이학석사 학위논문

Understanding the molecular mechanisms of seed number variation in *Arabidopsis* and *Lepidium*

애기장대와 다닥냉이의 열매 당 종자 수 다양성에 대한 분자적 메커니즘 이해

2022년 8월

서울대학교 대학원

생명과학부

김 동 우

Abstract

Understanding the molecular mechanisms of seed number variation in *Arabidopsis* and *Lepidium*

Dong Woo Kim

Department of Biological Sciences

The Graduate School

Seoul National University

Seeds, the reproductive organs that are the result of ovule fertility, cannot be separated from plant and human life. Seed production is a critical phase in the life history of plants and seeds are used as an essential resource of human history, such as food, feed and fuel. Especially, seed number is one of key traits for the plant fitness and crop yield. Thus, studying the genetic mechanism determining seed number is important, but still many unknown areas are there.

In this study, I identified significant variation in seed number per silique among 131 accessions of *Arabidopsis thaliana*. Ten ecotypes

with the lowest number of seeds per silique showed low fertility. Ten ecotypes with the highest number of seeds per silique showed large number of formed ovules. These results suggest that the variation in seed number per silique among 131 accessions is regulated by different fertility and formed ovule number.

To identify the genes regulating seed number per silique, Genome-wide Association Study (GWAS) was performed. The analysis characterized 107 Single Nucleotide Polymorphisms (SNPs) that are highly correlated with the observed variation in the seed number per silique among 131 accessions. Ten SNPs were selected for candidate genomic loci statistically supported. I further identified the candidate genes that are closely located or include the selected SNPs. Phenotyping T-DNA insertion knockout mutants of the candidate genes revealed that the loss of function of AT1G61890, *PUP15*, AT4G23030, and AT4G38710 caused significant different seed number per silique values. AT1G61890 and AT4G23030 encode MATE efflux family protein and the two genes decrease the seed number per silique. *PUP15* encodes a member of purine permeases family and *pup15* mutants displayed increased seed number per silique. Moreover, I further identified increase in the seed number per silique in the knockout mutants of AT4G38710 that encodes glycine-rich protein.

To complement the GWAS, I created a genetic population for Quantitative Trait Loci (QTL) analysis by crossing Col-0 with Di-G that showed consistent seed number phenotype between two independent studies. To confirm the crossing of Col-0 and Di-G, I produced genetic markers. The markers displayed different nucleic acid sequence in Col-0, Di-G, and their crossing. These results indicated success of crossing Col-0 and Di-G for QTL genetic population.

In addition to the study on the intra-species variation among *A. thalina* ecotypes, I tried to extend my study to the inter-species variation in seed number per silique between *Arabidopsis* and *Lepidium*. *Lepidium* and *Arabidopsis* are closely related species in the family *Brassicaceae* while they displayed distinct patterns of seed production in silique. *Lepidium* forms one seed per locule while *Arabidopsis* does around twenty seeds per locule.

To identify the collected *Lepidium* species, ITS part sequence of *Lepidium* species in NCBI is used. Used ITS primers to get ITS sequence of the collected *Lepidium*. Afterwards, I performed MEGA 11 software to compare the ITS sequence of the *Lepidium* samples and *Lepidium* species in NCBI. In MEGA 11, the ITS sequence of the *Lepidium* sample and *Lepidium* species are aligned by ClustalW option then construct phylogenic tree applying neighbor-joining

method. The phylogenic tree with 1000 Bootstrap replication, displayed that the *Lepidium* species is *Lepidium virginicum*.

To characterize the auxin accumulation patterns of *Lepidium* during ovule formation, I performed DR5 transformation to *Lepidium*. I used two plasmids pC3300:pDR5-ntdTOMATO and pC3300:pDR5-GUS for agrobacterium transformation. Afterwards, I performed floral dipping to *Lepidium*. The selection of *Lepidium* transformation sample is in progress.

In this dissertation, I investigated the genetic basis of intra- and inter-species divergence in seed production per silique. In intra-species study, I identified genes regulating seed number per silique through GWAS and phenotyping in *A. thaliana*. Furthermore, I did cross Di-G and Col-0 for QTL to overcome GWAS disadvantages. In Inter-species study, I dissected *Arabidopsis* and *Lepidium* fruits to characterize seed phenotype difference between them. I further identified collected *Lepidium* species using ITS sequences. In addition, I performed DR5 transformation to *Lepidium* to characterize auxin accumulation patterns for studying. I expect that my study will increase and yield control of important crop specie, such as *Brassica napus*.

Keyword : *Arabidopsis thaliana*, Natural variation, Seed number, Fertility, Formed ovule number, *Lepidium*

Student Number : 2020-29132

Table of Contents

List of Figures

List of Tables

Abbreviations

CMM	Carpel margin meristem
BR	Brassinosteroid
CK	Cytokinin
GA	Gibberellin
IAA	Indole-3-acetic acid
ABA	Abscisic acid
CRF	Cytokinin response factors
SNS	Seed number per silique
GWAS	Genome-wide association study
UONS	Unfertilized ovule number per silique
FONS	Formed ovule number per silique
ANT	AINTEGUMENTA
MP	MONOPTEROS
CUC	CUP-SHAPED COTYLEDON
BIN	BRASSINOSTEROID INSENSITIVE
BZR	BRASSINAZOLE RESISTANT
BES	BRI1-EMS-SUPPRESSOR
SNP	Single nucleotide polymorphisms
NAS	Nucleic acid sequence
MATE	Multidrug and toxic compound extrusion

1. Introduction

Seeds are reproductive organs that were produced when ovules in placenta of pistil were fertilized by pollens in anther of stamen. When pollens attach the stigma of pistil, they develop pollen tubes and deliver two sperm cells to ovules through pollen tubes. After two sperm cells reach the female gametophytes, ovule is fertilized. Seed production is a critical phase in the life history of plants. Seed production associated phenotypes such as seed number, sizes, shapes, nutrients, hardness of the seed coat and seed dispersal methods highly affect plants reproduction and survive (Sreenivasulu & Wobus, 2013). Seeds are not only important biologically but also other areas. Seeds are an essential source of food, feed, raw materials, and energy (Kern, 2020). Seeds, especially seed number is important. Seed number is a key trait tightly related to the plant fitness and crop domestication and improvement (Zhu *et al.*, 2020). Since seed number is phenotype that affect many areas, there are many studies what can affect it. For example, increasing the number of seeds per silique is a one of way to increase the total number of plant seeds (Huang *et al.*, 2013).

Ovule is an organ locate in ovary which is a part of the gynoecium

and it is initiated from the carpel margin meristem (CMM) during floral development. In *Arabidopsis*, ovule primordia are initiated at the same time and initiated where the auxin maxima that cells have high concentration of auxin are existed. These auxin maxima are fundamental for the primordia formation of plant lateral organs, such as lateral root, flower, and ovule (Benková *et al.*, 2003, 2009; Yamaguchi *et al.*, 2013). After initiated, ovule can be fertilized when pollination is occurred.

The irregular of floral organs, ovules, male and female gametes development can affect the ovules fertilization and development and cause decreasing seed number per silique in plant (Jiang *et al.*, 2020). These seed numbers of plants are determined by many factors such as formed ovule numbers and ovules fertility. The numbers of formed ovules that emerge from placenta determine the maximum possibility of seed number and ovule fertility (Yuan & Kessler, 2019). The ovules fertility can increase final seed number if developmental process were occurred normally. These showed that the complexity of seed number of plants. For instance, there were possibility that same seeds number that one was caused by high fertility and low ovule forming number, the other was caused by low fertility and high ovule forming number. For example, plant hormones such as auxins, brassinosteroids (BR), cytokinins (CK)

and gibberellins (GA) affect the seed number per silique through regulate ovule forming number.

Plant hormones affect many areas of plants such as development and growth. The representative four plant hormones auxins, BRs, CKs and GAs were also known to affect the seed number per silique both directly or indirectly. Auxins are the plant hormone that usually occurred as indole-3-acetic acid (IAA) in plant and affect plant development. It is known that auxins affect embryo, endosperm, seed coat and silique development during seed development (Nemhauser *et al.*, 2000; Cao *et al.*, 2020). Auxin triggers AINETEGUMENTA (ANT) and MONOPTEROS (MP) expression that induce ANT, CUP-SHAPED COTYLEDON1 (CUC1) and CUC2 during the early stage of placenta and ovule primordia development (Cucinotta *et al.*, 2014). Since auxins act on the gynoecium morphogenesis, seed can t develop normally when auxin expression is decrease or not work while seed development and cause decrease of seed number per silique. Also, ovule initiation is related with auxin maxima that cells have high concentration of auxin in *Arabidopsis*. BRs are a plant hormone that has functions to promote cell expansion, division, elongation and reproduction. BRs are also known positive regulator of the seed number per silique. When BRs are active, BRASSINOSTEROID

INSENSITIVE2 (BIN2) is inhibited by BRs signal and BRASSINAZOLE RESISTANT 1 (BZR1) and BRI1-EMS-SUPPRESSOR 1 (BES1) are released to regulate downstream (Li *et al.*, 2001; He *et al.*, 2005). BZR1 directly regulate ANT that controls cell proliferation in the placenta and ovules. BZR1 is the positive regulator and BIN2 is the negative regulator of the BRs signaling pathway. bzr1-1D that is gain of function mutant in BRs increase the seed and ovule number per silique while bin2 that is loss of function mutation in BRs decrease the seed and ovule number per silique (Huang, 2013). CKs are also one of hormones that affect ovule initiation and development. CKs have a critical role in the regulation of the plant cell cycle and many developments such as seed germination, apical dominance, and flower development (Akhtar, 2020). CKs positively regulate ovule formation and pistil development via auxin efflux carrier PIN-FORMED1 (PIN1). For instance, cytokinin response factors (CRFs) triple mutant indicate low PIN1 expression also crf2,3,6 triple mutant crf2,3,6 are known that decrease the ovule number per silique, placenta length and ovule density (Cucinotta *et al.*, 2016). GAs are known that negatively regulate ovule number separately of auxin signaling and transport (Gomez, 2018). GAs are plant hormones that affect plant development process such as seed germination, stem and root

elongation, leaf shape and de-etiolation, flowering time and fruit development (Swain & Singh, 2005). These effects are appeared because GAs induce the degradation of DELLAs. DELLAs are a subset of the plant-specific GRAS family of transcriptional regulators that act as GA-signaling repressors (Sun, 2010). DELLAs act as a node for crosstalk during nutrient responses, abiotic stress, light perception and signaling of some hormones interacting (Yu *et al.*, 2014). DELLA proteins are also known to affect the ovule forming of placenta as positive regulator in *Arabidopsis*. The gain of function DELLA mutant gai-1 indicated significant increasing ovule number while GA treated plants indicated reducing ovule number like null *della* mutant in *Arabidopsis* (Gomez, 2018).

This study tries to understand the molecular mechanisms of seed number variation in *Arabidopsis*. To do this, collect A. thaliana 131 accessions seed number per silique data. To measure precisely, only use siliques of main stem for data. After that do genome-wide association study (GWAS) analysis to find genes that regulate seed number per silique (SNS) GWAS result show the SNPs that highly relate with phenotype seed number per silique. The genes near SNPs are candidate genes. Compare the T-DNA insertion of candidate genes phenotype seed number per silique to wild type

Col-0 and find genes AT4G23030, *PUP15*, and AT4G38710. AT4G23030, *PUP15*, and AT4G38710 showed significantly different phenotype seed number per silique when compare to wild type. The results of GWAS analysis and phenotyping represent that AT4G23030, *PUP15*, and AT4G38710 affect the seed number per silique in *A. thaliana*.

2. Methods and Materials

2.1 Plant materials and growth conditions

178 *Arabidopsis thaliana* accessions, T-DNA insertion mutants, and *Lepidium virginicum* were grown under identical environmental conditions to prevent the environment from affecting the results.

178 ecotypes seeds were sowed for three days. 63, 63, and 54 ecotypes were sowed in the soil on the 1st, 2nd and 3rd day and each sowing included Col-0 for the control group. Seeds were moved to the germination chamber (dark room, 4℃) for five days and then transferred to the growth room for one week for germination and growth. Afterward, I transferred three plants per ecotype to pots. The pot shape was a reverse quadrangular pyramid, with six pots connected and ordered in two rows with three columns. The pots were plastic material and 4x4cm (top), 2x2cm (bottom) and 5.7cm (height) in size. To prevent tangling by nearby plants, I used half spaces of the pot, and four pots were placed in a zigzag configuration in one plastic tray. The size of the tray was 27cm x 59cm x 5cm. Many *Arabidopsis* accessions require to treat vernalization for their flowering. Thus, all plants were treated with vernalization for five weeks to ensure similar flowering. Afterward,

I returned the plants to the growth room, and I grew them until their meristems were terminated. When the plants were bolting, plant supports were used to prevent falls or crookedness. Each plant was incubated in the same type of soil (Sungro Sunshine® #5 Natural & Organic) to ensure identical nutrition, with an identical day/night cycle (18 hours light /6 hours dark) and a temperature of 22−23℃. For the light condition, six lamps were positioned 62 cm above the plants, and daylight color lamps were used. The intensity of the light from these lamps was 82.6 lm/W.

