Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2020:59
Faculty of Landscape Architecture, Horticulture
and Crop Production Science

Genomic tools and molecular breeding approaches for the domestication of field cress (*Lepidium campestre* L.)

Zeratsion Abera Desta

Genomic tools and molecular breeding approaches for the domestication of field cress (*Lepidium campestre* L.)

Zeratsion Abera Desta
Faculty of Landscape Architecture, Horticulture and Crop Production Science
Department of Plant Breeding
Alnarp

SLU
SWEDISH UNIVERSITY
OF AGRICULTURAL
SCIENCES

DOCTORAL THESIS
Alnarp 2020

Acta Universitatis agriculturae Sueciae
2020:59

Cover: Field cress plant from field experiment
(Photo: by Zeratsion Abera Desta)

ISSN 1652-6880
ISBN (print version) 978-91-7760-636-9
ISBN (electronic version) 978-91-7760-637-6

Alnarp
Print: SLU Service/Repro, Alnarp 2020

Genomic tools and molecular breeding approaches for the domestication of field cress *(Lepidium campestre* L.*)*

Abstract

Field cress (*Lepidium campestre* L.) is a biennial self-pollinated plant with a small genome size. The ever-increasing global population alongside climate change prompts urgent actions to save the ecosystem. Domesticating multi-purpose species such as field cress could be considered as part of the solution to mitigate the challenges posed by climate change and population growth. In addition to the oil producing potential, the domestication of field cress in arable lands has multitude effects – such as protecting environmental contamination and contributing as food and feed uses. In clues of these potentials, identifying the genomic variation underlying important traits using genomic tools is pivotal approach in field cress domestication.

The main goal of the research in this thesis was to develop genomic tools for field cress domestication, specifically aiming at constructing the genetic linkage map, identifying the quantitative trait loci (QTL) underpinning domestication traits, and elucidating the common genetic variants associated with the seed yield as well as seed oil, protein, and moisture contents in field cress. An integrated mapping approach were performed to developing the first genetic linkage map for field cress. Relying on the linkage map, the identification of domestication QTL using linkage analysis as well as common variants using genome-wide association study (GWAS) were succeeded. Furthermore, the developed linkage map will be used in guiding to develop the reference genome using whole-genome sequencing (WGS) in field cress. Given further functional genomic efforts, the identified QTL and single variants could facilitate the process of domestication and genomics-assisted breeding in field cress, including the use of evolving approaches such as genome-wide prediction in the field cress.

Keywords: Cytogenetics, Domestication, GEBV, Genomic selection, GWAS, LD, Oil, Protein, QTL, WGS

Author’s address: Zeratsion Abera Desta, Swedish University of Agricultural Sciences, Department of Plant Breeding, Alnarp, Sweden

Preface

The sustainability of agriculture and food security goals are influenced by the intensified threats of climate change alongside the rising human population sharply. Domesticating new plant species – such as field cress – could counteract the impacts of environmental contamination while meeting the food demands of the growing population. Field cress can be grown as a green cover crop in the arable lands, which could combat the environmental pollution (e.g. preventing leaching sensitive minerals). Given the tremendous advances in whole-genome sequencing (WGS) along with high-throughput phenotyping, it could be more feasible to explore the underlying genes of complex quantitative traits in the field cress domestication.

In addition to the published papers and manuscript, this thesis describes the focal genomic approaches for unravelling genomic variations associated with key traits of field cress. It begins to explain the construction of genetic linkage map intertwining with different mapping approaches as a hallmark to the rest of the genome-wide analyses. Next, it explores the identification of quantitative trait loci (QTL) underlying domestication traits of field cress using genome-wide linkage analysis. Following this, it continues with the detection of common variants associated with agronomic, physiological, and economic important traits, particularly with the detail illustrations using seed yield genome-wide association study (GWAS) in the field cress.

In this GWAS analysis, the use of naive linear model (i.e., in the absence of both the population stratification and kinship components), using population stratification alone, using kinship alone, and simultaneous use of population stratification and kinship are described. Additionally, the inability of the current GWAS standard procedures with low minor allele frequency and rare loci as well as possible alternatives are discussed. Finally, the thesis ends summarizing with the major findings along with vital perspectives in field cress domestication and breeding.

In moving from SNP-trait association to function, these identified QTL and single variants could plausibly guide the hypothesis of finding trajectory genetic mechanisms controlling domestication and desirable agronomic traits in field cress. Furthermore, these QTL and single variants could serve as templates to the still-growing field; i.e., genomic selection. Certainly, these QTL and common variants could lay a foundational framework of genomics-led domestication and breeding in field cress.

The overall work of this thesis was extended beyond the expected time schedule. In addition to the late delivery (at the end of the PhD study) of genotyping, developing the final genetic linkage map by combining multiple mapping approaches in a new species, mainly without reference genome, was not only challenging and cumbersome but also led to unexpected delays. Furthermore, implementing field phenotyping across three environmental sites of Sweden along with developing high-throughput protocols in field cress were some of the ambitious performances, contributing enormous workloads in this PhD study.

All these and other unexplained challenges were confronted and resolved. However, I was not alone, but rather the final successful achievements of this research thesis were the combined efforts and enthusiasms with my supervisors, highly respected scientists across the world (i.e., from the United Kingdom, Poland, and Canada), courageous employees of the SLU branch campuses (i.e., Uppsala, Lanna, and Alnarp) as well as with the kind supports from MISTRA-Biotech project.

Thank you so much for all the lessons and passions I have acquired throughout my studies.

Zeratsion Abera Desta
Sweden, 2020

Dedication

Dedicated to all my family members and friends

Contents

List of publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I. **Desta, Z.A.**, Kolano, B., Shamim, Z., Armstrong, S.J., Rewers, M., Sliwinska E., Sandeep, K., Parkin, I.A.P., Rodomiro, O., & de Koning, D.J. (2019). Field cress genome mapping: Integrating linkage and comparative maps with cytogenetic analysis for rDNA carrying chromosomes. *Scientific Reports 9*, 17028.

II. **Desta, Z.A**., de Koning, D.-J., & Ortiz, R. (2020). Molecular mapping and identification of QTL for domestication traits in field cress (*Lepidium campestre* L.). *Heredity 172*, 135–142.

III. **Desta, Z.A**., Ortiz, R., & de Koning, D.-J. (2020). Genome wide association mapping of quantitative trait loci for seed yield, oil, protein, and moisture content of field cress (*Lepidium campestre* L.). (Manuscript)

IV. **Desta, Z.A.**, & Ortiz, R. (2014). Genomic selection: genome-wide prediction in plant improvement. *Trends in Plant Science 19, 592–601*.

Paper I and II are open access published articles, and paper IV is reproduced with the permission from Elsevier publisher

The contribution of Zeratsion Abera Desta to the papers included in this thesis was as follows:

I. Initiated the integration of different mapping techniques and planned the experiment, performed the genetic experimental and practical works, data analysis, and wrote the manuscript together with co-authors.

II. Planned the experiment together with supervisors, performed the experimental and field works, data analysis and wrote the manuscript together with co-authors.

III. Planned the experimental and environmental fields together with supervisors, performed data analysis and wrote the manuscript together with co-authors.

IV. Planned together with a supervisor and wrote the manuscript together with a co-author.

List of tables

List of figures

Abbreviations

BLASTN	Basic local alignment search tool nucleotide
BLUE	Best linear unbiased estimation
BLUP	Best linear unbiased prediction
CRISPR	Clustered regularly interspaced short palindromic repeat
CTAB	Cetyltrimetheylammonium bromide
DNA	Deoxyribonucleic acid
eQTL	Expression quantitative trait loci
FDR	False-discovery rate
GBLUP	Genomic best linear unbiased prediction
GEBV	Genomic estimated breeding value
GWAS	Genome-wide association study
GWS	Genome-wide selection
kg/ha	Kilogram per hectare
LD	Linkage disequilibrium
LG	Linkage group
MAF	Minor allele frequency
MLM	Mixed linear model
MSc	Master of Sciences

NAM	Nested-association mapping
PCA	Principal component analysis
PCoA	Principal coordinate analysis
PhD	Doctor of Philosophy
QC	Quality control
QQ	Quantile-quantile
QTL	Quantitative trait loci
RAD	Restricted amplified DNA
RFLP	Restricted fragment length polymorphism
RNA	Ribonucleic acid
rrBLUP	Ridge regression best linear unbiased prediction
Seg.D	Segregation distortion
SNP	Single nucleotide polymorphism
SSR	Simple-sequence repeats
WGS	Whole-genome sequencing

1. Introduction

1.1 Botanical description of field cress

Field cress (*Lepidium campestre* L.), a highly self-pollinated biennial plant, is a member of Brassicaceae family (Warwick *et al.*, 2006). The basal (lower) leaves form rosette (i.e., usually the leaves arranged circularly at the base) shape (Figure 1a–b, d), from which later emerge branched stems (Figure 1d). The margins of young basal leaves are lobed (dissected), toothed or entire shape (Figure 1a–b). The stem leaves are small arrow shapes auriculate-clasping (i.e., extending growth of leaf sheath towards the stem) at their base (Figure 1c), with partially toothed or entire leaf margins. The flowering buds are tiny grown like “broccoli” shape (Brill & Dean, 1994), and the white flower of field cress consists of four white petals (Elias & Dykeman, 2009). The time of flowering in field cress usually occurs from May to June, and it can sporadically extend up to July or August depending on accession and altitude (Seebeck, 1998) effects.

The fruit pod, which consists of two seeds per pod, has a trapezoid-like shape, with a wing-like shape at the apex (Figure 1F). Field cress has variety of seed colours (Figure H–J) – such as brown, dark or brown dark. The anthocyanin pigments can be expressed in the leaves, including in the fruit pods of the plant (practical observations). Although field cress grows in a variety of soil types, mostly it is found in non-cultivated rocky area (i.e., xerophytic plant) around roadsides, railways, and in marginalized sites (Brill & Dean, 1994). Field cress is an introduced weed to United States of America and Canada (Elias & Dykeman, 2009), where it is endemic to Europe (Bandara *et al.*, 2007).

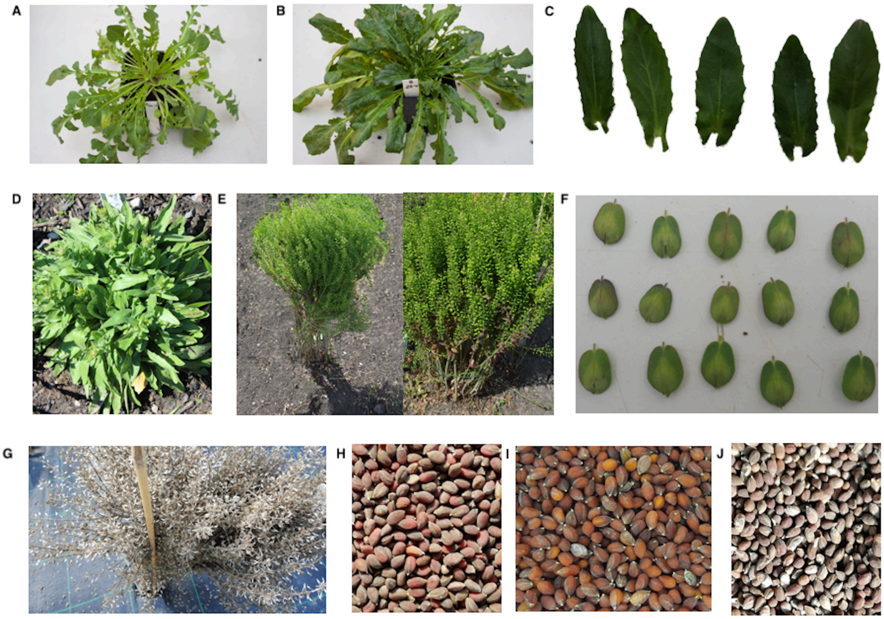

Figure 1. Morphological descriptions of field cress. A. Young basal leaves of deeply lobed or dissected leaf margins of field cress B. Young basal leaves of serrated leaf margins of field cress. C. Stem leaves in field cress plant D. The basal leaves in the emergence of stem branches. E. Physiological maturity stage in field cress plants. F. Fruit pods of filed cress. G. Field cress after harvest maturity revealing pod shattering. H–J, relating the brown dark, brown, and dark colour of field cress seeds, respectively. (Photos: by Zeratsion Abera Desta)

Both species have the same chromosome number ($2n = 2x = 16$), and similar small genome sizes (Desta *et al.*, 2019). The morphological descriptors for *L. heterophyllum* resemble those of field cress. In addition to their differences in the growth habit, in which the field cress is a biennial whereas the *L. hetrophyllum* is a perennial plant (Gilkey *et al.*, 1967), field cress shows more vigorous growth and produces more seeds per plant than *L. heterophyllum*. The stem position of the *L. heterophyllum* is usually prostrate or semi-prostrate plant, whereas field cress is an erect (Figure 1E) or a semi-erect plant position (Desta *et al.*, 2020).

