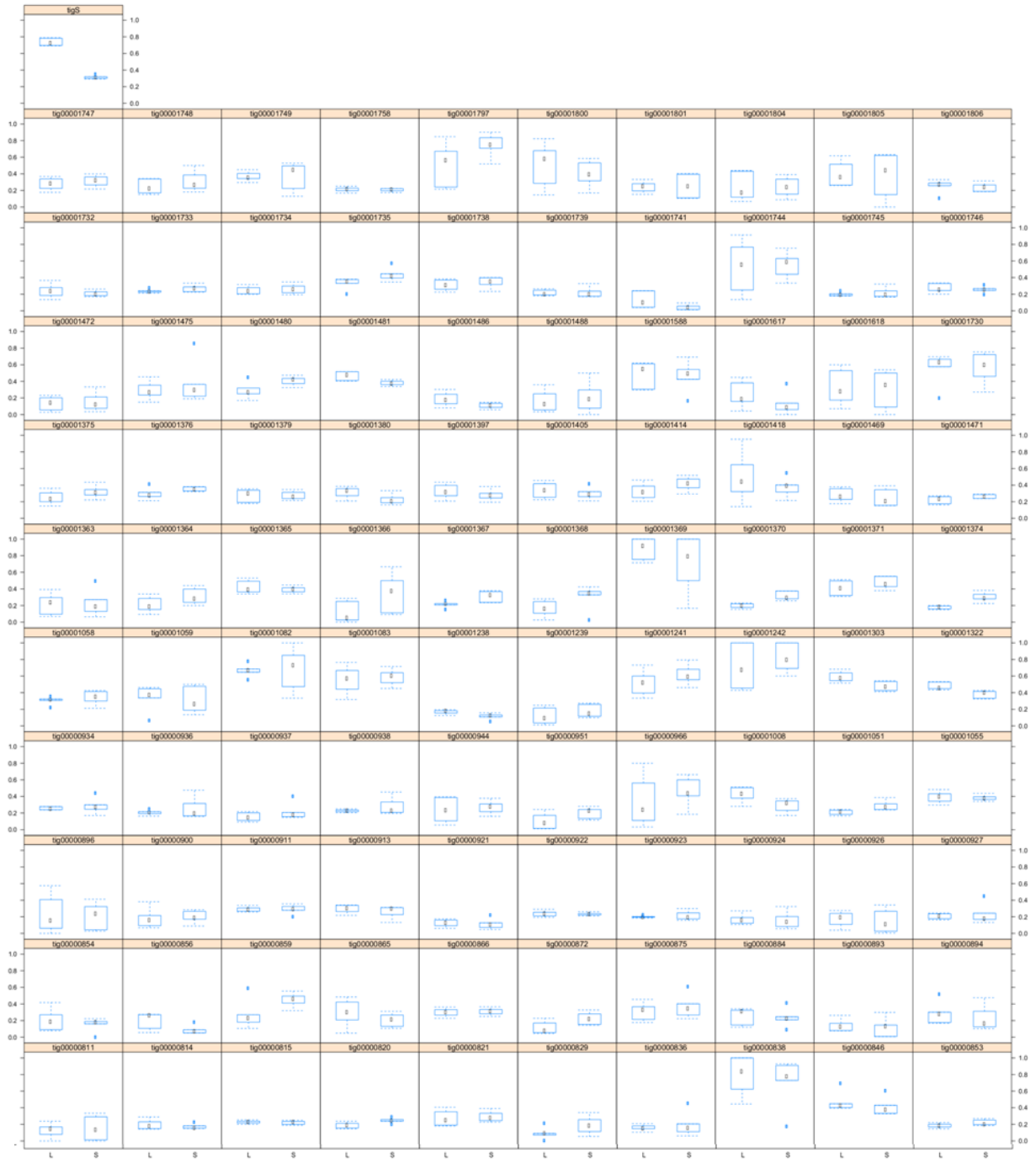


Fig. S5 Homozygosity plot for all selected PacBio assembly contigs I

Selection was based on contigs, where mapped RNA-seq reads could be detected and contigs containing morph-specific variances (quality score 900-999 and a minimal heterozygosity of at least three individuals with heterozygous call)