Growth conditions of the T−DNA insertion mutants were the same as the descriptions of the growth conditions of the 178 *Arabidopsis* accessions. However, there were some differences. Seeds of the T−DNA insertion mutants were sterilized in ethanol and sowed in Murashige and Skoog (MS−0) media plates, moved for five days to the germination chamber, and then brought to the growth room for one week for germination. Afterward, I transferred the plants to pots as described for the growth conditions of 178 *Arabidopsis* accessions. However, I used all six spaces of the pot while the 178 *Arabidopsis* accessions used three spaces of the pot. Also, the T−DNA insertion mutants did not need vernalization because their base ecotype was Col−0 that the ecotype could occur bolting without vernalization. After I transferred the mutants, the

process was the same as described for the growth conditions of the 178 *Arabidopsis* accessions.

Lepidium virginicum growth conditions were the same as the descriptions of the growth conditions of the 178 *Arabidopsis* accessions. However, there were two differences. The first difference was that they used different pots which were square shape and 7.5cm x 7.5cm x 7.7cm size. The second difference was that vernalization was conducted for three weeks while 178 *Arabidopsis* accessions was conducted for five weeks.

2.2 Seed number phenotyping

A statistical method was used to count the average seed number per silique of 178 *Arabidopsis thaliana* accessions and T-DNA insertion mutant plants. When siliques were matured and the meristem of the main stem was terminated, three plants of each accession and mutant line were chosen for counting seed number. Buds of the main stem and lateral branches had different concentrations of plant hormones such as IAA, CK, and abscisic acid (ABA) during development, which causes weak or fail development of lateral buds when compared to buds of the main stem (Emery *et al.*, 1998). For this reason, the first five siliques of

the *Arabidopsis thaliana* main inflorescence axis were not counted because of incomplete development and for minimizing age-related variation (Yuan & Kessler, 2019; Jiang, 2020). To avoid this issue, the seed number of each ecotype's main stem 6^{th} to 15^{th} siliques was measured by image analysis. Damaged siliques such as those with opened valves and no valves were removed during the measurement to ensure high accuracy. Then, I placed matured silique on A4 paper and spread the seeds. I photographed the seeds using a digital camera (SONY α6400) and ImageJ software to analyze the images and to count the number of seeds.

The number of unfertilized ovule and the seed number per silique of the ecotypes with the lowest and highest number of seeds per silique were measured by dissection with the aid of microscopes and forceps. Five plants of each ecotype were selected for measuring the phenotype. The first five silique of the main inflorescence axis were not used and the 9^{th} to 11^{th} siliques were used instead to avoid the issue as described above.

2.3 Genome-Wide association study

In this study, single nucleotide polymorphisms (SNPs) of the 131 *Arabidopsis thaliana* accessions that were available to use on the

1001 Genomes Project website dataset which TAIR 10 were applied. The genome-wide association study (GWAS) was performed by genome-wide efficient mixed-model association (GEMMA) which was a software toolkit used R and Python for linear mixed model (LMM), GWAS and other large scales data sets (Zhou & Stephens, 2012). LMM was used to generate a Manhattan plot and a Q-Q plot by using GEMMA. SNPs with p-values $\leq$ 5.86 X 10^{-6} were considered as candidate loci that regulate the seed number per silique.

2.4 Identification of T-DNA insertion mutants

Each T-DNA insertion had its own resistance of an antibody. SALK and SAIL had kanamycin resistance. T-DNA insertion was checked by primers consisting of T-DNA insertion sequences and *A. thaliana* sequences. When using primers, these mutant plants showed a different length PCR product depending on the existence of the T-DNA insertion.

2.5 Identification of *Lepidium* samples species

Sequences of whole internal transcribed spacer (ITS) parts were

used to identify the *Lepidium* samples species. To figure out the sequences of whole ITS parts of *Lepidium* samples, F-primer N-nc18S10 and R-primer C26A that were known as primers of ITS were used (Table 2).

2.6 Phylogenetic analysis

The phylogenic tree in which the nucleotide distance and neighbor-join tree file were calculated in this study was generated on Mega 11. The sequences of *Lepidium* ITS data of NCBI were used. I aligned the data by the ClustalW option of the ITS sequence of *Lepidium* samples and NCBI. The phylogenic tree of the *Lepidium* sample was drawn using the neighbor-joining method in Mega 11 and a bootstrap method phylogeny test with 1000 replications was applied.

2.7 Floral dipping transformation

To know auxin accumulation patterns during ovule development of *Lepidium virginicum*, auxin reporter gene DR5 with fluorescent protein ntdTOMATO and GUS were injected to agrobacterium strain GV3101 (Table 3).

Incubated agrobacterium GV3101 with DR5::GUS and DR5::ntdTOMATO vector, chose a colony, placed the colony into a 5ml lysogeny broth (LB) with kanamycin, and seed culture for 48 h at 28℃. Then, put 2ml of seed culture result to 200ml LB with kanamycin and rifampicin which had function to filter agrobacterium did not have the vector and overnight at 28℃. Afterward, centrifuged the agrobacterium, and combined the agrobacterium with transformation solution which contained 0.22g MS / 100ml, 5g sucrose / 100ml, 0.044, 1 ul benzyl amino purine (BAP) / 100ml and Silwet L−77 0.5 ul / 100ml. Then, soaked the meristem in solution for 30s, sealed the plants with aluminum foil to inhibit light, and left the meristem overnight in the growth room.

Table 1. List of 131 *A. thaliana* ecotype.

Number	Name	Abbrev name	Accession ID	Country
1	Graz	Gr-1	430	Austria
2	Karakol	Kar-1	763	Kyrgyzstan
3	Susamyr river	Sus-1	765	Kyrgyzstan
4	Djarly	Dja-1	766	Kyrgyzstan
5	Tchong-Kemin Valley	Zal-1	768	Kyrgyzstan
6	Jawshangoz Village	Neo-6	772	Tajikistan
7	Borky	Bor-1	5837	Czech Republic
8	Antwerpen	An-1	6898	Belgium
9	Brunn	Br-0	6904	Czech Republic
10	Col-0	Col-0	6909	USA
11	Eifel	Ei-2	6915	Germany
12	Gabelstein	Ga-0	6919	Germany
13	Gückingen	Gu-0	6922	Germany
14	Kindalville	Kin-0	6926	USA
15	Kondara	Kondara	6929	Tajikistan
16	Moscow	Ms-0	6938	Russia
17	Merzhausen	Mz-0	6940	Germany
18	Noordwijk	Nok-3	6945	Netherlands
19	Prudka (CS76579)	Pu2-23	6951	Czech Republic
20	Randan	Ra-0	6958	France
21	San Eleno	Se-0	6961	Spain
22	Tammisari	Tamm-2	6968	Finland
23	Tossa del Mar	Ts-1	6970	Spain
24	Ottenhof	Uod-1	6975	Austria
25	Weiningen	Wei-0	6979	Switzerland

26	Wassilewskija	Ws-2	6981	Russia
27	Wietze	Wt-5	6982	Germany
28	Zdarec	Zdr-1	6984	Czech Republic
29	Aberdeen	Abd-0	6986	United Kingdom
30	Achkarren	Ak-1	6987	Germany
31	Alston	Alst-1	6989	United Kingdom
32	Ameland	Amel-1	6990	Netherlands
33	Angleur	Ang-0	6992	Belgium
34	Appeltern	Appt-1	6997	Netherlands
35	Aua	Aa-0	7000	Germany
36	Baarlo	Baa-1	7002	Netherlands
37	Bennekom	Benk-1	7008	Netherlands
38	Berlin	Bd-0	7013	Germany
39	Blackmount	Ba-1	7014	United Kingdom
40	Bologna	Bl-1	7025	Italy
41	Boot, Eskdale	Boot-1	7026	United Kingdom
42	Büchen	Bch-1	7028	Germany
43	Buchschlag	Bsch-0	7031	Germany
44	Burghaun	Bu-0	7036	Germany
45	Calver	Cal-0	7061	United Kingdom
46	Camberg	Ca-0	7062	Germany
47	Cerveteri	Cerv-1	7068	Italy
48	Chateaudun	Chat-1	7071	France
49	Chisdra	Chi-0	7072	Russia
50	Coimbra	Co-1	7077	Portugal
51	Compiegne	Com-1	7092	France

52	Dijon	Di-G	7096	France
53	Donsbach	Do-0	7102	Germany
54	Drahonin	Dra-0	7103	Czech Republic
55	Dresden	Dr-0	7106	Germany
56	Durham	Durh-1	7107	United Kingdom
57	East Malling	Ema-1	7109	United Kingdom
58	Ellershausen	El-0	7117	Germany
59	Enkheim	En-2	7119	Germany
60	Erlangen	Er-0	7125	Germany
61	Espoo	Es-0	7126	Finland
62	Estland	Est	7127	Germany
63	Frankfurt	Fr-2	7133	Germany
64	Frickhofen	Fi-0	7138	Germany
65	Geleen	Gel-1	7143	Netherlands
66	Gieben	Gie-0	7147	Germany
67	Gudow	Gd-1	7161	Germany
68	Hannover	Ha-0	7163	Germany
69	Hennetalsperre	Hn-0	7165	Germany
70	Jamolice	Jm-0	7177	Czech Republic
71	Jena	Je-0	7181	Germany
72	Killean	Kil-0	7192	United Kingdom
73	Koeln	Kl-5	7199	Germany
74	Kronberg	Kb-0	7202	Germany
75	Krottensee	Krot-0	7203	Germany
76	Kyoto	Kyoto	7207	Japan
77	Lanark	Lan-0	7208	United Kingdom

78	Landsberg	La-0	7209	Germany
79	Le Mans	Lm-2	7217	France
80	Leiden	Le-0	7218	Netherlands
81	Limburg	Li-2:1	7223	Germany
82	Litva	Litva	7236	Lithuania
83	Mechthausen	Me-0	7250	Germany
84	Mühlen	Mh-0	7255	Poland
85	Neuweilnau	Nw-0	7258	Germany
86	Nieps	Np-0	7268	Germany
87	Oberursel	Ob-0	7276	Germany
88	Ovelgoenne	Ove-0	7287	Germany
89	Pitztal	Pi-0	7298	Austria
90	Point Grey	Pog-0	7306	Canada
91	Ravensglass	Ragl-1	7314	United Kingdom
92	Rome	Rome-1	7319	Italy
93	Rouen	Rou-0	7320	France
94	Rubezhnoe	Rubezhnoe-1	7323	Ukraine
95	Seattle	Seattle-0	7332	USA
96	Seis am Schlern	Sei-0	7333	Italy
97	Siegen	Si-0	7337	Germany
98	Southport	Su-0	7342	United Kingdom
99	Berlin / Spandau	Sp-0	7343	Germany
100	St.Georgen	Sg-1	7344	Germany
101	Stobowa	Stw-0	7347	Russia
102	Tabor	Ta-0	7349	Czech Republic
103	Tacoma	Tac-0	7350	USA
104	Toledo	Tol-0	7356	USA
105	Tsagguns	Tscha-1	7372	Austria
106	Umkirch	Uk-1	7378	Germany

107	Vancouver	Van-0	7383	Canada
108	Vindolanda	Vind-1	7387	United Kingdom
109	Warschau	Wa-1	7394	Poland
110	Westercelle	Wc-1	7404	Germany
111	Wildbad	Wl-0	7411	Germany
112	Vranov u Brna	Jl-3	7424	Czech Republic
113	Ville-en-Vermois	Nc-1	7430	France
114	Da(1)	Da(1)-12	7460	Czech Republic
115	CS76588_Rld	RLD-1	7471	Unknown
116	North Liberty	RRS-7	7514	USA
117	Lipovec	Lp2-2	7520	Czech Republic
118	Blanes	Bla-1	8264	Spain
119	Isenburg	Is-0	8312	Germany
120	Miramare	Mir-0	8337	Italy
121	Nantes	Na-1	8343	France
122	Perm	Per-1	8354	Russia
123	Playa de Aro	Pla-0	8357	Spain
124	Rodenbach	Rd-0	8366	Germany
125	Kelsterbach	Kelsterbach-4	8420	Germany
126	Kashmir	Kas-2	8424	India
127	Altai	Altai-5	9758	China
128	Bandar	Anz-0	9759	Iran
129	Bikfaiya	Bik-1	9761	Lebanon
130	Qartaba	Qar-8a	9764	Lebanon
131	West Karakol	WestKar-4	9766	Kyrgyzstan

Table 2. List of candidate SNPs and genes from GWAS result.