1.2 Domestication of field cress

Considering the repercussion of global warming along with the alarmingly ever-increasing world population, the domestication of new multi-purpose species that could alleviate these negative consequences while fulfilling the need for food and feed security is crucially important. The leaching effects of nitrogen and phosphorous minerals have series agricultural and environmental impacts. Domesticating multi-purpose species (such as field cress) could countervail these threats imposed by climate change (Kanter, 2018; Howarth, 2008). Furthermore, the presence of field cress in cultivated lands could be used to maintain the soil nutrient cycle including the soil microbes. In addition to its protein content (~ 24%), the winter-hardy field cress has oil (~19%) producing potential (Nilsson *et al.*, 1998; Merker & Nilsson, 1995). Furthermore, intercropping field cress with cereals could help to manage soil minerals, thereby circumventing use of chemical fertilizers in agricultural soils.

Traits related to domestication, agronomy, physiology, and economic importance in field cress are the plant height, inflorescence height, number of inflorescences, pod (seed) shattering, number of stem branches, plant architecture, rooting depth, leaf architecture, number of pods, flowering time, pod maturity, thousand seed weight, seed size, seed colour, seed yield, oilseed content, seed protein content, and seed moisture content. However, detailed research on the underlying genetic mechanisms of these traits lags behind. To facilitate the process of field cress domestication, as well as to enhance and modify its desirable traits, the use of genomic toolboxes and genomics-assisted breeding approaches is instrumental.

Domestication is a systematic process of changing or shaping the morphological and physiological traits of cultivated plants compared to their wild progenitors (Diamond, 2002) for the sake of human benefit. Plant domestication is an intricate and a multi-stage process to reshape the domestication genes from their ancestors. Large seed size, reduced shattering, high germinability, and upright (erect plant) position are some of the main target traits of field cress domestication.

1.3 High-throughput phenotyping

Phenotyping is not only used to measure various traits in a vast germplasm collection, but it is also used to deliver fast and reliable data that can be used to link the phenotype with genotype in complex trait analysis. However, the classical phenotyping approach is technically cumbersome, expensive, and lacks accuracy (Furbank & Tester, 2011). Notably, with traits that demand repeated recording, underground phenotyping (e.g. root phenotyping) (Yazdanbakhsh & Fisahn, 2012; Singh *et al.*, 2010), various physiological analysis, as well as with traits that express at different seasons are unable to cope with the performances in conventional phenotyping techniques. It is, therefore, essential to consider the high-throughput phenotyping (Berger *et al.*, 2012; Cole & Chory, 2012).

In some trait analyses, phenotyping obliges the destruction (e.g. oil and protein recording) of the germplasm (Berger *et al.*, 2012). However, this problem can be solved by using the non-destructive, accurate, and rapid technique (Ecarnot *et al.*, 2013) – e.g. near-infrared (NIR) spectroscopic instruments. High-throughput approach employs methods such as digital imaging, spectroscopy or robotics-assisted phenotyping that bridge the gap arising from the poor performances of the field, laboratory or greenhouse phenotyping. The intensive and detail data through phenomics research speed-up the analyses of genomics-assisted domestication and breeding of plants (Mir *et al.*, 2019).

1.4 Linkage, cytogenetic, and comparative maps

Over the last two decades, various types of markers – such as restricted fragment length polymorphism (RFLP), simple-sequence repeats (SSR) and single nucleotide polymorphism (SNP) markers – have been developed as one of the tools for the genome mapping. SNPs are the single base-pair changes in DNA while comparing between the genomes of individuals. In comparison to simple SSR markers, SNPs are cost effective (Ersoz *et al.*, 2007) and more abundant (Li *et al.*, 2017; Edwards *et al.*, 2007) across the genome of plants. The rapid advancement in using genome-wide markers, predominantly with the recent discovery of SNPs from high-throughput sequencing, has facilitated exploring the existing genomic variation of various wild and cultivated plants.

As a benchmark tool, a genetic map is used to solve complex questions in biological sciences. The two most broadly used methods are those leading to genetic and physical maps. In addition to these maps, the joint use of linkage map with cytogenetic and comparative maps is also popular, particularly, in mapping the wild and novel species such as field cress (Desta *et al.*, 2019). Genetic map utilizes the inherited genetic loci in a chromosome (Van Ooijen & Jansen, 2013). Hence, genetic mapping is the relative position of marker loci based on the estimates of recombination along chromosomes.

A linkage map is liable to various flaws such as over or under representation of crossovers or erroneous realignment of markers, thus deviating or inflating the estimates of a linkage map (Cartwright *et al.*, 2007; Howell *et al.*, 2002). Using sequencing technique, the physical map represents a more reliable position of the DNA bases. However, a predefined linkage map has a complementary assistance to order the assembled contigs in a chromosome while using physical mapping (Fierst, 2015; Troggio *et al.*, 2007), implying that the construction of a linkage map is a prerequisite for genetic analysis.

As there are pros and cons with each of the mapping techniques, combining linkage with less error-prone maps (e.g. integrating linkage, cytogenetic, and physical maps) (Desta *et al.*, 2019; Howell *et al.*, 2002) can provide a linkage map that is more reliable than either of these approaches alone. In addition to the substantial contributions to genome evolutionary research, cytogenetic mapping (i.e., dealing with specific probes and staining of karyotypes) is a promising technique to distinguish the inert and hot spots of recombination in a chromosome (Wang *et al.*, 2006). To identify the linked loci, comparative mapping or comparative genomics also explores the evolutionarily conserved genomic regions across the ancestors of related taxa.

The final construction of genetic linkage map has enormous benefits and some of these are: to locate and clone various traits (e.g. those with the domestication, physiological or agronomic traits), annotate and orient the DNA sequences (Fierst, 2015; Troggio *et al.*, 2007), impute the unobserved genotypes from reference maps (Xu *et al.*, 2015), identify the expression quantitative trait loci (eQTL) in gene expression analysis, investigate homology and whole-genome duplications between crop species, deploy and localize promising candidate genes (e.g. host plant resistance genes to pathogens and pests) in widely cultivated species (Hu *et al.*, 2019), assist in genome editing, and conduct genome-wide fine map analysis.

1.5 Quantitative trait loci (QTL) map analysis

Linkage and QTL map analyses rely on the cooccurrence of genome polymorphisms with loci affecting phenotypic traits (Astle & Balding, 2009), after which extra validation efforts are invested to identify the responsible genes and mutations. These genes (i.e., Mendelian genes) often can be detected in known pedigrees or within a relatively large family of relatives. These traits usually have major effects (e.g. monogenic traits) that can be cloned through functional genomics (Wang, 2011). Contrarily, many of the quantitative traits (e.g. agronomic traits) are governed by complex genetic architecture that can be affected by many genes (including gene-environment interactions), each gene contributing small effects. The nature and function of highly complex quantitative traits are poorly designated (Ahmadizadeh *et al.*, 2016) in pedigree populations of linkage studies.

Owing to the negligible influences of the environment, the major QTL are often suitable and stable (Li *et al.*, 2020; Wang *et al.*, 2020) not only for cloning their underlying genes but also incorporating their functional alleles in marker-assisted selection (MAS). However, the precision of finding the exact location of QTL is affected by, but not limited to, population size, population design (e.g. F_2 versus backcross population), method of analysis, density of markers, and trait heritability.

In addition to the poor mapping resolution and estimation bias, QTL analysis is population specific, and becomes infeasible when the population size is small (Teh *et al.*, 2020). Despite the limitations in QTL mapping, major QTL were successfully cloned in numerous plant species, for instance for dormancy and flowering time in *Arabidopsis* (Bentsink *et al.*, 2006; Werner *et al.*, 2005), fruit shape in tomato (Liu *et al.*, 2002), glume architecture in maize (Wang *et al.*, 2015), and seed shattering in rice (Konishi *et al.*, 2006). Hence, given the effective recombination in a large population size, simple traits can be managed appropriately in linkage studies and can further be cloned for the purpose of gene pyramiding and MAS in plants.

1.6 Genome-wide association study (GWAS)

Although enormous advances have been made through linkage studies, most of the detected variation seems to be affected by numerous very small effect genes (Holland, 2007). Although linkage analysis is a suitable approach for the large effect loci, the substantial effects of quantitative traits become undetected because many loci each having tiny effect alleles. Linkage study is neither extrapolated in another mapping population nor is always noticed in natural population. Furthermore, capturing of few effective recombination events in known pedigrees or relatives of large family (Mohammadi *et al.*, 2020; Huang & Han, 2014; Flint-Garcia *et al.*, 2003) shifted the trend of human genetic analysis towards the use of genome-wide association study (GWAS). Additionally, the availability of common SNP variation through high-throughput sequencing in GWAS has also led to identifying causal variants regulating complex quantitative traits (Tam *et al.*, 2019).

GWAS analysis utilizes the linkage disequilibrium (LD) that reveals linked genomic regions coinherited together more often than expected by chance alone (Ersoz *et al.*, 2007; Flint-Garcia *et al.*, 2003). Consequently, the variance explained by a marker can be used to explain the variance of another marker.

However, two unlinked loci either in the same or different chromosomes can also be in LD if they have similar allele frequencies, indicating that LD can also exist without linkage; for instance due to selection for a particular trait (Joukhadar *et al.*, 2019). The faster the decay in LD, the higher the density of markers are required (Mohammadi *et al.*, 2020; Abdurakhmonov & Abdukarimov, 2008; Ersoz *et al.*, 2007); in turn, the better the precision in locating the genes regulating the trait. The power to detect common variants in GWAS depends on the spectrum of allele frequencies, the extent of LD, the magnitude of additive genetic variation, and the resolution of the phenotyping.

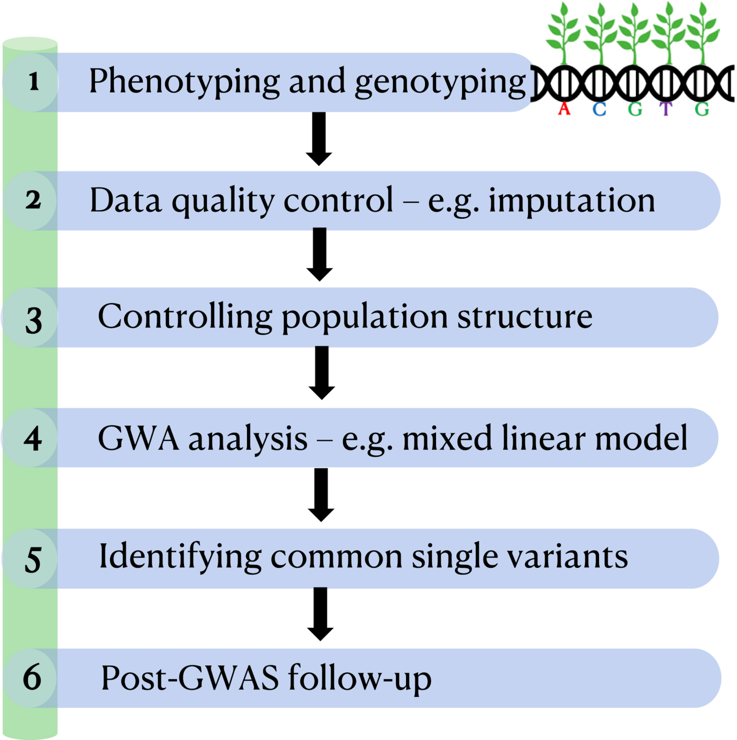
1 Phenotyping and genotyping
A C G T G
2 Data quality control – e.g. imputation
3 Controlling population structure
4 GWA analysis – e.g. mixed linear model
5 Identifying common single variants
6 Post-GWAS follow-up

Figure 2. Genome-wide association study (GWAS) steps. The initial step is performing detailed phenotyping and genotyping using high density SNPs on diverse germplasm collection. Next, quality control such as removing high missing data and imputing the moderately missing data, and adjusting for stratification to control false-positives in GWAS will be implemented. Next to this, analyzing using GWAS models, and identification of common variants associated with complex traits will be performed. The identified variants and causal genes need to be verified with post-GWAS pipelines (e.g. testing in a new population as well as use of mutation-based approaches).

Each of the steps in the GWA approach, from pre-analysis until identifying associated causal variants, need to be followed and implemented properly

(Figure 2). Important for succeeding in GWAS are multi-environment trials, accurate phenotyping (e.g. use of high-throughput methods), genotyping with super-dense polymorphic SNPs, quality control (QC) (e.g. removal of markers with minor allele frequencies (MAF); i.e., < 0.05 or < 0.1) and imputing missed dataset or unobserved genotypes. Subsequently, defining the population structure, and applying the proper model in the GWA analysis to identify the common variants controlling complex quantitative traits. Nevertheless, finding the actual gene demands to explore further verification steps (Huang & Han, 2014); i.e., a post GWAS follow-up (Figure 2) such as evaluating in independent population and use of gene cloning or isolation techniques.

1.7 Genomic selection

The use of GWAS has been continuing in plants, livestock and human genetics; however, each identified locus explains only a tiny fraction of the trait heritability. Moreover, many of the GWAS hits are in non-coding regions of the genome, thus complicating the efforts of post-GWAS verification in humans (Edwards *et al.*, 2013). To account the infinitesimal effects of highly multigenic variants, genomic selection (GS) or genome-wide selection (GWS) was first proposed in livestock (Meuwissen *et al.*, 2001), accounting variation across the whole-genome.