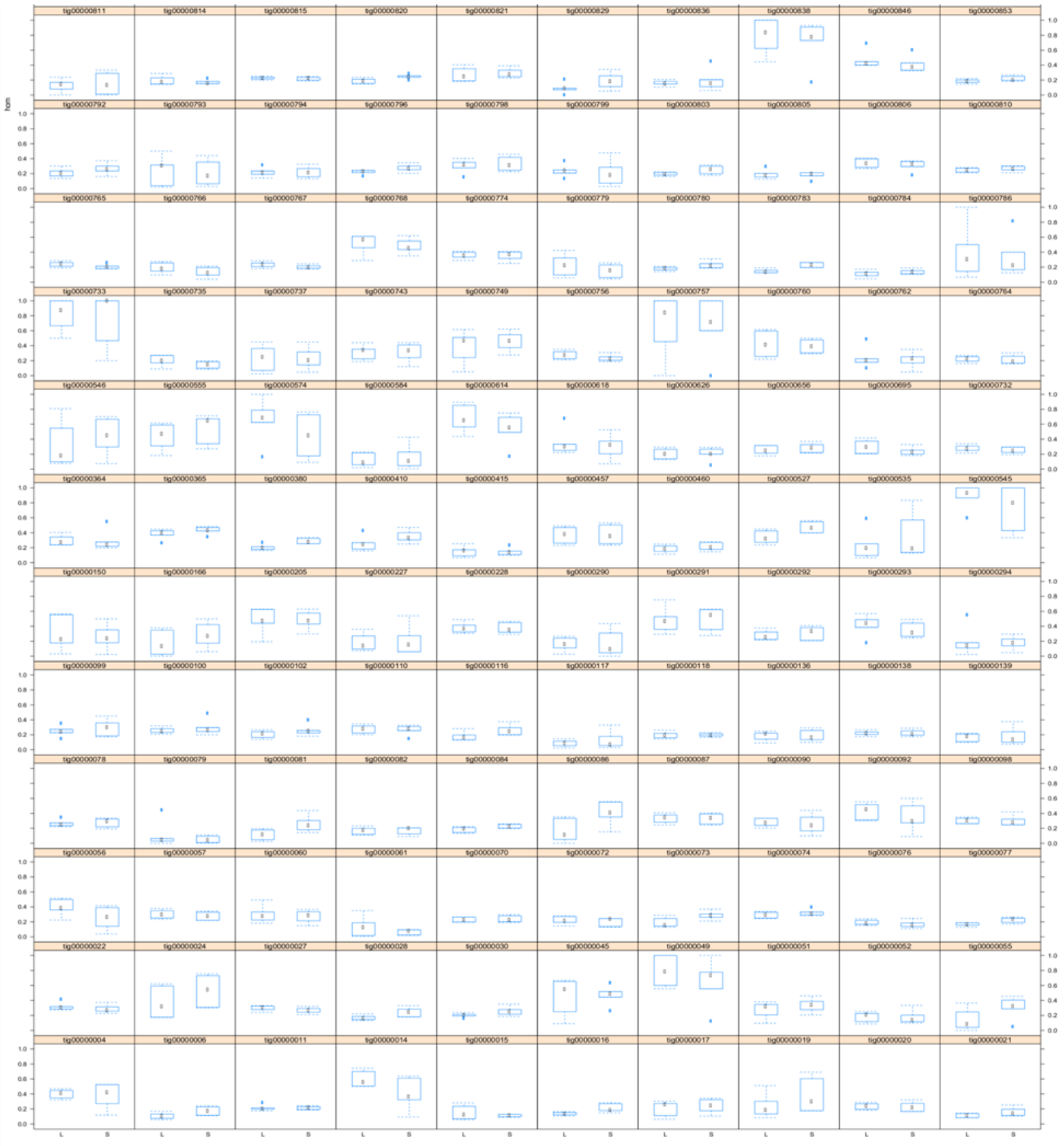


Fig. S6 Homozygosity plot for all PacBio assembly contigs II

Selection was based on contigs, where mapped RNA-seq reads could be detected and contigs containing morph-specific variances (quality score 900-999 and a minimal heterozygosity of at least three individuals with heterozygous call)