CHR	Position	minor allele	major allele	P-value	AT Number	Location	Gene name	Description
1	26035362	G	A	$8.17*10^{-09}$	AT1G69250	26,032,637.. 26,035,515		Nuclear transport factor 2 (NTF2) family protein with RNA binding (RRM-RBD-RNP motifs) domain-containing protein;(source:Araport11)
4	12070420	G	T	$1.77*10^{-07}$	AT4G23030	12,072,760.. 12,074,580		MATE efflux family protein
3	19737008	G	C	$1.92*10^{-07}$	AT3G53240	19,733,766.. 19,739,232	*RLP45*	receptor like protein 45;(source:Araport11)
1	28333180	C	T	$3.38*10^{-07}$	AT1G75470	28,330,054.. 28,331,300	*PUP15*	purine permease;(source:Araport11),Member of a family of proteins related to PUP1, a purine transporter. May be involved in the transport of purine and purine derivatives such as cytokinins, across the plasma membrane.
4	18077780	A	G	$5.70*10^{-07}$	AT4G38710	18,077,790.. 18,080,228		glycine-rich protein;(source:Araport11)
1	22871393	A	T	$7.15*10^{-07}$	AT1G61890	22,867,689.. 22,871,358		MATE efflux family protein;(source:Araport11)
1	26040242	A	C	$1.11*10^{-06}$	AT1G69260	26,038,905.. 26,040,720	*AFP1*	ABI five binding protein;(source:Araport11)
1	6659053	A	G	$1.14*10^{-06}$	AT1G19260	6,657,260.. 6,659,569	*LOH3*	Encodes a ceramide synthase that uses very-long- chain fatty acyl-CoA and trihydroxy LCB substrates.
2	6923139	C	A	$1.88*10^{-06}$	AT2G15890	6,920,369.. 6,922,161	*MEE14*	maternal effect embryo arrest 14;(source:Araport11),Encodes CBP1, a regulator of transcription initiation in central cell-mediated pollen tube guidance.

2	6949579	T	C	$1.88*10^{-06}$	AT2G15970	6,949,851.. 6,951,192	*COR41 3-PM1*	cold regulated 413 plasma membrane 1;(source:Araport11), encodes an alpha form of a protein similar to the cold acclimation protein WCOR413 in wheat. Expression is induced by short-term cold-treatment, water deprivation, and abscisic acid treatment. The mRNA is cell-to-cell mobile.

Table 3. Primers to identify *Lepidium* sample species.

Primer Name	Primer Sequence	Target	Primer Direction
N-nc18S10	5`-AGGAGAAGTCGTAACAAG-3`	ITS	Front
C26A	5`-GTTTCTTTTCCTCCGCT-3`	ITS	Reverse

Table 4. Plasmids of DR5 to identify *Lepidium* auxin accumulation patterns.

Plasmid name	Fluorescent protein	Bacteria resistance	Plant resistance
pC3300:pDR5	GUS	Kanamycin	BASTA
pC3300:pDR5	ntdTOMATO	Kanamycin	BASTA

3. Results

3.1 Genetic factors determining the seed number per silique in *A. thaliana*

3.1.1 Variation in seed numbers per silique among 131 *A. thaliana* ecotypes

The seed number is an important determinant of the crop yield. However, the mechanism of determining the seed number per silique (or fruit) remains unclear. To address this issue, the variation in the number of seeds per silique among *A. thaliana* ecotypes was examined (Figure 1). There were 178 accessions of *A. thaliana* at the beginning, but 47 accessions of A. thaliana were removed because of physical damages, diseases, and late flowering, so 131 accessions were left (Table 1). Under identical environmental conditions, the 131 accessions of *A. thaliana* examined here showed considerable variety in the number of seeds per silique, suggesting that there are genes that regulate the seed number per silique.

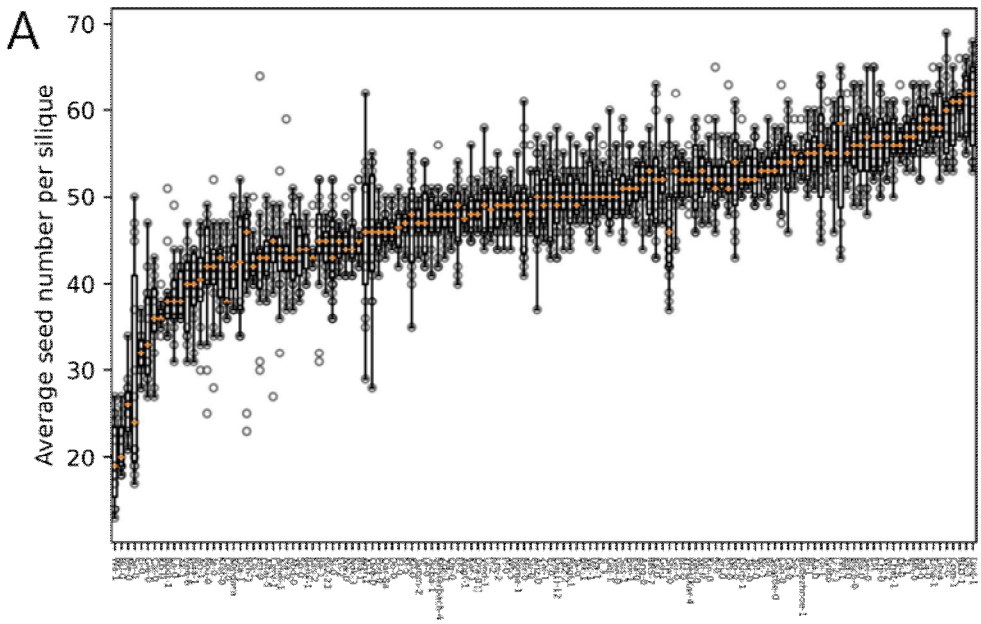
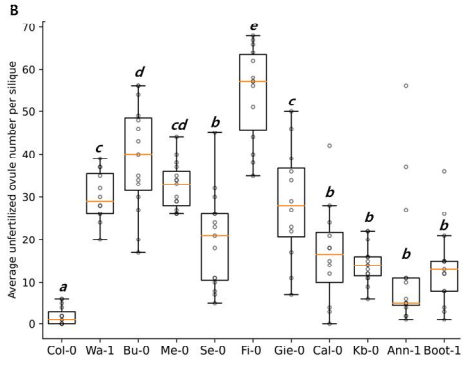
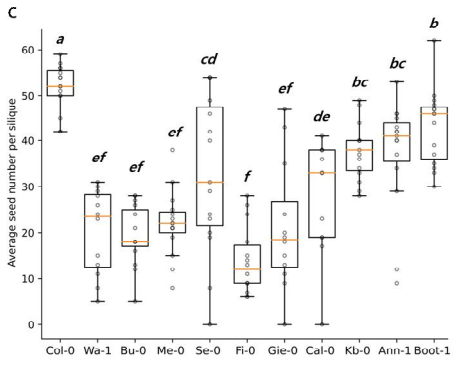


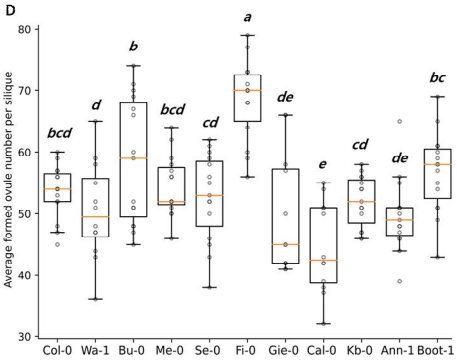
Figure 1. *Arabidopsis* accessions display natural variation of the seed number per silique. Boxplot of the seed number per silique among 131 accessions of *A. thaliana*. Orange color dots indicate median value. Bars show maximum and minimum observation in the fence.

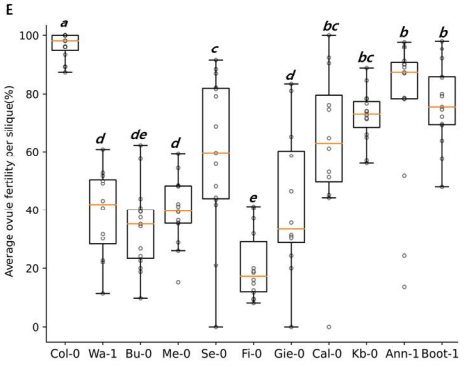
3.1.2 Ten ecotypes with the lowest number of seeds per silique have low fertility

Some ecotypes, especially the lowest ten ecotypes, showed a low number of seeds per silique in compared to the others. To analyze this result, the siliques of the lowest ten ecotypes were dissected (Figure 2A). The box graph of lowest ten ecotypes phenotype indicated average unfertilized ovule number per silique, average seed number per silique, average formed ovule number per silique and ovule fertility per silique (Figure 2B,2C,2D and 2E). Graph of average unfertilized ovule number per silique, average seed number per silique and average ovule fertility per silique showed that the lowest ten ecotypes had significantly different value than control ecotype Col-0 also, except two ecotype Frickhofen (Fi-0) and Calver (Cal-0), the lowest ecotype had non-significant different average formed ovule number per silique than control ecotype Col-0 (Figure 2F). These results indicated that the lowest ten ecotypes had similar formed ovule number per silique with control ecotype Col-0 except two ecotype Fi-0 and Cal-0, but had significantly low ovule fertility per silique compared to control ecotype Col-0 and this low ovule fertility caused low seed number per silique.