Unlike to GWAS, SNP markers are not treated separately in GS models but employed simultaneously. GS or GWS models – such as penalized or Bayesian regressions – are used to drag back the residual variances towards the population mean (e.g. genomic best linear unbiased prediction, GBLUP) or use of both shrinkage and selection methods such as BayesB model (Meuwissen *et al.*, 2001). The success of GS relies on designing a relevant training population, which in turn depends on the outcome of the breeding population (Desta & Ortiz, 2014; Lian *et al.*, 2014). Additionally, GS is also affected by other interacting comprehensive factors such as population size, the genomic architecture of the trait, the density of polymorphic markers, and the types of model employed.

Contrary to GWAS, which emphasizes on estimating SNP effect separately on diverse individuals, GS depends on the allele sharing among individuals to eventually predict the genomic estimated breeding values (GEBVs) in unobserved phenotypes. Hence, relatedness between the training and breeding population is a pivotal backbone to maintain prediction accuracy (Desta & Ortiz, 2014), whereas unrelated individuals are a nuisance in GS. Nevertheless, GS is not in place to replace the GWAS, nor is for linkage study; but it rather complements to either with one these approaches (Sawitri *et al.*, 2020; Spindel *et al.*, 2016). Most likely, with the advances and the opportunities of rapidly declining cost of whole-genome sequencing (WGS), along with the use of high-throughput phenotyping, GS is widely to hold more promise (Müller *et al.*, 2017) than employing either GWAS or linkage analysis for mapping variants contributing to complex traits.

2. Goals and objectives of the research thesis

The principal goal of this research was to develop genomic tools as well as to highlight the molecular breeding approaches for field cress domestication. The specific aims were to:

- develop genetic linkage map using integrated approaches of linkage, cytogenetic and comparative maps (Paper I),
- elucidate the genomic regions associated with domestication QTL in field cress (Paper II),
- estimate the LD and identifying the genetic variants in seed oil, protein, and moisture contents, as well as in the seed yield of field cress using GWAS approach (Paper III), and
- Assess the progresses, opportunities, and prospects of genomic selection in plants (Paper IV), which may become a useful tool to further accelerate field cress domestication.

3. Methodology

3.1 Mapping population and Genotyping field cress

The 503 offspring in the F_2 population; i.e., half-sibs of two subpopulations, were obtained from a cross between *L. campestre* and *L. heterophyllum* species. Genomic DNA was isolated from the leaves of F_2 individuals with some modification of cetyltrimetheylammonium bromide (CTAB) method (Saghai-Maroof *et al.*, 1984). DNA samples from 503 individuals were sent to Edinburgh Genomics (https://genomics.ed.ac.uk) for genotyping using iSelect Illumina Infinium method. SNP chips were custom designed from restricted amplified DNA (RAD) sequences that were implemented on diverse *Lepidium* samples (Lopes-Pinto *et al.*, 2016). Removing the unfit quality markers, 7624 SNPs were further employed to genotype the 503 F_2 individuals.

3.2 Genetic linkage map construction

In the preliminary quality control (e.g. missing values ≥ 5% and highly skewed segregation with $P < 0.0001$), 3623 SNPs and 13 F_2 individuals were excluded, while 4001 loci were retained in 490 individuals (Figure 3). Further quality inspection yielded 1517 SNPs and 482 individuals out of the 7624 SNPs and 503 individuals, respectively. From these loci (1517 SNP loci), 1401 and 116 corresponding to commonly and uniquely segregating loci, respectively were distributed across the eight linkage groups (LGs) of field cress (Figure 3). Genetic mapping was performed with JoinMap 4.1 version software (Van Ooijen, 2006), using maximum likelihood of mapping approach. We used the Haldane's (Haldane, 1919) mapping function, and LG map was visualized in 'Synbreed' package (Wimmer *et al.*, 2012). The extent and patterns of SNPs across LGs were illustrated using 3-dimensional (3d) principal coordinate analysis (PCoA) in R version 3.6.3 software package (R Core Team, 2019).

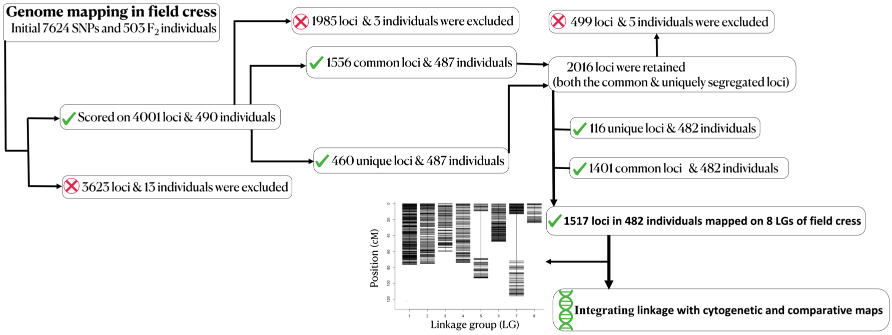
Genome mapping in field cress
Initial 7624 SNPs and 503 F_2 individuals
1985 loci & 3 individuals were excluded
499 loci & 5 individuals were excluded
1556 common loci & 487 individuals
2016 loci were retained
(both the common & uniquely segregated loci)
Scored on 4001 loci & 490 individuals
116 unique loci & 482 individuals
460 unique loci & 487 individuals
1401 common loci & 482 individuals
3623 loci & 13 individuals were excluded
1517 loci in 482 individuals mapped on 8 LGs of field cress
Position (cM)
Linkage group (LG)
Integrating linkage with cytogenetic and comparative maps

Figure 3. The multi-stage procedures of genome mapping in field cress. In total,7624 SNPs were used to genotype 503 two F_2 half-sib subpopulations. In the initial quality control, 3623 SNP loci were excluded, and 4001 SNPs in 490 F2 individuals were retained. After subsequent quality inspection, 1556 commonly segregating loci and 460 uniquely (loci segregating in either of the two half-sib populations) segregating loci remained for 487 individuals. Further analysis generated linkage map in field cress, resulting in 1517 mapped on eight linkage groups (LGs). The generated linkage map was combined with cytogenetic and comparative maps to construct the final genetic linkage map in field cress.

The basic local alignment search tool nucleotide (BLASTN) search versus the coding sequences of *A. thaliana* genome (TAIR; www.arabidopsis.org) were implemented to detect the sequence similarities in field cress genome. Conserved synteny regions between field cress and *Arabidopsis* were annotated as described in Gao *et al.* (2017).

3.3 QTL map analysis

Genotyping of 428 F_2 hybrids (two half-sib families) of field cress were implemented to generate the linkage map, while phenotyping on seven domestication traits was performed with the $F_{2:3}$ individuals in a field trial. These domestication traits were: plant height, inflorescence height, pod shattering, stem number, plant architecture, and leaf morphology. To identify QTL associated in field cress genome, we employed the multiple QTL mapping (MQM) algorithm in MapQTL version 6 software (Van Ooijen, 2009). Both the genotyping and linkage map construction was accomplished as described above in linkage map section.

3.4 Genome-wide association study (GWAS)

A total of 61 ecotypes (including three samples of *L. heterophyllum*) of field cress, most of which were from Sweden, were phenotyped across three sites; i.e., Alnarp in southern Sweden, Lanna in western Sweden, and Uppsala in central Sweden. Each of the ecotypes were planted with four replications across these sites. Phenotyping for agronomic (seed yield, kilogram per hectare, kgha^{-1}), economic-valued (seed oil and protein contents, percent, %), and physiological (seed moisture content, %) traits was undertaken. Following the described methods, the seed moisture (ISO665:, 2000), oil (Troëng, 1955), and protein (ISO/TS_16634_2:, 2009) contents were analysed using high-throughput infra-red spectroscopic technique, while phenotyping of seed yield was implemented conventionally.

The ecotypes of field cress were genotyped using 1517 SNPs. The association of trait and marker loci was performed as illustrated in Yu *et al.* (2006) using mixed linear model (MLM) equation (Figure 4):

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{S\alpha} + \mathbf{Qv} + \mathbf{Zu} + \mathbf{e}$$

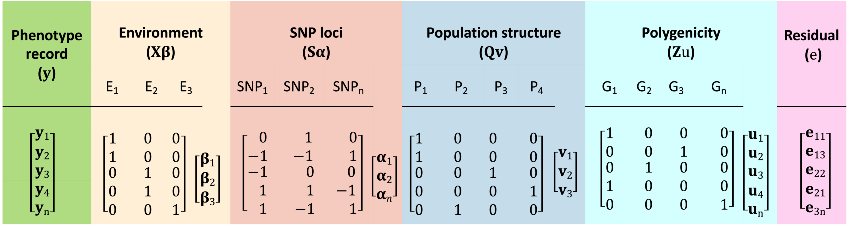

Figure 4. The mixed linear model (MLM) employed for genome-wide association study (GWAS) in field cress. The **y** is the phenotypic record (e.g. seed yield, Kgha^{-1}), the E_1, E_2, and E_3 was corresponding to the three environments i.e., Alnarp, Uppsala, and Lanna, respectively. The SNP loci are the allelic dosages with 1, -1, and 0 representing the two homozygote (i.e., 1 and -1) and the heterozygote (i.e. designated as 0) genotypes, respectively. The 371 of the 1517 SNP loci were employed in field cress GWA study. The four subpopulations were represented $P_1 – P_4$ in the **Q** matrix. The **v** is the mean effects of subpopulations while the G_1 to G_n in the **Z** matrix were random genomic background indicating each of the ecotypes in the field cress, and the **e** is the residual effect of the association mapping experiment.

where ***y*** is a vector of the observed phenotypic records (e.g. seed yield in Kgha^{-1}). The **X**, **S**, and **Z** represent incidence matrices for the environment, allelic dosage, and polygenic background that link the **y** to the vectors of environmental effects (**β**), SNP effects (**α**), and polygenic background effects (**u**)**,** respectively. The subpopulations are designated in **Q** matrix mapping the **y** to the vector of population effects (**v**). In this study, the principal component analysis (PCA)-based stratification was implemented. The values of **Q** matrix vary depending on the employed approaches; i.e., either with the PCA results (Price *et al.*, 2006) or STRUCTURE (Pritchard *et al.*, 2003), which groups individuals based on the likelihood inferences on multilocus genotypes.

The mixed model (Henderson, 1984) combines the components $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, and $\mathbf{Q}$ as fixed effects and the $\mathbf{u}$ as random effects corresponding to the best linear unbiased estimates (BLUE) and the best linear unbiased predictors (BLUP), respectively. In this analysis, the environment, population structure, SNP effects were treated as fixed effects while the kinship, i.e. the covariance between the ecotypes, was considered as random effects. The variance of the random effect $\mathbf{u}$ is $Var(\mathbf{u}) = \delta_{\mathrm{g}}^2\mathbf{K}$; where δ_g^2 is the genetic variance and $\mathbf{K}$ is the kinship matrix inferred from ridge regression best linear unbiased prediction (rrBLUP) package (Endelman & Jannink, 2012). The variance for the random effect $\mathbf{e}$ is $\mathrm{Var}(\mathbf{e}) = \delta_{\mathrm{e}}^2\mathbf{I}$, where δ_{e}^2 is the residual variance and $\mathbf{I}$ is an identity matrix.

To select the significantly associated variants, the false discovery rate (FDR) of the q −values (i.e., the adjusted p −values) at $q < 0.05$ (Storey & Tibshirani, 2003) and the unadjusted p −values at $p < 0.01$ were considered. The extent of LD , r^2, across the genome of field cress was estimated with the ‘genetics’ package (Warnes *et al.*, 2013). A plot for pairwise r^2 between SNPs within LG map was implemented using ‘LDheatmap’ package (Shin *et al.*, 2006). The LD plot was visualized in ‘Synbreed’ package (Wimmer *et al.*, 2012). The significantly associated common variants underlying the complex traits were visualized with Manhattan plot in ‘qqman’ package (Turner, 2014). All the analyses described including the MLM were implemented in R version 3.6.3 software package (R Core Team, 2019).

4. Results and discussions

4.1 Genetic linkage map construction (paper I)

As the mapping population consisted of two half-sib subpopulations, first the commonly segregating loci in both subpopulations were analysed. Next, these results were combined with the uniquely segregating loci, generated a total of 2016 loci (Figure 3). The final linkage analysis using genome-wide SNP arrays resulted in 1517 loci, distributed across the eight LGs of field cress (Figure 3; Figure 5a). The total map length was 566.07 cM, with a maximum of 115.90 cM in LG7 (Figure 6). The number of conserved synteny and skewed segregation was 866 and 483 of the 1517 loci in field cress (Figure 6), respectively. In addition to ascertain mapping quality (Zuo *et al.*, 2019), the segregation distortions have biological significance, for instance in QTL mapping (Xu, 2008); however, not all skewed segregations are employed to improve quality in linkage mapping.