7.4 Marker positions in PacBio assembly

Transcriptome ID	Contig origin	Subject position in PacBio assembly	Subject position on tigS	Putative gene (blastn)
k141_215832	tig00000125	1127551	1127707	Gibberellin regulated protein,snakin-2-like
k141_140146	tig00000723	20257424		40s ribosomal protein S11
k141_103992	tig00000723	8863686		beta-1,3-galactosyltransferase 7-like
k141_309157	tig00000723	16925353	19051984	ATP-citrate synthase alpha chain protein 1-like
k141_196847	tig00000723	14640845	16766643	copper transporter 2-like
k141_250207	tig00000125	1073207		No hit
k141_260889	tig00000723	16923445		Phosphatidylinositol N-acetylglucosaminyltransferase subunit P
k141_333738	tig00000723	2302291		annexin D2-like
k141_71391	tig00000723	5187251	7313157	arabinogalactan peptide 23
k141_841	tig00000723	10653406		acetyl-CoA-benzylalcohol acetyltransferase-like
k141_76744	tig00000723	20436021		30S ribosomal protein S6B
k141_205251	tig00000723	16766916		No hit
k141_258212	tig00000723	17484033		tubulin beta-2 chain-like isoform
k141_289497	tig00000723	14114417		40S ribosomal protein S16-like
k141_31321	tig00000723	6586525		probable indole-3-pyruvate monooxygenase YUCCA10
k141_314453	tig00000723	13759365		translation machinery-associated protein 22 isoform X1
k141_349598	tig00000723	5985017	8111108	tubulin beta-1 chain
k141_374145	tig00000723	15513017		protein BOBBER 1
k141_42659	tig00000723	1125396		palmitoyl-protein thioesterase 1-like
k141_73500	tig00000723	11365133	13491224	40S ribosomal protein S13
k141_338387	tig00000723	30536095		annexin
k141_254140	tig00000723	18512165	20638077	

Tab. S2 Individual marker positions in PacBio contigs and in tigS

7.5 TCP14 class I PacBio sequences

>TCP14_first_tigS_8440818-8441840
ATGGAAGGCGAAAATATTCATGTTCATGCCATGTCAAGTAGGTCACCGAATTTCCCA
TTACAGTTGCTAGAGAAAAGAGATCAAGAAGCAGTTTGTTCCAGCTCCACTTACCCT
TCTTCTTCTGAGATCTCTAGAAAAGATGAAGCTCTCAATTCAGCTGATCAGTCCATT
ATCAAGAAGCCTCTACCAAAACGCCCAACAACAAAGGATAGGCATACTAAAGTAGA
TGGTCGTGGCCGTAGGATTCGCATGCCGGCTGCATGTGCTGCTAGGGTTTTTCAGCT
CACCAAAGAATTAGGCCATAAGTCCGACGGCGAAACCGTCGAATGGCTGCTTCAAC
AAGCGGAGCCGTCTGTTATCGCCGCCACGGGGACGGGAACCATCCCGGCTAACTTC
ACTTCTCTGAATATATCCTTGAGAAGTTCTGGCTCCACCATCTCGGCTCCTTCATATT
TTAGGCACAACAACTATTATAACCAGAGTTTCATGGCGTCCCAACTAAGAGTTTTTG
AGGAGTCACAACGGCGTGTAATGTTCAATCAAATTGGTAATTTATCATCTGAAAATT
CGTCTCTAGGGTTGAATTTTAGTGGGAATGTCAATAATTTGAATGGACTAATGTTAC
AAGCTAAGCAAGAGTTACATGGTGGCACGGGCCTAGATATGGCGGAAATGGAGGCT
AGTATTGGGAGCAGAAAGAGGAGGTCGGAGGATGATAATCTACTTCAATTGCAAAA
TCATCAATGGAATTATATGTTACAATCAAGTGCTGGCTCAATTCCAGCTACTCATCA
AGGTCAGGTTCCAGCTACGGCATTGTTTATGGTAGCTAGCCCTAGTAGTAACAATCA
GATGGCTAGCGGAGACTCTTTGTGGACTAGTCCGAATGTTGGAAATAGAGAGAGTTT
AGTGCCAAGCAGTGGCTTAAATTTCTTAAATTTCCCTACTCAGTTATCACTGAGCAA
TAATAATAATGGATTGGGTAGTGGTGGTGGCGGCGGGGGCGGAATGGTGGAGGGGT
AA

