

Aus dem Institut für Pflanzenbau und Pflanzenzüchtung
der Christian-Albrechts-Universität zu Kiel

**Unraveling genetic mechanisms of flowering time control and other
agronomically important traits in quinoa (*Chenopodium quinoa*
Willd.)**

Dissertation

zur Erlangung des Doktorgrades

der Agrar- und Ernährungswissenschaftlichen Fakultät

der Christian-Albrechts-Universität zu Kiel

vorgelegt von

M.Sc. Dilan Sarange Rajapaksha Patiranage

aus Walasmulla, Sri Lanka

Kiel, 2020

Dekan: Prof. Dr. Karl H. Mühling

1. Berichterstatter: Prof. Dr. Christian Jung

2. Berichterstatter: Prof. Dr. Georg Thaller

Tag der mündlichen Prüfung: 17. 02. 2021

Table of Contents

1	General introduction.....	1
1.1	Origin and agronomical importance of the pseudocereal crop quinoa (<i>Chenopodium quinoa</i> Willd.)	1
1.2	Genetics, genomics, and breeding of quinoa	2
1.3	Adaptation of quinoa to different geographical regions	7
1.4	Flowering time regulation in the model species <i>Arabidopsis</i>	9
1.5	Flowering time regulation in quinoa and related species of the Amaranthaceae plant family	10
1.6	Genome-wide association studies in crop plants	12
1.7	Aims, objectives, and scientific hypotheses	13
2	Genome-wide association study in the pseudocereal quinoa reveals selection pattern typical for crops with a short breeding history.....	14
2.1	Abstract.....	14
2.2	Introduction	14
2.3	Materials and Methods	15
2.3.1	Plant materials and growth conditions	15
2.3.2	Statistical analysis	16
2.3.3	Genome sequencing and identification of genomic variations	16
2.3.4	Phylogenetic analysis and population structure analysis	17
2.3.5	Genomic patterns of variations	18
2.3.6	Linkage disequilibrium analysis.....	18
2.3.7	Genome-wide association study.....	18
2.4	Results	19
2.4.1	Re-sequencing 310 quinoa accessions reveals high sequence variation	19
2.4.2	Linkage disequilibrium and population structure of the quinoa diversity panel.....	20
2.4.3	Genomic patterns of variations between Highland and Lowland quinoa	21

Table of contents

2.4.4	Mapping agronomically important trait loci in the quinoa genome.....	22
2.4.5	Candidate genes for agronomically important traits	23
2.5	Discussion	27
2.6	Acknowledgments	29
2.7	Supplementary data	29
2.7.1	Supplementary figures	29
2.7.2	Supplementary tables	31
2.8	References	31
3	Haplotype variations of major flowering time genes in quinoa unveil their role in the adaptation to different environmental conditions	37
3.1	Summary	37
3.2	Introduction	37
3.3	Results	39
3.3.1	Flowering time variation under different photoperiodic conditions	39
3.3.2	Expression analysis of putative <i>FT</i> and <i>COL</i> orthologs	39
3.3.3	Sequence variation in <i>FT</i> and <i>COL</i> orthologs are corresponding with geographical origin and flowering time.....	43
3.4	Discussion	48
3.5	Experimental procedures.....	51
3.5.1	Plant materials and growth conditions	51
3.5.2	<i>In-silico</i> identification of quinoa <i>FT</i> -like genes and <i>CONSTANS</i> like genes	51
3.5.3	Phylogenetic tree construction	51
3.5.4	RNA isolation and gene expression analysis	51
3.5.5	DNA sequencing and haplotype analysis.....	52
3.5.6	Statistical analysis	52
3.6	Acknowledgments	52
3.7	Supporting Information	53
3.8	References	54

Table of contents

4	Closing discussion.....	59
4.1	Variations in sequences and expression of quinoa <i>CO</i> and <i>FT</i> homologs influences flowering time and geographical adaptation	59
4.2	High marker density allows rapid identification of candidate genes for agronomically important traits	62
4.3	Perspectives for quinoa breeding by haplotype building.....	63
5	Summary	66
6	Zusammenfassung.....	68
7	Appendix	70
7.1	Supplementary tables.....	70
7.2	Supplementary figures	85
8	References	117
9	Supplementary data on CD/DVD.....	134
10	Curriculum Vitae and Publications	135
10.1.1	Curriculum vitae.....	135
10.1.2	Publications	135
11	Publications and declaration of own contribution.....	137
12	Acknowledgments.....	139

List of Abbreviations

°C	Celsius
π	Nucleotide diversity
%	Percent
AFLP	Amplified fragment length polymorphism
<i>AGL24</i>	<i>AGAMOUSLIKE24</i>
<i>API</i>	<i>APETALA1</i>
Asn	Asparagine
<i>At</i>	<i>Arabidopsis thaliana</i>
ATGC	Allotetraploid goosefoot complex
BBCH	Biologische Bundesanstalt, Bundessortenamt und CHEMISCHE INDUSTRIE
<i>BBX19</i>	<i>DOUBLE B-BOX TYPE ZINC FINGER 19</i>
BIC	Bayesian Information Criterion
BLUE	Best linear unbiased estimator
BLUP	Best linear unbiased predictor
<i>BTC1</i>	<i>BOLTING TIME CONTROL 1</i>
<i>Bv</i>	<i>Beta vulgaris</i>
BWA	Burrows-Wheeler Aligner
<i>C. ambrosoide</i>	<i>Chenopodium ambrosoide</i>
<i>C. berlandieri</i>	<i>Chenopodium berlandieri</i>
<i>C. hircinum</i>	<i>Chenopodium hircinum</i>
<i>C. pallidicaule</i>	<i>Chenopodium pallidicaule</i>
<i>C. quinoa</i>	<i>Chenopodium quinoa</i>
<i>C. suecicum</i>	<i>Chenopodium suecicum</i>
<i>C. album</i>	<i>Chenopodium album</i>
<i>C. ficifolium</i>	<i>Chenopodium ficifolium</i>
ca.	Circa
CCT	CONSTANS, CO-like, and TOC1
cm	Centi meters
<i>CO</i>	<i>CONSTANS</i>
<i>COL</i>	<i>CO</i> -like genes
CV	Cross-validation
<i>CYP76AD1</i>	<i>CYTOCHROME P76AD1</i>
DNA	Deoxyribonucleic acid
<i>DODA1</i>	<i>3,4-DIHYDROXYPHENYLALANINE 4,5-DIOXYGENASE</i>
DTB	Days to bolting
DTF	Days to flowering
DTM	Days to maturity
EMMAX	Efficient mixed-model association expedited
EMS	Ethyl methanesulfonate
EST	Expressed sequence tag
FAO	Food and Agriculture Organization
FAOSTAT	Food and Agriculture Organization Corporate Statistical Database
FKF1	FLAVIN KELCH F BOX 1

Table of contents

<i>FLC</i>	<i>FLOWERING LOCUS C</i>
F _n	n th generation after cross
<i>FRI</i>	<i>FRIGIDA</i>
F _{ST}	Fixation index
<i>FT</i>	<i>FLOWERING LOCUS T</i>
<i>FTL</i>	<i>FLOWERING LOCUS T-like</i>
<i>FUL</i>	<i>FRUITFULL</i>
GA	Gibberellic acids
GATK	Genome analysis tool kit
Gb	Giga basepair
GEC	Genetic type I Error Calculator
GI	GIGANTEA
<i>GID1</i>	<i>GIBBERELLIN INSENSITIVE DWARF1</i>
Gln	Glutamine
Glu	Glutamic acid
GWAS	Genome-wide association studies
HCSNP	High Confidence SNP
<i>HD1</i>	<i>Heading date 1</i>
InDel	Insertion and deletions
KASP	Kompetitive allele specific PCR
kb	Kilobase pair
LD	Long day/linkage disequilibrium
<i>LFY</i>	<i>LEAFY</i>
LTRs	long terminal repeats
m	Meter
MAF	Minor allele frequency
MAGIC	Multi-parent Advanced Generation Intercrosses
Mb	Mega basepair
miRNA	micro RNAs
ML	Maximum likelihood
mm	Millimeter
MTA	Marker-trait associations
NoB	Number of branches
PC	principal component
PCA	principal component analysis
PCA _(PHEN)	Phenotype based principle component analysis
PCA _(SNP)	SNP based principle component analysis
PD	Panicle density
PE	Paired-end
PEBP	Phosphatidyl ethanolamine-binding proteins
pg	Peco gram
PH	Plant height
PIF4	Phytochrome-Interacting Factor4

Table of contents

PL	Panicle length
<i>PRR</i>	<i>PSEUDO-RESPONSE REGULATOR</i>
RAPD	Random Amplified Polymorphic DNA
RNA	Ribonucleic acid
RT-qPCR	Real-time quantitative PCR
<i>SBPs/SPLs</i>	<i>SQUAMOSA PROMOTER-BINDING PROTEIN-LIKEs</i>
SD	Short day
<i>SEP3</i>	<i>SEPALLATA3</i>
SNP	Single Nucleotide Polymorphism
<i>SOCI</i>	<i>SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1</i>
<i>SPL</i>	<i>SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE</i>
SSD	Single-seed descent
SSR	Simple sequence repeat
STL	Stem lying
<i>TFL1</i>	<i>TERMINAL FLOWER1</i>
Trp	Tryptophan
TSW	Thousand seed weight
USDA	United States Department of Agriculture
VCF	Variant calling format

1 General introduction

1.1 Origin and agronomical importance of the pseudocereal crop quinoa (*Chenopodium quinoa* Willd.)

Quinoa is an important grain crop that has been grown for more than five millennia by the native South Americans. It is regarded as one of the world's most nutritious seed crops due to its outstanding protein composition and vitamin and mineral content. Quinoa has become a versatile crop as a result of its diversity and the ability to stand against numerous biotic and abiotic stresses. The form of consumption has many similarities with cereals. Therefore, quinoa is classified as a pseudocereal. The indigenous name for quinoa, "Mother Grain" implies its sacred place in the Incan civilization that stands next to maize (Bhargava and Srivastava 2013). Recent worldwide recognition of its potential for cultivation and consumption boosted the demand. Consequently, there is a newly opened market for producers and industry.

Chenopodium quinoa is an allotetraploid species ($2n=4x=36$), which belongs to the family Amaranthaceae. Some other economically important crops like sugar beet, red beet, and spinach belong to this family. Quinoa and spinach (*Spinacia oleracea*) were separated about 16 million years ago, according to the phylogenetic analysis (Zou et al. 2017) (Figure 1A). The allotetraploid quinoa genome arose from a spontaneous hybridization event that occurred 3.3 - 6.3 million years ago between the progenitors, most likely in North America (Jarvis et al. 2017). The diploid species *C. pallidicaule* (A genome) and *C. suecicum* (B genome) are the closest relatives of quinoa progenitors. The A genome ancestor of the hybridization event suggested being the maternal parent, hence providing the cytoplasm (Maughan et al. 2019). *C. berlandieri* and *C. hircinum* are two closely related tetraploid species, which also consist of A and B sub-genomes. Together with quinoa, they form an interfertile group, namely the allotetraploid goosefoot complex (ATGC) (Maughan et al. 2019) (Table 1). Indigenous people in the Andean regions took the first step of quinoa domestication ~7,000 years ago. Selection by ancient farmers changed morphological and physiological characters such as seed coat thickness, seed diameter, seed margin configuration, and seed thickness, which are the domestication traits for quinoa (Gómez-Pando et al. 2019).

The genome size of quinoa is estimated to be 1.45 - 1.5 Gb (nuclear DNA content $1C = 1.49$ pg) (Palomino et al. 2008; Kolano et al. 2012). Quinoa karyotypes are small, and chromosomes are mostly metacentric and sub-metacentric (Palomino et al. 2008; Bhargava et al. 2006). Recently two high-quality reference genomes were made available for the public. Those reference genomes represent both coastal (accession: PI 614886) (Jarvis et al. 2017) and highland (accession: Real) (Zou et al. 2017) ecotypes. In addition, chloroplast genomes of three quinoa accessions are available for the public. The B genome (660 Mb) is larger compared to the A genome (524 Mb), accounting for 55.74% of the quinoa genome, which is in accordance with the *C. suecicum* and *C. pallidicaule* genomes. Gene annotations along the genome predicted 44,776 gene models, of which 33,365 genes were classified as high confidence predictions. The quinoa genome consists of a substantial proportion of repetitive sequences (64%), and the most abundant repetitive elements were long terminal repeats (LTRs) (Jarvis et al. 2017).

General Introduction

Table 1: Ploidy level, chromosome number, and native habitats of six *Chenopodium* species, including allotetraploid goosefoot complex (ATGC) and diploids that are closely related to quinoa.

Species	Ploidy	Chromosome number	Native distribution	Reference
<i>C. pallidicaule</i>	2x	18	South America	(Bhargava et al. 2006)
<i>C. suecicum</i>	2x	18	North America and Eurasia	(Uotila 1972)
<i>C. ficifolium</i>	2x	18	Eurasia	(Bhargava et al. 2006)
<i>C. hircinum</i>	4x	36	South America	(Wang et al. 1993)
<i>C. quinoa</i>	4x	36	South America	(Wang et al. 1993)
<i>C. berlandieri</i>	4x	36	South Texas to Mexico	(Wang et al. 1993)

Since domestication occurred many thousand years ago, quinoa has become an essential component of the diet of the Andean culture. The Andean region is one of the most hostile environments for crop cultivation due to the presence of areas with arid, semi-arid, and high saline conditions. Thanks to its substantial diversity and ecological plasticity, quinoa is capable of thriving in harsh conditions (Ruiz et al. 2014). Also, quinoa is highly efficient in utilizing agricultural resources such as water and fertilizer (Gómez-Pando et al. 2019). Therefore, quinoa is an attractive crop for small farmers and farmers from marginal lands. On the other hand, quinoa provides high-quality nutrition, unlike any other cereals. For example, quinoa provides high-quality gluten-free proteins with all the essential amino acids, and the fatty acid profile of quinoa matches the best with human nutrition requirements (González et al. 2015). Therefore, the demand for quinoa grows steadily.