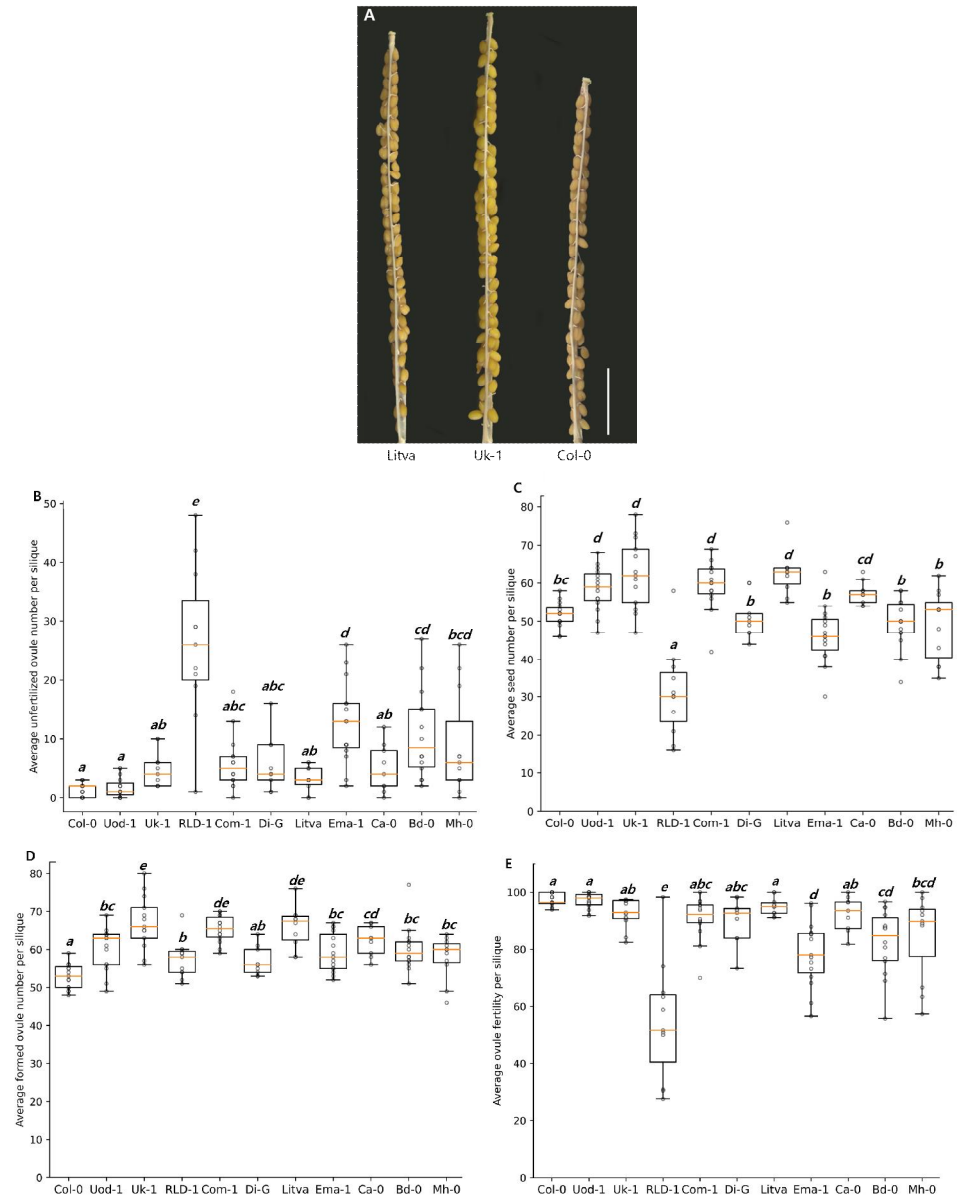
F

Ecotype	Average unfertilized ovule number	One-way ANOVA p-value	Duncan Groups	T-test P-value	Average seed number	One-way ANOVA p-value	Duncan Groups	T-test P-value	Average formed ovule number	One-way ANOVA p-value	Duncan Groups	T-test P-value	Average ovule fertility(%)	One-way ANOVA p-value	Duncan Groups	T-test P-value
Col-0	1.9333		a	-	51.66667		a	-	53.6		a	-	96.308		a	-
Ann-1	12.467		b	P<0.05(*)	37.26667		bc	P<0.001(***)	49.73333		de	P>0.05	76.72933		b	P<0.01(**)
Boot-1	13.6		b	P<0.001(***)	43.13333		b	P<0.01(**)	56.73333		bc	P>0.05	76.44467		b	P<0.001(***)
Kb-0	14.2		b	P<0.001(***)	37.8		bc	P<0.001(***)	52		cd	P>0.05	72.51067		bc	P<0.001(***)
Cal-0	16.583		b	P<0.001(***)	27.91667		de	P<0.001(***)	44.5		e	P<0.01(**)	62.74417		bc	P<0.001(***)
Se-0	19.8	$5.32*10^{-31}$	c	P<0.001(***)	32.86667	$5.75*10^{-23}$	cd	P<0.001(***)	52.66667	$2.52*10^{-13}$	cd	P>0.05	59.54067	$3.51*10^{-23}$	c	P<0.001(***)
Gie-0	28.417		c	P<0.001(***)	21.16667		ef	P<0.001(***)	49.58333		de	P>0.05	42.28083		d	P<0.001(***)
Wa-1	30.167		c	P<0.001(***)	20.25		ef	P<0.001(***)	50.41667		d	P>0.05	38.88083		d	P<0.001(***)
Me-0	32.533		cd	P<0.001(***)	54.46667		bcd	P>0.05	54.46667		bcd	P>0.05	39.87467		d	P<0.001(***)
Bu-0	39.2		d	P<0.001(***)	19.33333		f	P<0.001(***)	56.53333		b	P>0.05	34.202		de	P<0.001(***)
Fi-0	54.5		e	P<0.001(***)	13.92857		f	P<0.001(***)	68.42857		a	P<0.001(***)	20.61714		e	P<0.001(***)

Figure 2. Ten ecotypes with the lowest number of seeds per silique show low ovule fertility. (A) Dissection images of Fi-0 (Frickhofen), Bu-0 (Burghaun) and Col-0 in lowest ten ecotypes. Scale bar = 1.0mm (B-E) Box graph of lowest ten ecotype with group from Duncan test. Each box graph shows phenotype (B) Average unfertilized ovule number per silique. (C) Average seed number per silique. (D) Average formed ovule number per silique. (E) Average ovule fertility per silique. (F) Table about Col-0 and lowest ten ecotype name, phenotype value, group from Duncan test and p-value of t-test.

3.1.3 Ten ecotypes with the highest number of seeds per silique have many formed ovule number

Some ecotypes, especially the highest ten ecotypes, showed a high number of seeds per silique in compared to the others. To analyze this result, the siliques of the highest ten ecotypes were dissected (Figure 3A). The box graph of highest ten ecotypes phenotype indicated average unfertilized ovule number per silique, average seed number per silique, average formed ovule number per silique and average ovule fertility per silique (Figure 3B,3C,3D and 3E). Graph of average formed ovule number per silique showed that the highest ten ecotypes had significantly different value than control ecotype Col-0. Furthermore, the highest ten ecotypes excepted RLD-1, Di-G (Dijon), Ema-1 (East Malling), Bd-0 (Berlin), and Mh-0 (Muehlen) indicated significant high seed number value than Col-0, and their ovule fertility was high such as more than 90% fertility (Figure 3F). These results indicated that large number of formed ovules per silique and high ovule fertility caused many seed number per silique of highest ten ecotypes.



F

Ecotype	AUON	One-way ANOVA p-value	Duncan Groups	T-test P-value	ASN	One-way ANOVA p-value	Duncan Groups	T-test P-value	AFON	One-way ANOVA p-value	Duncan Groups	T-test P-value	AOF(%)	One-way ANOVA p-value	Duncan Groups	T-test P-value
Col-0	1.4		a	-	51.53333		bc	-	52.93333		a	-	97.35718		a	-
Uod-1	1.8		a	P>0.05	58.46667		d	P<0.05(*)	60.26667		bc	P<0.001(***)	97.01492		a	P>0.05
Uk-1	4.615385		ab	P<0.01(**)	62.38462		d	P<0.01(**)	67		e	P<0.001(***)	92.77167		ab	P<0.01(**)
RLD-1	26.27273		e	P<0.001(***)	31.09091		a	P<0.001(***)	37.36364		b	P<0.05(*)	54.61773		e	P<0.001(***)
Com-1	6		cde	P<0.01(**)	59.35714		d	P<0.001(***)	65.35714		de	P<0.001(***)	90.70639		abc	P<0.01(**)
Di-G	5.777778	$1.62*10^{-17}$	abc	P<0.05(*)	51.11111	$2.02*10^{-26}$	b	P>0.05	56.88889	$4.76*10^{-31}$	ab	P<0.05(*)	89.91355	$9.73*10^{-20}$	abc	P<0.05(*)
Litva	3.3		ab	P<0.05(*)	62.6		d	P<0.001(***)	65.9		de	P<0.001(***)	94.99445		a	P<0.05(*)
Ema-1	12.66667		d	P<0.001(***)	46.46667		b	P<0.05(*)	59.13333		bc	P<0.001(***)	78.64657		d	P<0.001(***)
Ca-0	4.888889		ab	P<0.05(*)	57.44444		cd	P<0.001(***)	62.33333		cd	P<0.001(***)	92.39941		ab	P<0.05(*)
Bd-0	10.85714		cd	P<0.001(***)	49.28571		b	P>0.05	60.14286		bc	P<0.001(***)	92.38504		cd	P<0.001(***)
Mh-0	9		bcd	P<0.05(*)	48.90909		b	P>0.05	57.90909		bc	P<0.05(*)	84.7952		bcd	P<0.05(*)

Figure 3. Ten ecotypes with the highest number of seeds per silique show large number of formed ovules. (A) Dissection images of Litva, Uk-1 (Umkirch) and Col-0 in highest ten ecotypes. Scale bar = 1.0mm (B-E) Box graph of highest ten ecotype with group from Duncan test. Each box graph shows phenotype (B) Average unfertilized ovule number per silique. (C) Average seed number per silique. (D) Average formed ovule number per silique. (E) Average ovule fertility per silique. (F) Table about Col-0 and highest ten ecotype name, phenotype value, group from Duncan test and p-value of t-test.

3.1.4 GWAS identifies candidate loci in seed number per silique among 131 *A. thaliana* ecotypes

To find genes had high association with seed number per silique, I did GWAS. The manhattans plot of GWAS showed 107 SNPs showed high association with seed number per silique (Figure 4A). Based on p-value, 10 SNPs with the lowest p-value were selected for candidates (Table 2). Among candidate SNPs, 1:22871393, 1:28333180, 4:1270420, and 4:18077780 showed significant seed number per silique difference compared to the Col-0 control(Figure 6E). The candidate genes were located near candidate SNPs or included SNPs. AT1G61890 was the candidate genes of SNP 1:22871393, *PUP15* (AT1G75470) was the candidate genes of SNP 1:2833318, AT4G23030 was the candidate genes of SNP 4:12070420, and AT4G38710 was the candidate genes of SNP 4:18077780 was (Figure 4B,C,D,E). Among four candidate genes, both AT1G61890 and AT4G23030 were MATE efflux protein (Figure 4A). The QQ-plot indicated that GWAS results rejected the null hypothesis that there were no relationships between ecotypes and seed number per silique (Figure 5). These GWAS results indicated candidate loci in seed number per silique among 131 *A. thaliana* ecotypes.

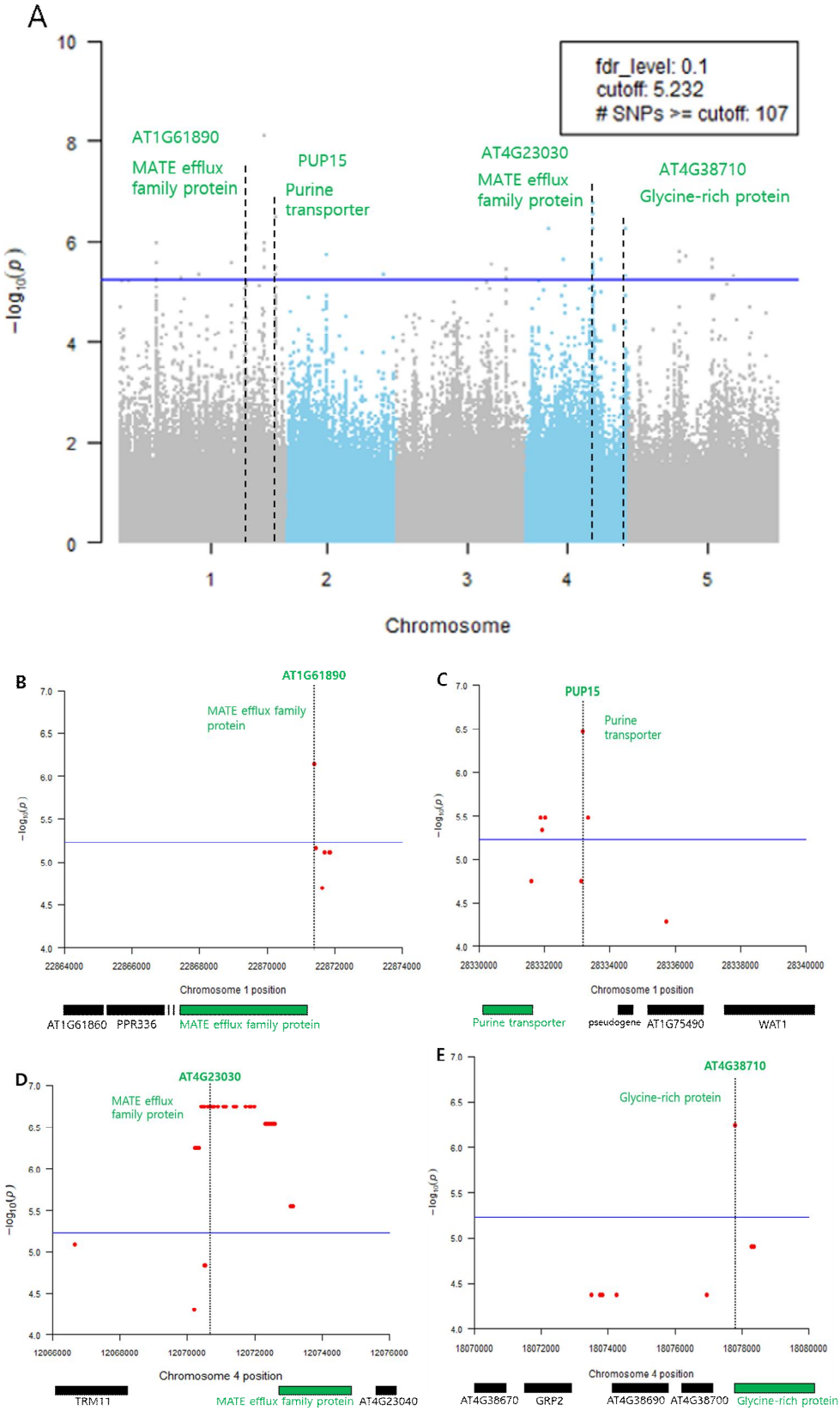

A
fdr_level: 0.1
cutoff: 5.232
SNPs >= cutoff: 107
AT1G61890
MATE efflux family protein
PUP15
Purine transporter
AT4G23030
MATE efflux family protein
AT4G38710
Glycine-rich protein
−log10(p)
Chromosome
B
AT1G61890
MATE efflux family protein
Chromosome 1 position
AT1G61860
PPR336
MATE efflux family protein
C
PUP15
Purine transporter
Chromosome 1 position
Purine transporter
pseudogene
AT1G75490
WAT1
D
AT4G23030
MATE efflux family protein
Chromosome 4 position
TRM11
MATE efflux family protein
AT4G23040
E
AT4G38710
Glycine-rich protein
Chromosome 4 position
AT4G38670
GRP2
AT4G38690
AT4G38700
Glycine-rich protein

Figure 4. GWAS of variation in seed number per silique among 131 *A. thaliana* ecotypes. (A) manhattans plot of GWAS. X-axis indicated the location of A. thaliana chromosomes and y-axis showed $-\log_{10}$ (p values). Each dot showed SNPs and the horizontal blue dashed line corresponds to a $-\log_{10}$ (p values) ≥ 5.232. (B) The genomic region of significant GWAS peaks of 1:22871393(C) 1:28333180, (D) 4:12070420, (E) 4:18077780. (G) Table of the 10 SNPs that had lowest p-value indicated SNP position, minor and major allele, p-value, nearest gene, gene location, gene name and gene description.