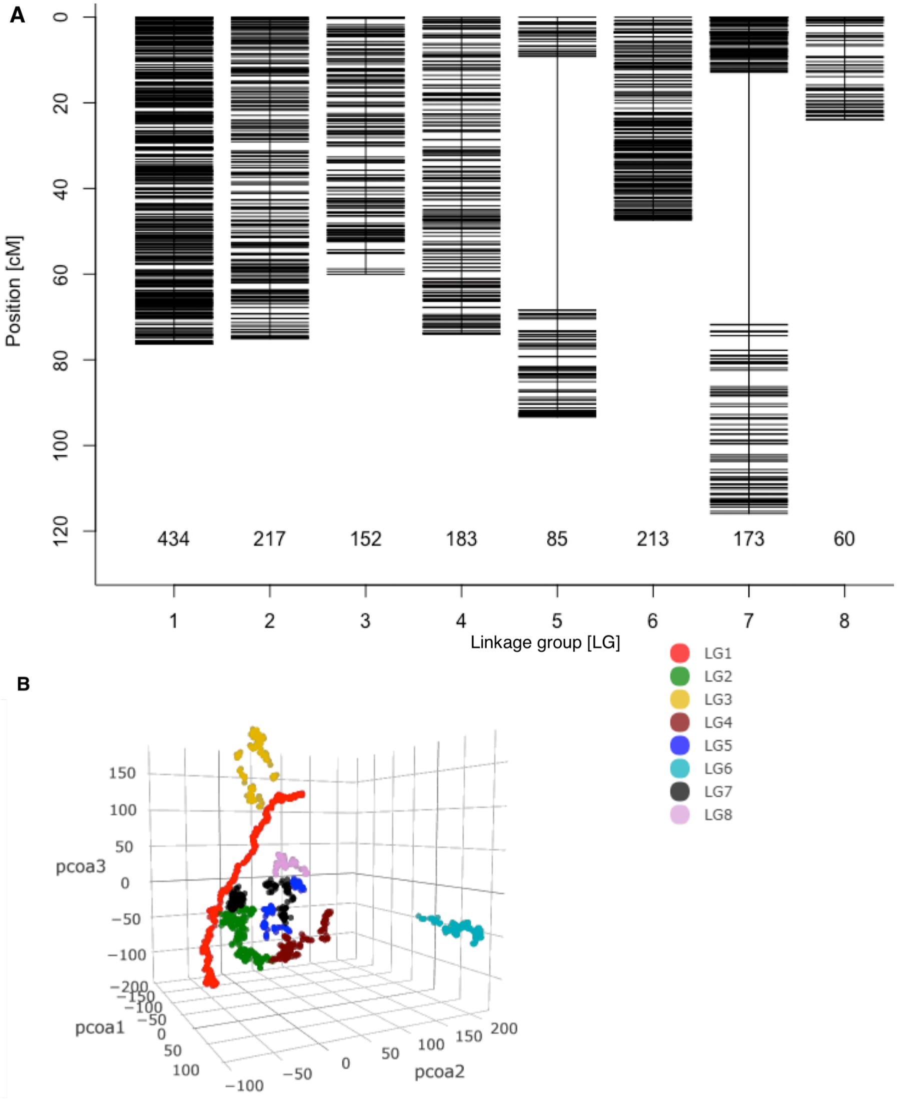

Figure 5. Linkage map construction in field cress. A. The distribution of loci across the eight linkage groups (LGs) of field cress. Numbers above the x-axis indicate the number of markers in each LG. B. Visualizing the pattern and distribution of loci across the eight LGs of field cress using 3D principal coordinate analysis (PCoA).

To reveal the pattern and distribution of loci within and between LGs, we used the 3D PCoA (Figure 5b) as it assists to visualize spuriously associated loci in linkage mapping (Farré *et al.*, 2011). Similarly, along with the distribution of loci, the split gaps generated in LG5 and LG7 were clearly visualized (Figure 5a; Figure 5b). However, the fragmentation of LGs was not because of the occurrence of independent LGs. Instead, as suggested with additional cytogenetic analysis (Desta *et al.*, 2019), the split regions or the heterochromatic sites existed in both field cress and *L. hetereophyllum* ancestor, and these gaps have been continued to exist after their evolutionary divergence these species. Nevertheless, developing reference sequence and future coordinated research in field cress are very important to fully resolve the issues of fragmentation in LG5 and LG7.

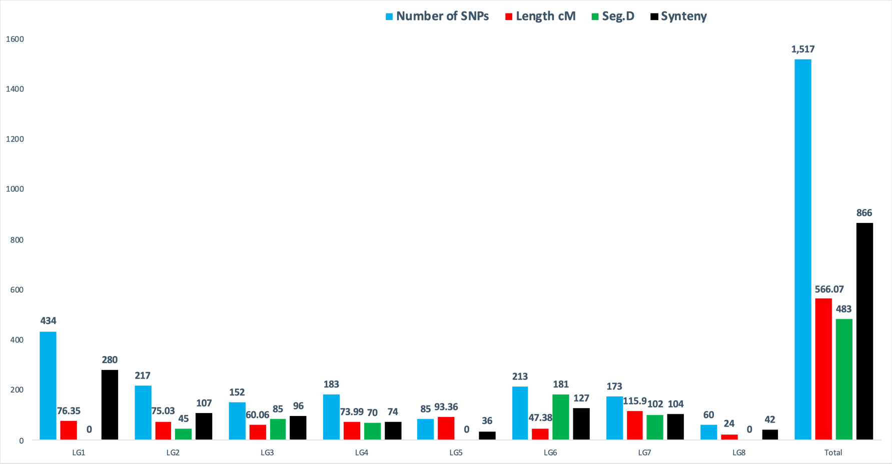

Figure 6. The overall information on the linkage mapping. This information includes the number of markers, linkage group (LG) map length (*cM*), segregation distortion (Seg.D), and conserved synteny in field cress.

4.2 Identifying domestication QTL in field cress (paper II)

The linkage map developed from 428 F_2 individuals were combined with the seven domestication traits – phenotyped in $F_{2:3}$ segregating population– to identify the underpinning QTL in field cress. Associated QTL were identified in all LGs, totalling 27 QTL (Figure 7). Among the domestication traits and LGs, the least number of QTL was recorded in stem number (one QTL) and in LG4 (one QTL).

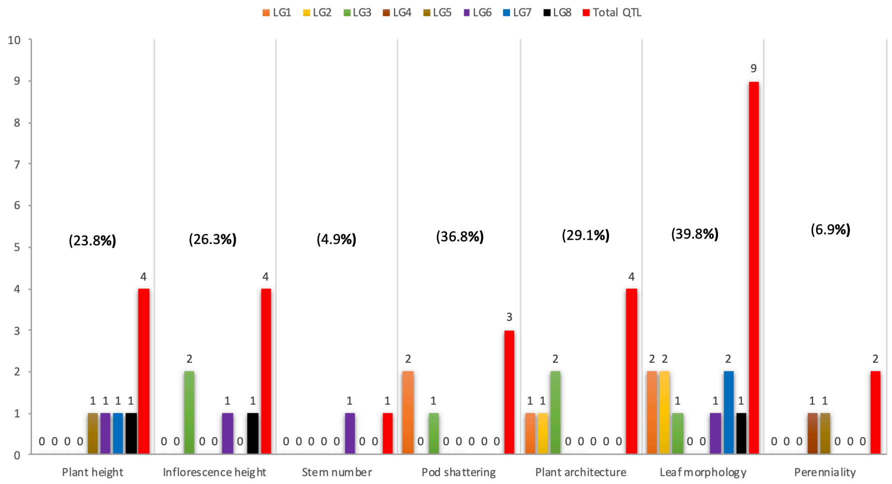

Figure 7. Identified QTL underlying domestication traits of field cress per linkage group (LG). The percentages inside parenthesis indicate the total phenotypic variation contributed by each of domestication traits in field cress.

The joint research on genetic and archaeobotany provides insights on the processes and timing of domestication in plants. The archaeobotany evidence shows that evolutionary rate of indehiscent seeds in cereals is slower than seed size (Fuller, 2007). In these crops, the evolution via domestication could take thousands of years until it gets fixed (Purugganan & Fuller, 2011; Purugganan & Fuller, 2009). Despite the gradual fixation of alleles in traits such as seed shattering, identifying and integrating underlying genes have been successfully achieved, e.g. the panicle spreading (*SPR3*) locus (Ishii *et al.*, 2013) and that for seed awning in rice (Bessho-Uehara *et al.*, 2016), and a Q major resistant gene to shattering in wheat (Simons *et al.*, 2006).

The *Btr1* and *Btr2* controlling brittleness in rachis (Pourkheirandish *et al.*, 2015) and *Nud* regulating free-threshing (the separation of chaffs from seeds) after harvest (Taketa *et al.*, 2008) are also typical examples of domestication genes that contribute in grain yield improvement of barley. Similar evidence in non-brittle rachis (*shp1* and *shp2* genes) has been illustrated in buckwheat (Matsui *et al.*, 2004; Matsui *et al.*, 2003). With the essence of translating these results altogether, further efforts of cloning and characterizing genes underlying domestication QTL could enhance not only the process of domestication, but also contribute to integrate promising alleles in field cress improvement.

4.3 Identifying variants using genome-wide association mapping in field cress (paper III)

The high-throughput phenotyping (for seed oil, protein, and moisture contents), as well as conventional phenotyping (for seed yield) was accomplished in the ecotypes of field cress that were planted and evaluated in Sweden across three agro-ecological sites. The half-decay distance – that is the distance at which the average pairwise squared correlation (r^2) is half of its maximum value – for the whole-genome in field cress was $\sim 37.8\ cM$, corresponding to a LD (r^2) of ~ 0.176. The patterns of LD within LG2 are visualized in Figure 8.

The nature and extent of LD in the study population has various evolutionary and agro-ecological benefits but not limited to: (1) determine the genetic drift or population bottleneck, (2) know the mapping resolution and experimental design of a population (Flint-Garcia *et al.*, 2003), (3) identify associated loci with a trait of interest, (4) detect gene mutation and signatures of selection (Kim & Nielsen, 2004) in study population, (5) define the population structure (Rossi *et al.*, 2009), and (6) shape and organize genetic variation in species population.

We found long stretch of LD estimate in field cress genome. As the extended segment of LD contains many linked genes, it is essential to shorten this into small pieces of LD while capitalizing both recent and ancient recombination events (Mohammadi *et al.*, 2020; Ersoz *et al.*, 2007) in field cress. To implement this, use of the nested association mapping (NAM) (Brachi *et al.*, 2011; Tian *et al.*, 2011) or multi-parent advanced generation inter-cross (MAGIC) (Huang *et al.*, 2015) population in field cress GWAS is very important. Despite the differences in the details of applications and benefits, the use of either NAM or MAGIC starts with proper selection of founder lines, followed by intermating and selfing to develop inbred lines, could potentially increase the effective recombination by reshuffling specific regions of the genome that was originally in LD (Brachi *et al.*, 2011) with a trait of interest.

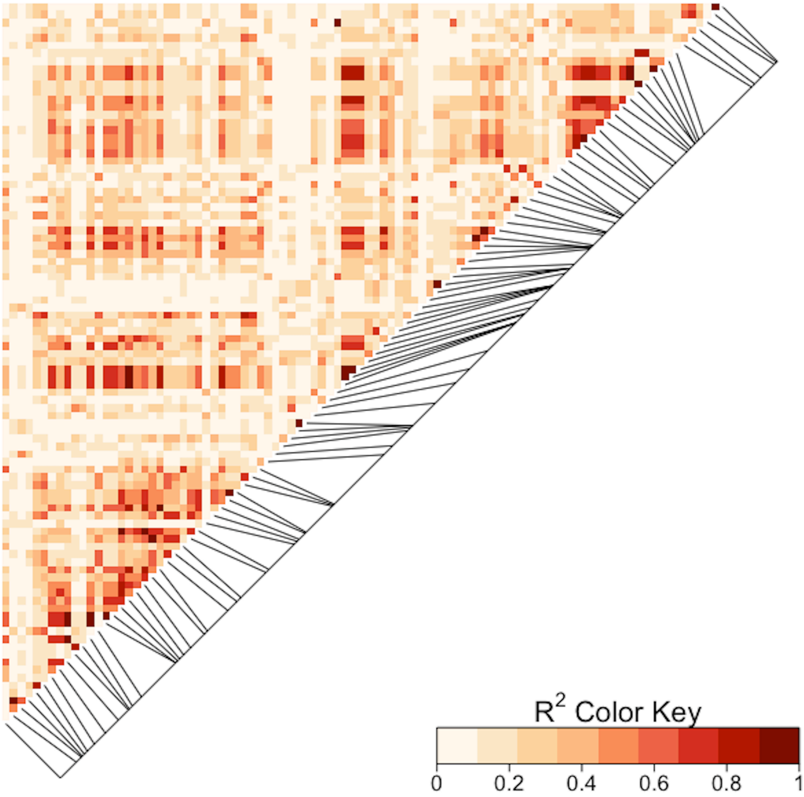

Figure 8. The LD plot in linkage group (LG)2. The pattern of LD among haplotypes of LG2 is expressed by the differences in the patterns of the colour from white (the lowest LD, i.e., $r^2 = 0$) to deeply intense brown colour (the highest LD i.e., $r^2 = 1$).