Currently, there are about 95 countries involved in quinoa cultivations or test trials, but most of the production is still coming from Bolivia and Peru (Bazile et al. 2016b). For instance, in 2018, quinoa's total worldwide production was 158,920 metric tons, from which 44.5% and 55.1% were produced in Bolivia and Peru, respectively (FAOSTAT: <http://www.fao.org/faostat/en/#home>). In the past two decades, the production of quinoa increased at an exponential pace. Nevertheless, this production improvement is mainly achieved by the expansion of the cultivation area. Despite the improvement of the production, the yield per hectare is stagnating (Figure 1 B, C, and D). This is mainly due to the lack of breeding and genetic improvement. Consequently, quinoa remains an underutilized crop (Gomez - Pando 2015). Expansion of quinoa towards higher latitudes and genetic enhancements based on modern breeding techniques are essential to explore its enormous potential as an alternative nutritional source.

1.2 Genetics, genomics, and breeding of quinoa

Numerous studies have been performed to characterize molecular genetic diversity in quinoa. The first generation morphometric and electrophoretic studies were initiated by Hugh Wilson in 1980. This pioneering work described representative germplasm from both Andean highland and coastal quinoa using 21 isozyme loci and leaf shape (Fuentes and Zurita-Silva 2013). Later, 148 accessions were characterized based on the seed protein electrophoretic mobility, solubility fractionation, and genetic variability. This study identified three polypeptides that are highly polymorphic among accessions. A high degree of heterogeneity within accessions was observed, mirroring that the accessions are unselected landraces (Fairbanks et al. 1990). Later, also isozymes were used but did not become popular due to the small number of available markers and the high genotyping costs.

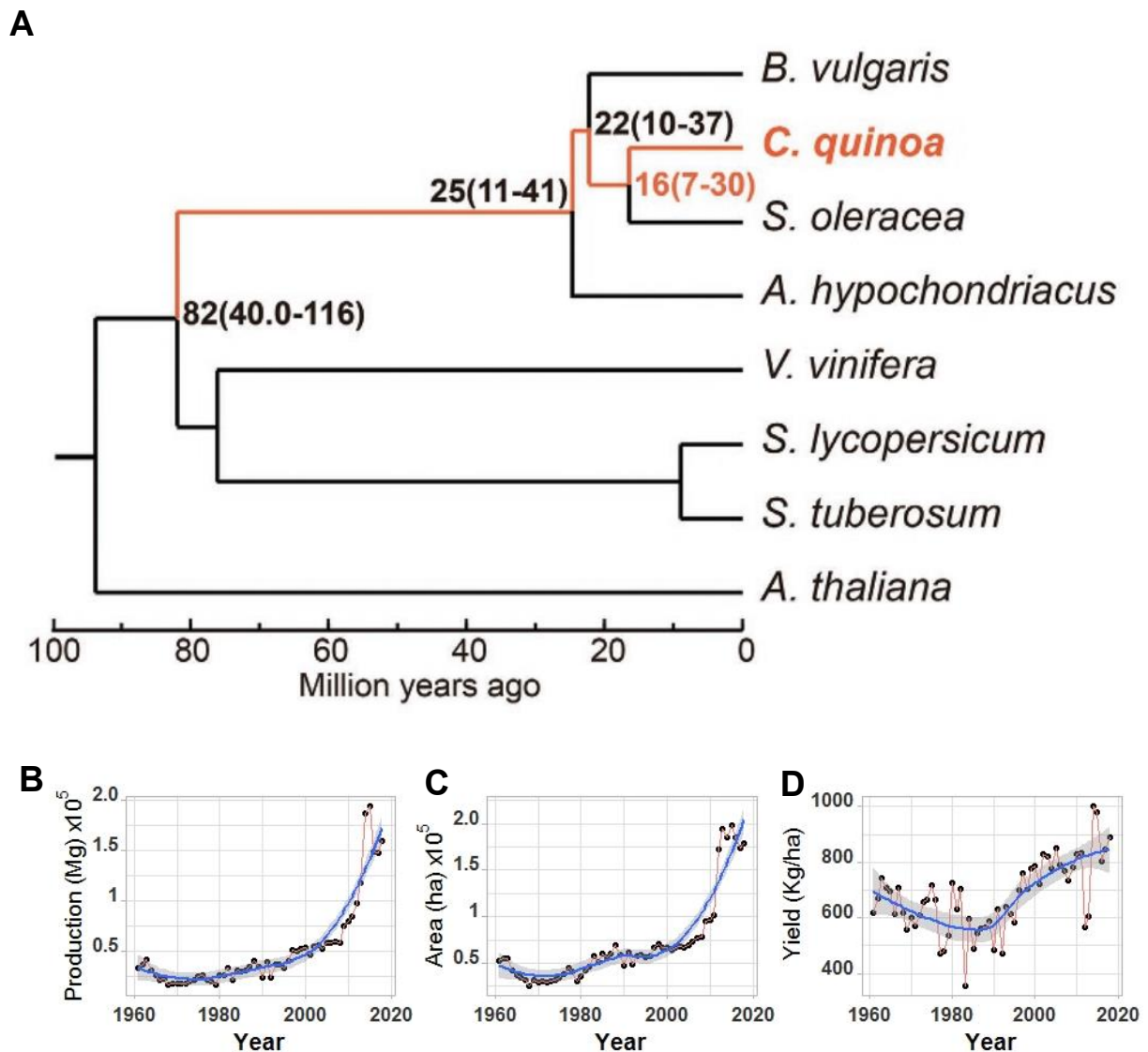


Figure 1: Phylogenetic analysis of quinoa and its production over the years. Phylogeny of quinoa and related crop species (A) (After: Zou et al. (2017)). Worldwide quinoa production (B), cultivation area (C), and yield (D) throughout six decades. The default loess (Local Polynomial Regression) function in the ggplot2 package was used to fit curves (geom_smooth) (FAOSTAT: <http://www.fao.org/faostat/en/#home>).

Molecular genetic studies were then mainly focused on DNA-based markers. The first DNA markers used in quinoa were Random Amplified Polymorphic DNA (RAPD). Thirty primer combinations were used to genotype 16 randomly obtained quinoa accessions, and 20 of them were polymorphic (Fairbanks et al. 1993). RAPDs were also used to identify the genetic relationship between 19 *Chenopodium* species, including quinoa (Ruas et al. 1999). Thirty-three primer combinations resulted in 399 polymorphic bands due to the broad diversity, and accessions were clustered into five groups depending on their species. A similar study was conducted by using cultivated and wild quinoa from the Bolivian altiplano. Eighty-seven accessions were characterized using ten primer combinations. Wild and cultivated accessions showed a significant but very low level of differentiation. Nevertheless, a strong structure was observed among eight populations highly correlating to the eco-geographic regions (del Castillo

General Introduction

et al. 2007). Identification of true F₁ hybrids was also achieved using RAPD markers (Bonifacio 1995; Maughan et al. 2004).

Amplified fragment length polymorphism (AFLP) markers were also employed in quinoa studies. An F₂ population consisting of 80 individuals was genotyped with 284 AFLPs using 68 primer combinations. The AFLP markers were proven to be polymorphic in other populations as well (Maughan et al. 2004), indicating the reliability of AFLPs in quinoa research. Using AFLP markers, the Chilean quinoa germplasm of 18 accessions, including *C. album* and *C. ambrosioides* were characterized by 130 polymorphic bands. As a result, wild *Chenopodium* species were differentiated from the cultivated quinoa. Eighteen quinoa accessions were grouped into three clades in accordance with the collection site (Rodríguez and Isla 2009).

Microsatellites are widely distributed across plant genomes. The simple sequence repeat (SSR) markers are highly reproducible, polymorphic, multi-allelic, and mostly co-dominant (Bhargava and Fuentes 2010). Therefore, an increasing number of studies started utilizing SSR markers. One of the first large-scale SSR marker studies in quinoa was conducted by Mason et al. (2005). They developed 208 markers to characterize 31 cultivated quinoa accessions. High heterozygosity among accessions was observed (Mason et al. 2005). 35 SSR markers evaluated the genetic diversity in a collection of 152 accessions obtained from the USDA and CIP FAO international nursery. Accessions from Argentina, Bolivia, Colombia, Chile, Ecuador, Peru, and the USA were included in this collection. This study was the first large-scale diversity analysis based on molecular markers. They observed a clear separation of highland and lowland quinoa (Christensen et al. 2007). Another study was carried out to assess the diversity of 59 Chilean accessions, which originated from altiplano and coastal areas. Like in the previous study, two distinct germplasms were identified based on cluster analysis and principal component analysis (PCA). Interestingly, they found that lowland quinoa is more diverse than highland quinoa, which is in contrast to other studies (Fuentes et al. 2009). Chilean germplasm was extensively subjected to genetic diversity analysis. Fuentes et al. (2012) used 118 SSR markers to analyze 34 quinoa accessions. They grouped Chilean quinoa into three populations (North, Center, and South). Genetic divergence among populations was correlated with geographical origin (Fuentes et al. 2009). Argentina is also a popular quinoa growing area globally, but its germplasm is underrepresented in genetic studies. Therefore, 35 accessions from north Argentina were characterized using 22 SSR markers. High genetic diversity among accessions belonging to four groups was identified. These four subpopulations were correlated with altitudinal distribution (Costa Tártara et al. 2012). SSR markers were also applied in linkage mapping in a RIL population by Jarvis et al. (2008). They used 200 SSR markers on 82 F₆ plants and constructed a linkage map that had 38 linkage groups covering 913 cM (Jarvis et al. 2008).

Single Nucleotide Polymorphism (SNP) markers are popular among researchers and breeders because of their low cost, high reproducibility, higher mapping resolution, and high throughput. Hence, quinoa researchers also started developing SNP markers. Coles et al. (2005) developed the first SNP markers for quinoa using EST libraries. They analyzed 20 ESTs and identified 51 SNPs (Coles et al. 2005). Later, a large set of SNPs was developed by Maughan et al. (2012). They used five bi-parental populations and performed combined genotyping by sequencing. More than 14,000 putative SNPs were reported in this study, from which 511 KASP markers were developed. KASP markers were then used for genotyping 113 quinoa accessions. In accordance with previous studies, two distinct quinoa subgroups of highland and lowland were identified. With a reference genome availability, the first whole-genome re-sequencing of 15 quinoa accessions identified 7.8 million SNPs (Jarvis et al. 2017). Another study also re-sequenced 11 quinoa accessions, and 8 million SNPs were identified (Zhang et al. 2017). They

General Introduction

also analyzed the InDel polymorphism among the accessions and identified 842,783 InDels. These InDels were used to develop 85 markers and genotyped 129 quinoa accessions together with 62 additional SSR markers. Population structure analysis of these accessions revealed two distinct populations, with an F_{ST} divergent of 0.70. Chilean lowland quinoa was less diverse than the highland quinoa (Zhang et al. 2017).

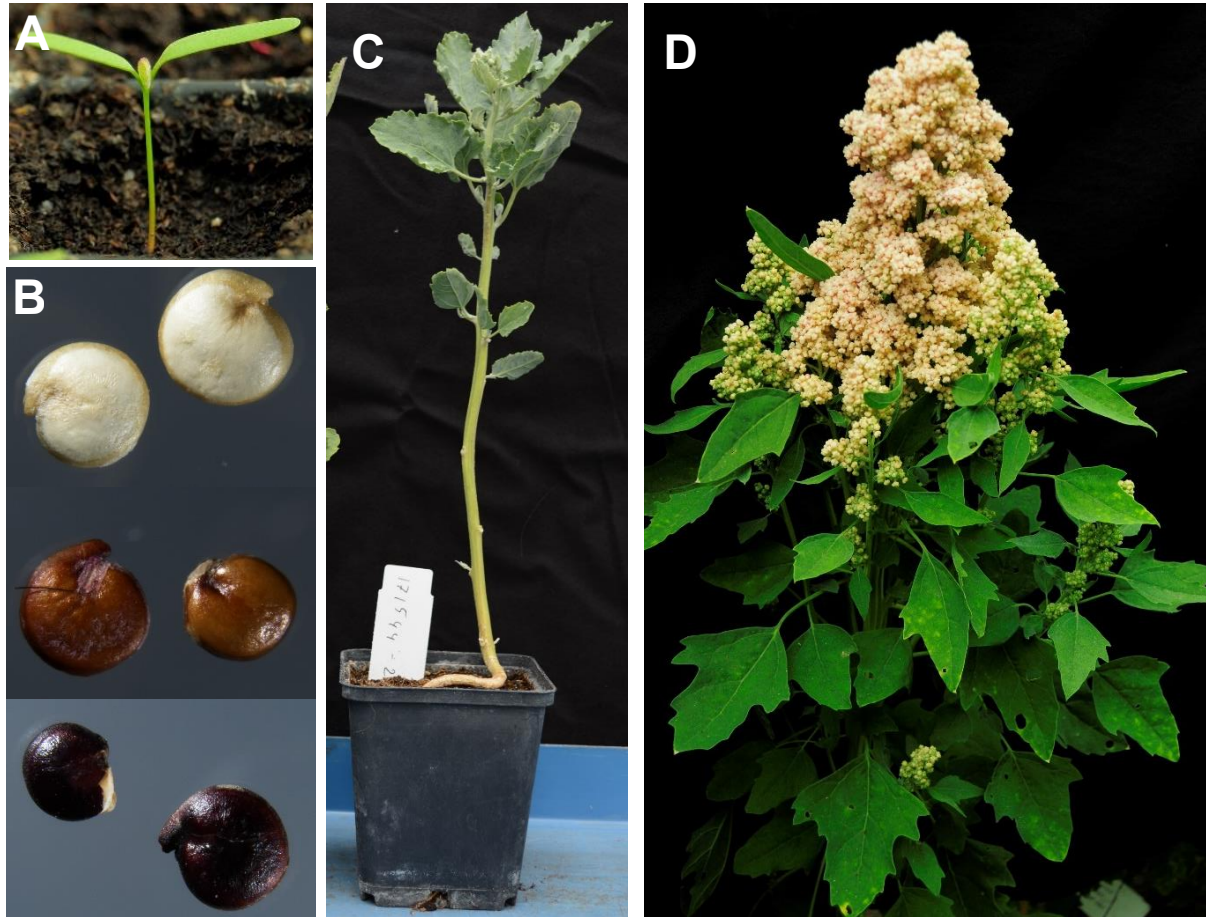


Figure 2: Representative images of quinoa seedling (A) seeds (B) whole plant (C) and panicle (D)

The inflorescence of quinoa is a panicle with a panicle axis. Panicles are often branched, resulting in the secondary and tertiary axis emerging from the central axis (Figure 2). Predominantly quinoa has hermaphrodite and female flowers. Hermaphrodite flowers are small and have a single pistil, five sepals, and five anthers but do not have petals, whereas female flowers are even smaller (Peterson et al. 2015). Quinoa is a predominantly self-pollinating crop. Several studies have reported outcrossing in quinoa; therefore, studies have been conducted to measure the outcrossing rate (Bhargava and Srivastava 2013). Outcrossing has been reported between plants located up to 1 m apart, but it could occasionally go up to 20 m. Crossing rates varied between 0.5 to 17.36 % amongst different accessions (Gandarillas 1979; Silvestri and Gil 2000). The likelihood of cross-pollination depends on multiple factors such as (1) the presence of wind and pollinators, (2) the ratio of female to hermaphrodite flowers, and (3) the presence of male sterility, and self-incompatibility (Peterson et al. 2015). Temperature also plays a significant role in pollen viability. Reduced pollen viability was observed mostly over 30 °C degree temperature conditions (Hinojosa et al. 2019). Accessions from the central highlands of Peru and Bolivia displayed male sterility. A male-sterile plant can be easily detected based on

General Introduction

the color of the anthers; typical fertile quinoa shows bright yellow anthers, whereas sterile anthers are white (Murphy et al. 2018). It has been reported that due to outcrossing, quinoa landraces are highly heterogeneous and heterozygous. Hence, maintaining pure lines is impossible for farmers who produce their own seeds (Lindhout and Danial 2006). However, this is highly subjective to accessions as they show varying degrees of outcrossing ability.