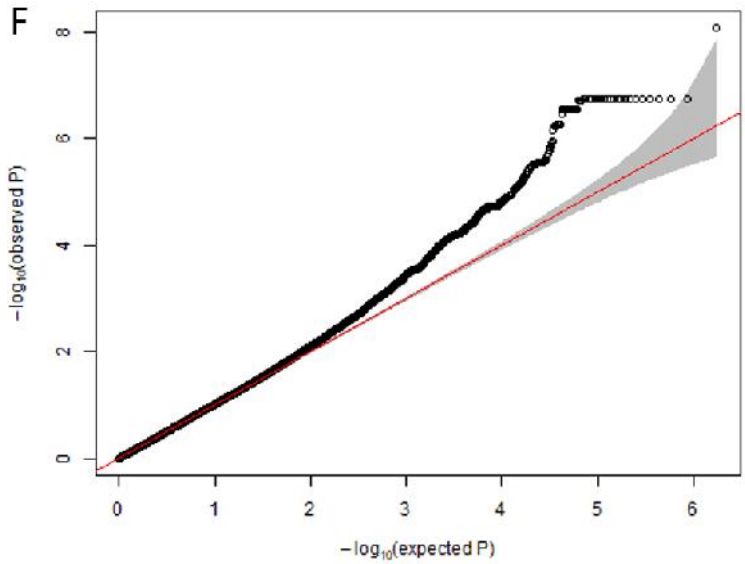
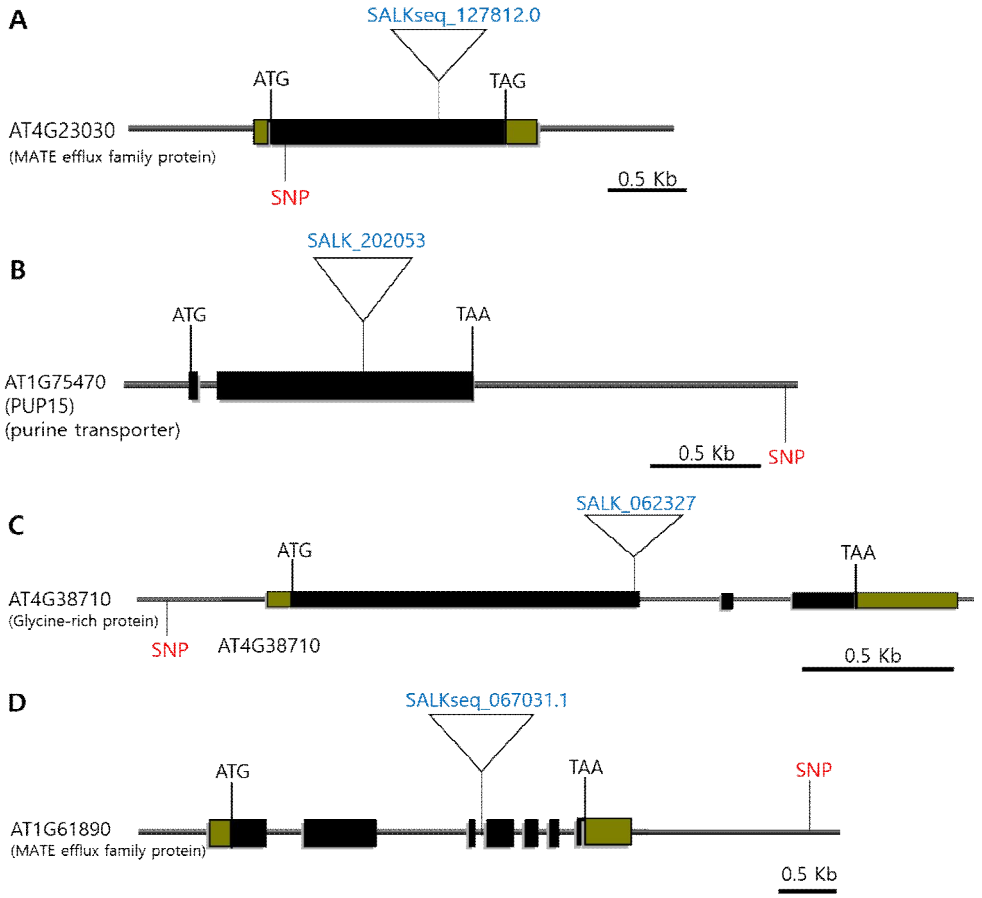


Figure 5. QQ-plot of GWAS result. The x-axis and y-axis indicated $-\log_{10}$ (expected p values) and $-\log_{10}$ (observed p values). Red line indicated null hypothesis that there were no relationships between ecotypes and seed number per silique.

3.1.5 Four genes that positive and negative regulator of seed number per silique in *A. thaliana*

To verify candidate genes function about seed number per silique, their T-DNA insertion mutants' phenotype were measured. One of candidate gene AT4G23030 had SALK_127812.0 T-DNA insertion in exon for mutant(Figure 6A). The other candidate genes *PUP15*, AT4G38710, and AT1G61890 each had SALK_202053, SALK_062327 and SALKseq_067031.1 T-DNA insertion for mutant (Figure 6B, C, D). AT4G23030 was a gene include candidate SNP while *PUP15*, AT5G38710, and AT1G61890 were genes near candidate SNP. Among candidate loci of GWAS result, T-DNA insertion mutants of four genes AT4G23030, AT1G7540 (*PUP15*), AT4G38710 and AT1G61890 indicated significant seed number per silique difference compared wild type background Col-0 phenotyping (Figure 6E). T-DAN insertion mutants of *PUP15* and AT4G38710 each indicated high and low seed number per silique difference compared to control ecotype Col-0. AT1G61890 and AT4G2303 that both MATE efflux protein indicated high significant seed number per silique difference, but the number of phenotyping AT1G61890 was one (Figure 6E). These results indicated that gene AT4G23030, *PUP15*, and AT1G61890 could be positive

regulator of seed number per silique and AT4G38710 could be negative regulator of seed number per silique.


A
SALKseq_127812.0
ATG
TAG
AT4G23030
(MATE efflux family protein)
SNP
0.5 Kb
B
SALK_202053
ATG
TAA
AT1G75470
(PUP15)
(purine transporter)
0.5 Kb
SNP
C
SALK_062327
ATG
TAA
AT4G38710
(Glycine-rich protein)
SNP
AT4G38710
0.5 Kb
D
SALKseq_067031.1
ATG
TAA
SNP
AT1G61890
(MATE efflux family protein)
0.5 Kb

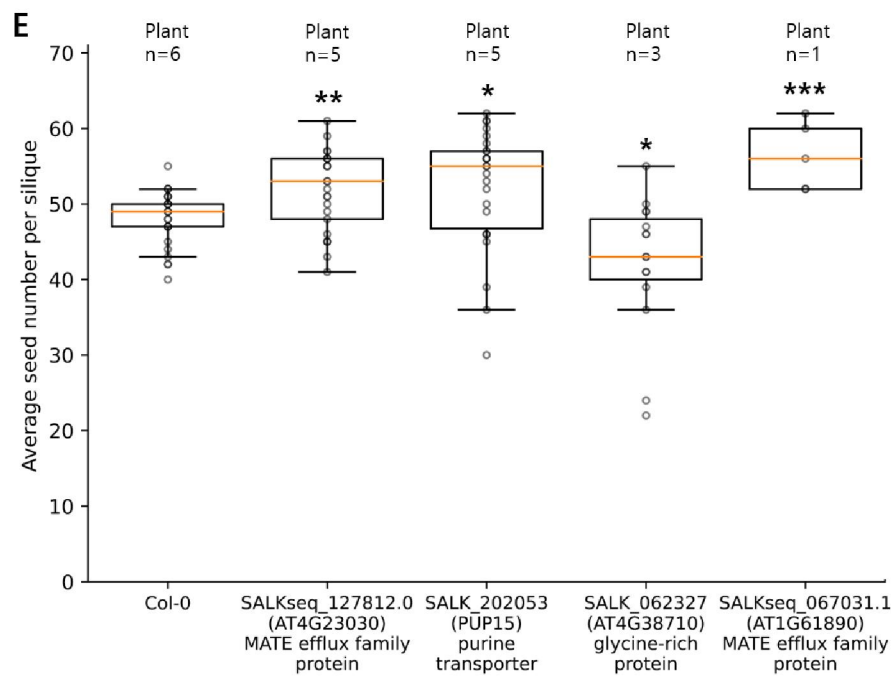

E
Plant
n=6
Plant
n=5
Plant
n=5
Plant
n=3
Plant
n=1
**
*
*
***
Average seed number per silique
70
60
50
40
30
20
10
0
Col-0
SALKseq_127812.0
(AT4G23030)
MATE efflux family
protein
SALK_202053
(PUP15)
purine
transporter
SALK_062327
(AT4G38710)
glycine-rich
protein
SALKseq_067031.1
(AT1G61890)
MATE efflux family
protein

Figure 6. AT4G23030, *PUP15*, AT4G38710, and AT1G61890 affect seed number per silique. (A) T-DNA insertion and SNP location of AT4G23030, (B) *PUP15*, and (C) AT4G38710. (E) Bar graph of average seed number per silique in Col-0, Salk_127812, Salk_202053, and Salk_062327. $n>3$. $p<0.05$ (*), $p<0.01$ (**). $n \geq 3$.

3.1.6 Ecotype Di-G had specific genes to increase seed number per silique.

Phenotyping of *A. thaliana* 131 ecotypes and dissection indicated that ten ecotypes with the highest number of seeds per silique had many formed ovule numbers. Among highest ten ecotypes, Di-G was the ecotype that known to have not only large seed number but also ovule number (Figure 7A). The average seed number per silique data was from my experiment data and the average ovule number per silique was from others data (Yuan & Kessler, 2019). Furthermore, Di-G showed larger seed number per silique than Col-0 that was control ecotype (Figure 7B). These two results indicated that there were genes in Di-G that made Di-G have more ovule and seed number than other ecotype such as control ecotype Col-0. To find the formed ovule number regulator genes, I crossed Col-0 and Di-G. Di-G had own SNPs, so that SNPs could be markers to separate Col-0 and Di-G (Figure 7C). To figure out Nucleic acid sequence(NAS) for SNPs, did sequencing to forward and reverse. In forward direction, Col-0 X Di-G had T and G which had red and black color while Col-0 had only T which had red color and Di-G had G which had black color. In reverse direction, Col-0 x Di-G had A and C which had green and blue color while Col-0

had only A which had green color and Di-G had only C which had blue color (Figure 7C).

This information will be base to find positive seed number per silique regulator genes in Di-G. F1 phenotyping of Col-0 and Di-G crossing will indicate the dominant and recessive features of seed number per silique regulator genes. Furthermore, the ratio of F2 phenotyping of Col-0 and Di-G will indicate that seed number per silique were affected by a single gene or multi genes. After the lines will be used as quantitative trait locus (QTL) to find positive seed number per silique regulator genes in Di-G .