The choice between NAM and MAGIC population depends on different factors, mainly on time and cost to develop the inbred population as well as the objective of the research. For example, to detect epistasis effects as well as to eliminate the influence of population structure in association mapping, MAGIC seems more desirable than NAM population (Holland, 2015). Moreover, MAGIC population is promising to shorten the extended chunks of LD in inbreeding species (Cockram & Mackay, 2018). Nevertheless, compared to NAM population, MAGIC approach utilizes larger population alongside the multiple crossings that are technically demanding and takes longer time (Arrones *et al.*, 2020; Huang *et al.*, 2015) to develop the inbred population. Furthermore, as various alleles are crossed and recombined, MAGIC population could introduce more bias and uncertainty relative to NAM population (Cockram & Mackay, 2018) when further gene cloning steps are applied.

GWAS (for seed yield along with seed oil, protein, and moisture contents) generated 13 associated variant loci across the LGs in field cress (Table 1). Unlike to linkage study, almost impossible to handle small size population (Risch & Merikangas, 1996), GWAS was able to identify associated QTL despite the small sample size and coarse mapping resolution (371 of the 1517 SNPs) in field cress. Similar GWA analyses were achieved with small population size such as in *Arabidopsis* (Atwell *et al.*, 2010) or in wheat (Bellucci *et al.*, 2015). Nevertheless, to detect the modest size effects of highly complex quantitative traits especially with the use of whole-genome association scans, a larger sample size is definitely crucial.

Table 1. Identified number of loci for seed yield as well as for seed moisture, oil, and protein contents of field cress.

Trait	Number of loci	Linkage group	Total R^2 (%)
Moisture content	5	2, 3, 6	35.13
Yield	5	2, 4, 6, 7	4.28
Oil content	1	6	5.80
Protein content	2	2	0.12
Total	13	-	-

The use of naïve (without controlling the population structure) linear model led to increasing the number of associated variants for seed yield GWAS (Figure 9a-b). In line with this, the use of kinship without considering the population structure delivered the same results of associating variants in field cress (Figure 9e-f), thus indicating that the covariance between ecotypes could have less significance (Astle & Balding, 2009) when population stratification is neglected. The inappropriate use of subpopulations (Figure 9c-d) generated as equivalent associated variants as in using naïve linear model (Figure 9a-b). In other words, the high inflations were mainly either because of the disregard of population structure or because of imprecise use of population stratification in field cress seed yield GWA analysis.

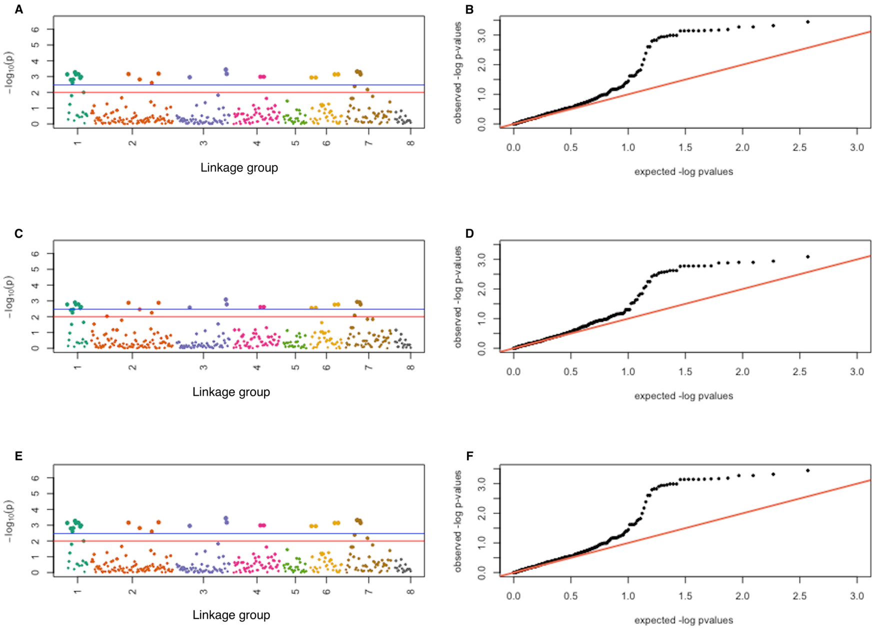

Figure 9. The Manhattan and QQ-plots for seed yield in field cress GWA analysis using polymorphic SNP loci. A–B. The naïve mixed linear model, which excludes both the population structure and kinship parameters, its corresponding QQ plot. C–D. A model considering two subpopulations in the GWA analysis, and its QQ plot. E–F. A model considering kinship alone in the GWA analysis. The horizontal red lines show the significant threshold level at $p < 0.01$ (unadjusted p −value), whereas the horizontal blue lines indicate the q − value (i.e., adjusted p −value) at p < 0.05.

The results in Figure 9 were reversed (i.e., reduced the association results) when considering proper subpopulations (Figure 10a-d), suggesting that population structure is a huge contributor of genetic variation (Brachi *et al.*, 2011; Ersoz *et al.*, 2007) for seed yield GWAS in field cress. After the proper adjustment of population structure (Figure 10a-c), none of the loci reached or surpassed the q − value (adjusted p − value) threshold level, instead significantly associated loci were detected while considering the unadjusted p − values. Correcting the population structure has weakened the power of association, thus likely to introduce false negatives (Astle & Balding, 2009). The use of proper stratification alone or in combination with a kinship delivered the same results (Figure 10a-d), demonstrating that the contribution of kinship was insignificant compared to controlling population structure in seed yield GWAS.

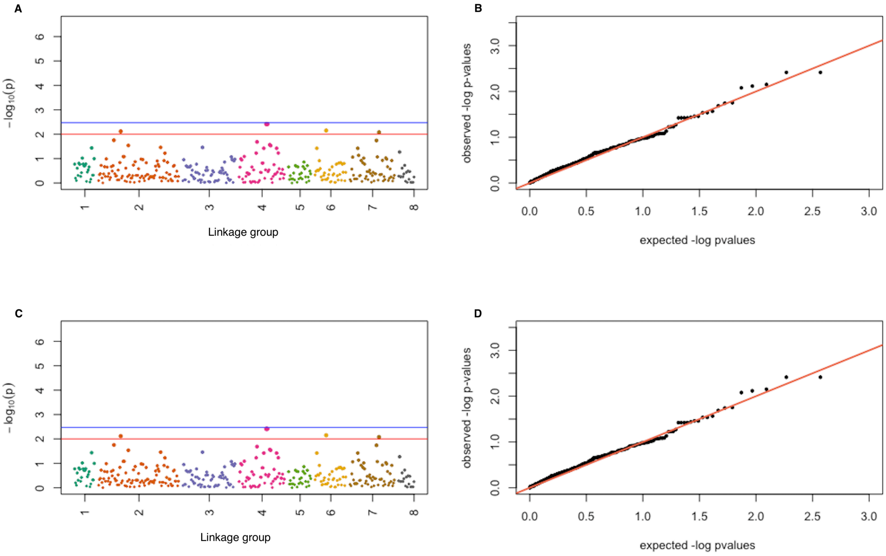

Figure 10. The Manhattan and QQ-plots for genome wide association study (GWAS) for seed yield in field cress using polymorphic SNP loci. A–B. GWAS model considering four population groups, and its corresponding QQ plot. C–D. GWAS considering both four subpopulation groups and kinship in the model, as well as the QQ plot. The fitness qualities were assessed using QQ plots in A and C. The horizontal red and blue lines in the Manhattan plots are used to indicate the significant threshold levels at $p < 0.01$ (unadjusted $p -$ value) and $q -$ value (adjusted $p -$ value), respectively.

The use of all loci – 1517 SNP loci– in field cress led to violate the assumption of the parametric analysis in the null hypothesis testing (Figure 11b), revealing that GWA approach is unable to recognize the rare and low frequency variants (Gibson, 2012; Brachi *et al.*, 2011). Consequently, these loci can mislead the final interpretation, thus underpowering the association mapping experiment in plants.

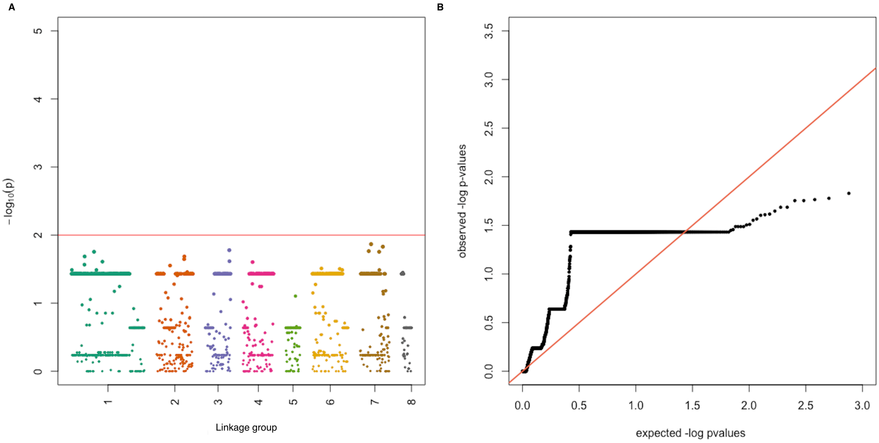

Figure 11. The Manhattan and QQ-plots for genome wide association study (GWAS) using 1517 SNP loci with naïve linear model. The polymorphic (371 loci) and non-polymorphic (1446 loci) loci were used simultaneously. A. Manhattan plot of the $-\log_{10}{}^{(\text{p-value})}$ in the y-axis against the linkage groups of field cress. The horizontal red line is the significant threshold value for the unadjusted p − values. B. The QQ plot of the expected versus the observed $-\log_{10}{}^{(\text{p-value})}$.

Despite genotyping with 1517 SNP markers, 1454 of the 1517 SNPs were excluded from further use by GWAS as these loci were low MAF (minor allele frequencies). Higher proportion of low MAF is not uncommon in plants, for example, over 40% contributed to MAF < 0.05 in the rice genome (Huang & Han, 2014). The presence of these SNP with low MAF does not, however, show their irrelevance to genetic analysis, instead they need very large sample size or require other proper approaches to have substantial power to detect the modest effect sizes of complex quantitative traits. Nevertheless, low allele frequency and rare variants can also be caused by genotyping error calls (Yan *et al.*, 2016) .

In spite of using broader sample size, these low MAF and rare variants may not be identified as these are only found in a specific population (Gibson, 2012; Brachi *et al.*, 2011). It is, therefore, important to adopt other complementary mechanisms of detecting variation controlling rare variants. Despite the ongoing progresses on the low frequency and rare variants, emphasizing on assaying the exome regions sheds light on capturing variation as demonstrated in human polygenic risk studies such as GWAS in hypertension analysis (Surendran *et al.*, 2016). Furthermore, direct sequencing of rare individuals separately (Gibson, 2012), and the use of GWAS with whole-exome sequencing or WGS (in place of SNP arrays) (Tam *et al.*, 2019) have been proposed as systematic approaches to ascertain the hidden variations underlying low MAF and rare loci. Additionally, switching from GWAS to GS may also alleviate the challenges associated with low frequencies loci in field cress.

5. Conclusion

Genetic linkage map is a benchmark tool that is not only assisting the location of crucial genes but it is also instrumental in guiding the development of reference sequence panels. Genetic linkage map was constructed using judicious combination of linkage with comparative and cytogenetic maps in field cress. Remarkably, the fragmented LGs (i.e., in LG5 and LG7) established in linkage mapping (Figure 5; Figure 6) were revealed by the less-error prone cytogenetic mapping, suggesting that these split LGs were part of their original haploids rather than separated LGs (Desta *et al.*, 2019). The complementarity of one over the other technique bridges the gap of ambiguity, particularly in the fragmented LGs, thus promising to develop final LG maps despite the bias due to the lack of reference sequence in field cress.

Dissecting the variance of quantitative traits and studying the underlying genetic mechanisms of domestication traits could play vital roles to facilitate field cress domestication and breeding. In response to genome-wide linkage study, 27 domestication QTL were estimated across all LGs in field cress (Figure 7). In this analysis, major and minor QTL effects were identified in all domestication traits except for both stem number and perenniality, which showed only QTL with small effects. With the aid of further evaluation and validation protocols (e.g. mutation-based screening or functional genetics approaches), the identified QTL could illuminate the process of domestication, as well as introgressing functional alleles into in field cress breeding programs.

GWAS with key agronomic, economic importance, and physiological traits – seed yield, as well as seed oil, protein, and moisture contents – in field cress identified 13 significantly ($p < 0.01$) associated variants across LGs of field cress (Table 1). Although small sample population size was employed, candidate variants were detected with all the traits presumably due to the high LD estimated in field cress genome. It is important to note that these identified variants were detected based on unadjusted $p -$value that is $p < 0.01$ (Figure

9; Figure 10), and these were unable to reach the adjusted p −value (the same as using q −value) at p −0.05. This was primarily owing to the small population size in seed yield GWA analysis. Possibly, ecotype seeds from different locations may also incorporate genetic heterogeneity (Brachi *et al.*, 2011), thus underpowering the association mapping experiment in plants.