>TCP14_second_tigS_8411318-8412340
ATGGAAGGCGAAAATATTCATGTTCATGCCATGTCAAGTAGGTCACCGAATTTCCCA
TTACAGTTGCTAGAGAAAAGAGATCAAGAAGCAGTTTGTTCCAGCTCCACTTACCCT
TCTTCTTCTGAGATCTCTAGAAAAGATGAAGCTCTCAATTCAGCTGATCAGTCCATT
ATCAAGAAGCCTCTACCAAAACGCCCAACAACAAAGGATAGGCATACTAAAGTAGA
TGGTCGTGGCCGTAGGATTCGCATGCCGGCTGCATGTGCTGCTAGGGTTTTTCAGCT
CACCAAAGAATTAGGCCATAAGTCCGACGGCGAAACCGTCGAATGGCTGCTTCAAC
AAGCGGAGCCGTCTGTTATCGCCGCCACGGGGACGGGAACCATCCCGGCTAACTTC
ACTTCTCTGAATATATCCTTGAGAAGTTCTGGCTCCACCATCTCGGCTCCTTCATATT
TTAGGCACAACAACTATTATAACCAGAGTTTCATGGCGTCCCAACTAAGAGTTTTTG
AGGAGTCACAACGGCGTGTAATGTTCAATCAAATTGGTAATTTATCATCTGAAAATT
CGTCTCTAGGGTTGAATTTTAGTGGGAATGTCAATAATTTGAATGGACTAATGTTAC
AAGCTAAGCAAGAGTTACATGGTGGCACGGGCCTAGATATGGCGGAAATGGAGGCT
AGTATTGGGAGCAGAAAGAGGAGGTCGGAGGATGATAATCTACTTCAATTGCAAAA
TCATCAATGGAATTATATGTTACAATCAAGTGCTGGCTCAATTCCAGCTACTCATCA
AGGTCAGGTTCCAGCTACGGCATTGTTTATGGTAGCTAGCCCTAGTAGTAACAATCA
GATGGCTAGCGGAGACTCTTTGTGGACTAGTCCGAATGTTGGAAATAGAGAGAGTTT
AGTGCCAAGCAGTGGCTTAAATTTCTTAAATTTCCCTACTCAGTTATCACTGAGCAA
TAATAATAATGGATTGGGTAGTGGTGGTGGCGGCGGGGGCGGAATGGTGGAGGGGT
AA

8. Supplementary: Chemicals and technical equipment

8.1 Chemicals

Chemicals were purchased from Biocat (Heidelberg, Germany), Biozym (Hessisch Oldendorf, Germany), Carl Roth (Karlsruhe, Germany), Duchefa Biochemie (Haarlem, Netherlands), Invitrogen (by Thermo Fischer Scientific, Waltham, Massachusetts), Macherey-Nagel (Düren, Germany), Quiagen (Hilden, Germany), Roche Diagnostics (Grenzach-Wyhlen, Germany), Serva (Heidelberg, Germany), Sigma-Aldrich (now Merck, St. Louis, Missouri, USA)., Enzymes were purchased from Bioline (London, UK), Invitrogen (by Thermo Fisher Scientific), NewEngland Biolabs (Frankfurt am Main, Germany), Roche Diagnostics (Grenzach-Wyhlen, Germany), TakaraBio USA Inc. (Mountain View, USA) and Thermo Fischer Scientific (Waltham, USA).