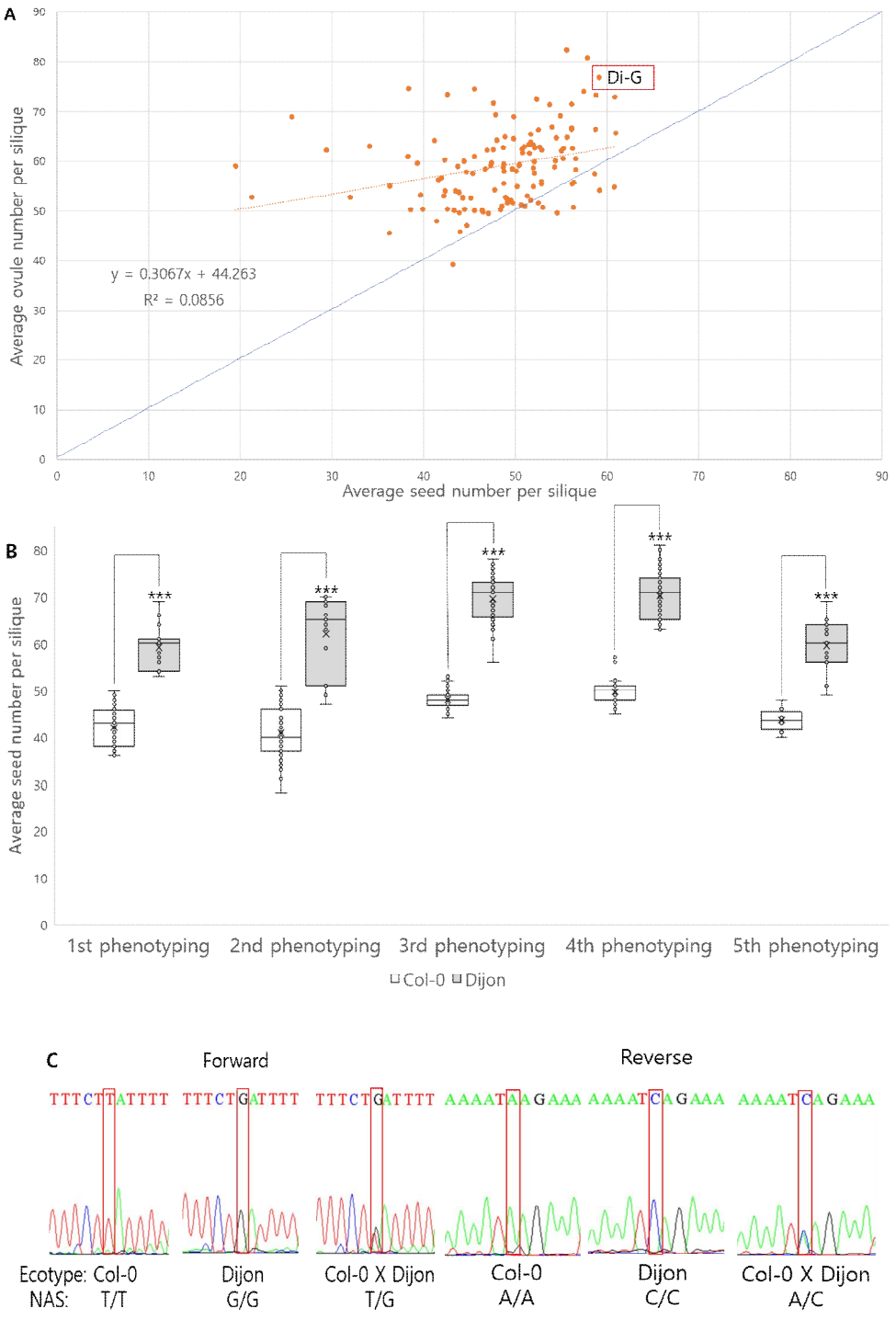

A
Average ovule number per silique
Average seed number per silique
Di-G
y = 0.3067x + 44.263
R² = 0.0856
B
Average seed number per silique
***
1st phenotyping
2nd phenotyping
3rd phenotyping
4th phenotyping
5th phenotyping
Col-0 Dijon
C
Forward
Reverse
TTTCTTATTTT
TTTCTGATTTT
TTTCTGATTTT
AAAATAAGAAA
AAAATCAGAAA
AAAATCAGAAA
Ecotype: Col-0
NAS: T/T
Dijon
G/G
Col-0 X Dijon
T/G
Col-0
A/A
Dijon
C/C
Col-0 X Dijon
A/C

Figure 7. Di-G had specific genes to increase seed number per silique. (A) Scatter graph of 131 accessions of *A. thaliana*. X-axis showed average seed number per silique of own data, Y-axis showed average ovule number per silique of others data. (B) Bar graph of Col-0 and Di-G average seed number per silique. X-axis indicated number of phenotyping (C) Nucleic acid sequence (NAS) and SNP of Col-0, Di-G, and Col-0 X Di-G.

3.2 Investigate the inter-species variation in seed number per silique between *Arabidopsis* and *Lepidium*

3.2.1 Genetic differences affected by evolution caused different seed number per silique between *Arabidopsis* and *Lepidium*

Lepidium showed interesting phenotype comparing *Arabidopsis thaliana*. their fruit shape and seed number per silique were different. *Lepidium* had round shape fruit, but *A. thaliana* had long stick shape fruit, also *Lepidium* had big size seeds but small number when *A. thaliana* had small size seeds but many number (Figure 8A,B). However, *Lepidium* and *A. thaliana* were close relationship because both were same linage of *Brassicaceae* even theses difference (Figure 8C). This result showed that there were genetic factors that caused these differences such as fruit shape, seed size and especially seed number per silique even their close relationship. The gene that regulate seed number per silique of *A. thaliana* also might affect *Lepidium*.


A
B

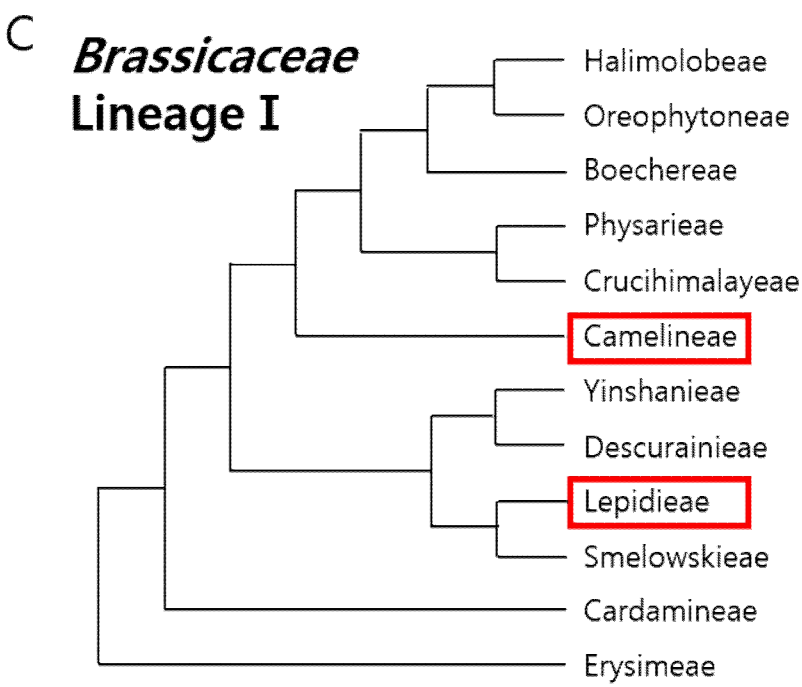

C
Brassicaceae
Lineage I
Halimolobeae
Oreophytoneae
Boechereae
Physarieae
Crucihimalayeae
Camelineae
Yinshanieae
Descurainieae
Lepidieae
Smelowskieae
Cardamineae
Erysimeae

Figure 8. Difference between *Lepidium* and *Arabidopsis thaliana*. (A) Silique images of *Lepidium* and *A. thaliana* silique. (B) Dissection image of *Lepidium* and *A. thaliana* silique. (A–B) Scale bar is 2 mm. (C) Phylogenetic tree of *Brassicaceae* Lineage I. *Arabidopsis* and *Capsella* are the geneus of *Camelineae* and *Lepidium* is the genus of *Lepidieae*.

3.2.2 Identify the *Lepidium* sample species using phylogenetic tree of *Lepidium* whole ITS parts

Lepidium samples that were collected in Seoul National University showed interesting phenotype comparing the *A. thaliana*. However, these *Lepidium* samples species were unknown, so identifying species of *Lepidium* sample was important for experiment detail. To identify *Lepidium* samples species, use entire ITS parts sequences that were highly conserved among plants such as *Lepidium*, *Arabidopsis*, and *Viola*. Used ITS F-primer N-nc18S10 of 18S and R-primer C26A of 25S, 26S and 28S for PCR (Table 2). After PCR of whole ITS parts of *Lepidium* sample, did sequencing the sample and made clustering of ITS parts of *Lepidium* species sequence included *Lepidium* sample sequence. Made phylogenetic tree of *Lepidium* species ITS parts sequence clustering with *Lepidium* sample (Figure 9). Red box showed what *Lepidium* specie was close and grouped with *Lepidium* sample. This result showed that *Lepidium* sample ITS sequence was matched with *Lepidium virginicum*, so *Lepidium* samples specie was *Lepidium virginicum*. Identifying this species could help find exact information about experiments and data.

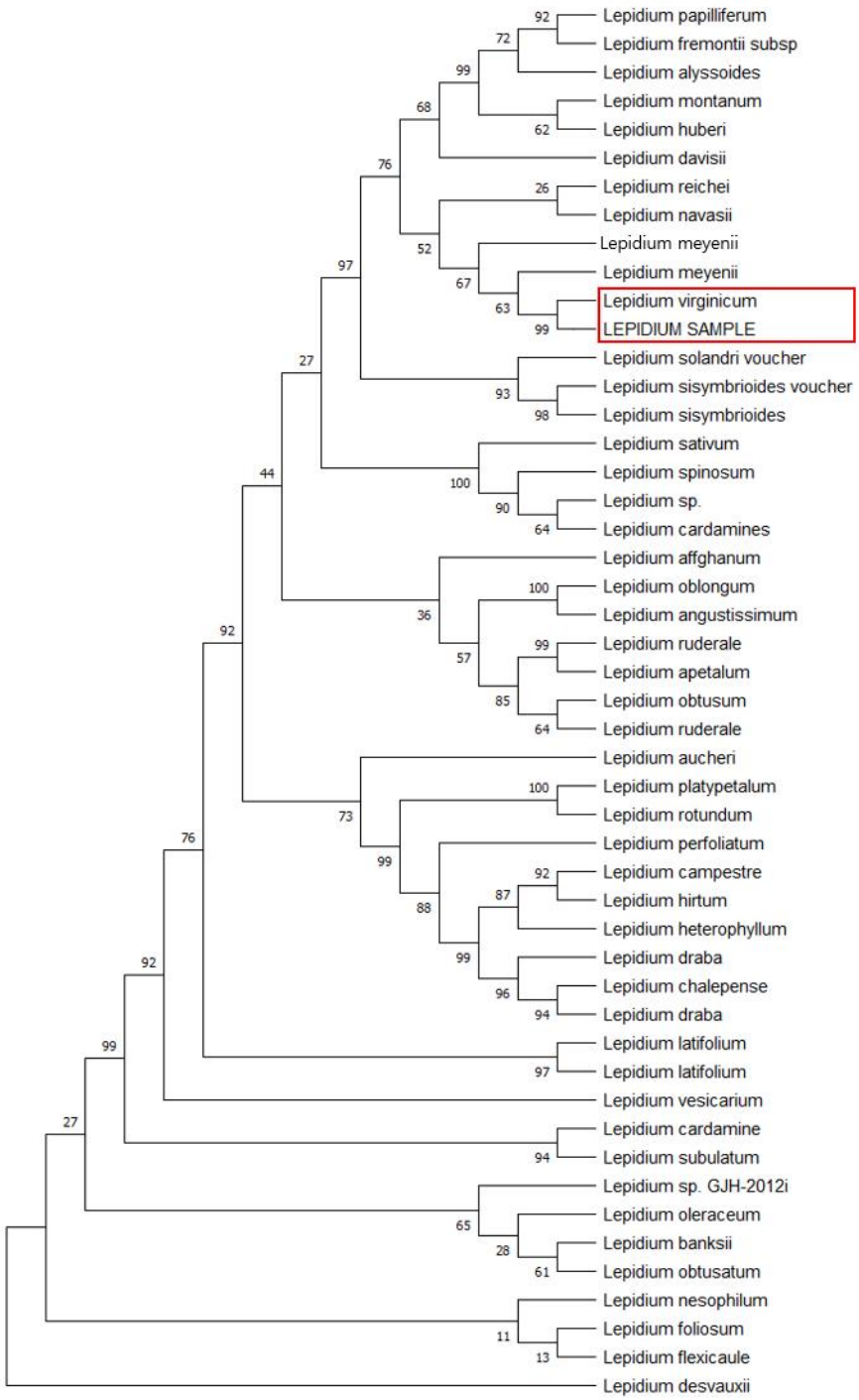


Figure 9. Phylogenetic tree of *Lepidium* species ITS sequences with *Lepidium* sample. Red box showed the closest group of a *Lepidium* specie and the *Lepidium* sample among *Lepidium* species.