Controlling the confounding effects of population stratification alone or in joint with kinship reduced the power of GWAS for seed yield in the field cress (Figure 10a–d). Furthermore, improper assigning of the subpopulations was as worse as neglecting the confounding effects of population stratification for seed yield GWAS in the field cress (Figure 9a–d). Considering the genome-wide SNPs, sizeable proportion of field cress genome is composed of low MAF and rare loci in field cress. It is also worth noting that the current GWAS procedure is unable to handle these lower MAF loci. Hence, boosting the sample size along with the use of other approaches (e.g. genomic selection) are central for accelerating both the domestication and genomics-assisted breeding program in field cress. In moving from association of common variants to function (i.e. toward genetic mechanisms) with post-GWAS pipelines, the associated loci could provide clues to find the target genes or transcripts underlying key traits in field cress.

6. Outlook

Constructing the linkage genetic map, identifying domestication QTL, and the associated single variants with the major economic, agronomic, and physiological traits are the milestones that could provide substantial insights for the domestication and breeding of field cress. The identified QTL and common variants in various traits of field cress become viable if further verification efforts (e.g. map-based cloning or mutation-based tools) are continued to validate and isolate the promising genes underpinning quantitative trait variants. The actual domestication of field cress can be facilitated if the expected obstacles and challenges are resolved on time. Determinants such as lack of germinability, vernalisation and flowering time, as well as issues related to over mineralization in field cress were described in Desta *et al.* (2020) as potential barriers that could hinder the process of domestication in field cress.

The current seed oil content is very low (~19%) compare to the other oil producing species. Strengthening the efforts in searching high oil content accessions followed by intermating among different oil producing lines to advance the oil content in field cress is an alternative to overcome this challenge. In parallel to this, functional genome editing tools (e.g. CRISPR-based approaches) could complement the ecotype-based oil improvement strategy in field cress. The challenges in oil composition of field cress was also noted previously, particularly the health risky erucic acid and the heat unstable linolenic acid (Andersson *et al.*, 1999; Nilsson *et al.*, 1998). Reducing both erucic acid and linolenic acid while increasing oleic acid using a genetic engineering approach was implemented in field cress (Ivarson *et al.*, 2016). Evidently, improving the oil content and composition using cutting-edge genome editing tools is highly likely to be achieved in field cress.

The members of Brassicaceae, including field cress, have constituted glucosinolate in their plant parts (Hopkins *et al.*, 2009). The detail analysis in glucosinolate is beyond the scope of this thesis. It is, however, one of the main concerns in field cress domestication. The presence of glucosinolate has advantages as it helps to defend against plant pathogens and pests (Hopkins *et al.*, 2009; Kim & Jander, 2007), as well as assists to thrive best in unfavourable environmental condition such as salt affected soils or extremely dryland environment (del Carmen Martínez-Ballesta *et al.*, 2013).

On the other hand, when glucosinolate is transported from shoot or root to seeds, it has demerits because the seeds become unpleasant for food consumption (Bisht & Augustine, 2019), nor is suitable as feed intake (Andersson *et al.*, 1999). In addition to finding naturally low glucosinolate ecotypes, the use of functional genomic approaches either to knockout (thus blocking) the transport of glucosinolate from other plant parts (the shoots or the roots) to the seeds or to reduce the accumulation of glucosinolate in the seeds is an important task of field cress domestication and breeding.

Detail plant phenotyping using conventional methods lack precision that misinform underlying causal variants (Mir *et al.*, 2019). It is, therefore, essential to develop various standard protocols to integrate the work of phenotyping across diverse environmental sites in field cress using a high-throughput phenotyping technique. The challenges and hurdles mentioned above are not impossible tasks to achieve, particularly with the current declining cost of WGS. Nevertheless, finding solution in each of these limitations could be reliable if the whole-genome reference panel in field cress is made available first (Desta *et al.*, 2020).

References

Abdurakhmonov, I.Y. & Abdukarimov, A. (2008). Application of association mapping to understanding the genetic diversity of plant germplasm resources. *International journal of plant genomics,* 2008.

Ahmadizadeh, M., Vispo, N.A., Calapit-Palao, C.D.O., Pangaan, I.D., Viña, C.D. & Singh, R.K. (2016). Reproductive stage salinity tolerance in rice: a complex trait to phenotype. *Indian Journal of Plant Physiology,* 21(4), pp. 528-536.

Andersson, A.A., Merker, A., Nilsson, P., Sørensen, H. & Åman, P. (1999). Chemical composition of the potential new oilseed crops Barbarea vulgaris, Barbarea verna and Lepidium campestre. *Journal of the Science of Food and Agriculture,* 79(2), pp. 179-186.

Arrones, A., Vilanova, S., Plazas, M., Mangino, G., Pascual, L., Díez, M.J., Prohens, J. & Gramazio, P. (2020). The Dawn of the Age of Multi-Parent MAGIC Populations in Plant Breeding: Novel Powerful Next-Generation Resources for Genetic Analysis and Selection of Recombinant Elite Material. *Biology,* 9(8), p. 229.

Astle, W. & Balding, D.J. (2009). Population structure and cryptic relatedness in genetic association studies. *Statistical Science,* 24(4), pp. 451-471.

Atwell, S., Huang, Y.S., Vilhjálmsson, B.J., Willems, G., Horton, M., Li, Y., Meng, D., Platt, A., Tarone, A.M. & Hu, T.T. (2010). Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. *Nature,* 465(7298), pp. 627-631.

Bandara, M., Savidov, N. & Driedger, D. The impact of selected abiotic stresses on glucoraphanin content in field pepperweed (Lepidium campestre L.). In: *Proceedings of II International Symposium on Human Health Effects of Fruits and Vegetables: FAVHEALTH 2007 841*2007, pp. 323-328.

Bellucci, A., Torp, A.M., Bruun, S., Magid, J., Andersen, S.B. & Rasmussen, S.K. (2015). Association mapping in scandinavian winter wheat for yield, plant height, and traits important for second-generation bioethanol production. *Frontiers in plant science,* 6, p. 1046.

Bentsink, L., Jowett, J., Hanhart, C.J. & Koornneef, M. (2006). Cloning of DOG1, a quantitative trait locus controlling seed dormancy in Arabidopsis. *Proceedings of the national academy of sciences,* 103(45), pp. 17042-17047.

Berger, B., de Regt, B. & Tester, M. (2012). High-throughput phenotyping of plant shoots. In: *High-throughput phenotyping in plants* Springer, pp. 9-20.

Bessho-Uehara, K., Wang, D.R., Furuta, T., Minami, A., Nagai, K., Gamuyao, R., Asano, K., Angeles-Shim, R.B., Shimizu, Y. & Ayano, M. (2016). Loss of function at RAE2, a previously unidentified EPFL, is required for awnlessness in cultivated Asian rice. *Proceedings of the national academy of sciences,* 113(32), pp. 8969-8974.

Bisht, N.C. & Augustine, R. (2019). Development of Brassica Oilseed Crops with Low Antinutritional Glucosinolates and Rich in Anticancer Glucosinolates. In: *Nutritional Quality Improvement in Plants* Springer, pp. 271-287.

Brachi, B., Morris, G.P. & Borevitz, J.O. (2011). Genome-wide association studies in plants: the missing heritability is in the field. *Genome biology,* 12(10), pp. 1-8.

Brill, S. & Dean, E. (1994). *Identifying and harvesting edible and medicinal plants in wild (and not so wild) places*: Hearst Books.

Cartwright, D.A., Troggio, M., Velasco, R. & Gutin, A. (2007). Genetic mapping in the presence of genotyping errors. *Genetics,* 176(4), pp. 2521-2527.

Cockram, J. & Mackay, I. (2018). Genetic mapping populations for conducting high-resolution trait mapping in plants. In: *Plant genetics and molecular biology* Springer, pp. 109-138.

Cole, B.J. & Chory, J. (2012). Image-based analysis of light-grown seedling hypocotyls in arabidopsis. In: *High-Throughput Phenotyping in Plants* Springer, pp. 1-7.

del Carmen Martínez-Ballesta, M., Moreno, D.A. & Carvajal, M. (2013). The physiological importance of glucosinolates on plant response to abiotic stress in Brassica. *International journal of molecular sciences,* 14(6), pp. 11607-11625.

Desta, Z.A., de Koning, D.-J. & Ortiz, R. (2020). Molecular mapping and identification of quantitative trait loci for domestication traits in the field cress (Lepidium campestre L.) genome. *Heredity,* 124(4), pp. 579-591.

Desta, Z.A., Kolano, B., Shamim, Z., Armstrong, S.J., Rewers, M., Sliwinska, E., Kushwaha, S.K., Parkin, I.A., Ortiz, R. & De Koning, D.-J. (2019). Field cress genome mapping: integrating linkage and comparative maps with cytogenetic analysis for rDNA carrying chromosomes. *Scientific reports,* 9(1), pp. 1-14.

Desta, Z.A. & Ortiz, R. (2014). Genomic selection: genome-wide prediction in plant improvement. *Trends in plant science,* 19(9), pp. 592-601.

Diamond, J. (2002). Evolution, consequences and future of plant and animal domestication. *Nature,* 418(6898), pp. 700-707.

Ecarnot, M., Bączyk, P., Tessarotto, L. & Chervin, C. (2013). Rapid phenotyping of the tomato fruit model, Micro-Tom, with a portable VIS–NIR spectrometer. *Plant physiology and biochemistry,* 70, pp. 159-163.

Edwards, D., Forster, J.W., Chagné, D. & Batley, J. (2007). What Are SNPs? In: *Association mapping in plants* Springer, pp. 41-52.

Edwards, S.L., Beesley, J., French, J.D. & Dunning, A.M. (2013). Beyond GWASs: illuminating the dark road from association to function. *The American Journal of Human Genetics,* 93(5), pp. 779-797.

Elias, T.S. & Dykeman, P.A. (2009). *Edible wild plants: a North American field guide to over 200 natural foods*: Sterling Publishing Company, Inc.

Endelman, J.B. & Jannink, J.-L. (2012). Shrinkage estimation of the realized relationship matrix. *G3: Genes, Genomes, Genetics,* 2(11), pp. 1405-1413.

Ersoz, E.S., Yu, J. & Buckler, E.S. (2007). Applications of linkage disequilibrium and association mapping in crop plants. In: *Genomics-assisted crop improvement* Springer, pp. 97-119.

Farré, A., Benito, I.L., Cistué, L., De Jong, J., Romagosa, I. & Jansen, J. (2011). Linkage map construction involving a reciprocal translocation. *Theoretical and applied genetics,* 122(5), pp. 1029-1037.

Fierst, J.L. (2015). Using linkage maps to correct and scaffold de novo genome assemblies: methods, challenges, and computational tools. *Frontiers in genetics,* 6, p. 220.

Flint-Garcia, S.A., Thornsberry, J.M. & Buckler IV, E.S. (2003). Structure of linkage disequilibrium in plants. *Annual review of plant biology,* 54(1), pp. 357-374.

Fuller, D.Q. (2007). Contrasting patterns in crop domestication and domestication rates: recent archaeobotanical insights from the Old World. *Annals of Botany,* 100(5), pp. 903-924.

Furbank, R.T. & Tester, M. (2011). Phenomics–technologies to relieve the phenotyping bottleneck. *Trends in plant science,* 16(12), pp. 635-644.

Gao, S., Yuan, Z., Li, N., Zhang, J. & Liu, Z. (2017). Determining Sequence Identities: BLAST, Phylogenetic Analysis, and Syntenic Analyses. In: Zhanjiang John Liu (ed.), Bioinformatics in Aquaculture: Principles and Methods. *Wiley Blackwell.*

Gibson, G. (2012). Rare and common variants: twenty arguments. *Nature Reviews Genetics,* 13(2), pp. 135-145.

Gilkey, H.M., Johnston, L. & Gilkey, H.M. (1967). Handbook of Northwestern plants.

Haldane, J. (1919). The combination of linkage values and the calculation of distances between the loci of linked factors. *J Genet,* 8(29), pp. 299-309.

Henderson, C.R. (1984). Applications of linear models in animal breeding models. *Univ Guelph.*

Holland, J.B. (2007). Genetic architecture of complex traits in plants. *Current opinion in plant biology,* 10(2), pp. 156-161.

Holland, J.B. (2015). MAGIC maize: a new resource for plant genetics. *Genome biology,* 16(1), p. 163.

Hopkins, R.J., van Dam, N.M. & van Loon, J.J. (2009). Role of glucosinolates in insect-plant relationships and multitrophic interactions. *Annual review of entomology,* 54.

Howarth, R.W. (2008). Coastal nitrogen pollution: a review of sources and trends globally and regionally. *Harmful algae,* 8(1), pp. 14-20.

Howell, E.C., Barker, G.C., Jones, G.H., Kearsey, M.J., King, G.J., Kop, E.P., Ryder, C.D., Teakle, G.R., Vicente, J.G. & Armstrong, S.J. (2002). Integration of the cytogenetic and genetic linkage maps of Brassica oleracea. *Genetics,* 161(3), pp. 1225-1234.

Hu, Q., Wang, H., Zaman, Q.U., Huang, W., Mei, D., Liu, J., Wang, W., Ding, B., Hao, M. & Fu, L. (2019). QTL and candidate genes identification for silique length based on high-dense genetic map in Brassica napus L. *Frontiers in plant science,* 10, p. 1579.