Small flowers and the tendency of self-pollination are major hurdles for quinoa hybridization. Nevertheless, emasculation methods have been developed, and multiple studies successfully obtained F₁ hybrids (Rea 1969; Gandarillas 1979; Gandarillas and Luizaga 1967; Bonifacio 1995; Peterson et al. 2015; Emrani et al. 2020). Both morphological and molecular markers have been employed in true hybrid identification. After creating F₁ hybrids from suitable parents with desirable characters, F₂ populations can be obtained.

In quinoa, selection methods are predominantly used to develop anticipated varieties. Bulk selection, pedigree selection, and single-seed descent (SSD) methods are the most common methods used in quinoa (Gomez - Pando 2015). To introduce new traits to existing cultivars, backcross breeding methods are used. New traits are also introduced from wild relatives using backcross programs (Bonifacio 2003). Hybrid breeding in quinoa has gained attention as heterosis was observed (Ward and Johnson 1993). Anecdotal evidence suggests that yield heterosis for quinoa ranges from 201 to 491% (Wilson 1990). However, further studies are needed to investigate heterosis in quinoa. Male sterility is an essential factor in hybrid breeding. Several studies have reported nuclear (Gandarillas 1969) and cytoplasmic (Simmonds 1971; Risi and Galwey 1984) male sterility, which can be exploited in future hybrid breeding.

Mutation breeding can be advantageous to introduce new variation and trait manipulation. Researchers in Peru have started gamma irradiation of quinoa aiming to remove undesired phenotypes such as lodging and extra tall plants (Gomez-Pando and Eguiluz-de la Barra 2013). Nevertheless, no varieties have been released so far after mutation induction. Currently, EMS mutagenized populations are under development for reverse genetic studies, which in the future could also be a beneficial resource for quinoa breeding (Jarvis 2020).

Scientists in the UK conducted a collection expedition through South America in 1970, and quinoa was then first introduced to Europe (Risi et al. 1989). In the UK, it was first grown as a cover crop for hunting games (Jacobsen 2017). Breeding and field trials of quinoa in four countries (UK, Denmark, Netherlands, and Italy) have started as a result of a European union-supported project (Galwey 1992). Quinoa as an organic food became popular in Denmark, especially among people who suffer from coeliac disease, hence, leading to breeding and production (Jacobsen et al. 2010). The first large-scale quinoa cultivation was started in 2009 by the French company AbbottAgra (Jacobsen 2017). Only accessions introduced from Chile have reached maturity in European conditions (Izquierdo et al. 2003). Quinoa was introduced to the temperate condition in the USA in early 1980 by Hugh Daniel Wilson. He had collected Chilean accessions and taken them to the USDA gene bank collections (Wilson 1988b, c, 1990). Therefore, most quinoa cultivars grown in temperate regions have roots from Chilean lowland accessions and have limited diversity, even though the global quinoa germplasm possesses a large diversity.

The main breeding objective of quinoa is to develop varieties that are locally adapted and stable. Since industries use quinoa for various food productions, breeding objectives are also diversified toward industrial usage. From the farmer's perspective, higher yield, short-plants (for mechanical harvesting), stalk strength, uniform maturity, tolerance to downy mildew, drought, and high-temperature are the major requirements that breeders should prioritize. Protein content

and composition, seed size, and saponin content are the main requirements of the industry (Gomez - Pando 2015).

1.3 Adaptation of quinoa to different geographical regions

Crop adaptation mainly occurs due to the exploitation of beneficial opportunities or mitigating damages (Sloat et al. 2020). Plant adaptation to local environments results from the accumulation of alleles that are best for the local environment, hence assuring reproductive success. Alleles best for one environment may not fit for another environment (Anderson et al. 2011). Phenotypic plasticity and having a broad diversity enable fast adaptation of crops to different environments. The flowering time and photoperiod adaptation have played a crucial role in local adaptation, especially into diverse latitudes and altitudes (Romero Navarro et al. 2017).

Quinoa is a remarkable example for a crop which possesses enormous diversity and phenotypic plasticity, resulting from broad adaptation. The Andean region, the diversity center of quinoa, shows a substantial eco-geographical variation. Quinoa has undergone selection to spread through the Andean region since it has been domesticated. The vast range of climatic conditions catalyzed quinoa adaptation, but the process has been relatively slow (Bertero et al. 2004). Present-day quinoa adaptation is much quicker due to modern breeding efforts compared to many other crops and extends towards new environments beyond its center of origin. The last four decades of breeding enabled the successful expansion of quinoa among non-native countries (Jacobsen and Bendevis 2013).

It has been reported that all growth stages of quinoa are sensitive to photoperiod (Bertero et al. 1999a). Nonetheless, depending on the accession, response to photoperiod varies. The reproductive stage, seed filling, and maturity are the most sensitive stages to photoperiod. (Bertero et al. 1999a; Bertero et al. 2004; Christiansen et al. 2010). A quantitative photoperiod effect on flowering has been reported, indicating that quinoa is a facultative short-day crop (Fuller 1949). Thus, reproductive success is not guaranteed in a long-day environment due to the hampered seed set. Therefore, the effect of photoperiod and its physiological relevance need to be thoroughly investigated to speed up quinoa adaptation, especially to long-day conditions.

Quinoa diversity has evolved after domestication. Ancient farmer societies moved along different regions of the Andes to find new areas for cultivations (Murphy et al. 2018). A wide diversity of altitude, temperature, rainfall, and soil types resulted in the requirement of complex and specific traits in order to obtain a stable harvest. Due to thousands of years of artificial and natural selection, quinoa germplasm is divided into multiple ecotypes to cope with the Andean climate. There are five broadly recognized ecotypes: Altiplano, Coastal (Sea-levels), Inter-Andean Valley, Salares, and Sub-Tropical. These ecotypes differ in their photoperiod response, altitude adaptation, drought and salinity tolerance, disease resistance, life cycle duration, and seed dormancy (Gomez - Pando 2015) (Table 2).

Table 2: Major ecotypes of quinoa and their features

Ecotype	Geographical distribution	Remarks
Altiplano	high altitudes around Lake Titicaca, altitude (3600 - 3800 m)	plant height varies between 0.5 - 1.5 m
		compact main panicle
		most accessions are susceptible to downy mildew
Coastal /Sea level	Chile, down to 36° South	plant height 1 - 1.4 m
		stalks are fragile, and plants are prone to mechanical damages
		branched growth
		cream/beige colored seeds
Inter Andean-Valley	found between 2500 - 3500 m	tolerant to downy mildew
		usually grow tall 2.5 m or higher
		branched and loose (lax) inflorescence
		intercropped with maize
Salares	grows in the salt flats of the high Andean plateau	drought tolerant (annual precipitation 200 - 300 mm)
		Salt tolerant
		large seeds, thick pericarp, and high saponin
Sub-tropical or Yungas	found in the eastern slope of Andes along with warm, rainy and humid forest, altitude 1500 - 2200 m	grow up to 2.2 m
		lightly branched, multiple harvests from a single plants

Interest in cultivation has increased among the importing countries. Hence adaptation to new environments should be considered through genetic improvement. Northern latitudes countries such as the Netherlands, Canada, Denmark, Spain, France, the UK, and the USA have started developing new varieties (Bazile et al. 2016a). Adaptation to northern latitudes specifically needs photoperiod adaption. Some accessions from southern Chile displayed photoperiod insensitivity. This allows the breeders to select suitable breeding materials for developing new varieties for higher latitudes (Jacobsen 2017).

1.4 Flowering time regulation in the model species *Arabidopsis*

A vegetative shoot apical meristem transition into a floral meristem is a crucial stage in the life cycle of a higher plant. Flowering time adaptation to the environment in which plants grow determines the evolutionary success of a plant species (Andres and Coupland 2012). Decades of research on flowering time control in the model plant *Arabidopsis* have delivered a well-understood yet complex regulatory network (Blümel et al. 2015; Bouché et al. 2015).

Arabidopsis is a dicot species with both winter and summer annual forms. It is a facultative long-day (LD) species that shows accelerated flowering when days become longer (Roden et al. 2002). Signals from multiple pathways are intricately involved in the regulations of flowering time. Depending on the initial stimuli, these pathways can mainly divide into two classes: external (photoperiod, vernalization, and ambient temperature) and internal (autonomous, age, circadian clock, gibberellin, and sugar pathway). In temperate climates, photoperiod and temperature are the major flowering induction signals (Andres and Coupland 2012). The convergence of main pathways into major transcription factors known as floral integrators is the most apparent cause of variation in flowering time. A plethora of studies unraveled novel genetic, epigenetic, and environmental factors and their interplay in flowering time regulation. Many studies dissected the multiple links between different pathways (Song et al. 2015). Consequently, flowering time regulation in *Arabidopsis* is only briefly reviewed herein.

The circadian clock genes sense the photoperiod and coordinate the expression of *CONSTANS* (*CO*), a zinc finger transcription factor. *CO* shows a rhythmic diurnal expression that reaches a peak in the evening. In the leaves, *CO* directly activates the transcription of *FLOWERING LOCUS T* (*FT*), the mobile flowering promoter, also known as florigen. *FT* encodes a small protein similar to phosphatidyl ethanolamine-binding proteins (PEBP). *CO* is not stable during the dark; consequently, *FT* is activated only under LD photoperiods (Valverde et al. 2004). The *CO* transcription depends on the activity of a protein complex formed by *GIGANTEA* (*GI*) and *FLAVIN KELCH F BOX 1* (*FKF1*). *GI* is specific to plants, and those two proteins are essential components of the circadian clock. *FKF1* is attached to a chromophore and perceives light signals. *GI* and *FKF1* interaction depends on the light (Andres and Coupland 2012). The long-distance florigenic signal is conveyed by moving *FT* from leaves via vascular tissues to the shoot apical meristem. From the same PEBP gene family, another gene acts agonistically to *FT*; *TERMINAL FLOWER1* (*TFL1*), also known as anti florigen (Bradley et al. 1997; Ohshima et al. 1997). *FT* and *TFL1* differ from each other by 39 non-conservative amino acids. The surface charge changes of *FT* changed its properties to a flowering repressor by acting as *TFL1*. This change is induced by mutations in vital amino acids at four sites (Glu-109, Trp-138, Gln-140, and Asn-152) (Ho and Weigel 2014a). *FT*, together with *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*) and *AGAMOUSLIKE24* (*AGL24*), activate four meristem identity genes (*LEAFY* (*LFY*), *APETALA1* (*API*), *SEPALLATA3* (*SEP3*), and *FRUITFULL* (*FUL*)). This marks the start of the transition of the vegetative meristem into a reproductive meristem (Blümel et al. 2015).

FLOWERING LOCUS C (*FLC*) is a MADS-box domain transcription factor and one of the major flowering repressor genes in the *Arabidopsis* vernalization pathway (Sheldon et al. 1999; Michaels and Amasino 1999). *FRIGIDA* is a major upstream element of *FLC*, and together, they contribute to a large proportion of flowering time variation in winter-annual *Arabidopsis* (Shindo et al. 2005). *FLC* directly binds to the MADS-box family genes promoters' whose members are activators or suppressors of floral transition. In winter-annual *Arabidopsis* accessions, flowering is facilitated by epigenetic silencing of *FLC*.

General Introduction

Flowering time control by the autonomous pathway is parallel to the vernalization pathway. Multiple genes have been identified to be involved in this pathway. Many of these genes are involved in RNA-mediated chromatin modification, and interactions of these genes are complex and non-linear. Interestingly their role in flowering converges to regulations of *FLC* (Wu et al. 2020).

In contrast to the flowering response to cold temperatures, moderately increased temperature also could induce flowering. A temperature change from 23°C to 37°C was adequate to induce flowering under non-inductive short day (SD) conditions (Balasubramanian et al. 2006). Phytochrome-Interacting Factor4 (PIF4), a transcription factor, promoted *FT* transcription at warm temperatures, consequently accelerating flowering (Kumar et al. 2012).