3.2.3 Characterize the patterns of auxin accumulation during ovule formation

Before identify the genes AT4G23030, *PUP15*, and AT4G38710 that regulate seed number per silique in *A. thaliana* expression pattern in *L. virginicum*, it needed to verified auxin accumulation patterns during ovule forming. As described above, *L. virginicum* made only one seed per locules while *A. thaliana* made around twenty seeds per locules, so auxin accumulation patterns during ovule forming might be different than *A thaliana*. To investigate auxin accumulation patterns during ovule formation, it needed to transformed auxin reporter gene DR5 to *L. virginicum*. I prepared two plasmids pC3300:pDR5-ntdTOMATO and pC3300:pDR5-GUS for transforming (Table 3). For transforming, *Lepidium* needed specific agrobacterium strain because each strain showed different efficiency when transformed. One of agrobacterium strain AGL1 showed 0%, LBA4404 showed 0.1% and GV3101 showed 0.19% efficiency of transforming (Teresa and Gunter, 2013). For this reason, prepared GV3101 and injected each DR5 plasmids doing electroporation. After finished to prepare transformation materials, did floral dip to *L. virginicum* with 3 weeks vernalization for transformation. This result will be a base to study specific genes

that affected ovule formation to regulate seed number per silique in *L. virginicum*.

4. Discussion

In this study, four candidate genes AT4G23030, *PUP15*, AT4G38710, and AT1G61890 were found as regulator genes of seed number per silique. As described above, seed number per silique is affected by two factors such as formed ovule number and ovule fertility. However, the seed number per silique of AT4G23030, *PUP15*, AT4G38710, and AT1G61890 data in this study cannot show the which factors caused these results. For this reason, search the gene expression level of AT4G23030, *PUP15*, AT4G38710, and AT1G61890 from TAIR10 to find what factors affect the seed number per silique (Figure 10). The locations and stages of gene expression are flowers, siliques, stamen, anther, and carpel that can affect the formed ovule number and ovule fertility. AT4G23030 showed high expression level in stamen filament of the mature flower and pod of the silique. Especially, pod of silique is the location which is related with ovule fertility. *PUP15* showed almost low expression level than other two gene. The location of *PUP15* highest expression was opened anthers which is related with ovule fertility. AT4G38710 expression level is higher than two genes. The highest expression location is axis of inflorescence which is related with formed ovule number. AT1G61890 showed high

expression level in flower stage 1 which is a mature flower and pod of silique stage 1. Flower stage relate with pistil and stamen development and pod of the silique stage can affect ovule fertility. The gene expression level and location of AT4G23030, *PUP15*, AT4G38710, and AT1G61890 from TAIR10 indicated that may affect formed ovule number to regulate seed number per silique.

AT4G23030 and AT1G61890 are a gene that is described MATE (multidrug and toxic compound extrusion) efflux family protein. MATE efflux family protein exist both animal and plant. MATE play important roles such as mediate multiple functions in plats through efflux of diverse substrates including organic molecules, specialized metabolites, hormones and xenobiotics (Nimmy, 2022). MATE proteins also regulate overall plant development by controlling phytohormone transport, tip growth processes, and senescence (Upadhyay *et al.*, 2019). There are 58 MATE transporters in *Arabidopsis thaliana*, which are also known as DETOXIFICATION (DTX) proteins and it grouped four MTAE I, II, III and IV (Upadhyay *et al.*, 2019; Wang *et al.*, 2016). AT4G23030 is a gene in MATE IV and AT1G61890 is a gene in MATE II. For example, DTX14 (AT1G71140) is a gene in MATE I group and representative MATE gene in *A. thaliana* as xenobiotic efflux pumps, which confer multidrug resistance against various antibiotics in

pathogens (Miyauchi *et al.*, 2017). MATE proteins that have function of secondary metabolite. DTX35(AT4G25640) is a gene in MATE II group and a MATE gene in *A. thaliana* as reported to transport mainly pro-anthocyanin that is proton coupled antiporter highly expressed in epidermal guard cells of anthers, stigmas, siliques, and nectaries (Marinova *et al.,* 2007). Anthocyanins are pigments that had red, blue, purple, or black colors and belong to flavonoids that regulate dormancy, biotic interaction, and transmit aroma and flavor in plant (Upadhyay *et al.*, 2019). DTX35 mutants indicated defects of reproductive developments in plant because changing of flavonoids composition affect lignification in the pollen coat, hence the viability of pollen, and dehiscence of anthers that can cause the differential reproductive success (Thompson *et al.*, 2010). MATE proteins that have function of regulating iron homeostasis. Iron(Fe) is an essential micronutrient of plant that perform as co-factor in the electron transport chain of respiration and photosynthesis, and affect root morphogenesis and flower development (Roschzttardtz et al., 2013). MATE transporter FRD3(AT3G08040) is a gene in MATE III that supports a role of Fe homeostasis in plant development is expressed in seeds and flowers in *Arabidopsis* and frd3 mutants indicated defective germination and sterility (Roschzttardtz *et al.*, 2013). MATE

proteins that regulate plant hormone levels and transport. ADP1(AT4G29140) is a MATE transporter in MATE IV and regulate local auxin levels in meristematic tissues of *Arabidopsis* (Li *et al.*, 2014). ABA is a plant hormone that affect seed maturation, seed dormancy, stomatal closure, drought responses, and lateral root formation and DTX50(AT5G52050) is a MATE transporter in MATE IV that regulate efflux of ABA from guard cells (Zhang *et al.*, 2014). MATE proteins act as anion channels and growth regulators. BIG EMBRYO1(BIGE1, AT1G71870) is a trans-Golgi localized MATE transporter in MATE IV that regulates embryo size, initiation, and lateral organ size in maize and Arabidopsis (Suzuki et al., 2015). Bige1a and bige1abige1b exhibit smaller seed size compared to the wild type (Suzuki et al., 2015). The detail function and pathway of MATE efflux family protein is still not known. May MATE efflux functions such as transfer secondary metabolite, regulate iron homeostasis, plant hormone levels and transport, and act as anion channels and growth regulators affect the number of seeds per silique of plant.

PUP15 is a gene that is described member of a family of proteins related to Phosphorus uptake1 (PUP1), a purine transporter. PUP1 is known as a major quantitative trait locus (QTL) located on rice (Oryza sativa) chromosome 12 that is associated with tolerance of

phosphorus (P) deficiency in soil (Wissuwa *et al.*, 2002). P is a key factor to plant growth and development. P in plant are used such as transfer energy, photosynthesis, and movement of nutrients (Wang *et al.*, 2018). P affect the seed yields of plants positively because plants with high P develop roots well and it causes more nutrients to plant from soil, hence P deficiency cause reduction of plant growth and development and it caused to decrease of seed yields of plants. For these reasons, *PUP15* that is related with P may affect the seed number per silique in *A. thaliana*.

AT4G38710 is a gene that is described glycine-rich protein. Glycine is an amino acid that is the simplest stable amino acid. The function of glycine-rich protein is still not known well. It is known that glycine-rich protein length positively associated with stress tolerance (Czolpinska & Rurek, 2018). The proper stress can affect the plants develop, so this function may affect the seed number per silique.

These known functions of four candidate genes AT4G23030, *PUP15*, AT4G38710, and AT1G61890 suggest that they can affect ovule forming number and ovule fertility to regulate seed number per silique. To confirm these effects, the repeat of phenotyping is required.

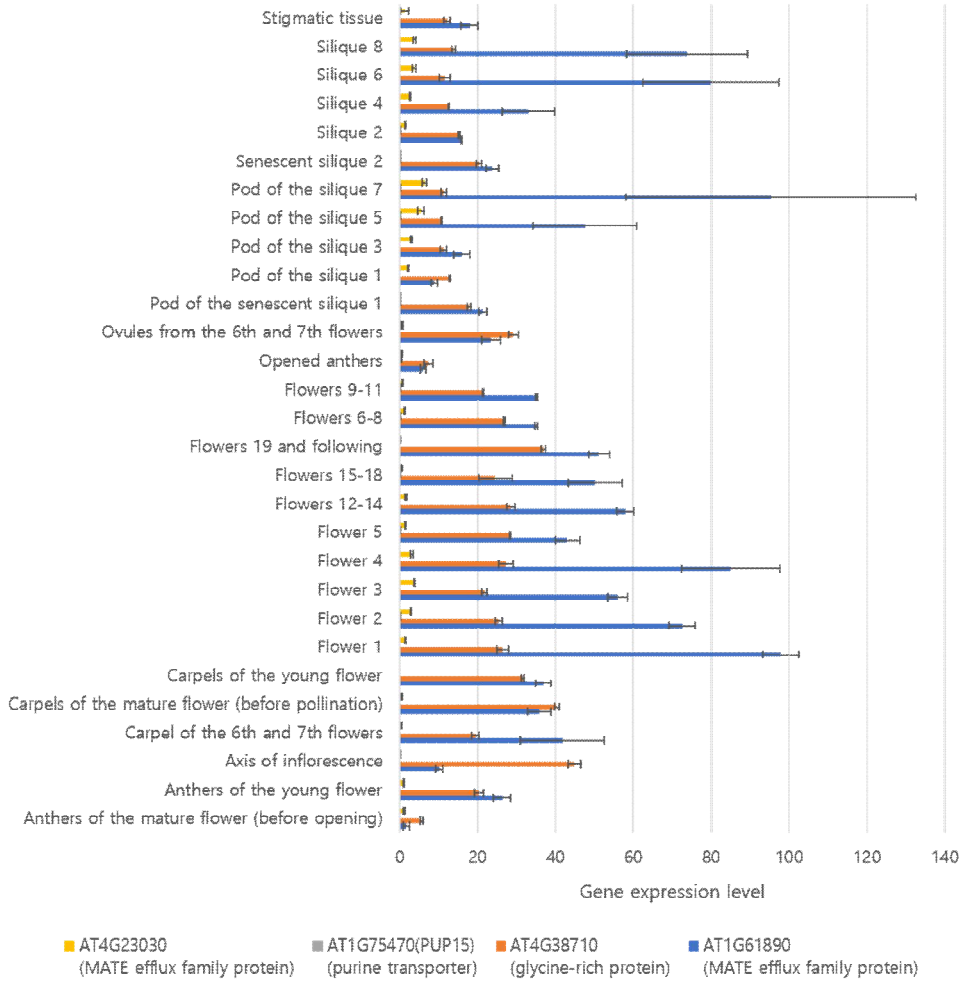


Figure 10. Gene expression level of four candidate genes in *A. thaliana*. The gene expression level of AT4G23030, *PUP15*, AT4G38710, and AT1G61890. X-axis indicated gene expression level and y-axis showed anther, carpel, flower and silique stages which could affect seed number. The gene expression level data is from TAIR10.

5. References

Benková, E., Michniewicz, M., Sauer, M., Teichmann, T., Seifertová, D., Jürgens, G., & Friml, J. (2003). Local, efflux-dependent auxin gradients as a common module for plant organ formation. Cell, 115 (5), 591-602.

Yamaguchi, S., Wu, S. I., Biswas, M., Yate, M., Aoki, Y., Barley, E. A., & Thornicroft, G. (2013). Effects of short-term interventions to reduce mental health-related stigma in university or college students: a systematic review. The Journal of nervous and mental disease, 201 (6), 490-503.

Sreenivasulu, N., & Wobus, U. (2013). Seed-development programs: a systems biology-based comparison between dicots and monocots. Annual review of plant biology, 64, 189-217.

Kern M. (2002). Food, Feed, Fibre, Fuel and Industrial Products of the Future: Challenges and Opportunities. Understanding the Strategic Potential of Plant Genetic Engineering. Journal of Agronomy and Crop Science, 188: 291-305.

Emery, R. N., Longnecker, N. E., & Atkins, C. A. (1998). Branch development in Lupinus angustifolius L. II. Relationship with endogenous ABA, IAA and cytokinins in axillary and main stem buds. Journal of Experimental Botany, 49 (320), 555-562.

Lenser, T., & Theißen, G. (2013). Conservation of fruit dehiscence pathways between *Lepidium campestre* and *Arabidopsis thaliana* sheds light on the regulation of INDEHISCENT. The Plant journal : for cell and molecular biology, 76 (4), 545–556.

Yuan, J., & Kessler, S. A. (2019). A genome–wide association study reveals a novel regulator of ovule number and fertility in *Arabidopsis thaliana*. PLoS genetics, 15 (2), e1007934.

Jiang, H. L., Hong, J., Jiang, Y. T., Yu, S. X., Zhang, Y. J., Shi, J. X., & Lin, W. H. (2020). Genome–Wide Association Analysis Identifies Candidate Genes Regulating Seed Number Per Silique in *Arabidopsis thaliana*. Plants (Basel, Switzerland), 9 (5), 585.