Huang, B.E., Verbyla, K.L., Verbyla, A.P., Raghavan, C., Singh, V.K., Gaur, P., Leung, H., Varshney, R.K. & Cavanagh, C.R. (2015). MAGIC populations in crops: current status and future prospects. *Theoretical and Applied Genetics,* 128(6), pp. 999-1017.

Huang, X. & Han, B. (2014). Natural variations and genome-wide association studies in crop plants. *Annual review of plant biology,* 65, pp. 531-551.

Ishii, T., Numaguchi, K., Miura, K., Yoshida, K., Thanh, P.T., Htun, T.M., Yamasaki, M., Komeda, N., Matsumoto, T. & Terauchi, R. (2013). OsLG1 regulates a closed panicle trait in domesticated rice. *Nature genetics,* 45(4), pp. 462-465.

ISO665: (2000). Oilseeds. Determination of moisture and volatile matter content.

ISO/TS_16634_2: (2009). Food products—Determination of the total nitrogen content by combustion according to the Dumas principle and calculation of the crude protein content—Part 2: Cereals, pulses and milled cereal products.

Ivarson, E., Ahlman, A., Lager, I. & Zhu, L.-H. (2016). Significant increase of oleic acid level in the wild species Lepidium campestre through direct gene silencing. *Plant cell reports,* 35(10), pp. 2055-2063.

Joukhadar, R., Daetwyler, H.D., Gendall, A.R. & Hayden, M.J. (2019). Artificial selection causes significant linkage disequilibrium among multiple unlinked genes in Australian wheat. *Evolutionary applications,* 12(8), pp. 1610-1625.

Kanter, D.R. (2018). Nitrogen pollution: a key building block for addressing climate change. *Climatic change,* 147(1-2), pp. 11-21.

Kim, J.H. & Jander, G. (2007). Myzus persicae (green peach aphid) feeding on Arabidopsis induces the formation of a deterrent indole glucosinolate. *The Plant Journal,* 49(6), pp. 1008-1019.

Kim, Y. & Nielsen, R. (2004). Linkage disequilibrium as a signature of selective sweeps. *Genetics,* 167(3), pp. 1513-1524.

Konishi, S., Izawa, T., Lin, S.Y., Ebana, K., Fukuta, Y., Sasaki, T. & Yano, M. (2006). An SNP caused loss of seed shattering during rice domestication. *Science,* 312(5778), pp. 1392-1396.

Li, H., Tsuchimoto, S., Harada, K. & Fukui, K. (2017). The Genome-Wide Association Study. In: *The Jatropha Genome* Springer, pp. 159-173.

Li, P., Fan, Y., Yin, S., Wang, Y., Wang, H., Xu, Y., Yang, Z. & Xu, C. (2020). Multi-environment QTL mapping of crown root traits in a maize RIL population. *The Crop Journal.*

Lian, L., Jacobson, A., Zhong, S. & Bernardo, R. (2014). Genomewide prediction accuracy within 969 maize biparental populations. *Crop Science,* 54(4), pp. 1514-1522.

Liu, J., Van Eck, J., Cong, B. & Tanksley, S.D. (2002). A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. *Proceedings of the national academy of sciences,* 99(20), pp. 13302-13306.

Lopes-Pinto, F., Vanhala, T., Geleta, M., Risse, J., Nichols, J., Karim Gharbi, K. & de Koning, D. RAD sequencing of diverse accessions of Lepidium campestre, a target species for domestication as a novel oil crop. In: *Proceedings of Plant and animal genome XXIV conference, plant and animal genome*, San Diego, CA, USA2016.

Matsui, K., Kiryu, Y., Komatsuda, T., Kurauchi, N., Ohtani, T. & Tetsuka, T. (2004). Identification of AFLP makers linked to non-seed shattering locus (sht1) in buckwheat and conversion to STS markers for marker-assisted selection. *Genome,* 47(3), pp. 469-474.

Matsui, K., Tetsuka, T. & Hara, T. (2003). Two independent gene loci controlling non-brittle pedicels in buckwheat. *Euphytica,* 134(2), pp. 203-208.

Merker, A. & Nilsson, P. (1995). Some oil crop properties in wild *Barbarea* and *Lepidium* species. *Swedish Journal of Agricultural Research*.

Meuwissen, T., Hayes, B. & Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics,* 157(4), pp. 1819-1829.

Mir, R.R., Reynolds, M., Pinto, F., Khan, M.A. & Bhat, M.A. (2019). High-throughput phenotyping for crop improvement in the genomics era. *Plant Science,* 282, pp. 60-72.

Mohammadi, M., Xavier, A., Beckett, T., Beyer, S., Chen, L., Chikssa, H., Cross, V., Moreira, F.F., French, E. & Gaire, R. (2020). Identification, Deployment, and Transferability of Quantitative Trait Loci from Genome-Wide Association Studies in Plants. *Current Plant Biology*, p. 100145.

Müller, B.S., Neves, L.G., de Almeida Filho, J.E., Resende, M.F., Muñoz, P.R., dos Santos, P.E., Paludzyszyn Filho, E., Kirst, M. & Grattapaglia, D. (2017). Genomic prediction in contrast to a genome-wide association study in explaining heritable variation of complex growth traits in breeding populations of Eucalyptus. *BMC genomics,* 18(1), p. 524.

Nilsson, P., Johansson, S.Å. & Merker, A. (1998). Variation in seed oil composition of species from the genera *Barbarea* and *Lepidium*. *Acta Agriculturae Scandinavica B—Plant Soil Sciences,* 48(3), pp. 159-164.

Pourkheirandish, M., Hensel, G., Kilian, B., Senthil, N., Chen, G., Sameri, M., Azhaguvel, P., Sakuma, S., Dhanagond, S. & Sharma, R. (2015). Evolution of the grain dispersal system in barley. *Cell,* 162(3), pp. 527-539.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics,* 38(8), pp. 904-909.

Pritchard, J.K., Wen, W. & Falush, D. (2003). Documentation for STRUCTURE software: Version 2.

Purugganan, M.D. & Fuller, D.Q. (2009). The nature of selection during plant domestication. *Nature,* 457(7231), pp. 843-448.

Purugganan, M.D. & Fuller, D.Q. (2011). Archaeological data reveal slow rates of evolution during plant domestication. *Evolution: International Journal of Organic Evolution,* 65(1), pp. 171-183.

R Core Team (2019). A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. *URL:(https://www.R-project.org)*.

Risch, N. & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science,* 273(5281), pp. 1516-1517.

Rossi, M., Bitocchi, E., Bellucci, E., Nanni, L., Rau, D., Attene, G. & Papa, R. (2009). Linkage disequilibrium and population structure in wild and domesticated populations of Phaseolus vulgaris L. *Evolutionary applications,* 2(4), pp. 504-522.

Saghai-Maroof, M.A., Soliman, K.M., Jorgensen, R.A. & Allard, R. (1984). Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proceedings of the national academy of sciences,* 81(24), pp. 8014-8018.

Sawitri, S., Tani, N., Na'iem, M., Widiyatno, W., Indrioko, S., Uchiyama, K., Suwa, R., Ng, K.K.S., Lee, S.L. & Tsumura, Y. (2020). Potential of Genome-Wide Association Studies and Genomic Selection to Improve Productivity and Quality of Commercial Timber Species in Tropical Rainforest, a Case Study of Shorea platyclados. *Forests,* 11(2), p. 239.

Seebeck, C.B. (1998). Best-tasting wild plants of Colorado and the Rockies.

Shin, J.-H., Blay, S., McNeney, B. & Graham, J. (2006). LDheatmap: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *Journal of Statistical Software,* 16(3), pp. 1-10.

Simons, K.J., Fellers, J.P., Trick, H.N., Zhang, Z., Tai, Y.-S., Gill, B.S. & Faris, J.D. (2006). Molecular characterization of the major wheat domestication gene Q. *Genetics,* 172(1), pp. 547-555.

Singh, V., van Oosterom, E.J., Jordan, D.R., Messina, C.D., Cooper, M. & Hammer, G.L. (2010). Morphological and architectural development of root systems in sorghum and maize. *Plant and Soil,* 333(1-2), pp. 287-299.

Spindel, J., Begum, H., Akdemir, D., Collard, B., Redoña, E., Jannink, J. & McCouch, S. (2016). Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement. *Heredity,* 116(4), pp. 395-408.

Storey, J.D. & Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the national academy of sciences,* 100(16), pp. 9440-9445.

Surendran, P., Drenos, F., Young, R., Warren, H., Cook, J.P., Manning, A.K., Grarup, N., Sim, X., Barnes, D.R. & Witkowska, K. (2016). Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nature genetics,* 48(10), pp. 1151-1161.

Taketa, S., Amano, S., Tsujino, Y., Sato, T., Saisho, D., Kakeda, K., Nomura, M., Suzuki, T., Matsumoto, T. & Sato, K. (2008). Barley grain with adhering hulls is controlled by an ERF family transcription factor gene regulating a lipid biosynthesis pathway. *Proceedings of the national academy of sciences,* 105(10), pp. 4062-4067.

Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G. & Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics,* 20(8), pp. 467-484.

Teh, C.-K., Sudirman, N.A., Rodzik, F.F.M., Ong, A.-L., Kwong, Q.-B. & Appleton, D.R. (2020). Genetic Dissecting Complex Traits via Conventional QTL Analysis and Association Mapping. In: *The Oil Palm Genome* Springer, pp. 131-140.

Tian, F., Bradbury, P.J., Brown, P.J., Hung, H., Sun, Q., Flint-Garcia, S., Rocheford, T.R., McMullen, M.D., Holland, J.B. & Buckler, E.S. (2011). Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nature genetics,* 43(2), pp. 159-162.

Troëng, S. (1955). Oil determination of oilseed. Gravimetric routine method. *Journal of the American Oil Chemists Society,* 32(3), pp. 124-126.

Troggio, M., Malacarne, G., Coppola, G., Segala, C., Cartwright, D.A., Pindo, M., Stefanini, M., Mank, R., Moroldo, M. & Morgante, M. (2007). A dense single-nucleotide polymorphism-based genetic linkage map of grapevine (Vitis vinifera L.) anchoring Pinot Noir bacterial artificial chromosome contigs. *Genetics,* 176(4), pp. 2637-2650.

Turner, S.D. (2014). qqman: an R package for visualizing GWAS results using QQ and manhattan plots. *Biorxiv*, p. 005165.

Van Ooijen, J. (2006). JoinMap 4. *Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen, Netherlands,* 33.

Van Ooijen, J. (2009). MapQTL 6. *Software for the mapping of quantitative trait loci in experimental populations of diploid species. Kyazma BV: Wageningen, Netherlands.*

Van Ooijen, J.W. & Jansen, J. (2013). *Genetic mapping in experimental populations*: Cambridge University Press.

Wang, C.-J.R., Harper, L. & Cande, W.Z. (2006). High-resolution single-copy gene fluorescence in situ hybridization and its use in the construction of a cytogenetic map of maize chromosome 9. *The Plant Cell,* 18(3), pp. 529-544.

Wang, H., Studer, A.J., Zhao, Q., Meeley, R. & Doebley, J.F. (2015). Evidence that the origin of naked kernels during maize domestication was caused by a single amino acid substitution in tga1. *Genetics,* 200(3), pp. 965-974.

Wang, Y.-H. (2011). Mapping and molecular breeding of monogenic traits. In: *Genetics, Genomics and Breeding of Cucurbits* CRC Press, pp. 247-259.

Wang, Z., Wang, S., Yu, C., Han, X. & Zou, D. (2020). QTL analysis of rice photosynthesis-related traits under the cold stress across multi-environments. *Euphytica,* 216(7), pp. 1-21.

Warnes, G., Gorjanc, G., Leisch, F. & Man, M. (2013). genetics: Population Genetics. R package version 1.3. 8.1. *The Comprehensive R Archive Network.*, pp. -.

Warwick, S., Francis, A. & Al-Shehbaz, I. (2006). Brassicaceae: species checklist and database on CD-Rom. *Plant Systematics and Evolution,* 259(2-4), pp. 249-258.

Werner, J.D., Borevitz, J.O., Uhlenhaut, N.H., Ecker, J.R., Chory, J. & Weigel, D. (2005). FRIGIDA-independent variation in flowering time of natural Arabidopsis thaliana accessions. *Genetics,* 170(3), pp. 1197-1207.

Wimmer, V., Albrecht, T., Auinger, H.-J. & Schön, C.-C. (2012). synbreed: a framework for the analysis of genomic prediction data using R. *Bioinformatics,* 28(15), pp. 2086-2087.

Xu, S. (2008). Quantitative trait locus mapping can benefit from segregation distortion. *Genetics,* 180(4), pp. 2201-2208.

Xu, Y., Wu, Y., Gonda, M.G. & Wu, J. (2015). A linkage based imputation method for missing SNP markers in association mapping. *Journal of Applied Bioinformatics & Computational Biology,* 4(1).

Yan, Q., Chen, R., Sutcliffe, J.S., Cook, E.H., Weeks, D.E., Li, B. & Chen, W. (2016). The impact of genotype calling errors on family-based studies. *Scientific reports,* 6(1), pp. 1-6.

Yazdanbakhsh, N. & Fisahn, J. (2012). High-throughput phenotyping of root growth dynamics. In: *High-Throughput Phenotyping in Plants* Springer, pp. 21-40.