The age pathway is another endogenous mechanism that affects floral transitions. In the last decade, age-dependent flowering studies mainly focused on micro RNAs (miRNA) (Hyun et al. 2017). miR156 and miR172 are vital factors in the aging pathway. It has been shown that the expression of these miRNAs is negatively correlated. (Teotia and Tang 2015). SQUAMOSA PROMOTER-BINDING PROTEIN-LIKEs (SBPs/SPLs) are transcription factors involved in the regulation of plant development. *SPL* promotes flowering even under the non-inductive SD photoperiod (Wang et al. 2009). There are 16 *SPL* genes identified in Arabidopsis, of which 11 genes have miRNA recognition sites (Guo et al. 2008). Both *SPL* and miR156/172 are expressed in leaves and meristems. The abundance of miR156 decreases with age, whereas miR172 increases. Expression patterns of *SPL* and miR156 are inversely related, indicating that *SPL* is under the regulation of miR156/172 (Zhou et al. 2013).

Under both inductive and non-inductive conditions, a group of hormones known as gibberellic acids (GA) promotes Arabidopsis flowering (Bao et al. 2020). There are more than 130 GAs described, of which some of the major bioactive GAs such as GA₁, GA₃, GA₄, GA₅, and GA₆ are associated with flowering (Yamaguchi 2008). For instance, the Arabidopsis *gal-3* mutant showed impaired flowering due to the suppression of *FT*. Artificial GA₃ treatment restores the *FT* expression and, consequently, plants flowered normally (Galvao et al. 2012). DELLA proteins are transcription factors that are involved in GA-mediated *FT* regulation. DELLA proteins can bind to *CO* directly to suppress its expression. As a result, the expression of the *FT* is suppressed (Yu et al. 2016a). DELLA proteins are degraded by a complex made from GA and GIBBERELLIN INSENSITIVE DWARF1 (GID1) (Murase et al. 2008). DELLA proteins are also involved in the regulation of *SOC1* and *LFY* directly and *API* indirectly through *SPL* genes (Bao et al. 2020). Consequently, GA plays an essential role in flowering regulations.

1.5 Flowering time regulation in quinoa and related species of the Amaranthaceae plant family

Species like quinoa, sugar beet (*Beta vulgaris*), spinach, *C. album*, and *C. rubrum* belong to the Amaranthaceae family. Sugar beet is an LD species with vernalization requirement. Two copies of *FT* homologs have been identified (*BvFT1* and *BvFT2*) (Pin et al. 2010). Interestingly *BvFT1* and *BvFT2* act antagonistically. *BvFT1* represses flowering, and *BvFT2* promotes it. Before vernalization, *BvFT1* is highly active; hence flowering is repressed. The activity of *BvFT1* reduces with the prolonged cold period and allows the expression of *BvFT2* (Pin et al. 2010). Vernalized sugar beet strictly requires LDs to flower. If plants are transferred to SDs after vernalization, flowering does not occur, indicating a devernalization response. Also, high-temperature exposure just after the vernalization results in devernalization (Lexander 1980). Both genes showed diurnal expression under LDs, where *BvFT1* shows the typical *FT* expression pattern with a peak before the dark. Interestingly, under non-inductive SDs, *BvFT2* expression

General Introduction

was undetectable, whereas *BvFT1* was expressed with a morning peak. At the protein level, BvFT1 and BvFT2 showed a high degree of sequence similarity. Three amino acid differences between BvFT1 and BvFT2 were identified, and these differences occurred due to SNPs in the external loop encoded from the exon 4. This resulted in the antagonistic functions of BvFT proteins (Pin et al. 2010).

No true *CO* homolog has been identified in sugar beet. Chia et al. (2008) investigated two *CO*-like genes (*BvCOL1* and *BvCOL2*). Both genes did not show typical *AtCO* diurnal expression patterns, but expression patterns resembled *AtCOL1*. *BvCOL1* complemented the *co* mutant phenotype in Arabidopsis (Chia et al. 2008). Two genes that were proven to be upstream of the *BvFT* genes were identified. *BOLTING TIME CONTROL 1 (BTC1)*, a pseudo-response regulator (*PRR*) gene, defines the life history of beets (Pin et al. 2010). The dominant *BTC1* allele is present in annual beets. In other words, *BTC1* is necessary for bolting when the vernalization is absent. In contrast, sugar beets carry the partial loss of function *btc1* allele; hence, they do not flower without vernalization. However, vernalization enhances the activity of *btc1*, resulting in the induction of bolting. It has been proposed that the late evening expression peak of both *BTC1* and *btc1* results in repression of *BvFT1* and activation of *BvFT2* (Pin et al. 2010). The other upstream regulator of the *BvFT* gene, which was identified based on map-based cloning is *BvBBX19*, a double B-box type zinc finger protein (Dally et al. 2014). *BvBBX19* is indispensable for bolting without vernalization; the mutants needed vernalization despite the presence of the dominant *BTC1* allele. Therefore, it has been proposed that *BTC1* and *BvBBX19* act epistatically to induce bolting in beets (Dally et al. 2014). The reason for this is evident because *BTC1* has a CCT domain and *BBX19* has two B-BOX domains. Together they form a protein similar to *CO*, which combines CCT and B-BOX domains. This hypothesis was tested by Dally et al. (2018), and they showed that *BTC1* and *BBX19* interact with each other. Hence they obtain the function of *CO* to upregulate *BvFT2* (Dally et al. 2018).

Scientists from the Czech Academy of Sciences studied flowering time regulation of three wild quinoa relatives: *C. rubrum*, *C. ficifolium*, and *C. suecicum*. First, they analyzed two *FT*-like homologs (*CrFTL1* and *CrFTL2*) in *C. rubrum*. *CrFTL2* expression was stable during light-dark cycles, where *CrFTL1* showed diurnal expression with a peak at noon. These results indicated that *CrFTL2* has no function in flowering, whereas *CrFTL1* could be the floral regulator. Moreover, flowering was abolished entirely when night breaks were introduced, indicating the true SD nature of *C. rubrum* (Cháb et al. 2008). Secondly, they confirmed the necessity of *CrFTL1* as a floral promoter by complementing Arabidopsis *ft-2* mutants (Drabešová et al. 2014b). In the same study, two *COL* genes (*CrCOL1* and *CrCOL2*) were investigated. Both genes have a CCT and two BBX domains like *CO*. However, unlike *BvCOL1*, none of them rescued the flowering phenotype of Arabidopsis *co* mutants, indicating that they are not true *CO* homologs. Expression patterns of those genes showed that light inversely affects the expression of *CrCOL* and *CrFTL* genes, unlike in Arabidopsis, in which *CO* and *FT* expression coincide with each other (Drabešová et al. 2014b). Transcriptome analysis by Drabešová et al. (2016) revealed a third *FTL* in *C. rubrum*. Having three copies of *FTL* genes is specific to chenopods/amaranths. Investigation of a *TFL* homolog, *CrTFL1*, showed that it might repress flowering (Drabešová et al. 2016). However, further investigation is needed to confirm the true function of this gene.

A recent study from the same research group investigated *FTL* genes in *C. ficifolium* and *C. suecicum* (Štorchová et al. 2019). They showed that *CfFTL1* and *CsFTL1* expressions were similar to *FT* expression and consequently correlated with flowering. Interestingly they showed that under LDs, expression of *CfFTL1* in the facultative LD *C. ficifolium* was negligible,

General Introduction

although plants flowered normally. They hypothesize that *C. ficifolium* has adapted specific flowering pathways to flower under long-day conditions (Štorchová et al. 2019).

The knowledge about flowering time regulation of quinoa is in its infancy. Five *FT* homologs were identified, and three of them are similar to *BvFT1*, whereas others show high sequence similarity to *BvFT2* (Jarvis et al. 2017). The functions of these genes are expected to be as floral repressors and activators, respectively. A recent study investigated quinoa genes homologous to known flowering genes from *Arabidopsis*. They identified 611 putative genes with homology to flowering genes, which may be involved in flowering time regulation. In addition, 459 quinoa-specific putative flowering time genes were identified based on their expression profiles (Golicz et al. 2020). However, genetic or functional evidence for quinoa flowering time candidate genes is still lacking.

1.6 Genome-wide association studies in crop plants

Crop domestication by humans started about 12,000 - 15,000 years ago, and diversification accelerated due to conscious and unconscious selection (Meyer and Purugganan 2013). The genetic diversity of crops at the present day was profoundly affected by diversification and breeding. The efficient use of germplasm resources can be achieved by improving our understanding of the genetic basis of phenotypic variation. The phenotypic variation of a crop is mainly caused by mutations that are then maintained during the selection process (Huang and Han 2014).

The identification of genetic variation of crops has advanced due to the introduction of high throughput genotyping methods. There are mainly five high output genotyping methods used in the identification of natural genetic variation in crops, namely (1) microarray-based genotyping, (2) sequencing-based genotyping, (3) genotyping by sequencing, (4) RNA-seq-based genotyping, and (5) exon-sequencing-based genotyping (Huang and Han 2014). Nevertheless, method selection is restricted by the availability of a reference genome and costs (Huang and Han 2014). Re-sequencing large diversity panels becomes more popular due to the advancement of computational resources, cost reduction, and the ability to achieve high resolution.

Genome-wide association studies (GWAS) are also known as linkage disequilibrium (LD) based association mapping. As the name suggests, it relies on the association of phenotypic variation with a sequence polymorphism. GWAS takes advantage of ancient recombination events. The general assumption is that a subset of individuals with the same phenotype carries the same genetic variation which is in LD with the causal mutation (Liu and Yan 2019). The first GWAS study on plants was conducted nearly two decades ago on maize, and it identified the *Dwarf8* gene, a repressor of gibberellin that affects flowering time and plant height (Thornsberry et al. 2001). Since then, there have been a plethora of studies conducted, and numerous genes have been identified. The power to identify significant association essentially relies on the resolution of markers, sample size (number of accessions), LD between the marker and causative SNP (Mackay and Powell 2007), phenotype variation, and trait architecture (Buckler et al. 2009; Alqudah et al. 2020). High-quality phenotype data is essential for gaining high power and accurate associations, and noisy data and extreme outliers should be excluded. Multi environment data are adjusted using advanced statistical models. The best linear unbiased estimator (BLUE) and the best linear unbiased predictor (BLUP) are the most common methods used for adjusting means (Alqudah et al. 2020). In principle, GWAS based on whole-genome re-sequencing analysis should reveal all possible allelic variation. However, it is far from reality, especially for complex traits, because in practice, only an essential subset of allelic variation underlying a trait is identified. The genetic architecture of a trait is influenced by the mating system, demography, and the breeding history of the crop. Multiple genes with small effects

General Introduction

control the genetic architecture of complex traits; nevertheless, some are expected to be controlled by major genes. For instance, flowering-time variation in plants is controlled by major genes with larger effects (Buckler et al. 2009).

Although GWAS studies are highly successful, it has some limitations. These limitations are especially more pronounced if the population structure is strong and causal alleles are represented at low-frequencies (Korte and Farlow 2013). For instance, flowering time is highly correlated with population structure, and it has been extensively shown that many false negatives arise due to the methods used for correcting population structure in GWAS (Huang and Han 2014). Scientists are continuously trying to solve these issues by improving statistical models to detect associations with rare sequence variants (Listgarten et al. 2013; Kaakinen et al. 2017) and designing artificial populations with multiple alleles but without population structure (Buckler et al. 2009; Navarro et al. 2011). Moreover, elucidating the complex genetic architecture of traits such as epistasis and pleiotropy is challenging for GWAS studies. For example, the detection of epistasis in unrelated populations leads to substantially low statistical power. Multiple computational algorithms and modeling approaches were developed to overcome this challenge (Liu and Yan 2019). Among many other usages, MAGIC populations have also been used to uncover epistasis interactions (Mathew et al. 2018). The identification of the pleiotropic effects was also challenging in GWAS studies. This was overcome by improved methods such as multivariate analysis (Solovieff et al. 2013). Currently, 26 different mixed model approaches have been proposed and used for GWAS analysis (Gupta et al. 2019).

1.7 Aims, objectives, and scientific hypotheses

This study is divided into two parts. The first part aimed to analyze the expression and haplotype variation of quinoa *FT* and *COL* genes. It was based on the hypothesis that quinoa and beet flowering time genes share common functions and that *CqFT* homologs, as in sugar beet, have antagonistic functions and act as floral activators and repressors.

The second part is a genome-wide association analysis using a quinoa diversity set of 310 accessions. This study aimed to analyze population structure, genomic patterns of variations, linkage disequilibrium decay, and marker-trait association for flowering time and other agronomically important traits.

I aimed to answer six initial questions:

1. Do *FT* and *COL* homologs regulate flowering time in quinoa?
2. Is the function of flowering time genes conserved between sugar beet and quinoa?
3. How are the flowering time genes expressed in quinoa under long-day and short-day conditions?
4. Do sequence variations within *CqFT* and *CqCOL* genes correspond with the geographical distribution of quinoa accessions?
5. What is the genetic structure of the diversity panel?
6. What is the degree of LD decay in the quinoa diversity panel?

2 Genome-wide association study in the pseudocereal quinoa reveals selection pattern typical for crops with a short breeding history

2.1 Abstract

Quinoa germplasm preserves useful and substantial genetic variation, yet it remains untapped due to a lack of implementation of modern breeding tools. We have integrated field and sequence data to characterize a large diversity panel of quinoa. Whole-genome sequencing of 310 accessions revealed 2.9 million polymorphic high confidence SNP loci. Highland and Lowland quinoa were clustered into two main groups, with F_{ST} divergence of 0.36. Diverged selection pressure was observed between two groups where Lowland quinoa underwent strict selection. Fast LD decay was identified, ranging between 6.5 and 49.8 kb, which is typical for crops with a short breeding history. A genome-wide association study uncovered 600 SNPs stably associated with 17 agronomic traits. A SNP density of 2.39 SNPs/kb enabled the identification of candidate genes controlling these traits. Two candidate genes are associated with thousand seed weight, and a resistance gene analog is associated with downy mildew resistance. We also identified pleiotropically acting loci for four agronomic traits that are highly responding to photoperiod hence important for the adaptation to different environments, e.g., long-day conditions in Europe and North America. This work demonstrates the use of re-sequencing data of an orphan crop, which is partially domesticated to rapidly identify marker-trait association and provides the underpinning elements for genomics-enabled quinoa breeding.