Koornneef M and Meinke D. (2010). The development of *Arabidopsis* as a model plant. The Plant Journal 61, 909–921.

Jiang, W. B. and Lin, W. H. (2013). Brassinosteroid functions in *Arabidopsis* seed development. Plant signaling & behavior, 8 (10).

Nemhauser, J. L., Feldman, L. J., and Zambryski, P. C. (2000). Auxin and ETTIN in *Arabidopsis* gynoecium morphogenesis. Development (Cambridge, England), 127 (18), 3877–3888.

Cao, J., Li, G., Qu, D., Li, X., & Wang, Y. (2020). Into the Seed: Auxin Controls Seed Development and Grain Yield. International journal of molecular sciences, 21 (5), 1662.

Li, L., Yu, X., Thompson, A., Guo, M., Yoshida, S., Asami, T., Chory, J., & Yin, Y. (2009). Arabidopsis MYB30 is a direct target of BES1 and cooperates with BES1 to regulate brassinosteroid-induced gene expression. The Plant journal : for cell and molecular biology, 58(2), 275-286.

Huang, H. Y., Jiang, W. B., Hu, Y. W., Wu, P., Zhu, J. Y., Liang, W. Q., Wang, Z. Y., & Lin, W. H. (2013). BR signal influences Arabidopsis ovule and seed number through regulating related genes expression by BZR1. Molecular plant, 6 (2), 456-469.

Bartrina, I., Otto, E., Strnad, M., Werner, T., and Schmülling, T. (2011). Cytokinin regulates the activity of reproductive meristems, flower organ size, ovule formation, and thus seed yield in *Arabidopsis thaliana*. The Plant cell, 23 (1), 69-80.

Gomez, M. D., Barro-Trastoy, D., Escoms, E., Saura-Sánchez, M., Sánchez, I., Briones-Moreno, A., Vera-Sirera, F., Carrera, E., Ripoll, J. J., Yanofsky, M. F., Lopez-Diaz, I., Alonso, J. M., and Perez-Amador, M. A. (2018). Gibberellins negatively modulate ovule number in plants. Development (Cambridge, England), 145 (13).

Cucinottam M. Colombo, L. and Roig-Villanova, I. (2014). Ovule

development, a new model for lateral organ formation. Frontiers in plant science, 5, 117.

Akhtar, S. S., Mekureyaw, M. F., Pandey, C., & Roitsch, T. (2020). Role of Cytokinins for Interactions of Plants With Microbial Pathogens and Pest Insects. Frontiers in plant science, 10, 1777.

Cucinotta, M., Manrique, S., Guazzotti, A., Quadrelli, N. E., Mendes, M. A., Benkova, E., & Colombo, L. (2016). Cytokinin response factors integrate auxin and cytokinin pathways for female reproductive organ development. Development (Cambridge, England), 143 (23), 4419–4424.

Swain, S. M., & Singh, D. P. (2005). Tall tales from sly dwarves: novel functions of gibberellins in plant development. Trends in plant science, 10 (3), 123–129.

Sun T. P. (2010). Gibberellin–GID1–DELLA: a pivotal regulatory module for plant growth and development. Plant physiology, 154 (2), 567–570.

Yu, N., Luo, D., Zhang, X., Liu, J., Wang, W., Jin, Y., Dong, W., Liu, J.,

Liu, H., Yang, W., Zeng, L., Li, Q., He, Z., Oldroyd, G. E., & Wang, E. (2014). A DELLA protein complex controls the arbuscular mycorrhizal symbiosis in plants. Cell research, 24 (1), 130–133.

Zhou, X., & Stephens, M. (2012). Genome–wide efficient mixed–model analysis for association studies. Nature genetics, 44 (7), 821–824.

Nimmy, M. S., Kumar, V., Suthanthiram, B., Subbaraya, U., Nagar, R., Bharadwaj, C., Jain, P. K., & Krishnamurthy, P. (2022). A Systematic Phylogenomic Classification of the Multidrug and Toxic Compound Extrusion Transporter Gene Family in Plants. Frontiers in plant science, 13, 774885.

Czolpinska M, Rurek M. (2018). Plant Glycine–Rich Proteins in Stress Response: An Emerging, Still Prospective Story. Front Plant Sci. 2018 Mar 8;9:302.

Miyauchi, H., Moriyama, S., Kusakizako, T., Kumazaki, K., Nakane, T., Yamashita, K., Hirata, K., Dohmae, N., Nishizawa, T., Ito, K., Miyaji, T., Moriyama, Y., Ishitani, R., & Nureki, O. (2017). Structural basis for xenobiotic extrusion by eukaryotic MATE transporter. *Nature*

communications, *8*(1), 1633.

Upadhyay, N., Kar, D., Deepak Mahajan, B., Nanda, S., Rahiman, R., Panchakshari, N., Bhagavatula, L., & Datta, S. (2019). The multitasking abilities of MATE transporters in plants. Journal of experimental botany, 70(18), 4643–4656.

Li, R., Li, J., Li, S., Qin, G., Novák, O., Pěnčík, A., Ljung, K., Aoyama, T., Liu, J., Murphy, A., Gu, H., Tsuge, T., & Qu, L. J. (2014). ADP1 affects plant architecture by regulating local auxin biosynthesis. PLoS genetics, 10(1).

Zhang, H., Zhu, H., Pan, Y., Yu, Y., Luan, S., & Li, L. (2014). A DTX/MATE-type transporter facilitates abscisic acid efflux and modulates ABA sensitivity and drought tolerance in *Arabidopsis*. Molecular plant, 7(10), 1522–1532.

Thompson, E. P., Wilkins, C., Demidchik, V., Davies, J. M., & Glover, B. J. (2010). An *Arabidopsis* flavonoid transporter is required for anther dehiscence and pollen development. Journal of experimental botany, 61(2), 439–451.

Roschzttardtz, H., Conéjéro, G., Divol, F., Alcon, C., Verdeil, J. L., Curie, C., & Mari, S. (2013). New insights into Fe localization in plant tissues. Frontiers in plant science, 4, 350.

Suzuki, M., Sato, Y., Wu, S., Kang, B. H., & McCarty, D. R. (2015). Conserved Functions of the MATE Transporter BIG EMBRYO1 in Regulation of Lateral Organ Size and Initiation Rate. The Plant cell, 27(8), 2288–2300.

Wissuwa, M., Wegner, J., Ae, N., & Yano, M. (2002). Substitution mapping of Pup1: a major QTL increasing phosphorus uptake of rice from a phosphorus-deficient soil. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik, 105(6-7), 890–897.

Wang, F., Deng, M., Xu, J., Zhu, X., & Mao, C. (2018). Molecular mechanisms of phosphate transport and signaling in higher plants. Seminars in cell & developmental biology, 74, 114–122.

Wang, L., Bei, X., Gao, J., Li, Y., Yan, Y., & Hu, Y. (2016). The similar and different evolutionary trends of MATE family occurred between rice and *Arabidopsis thaliana*. BMC plant biology, 16(1),

207.

국문초록

밑씨의 수정으로 생성되는 생식기관인 종자는 식물과 인류의 생활에서 떼어낼 수 없는 중요성을 가지고 있다. 종자의 생성은 식물 생활사의 매우 중요한 역할을 하고 있으며, 종자는 인류 역사상 음식, 사료 그리고 연료 등 중요한 자원으로써 사용되어져 왔다. 특히 종자 수는 번식의 성공과 수확량의 영향을 줄 수 있는 중요 형질들 중 하나이다. 그렇기에 종자수를 조절하는 인자의 연구는 중요한 가치를 가지고 있지만, 여전히 미지인 부분들이 남아있다.

본 연구에서는 애기장대에서 열매 당 종자 수 조절인자에 대해 집중을 하였으며, 이를 위해 131개의 애기장대 생태형을 이용하였다. 애기장대의 131 생태형 중 열매 당 종자 수 표현형이 가장 낮은 10개의 생태형은 밑씨 수정률에 영향을 받는다. 열매 당 종자 수 표현형이 가장 높은 10개의 생태형은 밑씨의 형성 수의 영향을 받는다. 이러한 결과들을 통해 본 실험에서 열매 당 종자 수는 밑씨의 수정률과 밑씨의 생성숫자에 의해 조절되고, 찾고자 하는 조절 유전 인자 또한 밑씨의 수정률과 밑씨의 생성숫자를 조절함으로써 열매 당 종자 수를 조절할 것임을 알 수 가 있다.

애기장대의 131 생태형의 열매 당 종자 수의 GWAS 분석은 열매 당 종자수와 연관이 있는 107 개의 SNP를 나타냈다. 그 중에서 p-value 가 가장 낮은 10개의 SNP를 후보로써 선정했으며 그 후 해당 후보 SNP 주변에 위치하거나 또는 SNP를 포함하고 있는 유전자를 대표

유전자로 선정한다. 선정된 10개의 대표 유전자들의 T-DNA 삽입 돌연변이들의 표현형을 측정하였고, 그 중 AT1G61890, *PUP15*, AT4G23030, AT4G38710 네 개의 대표 유전자의 T-DNA 삽입 돌연변이의 표현형이 비교용 생태형인 Col-0와 비교했을 시 유의미한 변화를 보여주었다. AT1G61890과 AT4G23030은 MATE 유출 계열(family) 단백질로 T-DNA 삽입 돌연변이에서 둘다 열매 당 종자 수가 증가됐다. *PUP15*는 푸린 투과효소 구성 계열 중 하나 로 T-DNA 삽입 돌연변이에서 열매 당 종자 수가 증가됐다. AT4G38710은 글리신 풍부 단백질로 T-DNA 삽입 돌연변이에서 열매 당 종자 수가 감소된다.

GWAS의 한계점을 극복하기위해 QTL을 수행하기로 하였고, 이를 위해 Col-0와 Di-G를 서로 교배시켰다. 애기장대의 생태형 중 하나인 Di-G가 본 실험 및 다른 실험에서 모두 열매 당 종자 수 및 밑씨 수가 크게 측정되었다. 이러한 교배가 잘 이뤄졌는지를 확인하기위해 유전자 표식을 제작하여 각 개체의 염기서열을 확인하였고, Col-0, Di-G, 그 둘의 교배종의 염기서열이 서로 다름을 통해 교배가 잘 이뤄졌음을 확인할 수 있었다.

열매 당 종자 수를 같은 종내 말고 종간 에서도 확인하기 위해 비교대상으로 애기장대와 같은 과인 십자화과에 속하며 가까운 유연관계를 가지고 있는 다닥냉이를 선정하였다. 열매당 종자 수에 대해서 다닥냉이는 씨방 당 한 개의 종자가 있는 반면 애기장대는 씨방 당 약 스무개의 종자가 있는 차이점을 보여준다. 이러한 열매 당 종자 수 차이점은 식물의 진화와 연관이 되어있으며 이런 변화에 앞서 발견한 애기장대에서 열

매 당 종자 수를 조절하는 유전자가 관련이 있는지를 알아보고자 그 실험의 기초를 세우고자 한다.

그를 위해, 먼저 수집한 다닥냉이의 종이 무엇인지를 ITS 염기서열을 이용해 확인하고자 하였고, NCBI에 있는 다닥냉이 종들의 ITS 염기서열과 수집한 다닥냉이의 ITS 염기서열을 MEGA 11 프로그램을 이용해 비교한 결과, 수집한 다닥냉이의 종이 콩 다닥냉이임을 확인할 수 있었다.

다닥냉이에서 밑씨 생성 시 옥신의 축적 패턴을 확인하기위해 DR5를 형질전환을 수행하였고, 꽃 담그기(floral dippig) 방식을 이용하여 DR5의 형질전환을 시행하였다. 현재 형질전환 된 종자를 구분하는 과정을 진행 중이다.

이번 실험에서, 종내 및 종간 분화에서 열매 당 종자 수 를 조절하는 유전적 메커니즘에 대한 연구를 실시하였다. 종내 연구에서, 나는 열매 당 종자수를 조절하는 유전자들을 GWAS 및 표현형 측정을 통해 찾았으며, GWAS의 한계점을 보완하기 위해 QTL을 이용해 Col-0와 Di-G에서 열매 당 종자 수를 조절하는 유전자를 찾고자 한다. 종간 연구에서는 다닥냉이와 애기장대의 차이점을 해부를 통해 확인하였으며, 다닥냉이의 종을 판별 하고 DR5를 형질전환하여 밑씨 생성 간 옥신패턴을 확인하여 열매 당 종 자 수 와 관련된 연구를 시작하고자 한다. 해당 연구의 결과들은. 유채 등 다른 식물들의 종자 수 연구에 이용될 수 있을 것으로 기대된다.