Yu, J., Pressoir, G., Briggs, W.H., Bi, I.V., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M. & Holland, J.B. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature genetics,* 38(2), pp. 203-208.

Zuo, J.-F., Niu, Y., Cheng, P., Feng, J.-Y., Han, S.-F., Zhang, Y.-H., Shu, G., Wang, Y. & Zhang, Y.-M. (2019). Effect of marker segregation distortion on high density linkage map construction and QTL mapping in Soybean (*Glycine max* L.). *Heredity*, pp. 1-14.

Popular science summary

The potential oilseed field cress (*Lepidium campestre* L.) has a small genome size and is a self-pollinated plant that is endemic to Europe. However, the environmental, agricultural, and industrial benefits of field cress are not explored yet. One of the most likely options to mitigate the ecosystem associated problems is identifying and domesticating multipurpose species (e.g. field cress), thus alleviating the consequences of climate change in parallel with achieving the food and feed security. Planting field cress together with cereals in an intercropping system reduces the leaching sensitive soil minerals to underground water or the ocean, thus maintaining the soil nutrients in the plant-soil system. Normal recycling of the soil nutrient is not only useful for plant growth but also beneficial for the survival of soil microorganisms.

The main aim of this thesis study was identifying key genomic variations associated with domestication, economical-valued, agronomic, and physiological traits using genomic tools in field cress. To implement this, the genetic linkage map was initially developed that serves as a benchmark, by which all the succeeding researches rely on identifying various genomic variants associated with key traits of field cress. Following linkage map construction, genomic regions underlying seven key domestication traits (e.g. pod shattering) were estimated, across all chromosomes of field cress. Furthermore, the genome association analysis in four major traits of field cress – the seed oil, protein, and moisture contents as well as seed yield – were performed. In this association study, a total of 13 common variants were identified. The overall noted candidate variants using both genome-wide linkage and association studies could accelerate the process of domestication and genomics-assisted breeding in field cress.

Populärvetenskaplig sammanfattning

Den potentiella oljeväxt fältkrassing (*Lepidium campestre* L.) har en liten genom storlek och är en självpollinerande växt som är endemisk i Europa. Samtidig är de miljömässiga, jordbruksmässiga och industriella fördelarna med fältkrassing ännu inte utforskade. Ett av de mest troliga alternativen för att mildra de ekosystemrelaterade problemen är att identifiera och domesticera mångsidiga arter (t.ex. fältkrassing) och därmed lindra konsekvenserna av klimatförändringar parallellt med ökad livsmedels- och fodersäkerhet. Att plantera fältkrassing tillsammans med spannmål i ett samodlingssystem minskar läckage av jordmineralerna till underjordiskt vatten eller havet, och därmed bibehåller jordens näringscykel i jordsystemet. Normal återvinning av jordnäringsämnen är inte bara nyttig för växttillväxt utan är också fördelaktigt för jordmikroorganismernas överlevnad.

Huvudsyftet med denna avhandling var att identifiera viktiga genomregioner som är kopplade till domesticerings-, ekonomiska, agronomiska och fysiologiska egenskaper med hjälp av genomverktyg i fältkrassing. För att genomföra detta utvecklades i första hand den genetiska kopplingskartan som fungerar som en referenspunkt, genom vilket alla efterföljande undersökningar förlitar sig på att identifiera olika genomegenskaper. Efter konstruktion av kopplingskartor uppskattades genomiska regioner som ligger till grund för sju viktiga domesticeringsegenskaper (t.ex. pod-splittring) över alla kromosomer i fältkrassing. Vidare utfördes helgenomanalysen i fyra huvudegenskaper av fältkrassing – fröolja-, protein- och fuktinnehåll samt fröutbyte. I denna associeringsstudie hittades totalt 13 olika regioner. De övergripande identifierade kandidatvarianterna som hittades både genom genomkopplings- och associeringsstudier kan påskynda processen för domesticering och genomisk förädling i fältkrassing.

Acknowledgement

I would like to start acknowledging myself first, as this PhD work was not an ordinary program, nor was clearly planned and fitted to achieve the program. Instead, it was broad, complicated, and entangled, leaning toward multi-challenges that need thorough decisions including the risks to manage. I took different risky initiatives at different hierarchical levels. I wish I could have mentioned all these responsibilities I played roles in, but for the sake of space, let me raise the major few duties. After the late arrival of data in SNP genotyping (at the end of my PhD study), I decided the first and most challenging work of field cress linkage mapping by combining with the broad disciplines of the cytogenetic and comparative maps, mainly sending the seeds to cytogenetic experts to United Kingdom and Poland from my private pocket money. Despite the challenges, I strongly believe that the errors in using linkage mapping alone can be improved and bolstered when additional mapping approaches are employed. Furthermore, I took the initiatives on performing three different environments across Sweden (instead of using only one site), developing high-throughput phenotyping protocols in field cress, and handling the top-notch published journal articles including writing this thesis while I was sustaining with my own private finance for more than a year.

I would like to thank my marvellous supervisors, Rodomiro Ortiz (a main supervisor) and Dirk-Jan de Koning (DJ, a co-supervisor). I had been through lots of challenges and complexities throughout my studies, many of which were focal lessens to my future career; however, it would not have been possible to face and resolve these difficulties without the involvement of my supervisors. Their wise patience, the way they promote me to be more creative, creating working environment in non-conducive and repellent environment, their all-time focus on scientific discussions, and their inspirations were some of the unforgettable memory that are not only saved in my heart but also tests and exercises in my future endeavour on how to implement given these magnetic qualities. Definitely, I will miss both Rodomiro and DJ, and let me thank you so much once again!

I gratefully acknowledge the MISTRA-Biotech (as a part of MISTRA), and SLU for funding my research and study. Thank you so much for the follow-up and supports from Sven Ove Hansson and Anna Lehrman. I thank you also for the motivation and kind efforts from Gunter Backes, who was my previous MSc supervisor at University of Copenhagen, Denmark.

Thank you, Tomas Bryngelesson who was my main supervisor during the early phase of this PhD study, handling with great care and follow-up until he handed in to Rodomiro. Despite the unfortunate difficulties working together, this is the opportunity for me to say thank you to Mulatu Geleta for familiarizing the overall situation in Alnarp at the beginning of my PhD program, as well as for offering the newly collected ecotype seeds including the hybrids of *Lepidium*.

Susan Armstrong, Elwira Sliwinska, Isobel Parkin, Bozena Kolano, Zeeshan Shamim, and Monika Rewers, many thanks for all of you, and I really have shortage of words to express from the time you accepted my kind request until achieving the fantastic jobs to alleviate the challenges in the first linkage map construction in field cress. Graham King, Fikret Isik, and Jim Holland, who deserve the highest praise for viewing and forwarding their valuable comments.

I want to thank Anna Zborowska for her assistance in the lab work and kind support in DNA isolation of field cress plants. Thank you so much Cecilia Gustafsson, for your enormous help in DNA isolation as well as, sending the DNA samples to Edinburgh Genomics Institute. I also want to thank you Richard Talbot for familiarizing the Infinium iSelect genotyping platform.

Vebho Hot, who is a very collaborative, friendly and focussed person, and thank you so much for your tremendous achievements in both the greenhouse and field experiments. Thank you, Nils-Ove Bertholdsson for your enormous efforts in finding the experimental sites both at Uppsala and Lanna SLU branches, as well as for your technical assistance and providing practical support in the experimental sites. I also want to thank you Inger Åhman for visiting and helping in my experimental site, and emailing valuable articles related to my field of study. I would like to thank Waheeb Heneen for his generous advices and help.

It is my great pleasure to thank both Mahbub Rahmatov and Mohamed Ahmed Omer Elsafy for their enthusiastic follow-up and advices
from the beginning till the end of this PhD program. I am very grateful for the sincere assistance, and highly friendly Sandeep Kushwaha. I thank you so

much for my office mates both Alex Koc and Michael Vagiri for creating the environment so attractive and enjoyable to work together. Thank you, Mohammed (head of Nordgene bank) and Per Snell for your awesome collaboration and allowing me to use the seed weighing equipment. Many thanks to Staffan Andersson, Niklas Olsson, and Frederik Reslov for their help and creating friendly environment.

Thank you so much Camilla for your tremendous supports with different issues from the beginning until the end of this PhD program such as primula follow-up, introducing how the overall system works well, and familiarizing the Swedish culture. Ann-Sofie, thank you very much for your most generous greetings and remarkable collaboration. I acknowledge Helle Turesson for her collaboration and assistance whenever I requested technical guidance in using the lab instruments. I would like to thank both Helene Lindgren and Annelie Ahlman for their carful help and guidance as well as for providing lab working equipment including chemicals on time. Thank you for Emma Linnér, Gunilla Andersson and Gunilla Eriksson for their administrative help during my study. I also want to thank to all members of IT section for providing technical maintenance and supports whenever requested.

I have got unforgettable and highly tremendous supports in all experimental sites (Alnarp, Lanna, and Uppsala), thank you very sincerely for your kind help and collaboration from Alnarp: Erik Rasmusson, Alexandra Nikolic, Lucy Clark, and Fredrik Alfsson; from Uppsala: Ewa Magnuski and Jannie Hagman; and from Lanna: Johan Roland, Sofia Delin, Matias, and Johan Fredriksson.

I am very thankful for Johan W. Ooijen (Kyazma, Netherlands) for your invaluable feedback, advices, and encouragement while I was working with linkage and QTL map analyses. Special thanks to Harry Wu in particular and the plant breeding platform in general for handling the unique and rare experts in teaching and capacity building programs at SLU.

I would like to thank to my mentor and teacher, and most importantly a friend and a brother Dereje Assefa (Mekelle University, Ethiopia), who passed away suddenly, and may your soul rest in peace! I would like to thank gratefully for my friends who always motivate me throughout my study, particularly Araya Aleme, Yemane Tsehaye, Merid Alemayaho, Zahra Ademnur, Tsehaye Kahsay, Negash Araya, Molla Mengesha, Afewerki Kebede, Fikadu Reta, Wondem Gizaw, Hagos Desta, and Kahsay Gebrehiwot. Thank you for the wonderful help and motivation of work from Elnura Torutaeva, Hailu Faris, Sergey Hegay, Faraz Muneer, Dennis Eriksson, Firuz Odilbekov, Hassen, as well as current and previous PhD candidates.

I also want to thank for försakringskassan and Malmö hospital medical experts for their kind hospitality and support during my sick leave. I would like to thank the union members: Anna Iliar, Andrus Kangro, Karin Thorasdotter, and Robert Andersson for their enormous efforts to solve the hostile environment. I highly appreciate Rikard Westbom, Eva Nordqvist, Bo Gertsson, and all staff members of Lantmännen company for their kind support and allowing me to work in their lab with high-throughput phenotyping protocol in field cress.

Thank you so much Haileslassie Yohanes and his beloved wife and children, for considering me as part of their family, and hosting me all the time with priceless hospitality, pretending as if I was with my parents back home.

Particular Thanks for Helena Persson Hovmalm and Helene Larsson Jönsson for their generous efforts in resolving my inquiries. I want also to thank Cecilia Aslved and Lotta Malmborg for their immediate replying and answering every time to my requests. I acknowledge Christina Johansson for helping to find a very good place of accommodation in the vicinity of SLU campus. I would also to thank Gary Nilsson for helping me in printing this thesis book.

Last, but by no means least, I thank to all members of my family for their love and support as well as, for their long awaiting with infrequent communication for all these years. Special thanks to Abera Desta, Elsabet Gebreselassie, Abay Gebrehiwot, Etay Desta, Gebremeskel (Flasfa), Zenebech Abebe, Genet, Hanna, Meron, Andenet, Helen, Berhane, Anteneh, Siabel, Ashenafi (Aklilu), and Alula. Thank you Simret, Buzualen, Frehiwot, Messay, Adey, Rigbe, Alem, Saba, Mitye, Tigest, Asmeret, and Lemlem for your wise advices and supports throughout my studies. All remaining mistakes or oversights are of course mine.

Acta Universitatis agriculturae Sueciae

Doctoral Thesis No. 2020:59

Field cress, an underutilized and not domesticated plant yet, has multiple roles in both climate change mitigation and agricultural production. Exploring its potential using genomic and molecular tools delivered key associated genomic variants underlying the domestication, agronomic, physiological as well as economically important traits in field cress. These identified genomic variants could provide insightful guidance of discovering the underlying genes affecting various traits, thus accelerating the process of domestication and genomics-assisted breeding in field cress.

Zeratsion Abera Desta received his PhD from the Department of Plant Breeding, Swedish University of Agricultural Sciences, Sweden. He received his MSc from the Faculty of Life Sciences, University of Copenhagen, Denmark.

Acta Universitatis agriculturae Sueciae presents doctoral theses from the Swedish University of Agricultural Sciences (SLU).

SLU generates knowledge for the sustainable use of biological natural resources. Research, education, extension, as well as environmental monitoring and assessment are used to achieve this goal.

Online publication of thesis summary: https://pub.epsilon.slu.se

ISSN 1652-6880
ISBN (print version) 978-91-7760-636-9
ISBN (electronic version) 978-91-7760-637-6