2.2 Introduction

Climate change poses a great threat to crop production worldwide. In temperate climates of the world, higher temperatures and extended drought periods are expected. Moreover, crop production in industrialized countries depends on only a few major crops resulting in narrow crop rotations. Therefore, rapid transfer of wild species into crops using genetic modification and targeted mutagenesis is currently discussed (Stetter et al. 2017; Li et al. 2018). Alternatively, orphan crops with a long tradition of cultivation but low breeding intensity can be genetically improved by genomics assisted selection methods. Quinoa (*Chenopodium quinoa* Willd.) is a pseudocereal crop species with a long history of cultivation. It was first domesticated about 5000-7000 years ago in the Andean region. Quinoa was a staple food during the pre-Columbian era, and the cultivation declined after the introduction of crops like wheat and barley by the Spanish rulers. Owing to diversity, biotic and abiotic stress tolerance, and ecological plasticity, quinoa can adapt to a broad range of agroecological regions (Ruiz et al. 2014; González et al. 2015). Due to a high seed protein content and a favorable amino acid composition, its biological value is even higher than beef, fish, and other major cereals (James 2009; Vega-Gálvez et al. 2010). These favorable characteristics contributed to the increasing worldwide popularity of quinoa among consumers and farmers.

A spontaneous hybridization event between two diploid species between 3.3 and 6.3 million years ago gave rise to the allotetraploid species quinoa ($2n = 4x = 36$) with a genome size of 1.45 - 1.5 Gb (nuclear DNA content $1C = 1.49$ pg) (Palomino et al. 2008; Kolano et al. 2012). A reference genome of the coastal Chilean quinoa accession PI 614886 has been published with 44,776 predicted gene models together with whole-genome re-sequencing of *C. pallidicaule* and *C. suecicum* species, close relatives of the A and B subgenome donor species, respectively (Jarvis et al. 2017). The organellar genomes are originated from the A-genome ancestor (Maughan et al. 2019).

Quinoa belongs to the Amaranthaceae, together with some other economically important crops like sugar beet, red beet, spinach, and amaranth. It reproduces sexually after self-pollination.

Chapter 2

Facultative autogamy was reported for plants in close proximity with outcrossing rates in a range of 0.5 to 17.36 % (Gandarillas et al. 1979; Silvestri and Gil 2000). Thus, quinoa accessions are typically homozygous inbred lines. Nonetheless, heterozygosity in some accessions has been reported, which indicates cross-pollination (Christensen et al. 2007). The inflorescences are panicles, which are often highly branched. Florets are tiny, which is a significant obstacle for hand-crossing. However, routine protocols for F₁ seed production in combination with marker-assisted selection have been developed recently (Peterson et al. 2015; Emrani et al. 2020).

Systematic breeding of quinoa is still in its infancy compared to major crops. Until recently, breeding has been mainly limited to Bolivia (Gandarillas 1979) and Peru (Bonifacio et al. 2013), which are the major growing areas of quinoa. Therefore, quinoa can be regarded as a partially domesticated crop. Many accessions suffer from seed shattering, branching, and non-appropriate plant height, which are typical domestication traits. Apart from these characters, grain yield and seed size, downy mildew resistance, synchronized maturity, stalk strength, and low saponin content are major breeding objectives (Gomez-Pando 2015). In the past years, activities have been intensified to breed quinoa genotypes adapted to temperate environments, for example, Europe, North America, and China (Murphy et al. 2018). Here, the major problem is the adaptation to long-day conditions because quinoa is predominantly a short-day plant due to its origin from regions near the equator.

There are only a few studies about the genetic diversity of quinoa. They were mainly based on phenotypic observations (Gandarillas 1979; Wilson 1988a) and low throughput marker systems like random amplified polymorphic DNA (Ruas et al. 1999), amplification fragment length polymorphisms (Rodríguez and Isla 2009), and microsatellites (Mason et al. 2005). A limited number of single nucleotide polymorphisms (SNP) based on expressed sequence tags were published (Coles et al. 2005). Maughan et al. (2012) used five bi-parental populations to identify ca. 14,000 SNPs, from which 511 KASP markers were developed. Genotyping 119 quinoa accessions gave the first insight into the population structure of this species (Maughan et al. 2012). Now, the availability of a reference genome enables genome-wide genotyping (Jarvis et al. 2017). Jarvis et al. (2017) re-sequenced 15 accessions and identified ca. 7.8 million SNPs. In another study, 11 quinoa accessions were re-sequenced, and 8 million SNPs and ca. 842 thousand indels were identified (Zhang et al. 2017).

Our study aimed to analyze the population structure of quinoa and patterns of variation by re-sequencing a diversity panel encompassing germplasm from all over the world. Using millions of markers, we performed a genome-wide association study using multiple-year field data. Here, we identified QTLs that control agronomically important traits important for breeding cultivars to be grown under long-day conditions. We are discussing the fundamental differences between an underutilized crop and crops with a long breeding history. Our results provide useful information for further understanding the genetic basis of agronomically important traits in quinoa and will be instrumental for future breeding.

2.3 Materials and Methods

2.3.1 Plant materials and growth conditions

We selected 350 quinoa accessions for phenotyping, and of these, 296 were re-sequenced in this study. Re-sequencing data of 14 additional accessions that had already been published (Jarvis et al. 2017) were also included in the study, together with the wild relatives (*C. belandieri* and *C. hircinum*) (Jarvis et al. 2017). These accessions represent different geographical regions of quinoa cultivation (Supplementary Fig. 1). Plants were grown in the field in Kiel, Northern Germany, in 2018 and 2019. Seeds were sown in the second week of April in 35x multi-tray

Chapter 2

pots. Then plants were transplanted to the field in the first week of May as single-row plots in a randomized complete block design with three blocks. The distances between rows and between plants were set to 60 cm and 20 cm, respectively. Each row plot contained seven plants per accession.

We recorded days to bolting (DTB) as BBCH51 and days to flowering (DTF) as BBCH60 twice a week during the growth period. Days to maturity (DTM) was determined when plants reached complete senescence (BBSHC94). If plants did not reach this stage, DTM was set as 250 days. In both years, plants were harvested in the second week of October. Plant height (PH), panicle length (PL), and the number of branches (NoB) were phenotyped at harvest. Stem lying (STL) (Supplementary Fig. 2) was scored on a scale from one to five, where score one indicates no stem lying. Similarly, panicle density was recorded on a scale from one to seven, where density one represents lax panicles, and panicle density seven represents highly dense panicles. Flower color and stem color were determined by visual observation. Pigmented and non-pigmented plants were scored as 1 and 0, respectively. Growth type (GT) was classified into two categories and analyzed as a dichotomous trait as well. We observed severe mildew infection in 2019. Therefore, we scored mildew infection on a scale from 1 to 3, where 1 equals no infection, and 3 equals severe infection.

2.3.2 Statistical analysis

We calculated the best linear unbiased estimates of the traits across years by fitting a linear mixed model using the lme4 R package (Bates et al. 2015). We used the following model:

$$Y_{ijk} = \mu + \text{Accession}_i + \text{Block}_j + \text{Year}_k + (\text{Accession} \times \text{Block})_{ij} + (\text{Accession} \times \text{Year})_{ik} + \text{Error}_{ijk}$$

Where μ is the mean, Accession_i is the genotype effect of the i -th accession, Block_j is the effect of the j -th Block, Year_k is the effect of the k -th year, $(\text{Accession} \times \text{Block})_{ij}$ is the Accession-Block interaction effect, $\text{Accession} \times \text{Year}_{ik}$ is the accession-year interaction effect, Error_{ijk} is the error of the j -th block in the k -th year. We treated all items as random effects for heritability estimation, and for best linear unbiased estimates (BLUE), accessions were treated as fixed effects. We analyzed the principle components of phenotypes using the R package FactoMineR (Lê et al. 2008).

2.3.3 Genome sequencing and identification of genomic variations

For DNA extraction, two plants per genotype were grown in a greenhouse at the University of Hohenheim, and two leaves from a single two-months old plant were collected and frozen immediately. DNA was subsequently extracted using the AX Gravity DNA extraction kit (A&A Biotechnology, Gdynia, Poland) following the manufacturer's instructions. The purity and quality of DNA were controlled by agarose gel electrophoresis and the concentration was determined with a Qubit instrument using SYBR green staining. Whole-genome sequencing was performed for 312 accessions at Novogene (China) using short-reads Illumina NovaSeq S4 Flowcell technology and yielded an average of 10 Gb of paired-end (PE) 2 x 150 bp reads with quality $Q > 30$ Phred score per sample, which is equivalent to $\sim 7X$ coverage of the haploid quinoa genome (~ 1.45 Gb). We then used an automated pipeline (<https://github.com/IBEXCluster/IBEX-SNPcaller/blob/master/workflow.sh>) compiled based on the Genome Analysis Toolkit. Raw sequence reads were filtered with trimmomatic-v0.38 (Bolger et al. 2014) using the following criteria: LEADING:20; TRAILING:20; SLIDINGWINDOW:5:20; MINLEN:50. The filtered paired-end reads were then individually mapped for each sample against an improved version of the QQ74 quinoa reference genome (CoGe id53523) using BWA-MEM (v-0.7.17) (Li and Durbin 2010) followed by sorting and indexing using samtools (v1.8) (Li et al. 2009). Duplicated reads were marked, and read groups

Chapter 2

were assigned using the Picard tools (<http://broadinstitute.github.io/picard/>). Variants were identified with GATK (v4.0.1.1) (McKenna et al. 2010) (Van der Auwera et al. 2013) using the “--emitRefConfidence” function of the HaplotypeCaller algorithm and “—heterozygosity” value set at 0.005 to call SNPs and InDels for each accession. Individual g.vcf files for each sample were then compressed and indexed with tabix (v-0.2.6) (Li 2011) and combined into chromosome g.vcf using GenomicsDBImport function of GATK. Joint genotyping was then performed for each chromosome using the function GenotypeGVCFs of GATK. To obtain high confidence variants, we excluded SNPs with the VariantFiltration function of GATK with the criteria: $QD < 2.0$; $FS > 60.0$; $MQ < 40.0$; $MQRankSum < -12.5$; $ReadPosRankSum < -8.0$ and $SOR > 3.0$. Then, SNP loci which contained more than 70% missing data, were filtered by VCFtools (Danecek et al. 2011) (v0.1.5), which resulted in our initial set of ~45 million SNPs for all the 332 accessions, including 20 previously re-sequenced accessions (Jarvis et al. 2017).

Our panel had three triplicates for quality checking and nine duplicates between Jarvis et al. 2017 and 312 newly re-sequenced accessions. To remove duplicates, as a preliminary analysis, we removed SNP loci with a minimum mean-depth < 5 across samples and SNP loci with more than 5% missing data. Then, we filtered SNPs with a minor allele frequency lower than 0.05 ($MAF < 0.05$). After these filtering steps, we obtained a VCF file that contained 229,017 SNPs. Then, we construct a maximum likelihood (ML) tree. First, we used the modelFinder (Kalyaanamoorthy et al. 2017) in IQ-TREE v1.6.619 (Nguyen et al. 2015) to determine the best model for ML tree construction. We selected GTR+F+R8 (GTR: General time-reversible, F: Empirical base frequencies, R8: FreeRate model) as the best fitting model according to the Bayesian Information Criterion (BIC) estimated by the software. We used 1000 replicates with ultrafast bootstrapping (UFboot)(Hoang et al. 2017) to check the reliability of the phylogenetic tree. To visualize the phylogenetic tree, we used the Interactive Tree Of Life tool (<https://itol.embl.de/>)(Letunic and Bork 2016). Then, based on the phylogenetic tree, we removed duplicate accessions and accessions with unclear identity. After the quality control, we retained 310 accessions (303 quinoa accessions and 7 wild *Chenopodium* accessions).

Then we used the initial SNP set and defined two subsets using the following criteria: (1) A base SNP set of 5,817,159 biallelic SNPs obtained by removing SNPs with more than 50% missing genotype data, minimum mean depth less than five, and minor allele frequency less than 1%. (2) A high confidence (HCSNP) set of 2,872,935 SNPs from the base SNP set by removing SNPs with a minor allele frequency of less than 5%. The base SNP set was used for the diversity statistics, and the HCSNPs set was used for GWAS analysis.

We annotated the HCSNP using SnpEff 4.3T (Cingolani et al. 2012) and a custom database (Cingolani et al. 2012) based on the QQ74 reference genome and annotation (CoGe id53523). Afterward, we extracted the SNP annotations using SnpSift (Ruden et al. 2012). Based on the annotations, SNPs were mainly categorized into five groups, (1) upstream of the transcript start site (5kb), (2) downstream of the transcript stop site (5kb), (3) coding sequence (CDS), (4) intergenic, and (5) intronic. We used SnpEff to categorize SNPs in coding regions based on their effects, such as synonymous, missense, splice acceptor, splice donor, splice region, start lost, start gained, stop lost, and spot retained.

2.3.4 Phylogenetic analysis and population structure analysis

For population structure analysis, we employed SNP subsets, as demonstrated in previous studies, to reduce the computational time (Wang et al. 2018). We created ten randomized SNP sets, each containing 50,000 SNPs. To create subsets, first, the base SNP set was split into 5000 subsets of an equal number of SNPs. Then, 10 SNPs from each subset were randomly selected,

Chapter 2

providing a total of 50,000 SNPs in a randomized set (randomized 50k set). We then repeated this procedure nine more times and finally obtained ten randomized 50k sets. Population structure analysis was conducted using ADMIXTURE (Version: 1.3) (Alexander et al. 2009). We ran ADMIXTURE for each subset separately with a predefined number of genetic clusters K from 2 to 10 and varying random seeds with 1000 bootstraps. Also, we performed the cross-validation (CV) procedure for each run. Obtained Q matrices were aligned using the greedy algorithm in the CLUMPP software (Jakobsson and Rosenberg 2007). Population structure plots were created using custom R scripts. We then combined SNP from the ten subsets to create a single SNP set of 434,077 unique SNPs for the phylogenetic analysis. We used the same method mentioned above to create the phylogenetic tree. Here we selected the model GTR+F+R6 based on the BIC estimates. For the principal component analysis (PCA) we used the HCSNP set and analysis was done in R package SNPrelate (Zheng et al. 2012). We estimated the top 10 principal components. The first (PC1) and second (PC2) were plotted using custom R scripts.