Sequencing materials were ordered from Illumina (San Diego, USA) and oligonucleotides were produced by Sigma-Aldrich (now Merck, Darmstadt, Germany).

Concentrations of antibiotics that were used for selective growth media are shown in Table S3.

Antibiotics	Dissolved in	Working concentration
Ampicillin	ddH2O	100 µg/ml
Kanamycin	ddH2O	50 µg/ml
Gentamycin	ddH2O	25 µg/ml
Rifampicin	DMSO	80 µg/ml
Hygromycin B	ddH2O	50 µg/ml

Tab. S3 Working concentrations of used antibiotics

8.2 Disposable equipment

Lab consumables were purchased from Greiner Bio-One (Kremsmünster, Austria), Kisker Biotech (Steinfurt, Germany), Sarstedt (Nümbrecht, Germany), Starlab (Hamburg, Germany) and VWR (Radnor, USA). Materials for plant cultivation were ordered from Fitz Kausek (Mittenwalde, Germany).

Affidavit

I confirm that the PhD thesis entitled “Impact of growth-related genes on petal size in *Arabidopsis thaliana* and the formation of two distinct floral morphs in *Amsinckia spectabilis*” is the result of my own work. I did not receive any aids or support other than stated. All sources and materials applied are listed and specified in the thesis. Furthermore I confirm that this thesis has not yet been submitted as part of another examination process, neither in identical nor in similar form.

Ich versichere hiermit, dass ich die Doktorarbeit mit dem Titel „Impact of growth-related genes on petal size in *Arabidopsis thaliana* and the formation of two distinct floral morphs in *Amsinckia spectabilis*“ selbstständig und ohne unzulässige fremde Hilfe erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Potsdam, 10.03.2021 Lisa Bartholomäus

Acknowledgements

First, I would like to thank Prof. Dr. Michael Lenhard, who gave me the opportunity to perform my PhD within his lab, for his constant support and advice over the years.

I would like to thank all members of the Lenhard group for the scientific discussions and for the very friendly and supportive working atmosphere. I am especially thankful to Dr. Adrien Sicard and Dr. Duarte Figueiredo for their constant and immediate help and advice, as well for introducing me into the confocal microscopy. Special thanks to Dr. Cuong Huu Nguyen, for being the best heterostyly supervisor, for his friendly guidance and helpfulness. I am especially thankful to our technicians: Melanie Teltow, Cindy Marona, Peggy Lange and Lisa-Marie Barthel for all their help and organization that allow the running system of the lab. Melli, thanks for taking me into your confidence and for always listening, I really enjoyed our tinkering times. Peggy, thanks for sharing silent hours as well as nice talks in the last months. I would like to thank Dr. Yunming Zhang for his positivity and kindness, Dr. Moritz Jöst for being a great office mate, for listening, for being non-judgmental and for his calm honesty and Thi Chi Tran for not leaving me alone in a male office, for lots of laughter and lightness. Special thanks to Dr. Christian Kappel, for being a fantastic office mate and of course, thanks for his great bioinformatics for this project and for his helpfulness; I could always rely on. Very special thanks go to Dr. Natalia Wozniak for being my best friend, my sports companion, for tons of great food, endless shared hours of series and most and for all for providing me a safe place. Working with all of you made everything a whole lot better, and filled my heart with lots of great experiences and memories, I will always be grateful for.

Thanks to Prof. Dr. Axel Himmelbach and Susanne König from the IPK in Gatersleben for their assistance with the PacBio sequencing.

Finally and foremost, I want to thank my family: My two perfect boys, for every smile, for their giggling and for being the way, they are and especially my husband Alex for being my rock, my constant believer, for making everything easier and for his love.

I would like to thank my parents, my sister and my brother for their unconditional support.