2.3.5 Genomic patterns of variations

Using the base SNP set, we calculated nucleotide diversity (π) for subpopulations and π ratios for Highland and Lowland population regions with the top 1% ratios of $\pi_{\text{Highland}}/\pi_{\text{Lowland}}$ candidate regions for population divergence. We also estimated Tajima's D values for both populations to check the influence of selection on populations. F_{ST} values were calculated between Highland and Lowland populations using the 10kb non-overlapping window approach. Nucleotide diversity, Tajima's D , and F_{ST} calculations were carried out in VCFtools (v0.1.5) (Danecek et al. 2011).

2.3.6 Linkage disequilibrium analysis

First, we calculated linkage disequilibrium in each population separately (Highland and Lowland). Then, LD was calculated in the whole population, excluding wild accessions. For LD calculations, we further filtered the HCSNP set by removing SNPs with >80% missing data (Varshney et al. 2019b). Using a set of 2,513,717 SNPs, we calculated the correlation coefficient (r^2) between SNPs up to 300kb apart by setting -MaxDist 300 and default parameters in the PopLDdecay software (Zhang et al. 2018). LD decay was plotted using custom R scripts based on the ggplot2 package.

2.3.7 Genome-wide association study

We used the best linear unbiased estimates (BLUE) of traits and HCSNPs for the GWAS analysis. Morphological traits were treated as dichotomous traits and analyzed using generalized mixed linear models with the lme4 R software package (Bates et al. 2015). We used population structure and genetic relationships among accessions to minimize false-positive associations. Population structure represented by the PC was estimated with the SNPrelate software (Zheng et al. 2012). Genetic relationships between accessions were represented by a kinship matrix calculated with the efficient mixed-model association expedited (EMMAX) software (Kang et al. 2010) using HCSNPs. Then, we performed an association analysis using the mixed linear model, including K and P matrices in EMMAX. We estimated the effective number of SNPs ($n=1,062,716$) using the Genetic type I Error Calculator (GEC) (Li et al. 2012). We set the significant P -value threshold (Bonferroni correction, $0.05/n$, $-\log_{10}(4.7e-08)=7.32$) and suggestive significant threshold ($1/n$, $-\log_{10}(9.41e-7)=6.02$) to identify significant loci underlying traits. We plotted SNP P -values on Manhattan plots using the qqman R package (Turner 2014).

2.4 Results

2.4.1 Re-sequencing 310 quinoa accessions reveals high sequence variation

We assembled a diversity panel made of 310 quinoa accessions representing regions of major geographical distributions of quinoa (Supplementary Fig. 1). The diversity panel comprises accessions with different breeding history (Supplementary Table 1). We included 14 accessions from a previous study, of which 7 are wild relatives (Jarvis et al. 2017). The sequence coverage ranged from 4.07 to 14.55, with an average coverage of 7.78. We mapped sequence reads to the reference genome V2 (CoGe id53523). Using mapping reads, we identified 45,330,710 single nucleotide polymorphisms (SNPs).

Table 3: Summary statistics of genome-wide single nucleotide polymorphisms identified in 310 quinoa accessions.

Parameter	Type	All genotypes (quinoa only)	Highland population	Lowland population
SNP	Total	2,872,935	2,590,907	1,938,225
	Intergenic	2,452,347	2,227,952	1,649,310
	Introns	251,481	101,546	172,692
	Exons	114,654	214,945	78,248
Nucleotide diversity			5.78×10^{-4}	3.56×10^{-4}
Tajima's <i>D</i>			0.884	-0.384
Population divergences	<i>F_{ST}</i> (Weighted_average)		0.36	

After filtering the initial set of SNPs, we identified 4.5 million SNPs in total for the base SNP set. We further filtered the SNPs for MAF >5 % (HCSNPs). We obtained 2.9 million high confident SNPs for subsequent analysis (Supplementary Table 2). Across the whole genome, SNP density was high, with an average of 2.39 SNPs/kb. However, SNP densities were highly variable between genomic regions and ranged from 0 to 122 SNPs/kb (Supplementary Fig. 3). We did not observe significant differences in SNP density between the two subgenomes (A subgenome 2.43 SNPs/kb; B subgenome 2.35 SNPs/kb). Then, we split the SNPs by their functional effects as determined by SnpEff (Cingolani et al. 2012). Among SNPs located in non-coding regions, 598,383 and 617,699 SNPs were located upstream (within 5kb from the transcript start site) and downstream (within 5kb from the stop site) of a gene, whereas 114,654 and 251,481 SNPs were located within exon and intron sequences, respectively (Table 3). We further searched for SNPs within coding regions. We found 70,604 missense SNPs and 41,914 synonymous SNPs within coding regions of 53,042 predicted gene models.

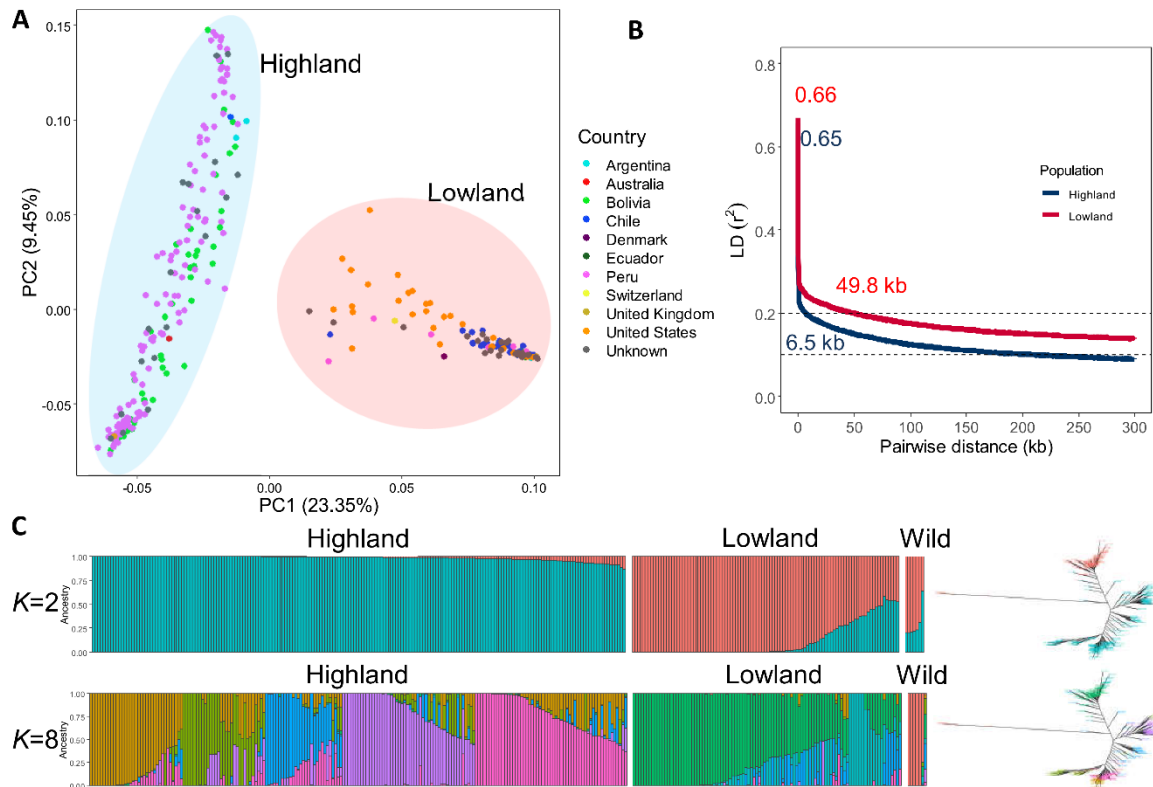


Figure 3: Genetic diversity and population structure of the quinoa diversity panel. (A) PCA of 303 quinoa accessions. PC1 and PC2 represent the first two components of analysis, accounting for 23.35 and 9.45% of the total variation, respectively. The colors of dots represent the origin of accessions. Two populations are highlighted by different colors: Highland (light blue) and Lowland (pink). (B) Subpopulation wise LD decay in Highland (blue) and Lowland population (red). (C) Population structure is based on ten subsets of SNPs, each containing 50,000 SNPs from the whole-genome SNP data. Model-based clustering was done in ADMIXTURE with different numbers of ancestral kinships ($K=2$ and $K=8$). $K=8$ was identified as the optimum number of populations. Left: Each vertical bar represents an accession, and color proportions on the bar correspond to the genetic ancestry. Right: Unrooted phylogenetic tree of the diversity panel. Colors correspond to the subpopulation.

2.4.2 Linkage disequilibrium and population structure of the quinoa diversity panel

Across the whole genome, LD decay between SNPs averaged 32.4 kb. We did not observe substantial LD differences between subgenome A (31.9 kb) and subgenome B (30.7 kb) (Supplementary Fig. 4C). The magnitude of LD decay among chromosomes did not vary drastically except for chromosome Cq6B, which exhibited a substantially slower LD decay (Supplementary Fig. 4A and B).

Then, we unraveled the population structure of the diversity panel. We performed principal component ($PCA_{(SNP)}$), population structure, and phylogenetic analyses. $PCA_{(SNP)}$ showed two main clusters consistent with previous studies (Christensen et al. 2007). The first and second principal components ($PC1_{(SNP)}$ and $PC2_{(SNP)}$) explained 23.35% and 9.45% of the variation, respectively (Figure 3). 202 (66.67%) accessions were assigned to subpopulation 1 (SP1) and 101 (33.33%) to subpopulation 2 (SP2). SP1 comprised mostly Highland accessions, whereas Lowland accessions were found in SP2. PCA demonstrated a higher genetic diversity of the Highland population (Figure 3A). We also calculated PCs for each chromosome separately. For 16 chromosomes, the same clustering as for the whole genome was calculated. Nevertheless, two

Chapter 2

chromosomes, Cq6B and Cq8B showed three distinct clusters (Supplementary Fig. 5). This is due to the split of the Lowland population into two clusters. We reason that gene introgressions on these two chromosomes from another interfertile group might have caused these differences. This is also supported by a slower LD decay on chromosome Cq6B (Supplementary Fig. 4B).

We also performed a population structure analysis with the ADMIXTURE software (Alexander et al. 2009). We used cross-validation to estimate the most suitable number of populations. Cross-validation error decreased as the K value increased, and we observed that after $K = 5$, cross-validation error reached a plateau (Supplementary Fig. 6B). We observed allelic admixtures in some accessions, likely owing to their breeding history. The wild accessions were also clearly separated at the smallest cross-validation error of $K=8$, except two *C. hircinum* accessions (Figure 3C). The reason for this could be that because *C. hircinum* is the closest crop wild relative, it also may have outcrossed with quinoa. The Highland population was structured into five groups, while the Lowland accessions were split into two subpopulations. The broad agro-climatic diversity of the Andean highland germplasm might have caused a higher number of subpopulations.

We analyzed the phylogenetic relationships between quinoa accessions using 434,077 SNPs. Constructing a maximum likelihood tree gave rise to five clades (Figure 4). We found that the placement of the wild quinoa species was concordant with the previous reports confirming that quinoa was domesticated from *C. hircinum* (Jarvis et al. 2017). However, we found that the *C. hircinum* accession BYU 566 (from Chile) was placed at the base of both Lowland and Highland clades, which is in contrast to (Jarvis et al. 2017), where this accession was placed at the base of coastal quinoa. As expected, accessions from the USA and Chile are closely related because the USDA germplasm had been collected at these geographical regions.

2.4.3 Genomic patterns of variations between Highland and Lowland quinoa

We were interested in patterns of variation in response to geographical diversification. We used principal component analysis derived clusters and phylogenetic analysis to define two diverged quinoa populations (namely Highland and Lowland). These divergent groups are highly correlated with South American highlands and lowlands geographical origin. We used the base SNP set to analyze diversity statistics. To detect genomic regions affected by the population differentiation, we measured the level of nucleotide diversity using 10 kb non-overlapping windows (Varshney et al. 2017). Then we calculated the whole genome-wide LD decay across the two populations (Highland vs. Lowland); LD decayed more rapidly in Highland quinoa (6.5 kb vs. 49.8 kb) (Figure 3B). To measure nucleotide diversity, we scanned the quinoa genome with non-overlapping windows of 10 kb in length in both populations separately. The nucleotide diversity of the Highland population (5.78×10^{-4}) was 1.62 fold higher compared to the Lowland population (3.56×10^{-4}) (Table 3 and Supplementary Fig. 7). We observed left-skewed distribution and negative Tajima's D value (-0.3883) in the Lowland populations indicating recent population growth (Table 3 and Supplementary Fig. 8). Genomic regions favorable for adaptation to Highlands should have substantially lower diversity in the Highland population than the Lowland population. Therefore, we calculated the nucleotide diversity ratios between Highland and Lowland to identify major genomic regions that are underlying the population differentiation. The F_{ST} value between populations was estimated to be 0.36, illustrating strong population differentiation. Concerning the regions of variants, the number of exonic SNPs is substantially higher in the Highland population (Table 3 and Supplementary Fig. 7).

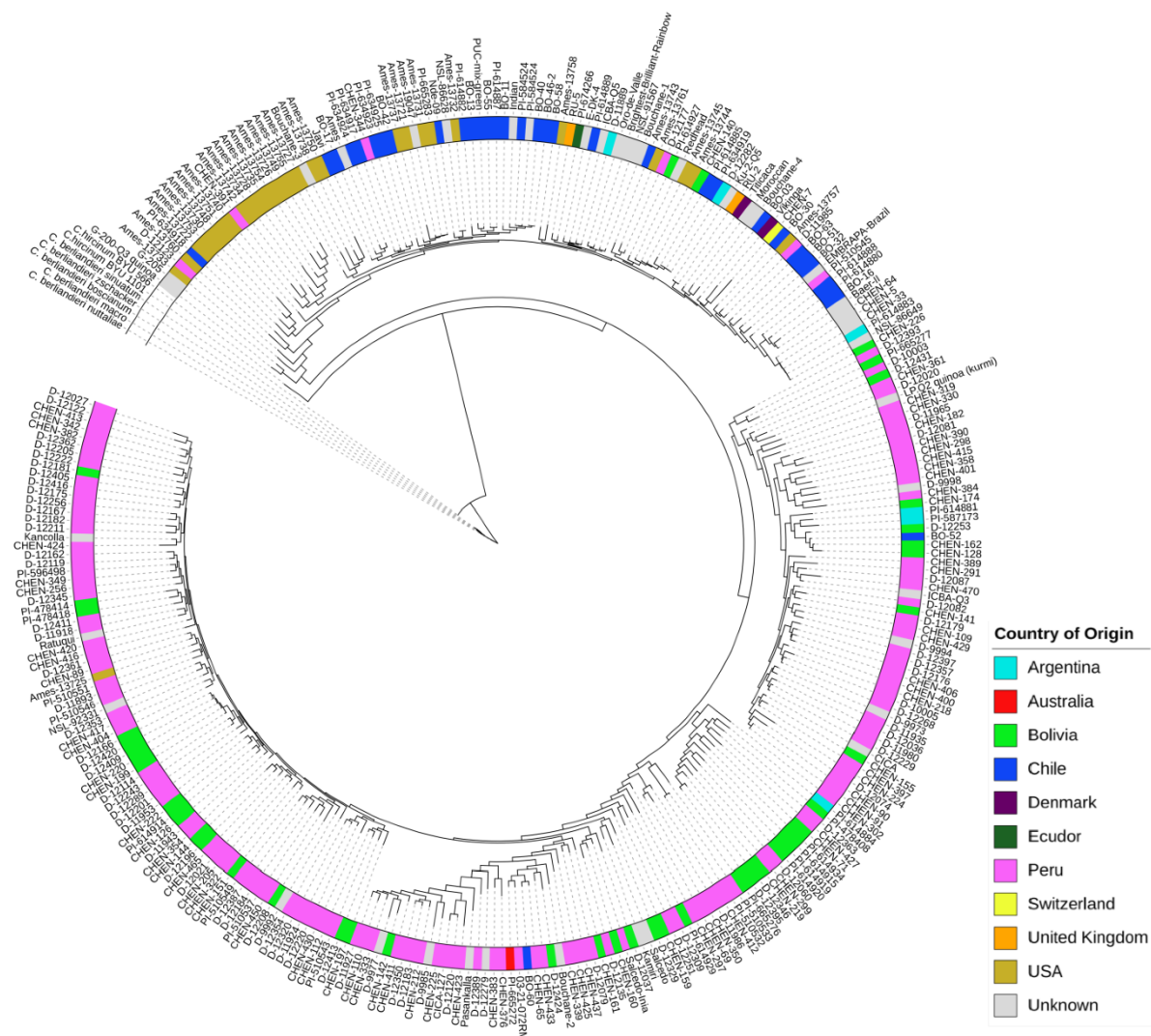


Figure 4: Maximum likelihood tree of 303 quinoa and seven wild Chenopodium accessions from the diversity panel. Colors are depicting the geographical origin of accessions.

2.4.4 Mapping agronomically important trait loci in the quinoa genome

We evaluated 13 qualitative and four dichotomous traits on 350 accessions across two different environments. At the time of the final harvest, 254 accessions did not reach maturity (senescence). All accessions produced seeds therefore used in seed analysis. For all traits, substantial phenotypic variation among accessions was found. High heritabilities were calculated for all quantitative traits except for NoB and STL, which indicates that the phenotypic variation between the accessions is mostly caused by genetic variation (Supplementary Table 3). Trait correlations between years were also high (Supplementary Fig. 10), which is in accordance with the heritability estimates. We found the strongest positive correlation between DTM and PL and PH and PL, whereas the strongest negative correlation was found between DTM and TSW (Supplementary Fig. 10). Then a principal component analysis was performed based on 12 quantitative traits ($PCA_{(PHEN)}$) to explore the phenotypic relationship among quinoa accessions. The first two principal components explained 62.12% of the phenotypic variation between the accessions. The score plot of the principal components showed a similar clustering pattern as the SNP based PCA analysis ($PCA_{(SNP)}$) (Figure 3A and Supplementary Fig. 11). $PCA_{(PHEN)}$ variables factor map indicated that most Lowland accessions were high yielding with high TSW and dense panicles. Moreover, these accessions are early flowering and early maturing, and they

are short (Supplementary Fig. 11B). Phenotype-based $PCA_{(PHEN)}$ also showed that the Lowland accessions are better adapted/selected for cultivation in long-day photoperiods compared to the Highland accessions. These results are in accordance with LD, nucleotide diversity, and Tajima's D estimations implying the Lowland accessions went through a stronger selection during breeding.

Then, we calculated the best linear unbiased estimates (BLUE) of the traits investigated. In total, 294 accessions shared the re-sequencing information and phenotypes out of 350 phenotypically evaluated accessions. For GWAS analysis, we used ~2.9 million high-confidence SNPs. In total, we identified 1480 significant ($P < 9.41e-7$) SNP-trait associations (MTA) for 17 traits (Supplementary Fig. 12). The number of MTAs ranged from 4 (STL) to 674 (DTM) (Supplementary Table 4). In agreement with previous reports, we defined an MTA as "consistent" when it was detected in both years (Varshney et al. 2019b). We identified 600 consistent MTAs across eleven traits. TSW and DTM showed the highest number of "consistent" associations. Among these, 143 MTAs are located within a gene, and 22 SNPs resulted in a missense mutation (Supplementary Table 5). MTA for DTB to DTF, NoB, Seed yield, STL, and GT were not "consistent" between years (Supplementary Table 4). This is also reflected by the low heritability estimations of these traits, indicating considerably higher genotype x environment interactions.

2.4.5 Candidate genes for agronomically important traits

First, we tested the resolution of our mapping study. We searched for major genes 50kb down- and upstream of significant SNPs for two qualitative traits in quinoa, flower color, and seed saponin content. We identified highly significant MTAs for stem color on chromosome Cq1B (69.72-69.76 Mb). There are two genes (*CqCYP76AD1* and *CqDODAI*) from the associated loci displaying high homology to betalain synthesis pathway genes *BvCYP76AD1* (Hatlestad et al. 2012) and *BvDODAI* (Bean et al. 2018) from sugar beet (Supplementary Fig. 14A and Supplementary Fig. 12). A significant MTA for saponin content on chromosome Cq5B between 8.85 Mb to 9.2 Mb harbored the two *BHLH25* genes which have been reported to control saponin content in quinoa (Jarvis et al. 2017) (Supplementary Fig. 14B and Supplementary Fig. 12). This demonstrates that the marker density is high enough to narrow down to causative genes underlying a trait.

Chapter 2

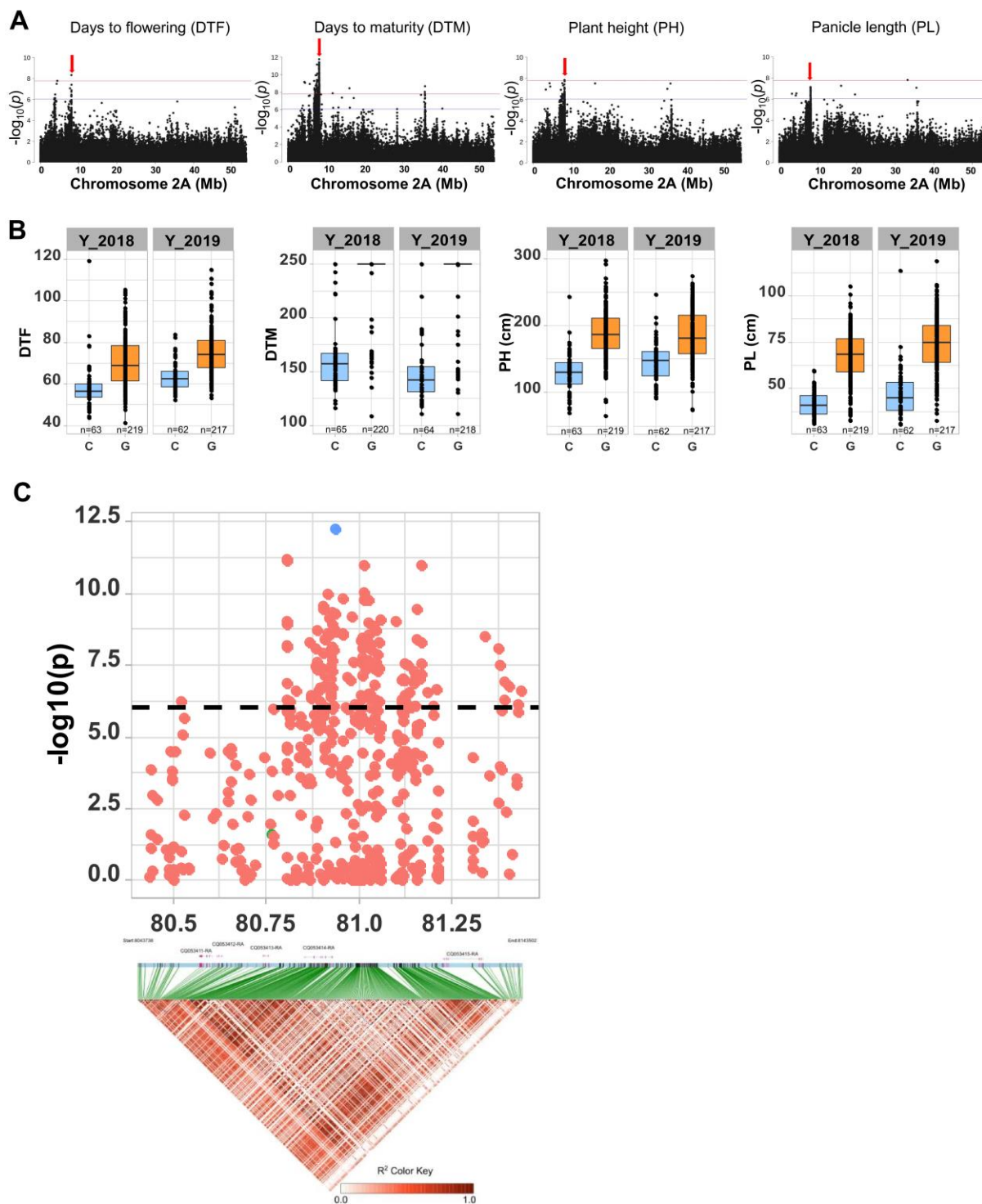


Figure 5: Genomic regions associated with important agronomic traits (A) Significant marker-trait associations for days to flowering, days to maturity, plant height, and panicle density on chromosome Cq2A. Red color arrows indicate the SNP loci pleiotropically acting on all four traits. (B) Boxplots showing the average performance for four traits over two years, depending on single nucleotide variation (C or G allele) within locus Cq2A_ 8093547. (C) Local Manhattan plot from region 80.40 - 81.43 Mb on chromosome Cq2A associated with PC1 of the DTF, DTM, PH and PL, and local LD heat map (bottom). The colors represent the pairwise correlation between individual SNPs. Green color dots represent the strongest MTA (Cq2A_ 8093547).

Chapter 2

Then, we examined four quantitative traits. We found seven MTA on chromosome Cq2A that are associated with DTF, DTM, PH, and PL (cross-phenotype association), indicating evidence for pleiotropic gene action (Figure 5 and Supplementary Table 6). For further confirmation and to investigate genes that are pleiotropically active on different traits, we followed a multivariate approach (Solovieff et al. 2013). First, we performed a PCA using the four phenotypes (cross-phenotypes). We found 89.94% of the variation could be explained by the first two principal components of the cross-phenotypes (PCA_(CP)) (Supplementary Fig. 15A). This indicates the adequate power of the PCA_(CP) to reduce dimensions for the analysis of the cross-phenotypes association. We observed similar clustering as in PCA (SNP). Therefore, these results indicate that in quinoa, DTF, DTM, PH, and PL are highly associated with population structure and thus, the adaptation to diverse environments. Then, we performed a GWAS analysis using the first three PCs as traits (PC-GWAS) (Supplementary Fig. 15C). We identified strong associations on chromosomes Cq2A, Cq7B (PC1), and Cq8B (PC2) (Supplementary Fig. 16). Out of 468 MTAs (PC1:426 and PC2:42) across the whole genome, 222 (PC1:211 and PC2:11) are located within 95 annotated genes. We found 14 SNPs that changed the amino acid sequence in 12 predicted protein sequences of associated genes (Supplementary Table 7). In the next step, we searched genes located within 50kb to an MTA. Altogether, 605 genes were identified (PC1:520 and PC2:85) (Supplementary Table 7).

We found the region 80.50 -81.50 Mb on chromosome Cq2A to be of special interest because it displays stable pleiotropic MTA for DTF, DTM, PH, and PL. The most significant SNP is located within the *CqGLX2-2* gene, which encodes an enzyme of the glyoxalase family (Figure 5). The Arabidopsis *GLX2-1* has been shown to be essential for growth under abiotic stress (Devanathan et al. 2014). The allele carrying a cytosine at the position with the most significant SNP resulted in early flowering, maturing, and short panicles and plants (Figure 5B). These traits are essential for the adaptation to long-day conditions.

Thousand seed weight is an important yield component. We found a strong MTA between 63.2 – 64.87 Mb on chromosome Cq8B. Significantly associated SNPs were localized within two genes (Figure 6). One gene displays homology to *PP2C* encoding a member of the phosphatase-2C (*PP2C*) protein family, which participates in Brassinosteroids signaling pathways and controls the expression of the transcription factor *BZR1* (Lu et al. 2017b). The second gene encodes a member of the RING-type E3 ubiquitin ligase family. These genes are controlling seed size in soybean, maize, rice, soybean, and Arabidopsis (Li et al. 2019). We then checked haplotype variation and identified 5 and 7 haplotypes for *CqPP2C* and *CqRING* genes, respectively. Accessions carrying *PP2C* hap3 and *RING* hap7 displayed larger seeds in both years (Figure 6 and Supplementary Fig. 17).

Downy mildew is one of the major diseases in quinoa, which causes massive yield damage. Notably, our GWAS identified strong MTA for resistance against this disease. The most significant SNPs are located in subgenome A (Supplementary Fig. 12). Thus, the A-genome progenitor seems to be the donor of downy mildew resistance. We identified a candidate gene within a region 38.99 - 39.03 Mb on chromosome Cq2A, which showed the highest significant association (Supplementary Fig. 14C). This gene encodes a protein with an NBS-LRR (nucleotide-binding site leucine-rich repeat) domain often found in resistance gene analogs with a function against mildew infection (Zhang et al. 2019).

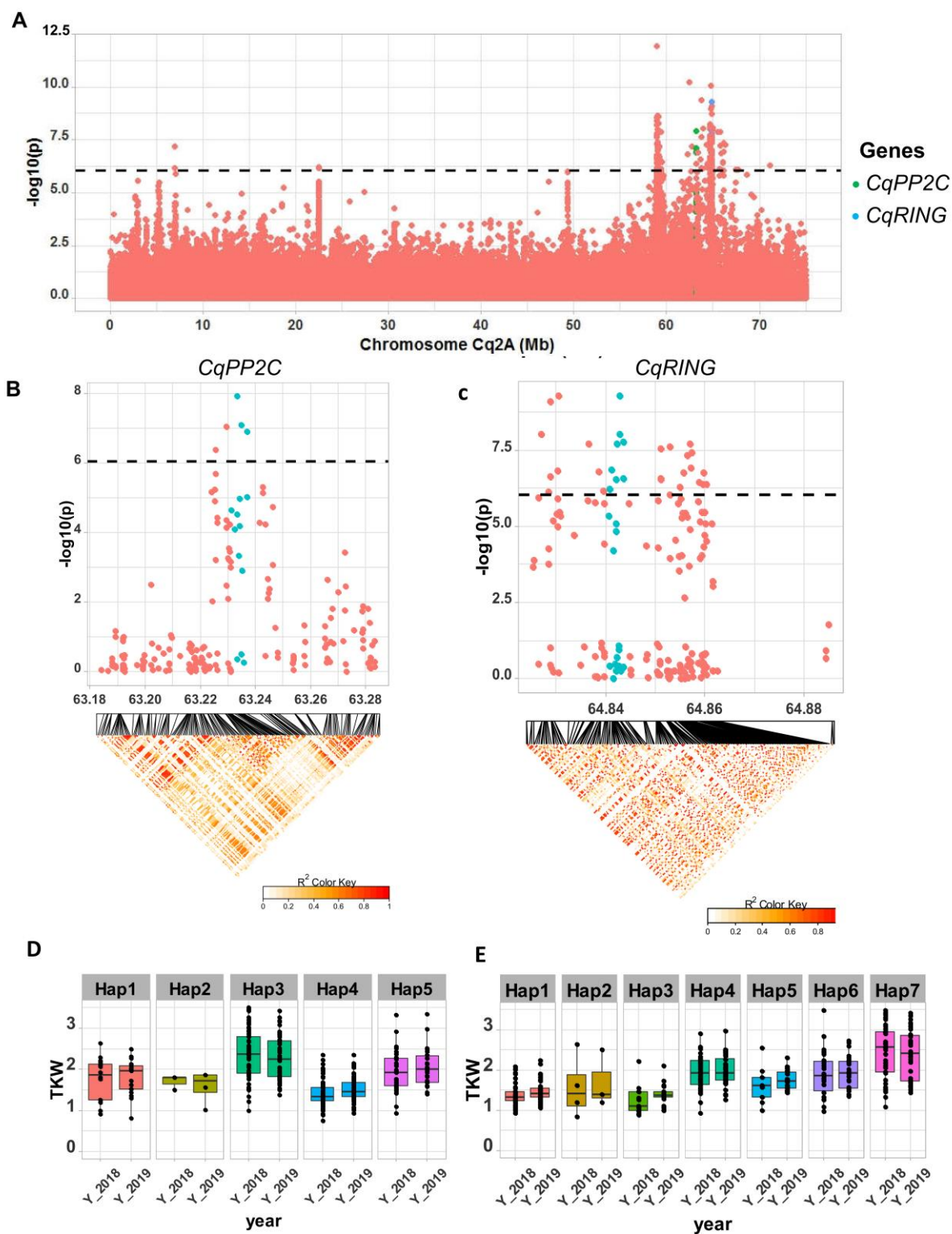


Figure 6: Identification of candidate genes for thousand seed weight. (A) Manhattan plot from chromosome Cq8B. Green and blue dots are depicting the *CqPP2C5* and the *CqRING* gene, respectively. (B) Top: Local Manhattan plot in the neighborhood of the *CqPP2C* gene. Bottom: LD heat map. (C) Top: Local Manhattan plot in the neighborhood of the *CqRING* gene. Bottom: LD heat map. Differences in thousand seed weight between five *CqPP2C* (D) and seven *CqRING* haplotypes (E).

2.5 Discussion

We assembled a diversity set of 303 quinoa accessions and seven accessions from wild relatives. Plants were grown under northern European conditions, and agronomically important traits were studied. In total, 2.9 million SNPs were found after re-sequencing. We found substantial phenotypic and genetic variation. Our diversity set was structured into two highly diverged populations, and genomic regions associated for this diversity were localized. Due to a high marker density, candidate genes controlling qualitative and quantitative traits were identified. The high genetic diversity and rapid LD breakdown are reflecting the short breeding history of this crop.

We were aiming to assemble the first diversity set, which represents the genetic variation of this species. Therefore, we established a permanent resource that is genotypically and phenotypically characterized. We believe that this collection is important for future studies due to the following reasons: We observed substantial phenotypic variation for all traits and high homogeneity within accessions. Moreover, low or absent phenotypic variation within accessions demonstrates homogeneity as expected for a self-pollinating species. Therefore, the sequence of one plant is representative of the whole accession, which is important for the power of the GWAS.

Today, over sixteen thousand accessions of quinoa are stored *ex-situ* in seed banks in more than 30 countries (Rojas et al. 2015). Despite the enormous diversity, only a few accessions have been genotyped with molecular markers. We found a clear differentiation into Highland and Lowland quinoa. In previous studies, five ecotypes had been distinguished: Valley type, Altiplano type, Salar type, Sea level type, and Subtropical type (Murphy et al. 2018). Adaptation to different altitudes, tolerance to abiotic stresses such as drought and salt, and photoperiodic responses are the major factors determining ecotypes (Gomez-Pando 2015). In our study, we could further allocate the quinoa accessions to five Highland and two Lowland subpopulations. This demonstrates the power of high-density SNP mapping to identify finer divisions at higher K . The origin of accessions and ecotype differentiation could be meaningfully interpreted by combining the information from phylogenetic data and population structure. As we expected, North American accessions (accessions obtained from USDA) were clustering with Chilean accessions, suggesting sequence-based characterization of ecotypes would be more informative and reproducible. Moreover, high-density SNP genotyping unveiled the origin of unknown or falsely labeled gene bank accessions, as recently proposed by Milner et al. (2019). The geographical origin of 52 accessions from our panel was unknown. We suggest using phylogenetic data and admixture results to complement the available passport data. For instance, two accessions with origin recorded as Chile are closely related to Peruvian and Bolivian accessions, which suggests that they are also originating from Highland quinoa.

What can we learn about the domestication of quinoa and its breeding history by comparing our results with data from other crops? LD decay is one parameter reflecting the intensity of breeding. LD decay in quinoa (32.4 kb) is faster than in most studies with major crop species, e.g. rapeseed (465.5 kb) (Wu et al. 2019), foxtail millet (*Setaria italica*, 100 kb) (Jia et al. 2013), pigeonpea (*Cajanus cajan*, 70 kb) (Varshney et al. 2017), soybean (150 kb) (Zhou et al. 2015) and rice (200 kb) (Mather et al. 2007). Although comparisons must be regarded with care due to different numbers of markers and accessions, different types of reproduction, and the selection intensity, the rapid LD decay in quinoa reflects its short breeding history and low selection intensity. Moreover, quinoa is a self-pollinating species where larger linkage blocks could be expected. However, cross-pollination rates in some accessions can be up to 17.36 % (Silvestri and Gil 2000), which is exploited by small Andean farmers who grow mixed quinoa accessions

Chapter 2

to ensure harvest under different biotic and abiotic stresses. This may facilitate a certain degree of cross-pollination and admixture.

Interestingly, the LD structure between Highland and Lowland populations is highly contrasting (6.5 vs. 49.8 kb), indicating larger LD blocks in the Lowland population. Low nucleotide diversity and negative Tajima's D were also observed in the Lowland population compared to Highland quinoa. The population differentiation index and LD differences have been used to test the hypothesis of multiple domestication events. As an example, different domestication bottlenecks have been reported for japonica (LD decay: 65 kb) and indica rice (LD decay: 200 kb) (Xu et al. 2012). The estimated F_{ST} value from this study (0.36) is in the similar range of F_{ST} estimates in rice subspecies *indica* and *japonica* (0.55) (Huang et al. 2010) and melon (*Cucumis melo*) subspecies *melo* and *agrestis* (0.46) (Zhao et al. 2019). Two hypotheses have been proposed for the domestication of quinoa from *C. hircinum*; (1) one event that gave rise to Highland quinoa and subsequently to Lowland quinoa and (2) two separate domestication events giving rise to Highland and Lowland quinoa independently (Jarvis et al. 2017). However, our study is not strictly following the second hypothesis because *C. hircinum* accession BYU 566 was basal to both clades of the phylogenetic tree (Highland and Lowland). Moreover, our wild *Chenopodium* germplasm does not represent enough diversity for in-depth analysis of domestication events. Therefore, we propose three possible scenarios to explain strong differences in LD structure, nucleotide diversity, Tajima's D and F_{ST} , (1) two independent domestication events with a strong bottleneck on lowland populations, (2) a single domestication but strong population growth after adaptation of lowland quinoa or (3) strong adaptive selection after domestication. To understand the history and genetics of domestication, it will be necessary to sequence a large representative set of outgroup species such as *C. berlandieri*, *C. hircinum*, *C. pallidicaule*, and *C. suecicum*.

Apart from marker density and sample size, the power of GWAS depends on the quality of the phenotypic data. Plants were grown in Northern Europe. Therefore, the MTAs are, first of all, relevant for temperate long-day climates. The share of genetic variances and thus, the heritabilities were high across environments. We expect higher genotype x environment interaction for flowering time, days to maturity, plant height, and panicle length if short-day environments will be included because many accessions have a strong day-length response (data not shown). Furthermore, the positions of genes controlling Mendelian traits were precisely coinciding with significant SNP positions, as exemplified by the genes associated with saponin content and flower color. Hence, the diversity panel provides sufficient power to identify SNP-trait associations for important agronomic traits such as TSW and downy mildew tolerance. In different plant species, seed size is controlled by six different pathways (Li et al. 2019). We found two important genes controlling seed size from the Brassinosteroid (*CqPP2C*) and the ubiquitin-proteasome (*CqRING*) pathway. The non-functional allele of soybean *PP2C1* resulted in small seeds (Lu et al. 2017b). We detected a superior haplotype (*PP2C_hap3*), which results in larger seeds. *CqRING* encodes an E3 ubiquitin ligase protein. There are two RING-type E3 ubiquitins known as *DA1* and *DA2*, which are involved in seed size controlling pathway. They were found in Arabidopsis rice, maize, and wheat. Downy mildew is the most acute disease for quinoa, caused by the fungus *Peronospora variabilis* (Choi et al. 2010). A recent study attempted identification of genes based on a GWAS analysis. However, no significant associations were found, probably due to the lack of power because of the small number of accessions used (61 and 88) (Colque-Little et al. 2020). In our study, a strong MTA suggests that the NBS-LRR gene on chromosome Cq2A contributes to downy mildew resistance in quinoa. We propose using this sequence for marker-assisted selection in segregating F₂ populations produced during pedigree breeding of quinoa.

Chapter 2

In this study, the advantage of multivariate analysis of cross-phenotype association became obvious. We could identify candidate genes with a pleiotropic effect on days to flowering, days to maturity, plant height, and panicle length. Interestingly, the most significant SNP was residing within a putative *GLX-2* ortholog. *GLX* genes, among other functions, have been shown to impact cell division and proliferation in *Amaranthus paniculatus* (Chakravarty and Sopory 1998). Therefore, the *CqGLX-2* gene is one candidate for controlling day length response.

This study also has a major breeding perspective. We aimed to elucidate the potential of quinoa for cultivation in temperate climates. Evidently, many accessions are not adapted to northern European climate and photoperiod conditions because they flowered too late and did not reach maturity before October. Nevertheless, 48 accessions are attractive as crossing partners for breeding programs because they are insensitive to photoperiod or long-day responsive. Moreover, they are attractive due to their short plant height, low tillering capacity, favorable inflorescence architecture, and high TSW. These are important characters for mechanical crop cultivation and combine harvesting. The MTA found in this study offers a perspective to use parents with superior phenotypes in crossing programs. We suggest a genotype building strategy by pyramiding favorable alleles (haplotypes). In this way, also accessions from our diversity set, which are not adapted to long-day conditions but with favorable agronomic characters, will be considered. Then, favorable genotypes will be identified from offspring generations by marker-assisted selection using markers in LD with significant SNPs. Furthermore, the MTA from this study will be useful for allele mining in quinoa germplasm collections to identify yet unexploited genetic variation.

2.6 Acknowledgments

We thank Prof.Dr. Mario Hasler for assisting with statistical analysis. We thank Monika Bruisch, Brigitte Neidhardt-Olf, Elisabeth Kokai-Kota, Verena Kowalewski, and Gabriele Fiene for technical assistance. The financial support of this work was provided by the Competitive Research Grant (Grant No. OSR-2016-CRG5- 466 2966-02) of the King Abdullah University of Science and Technology, Saudi Arabia and baseline funding from KAUST to Mark Tester.

2.7 Supplementary data

2.7.1 Supplementary figures

Supplementary Fig. 1: Geographical origin of the accessions forming the quinoa diversity panel.

Supplementary Fig. 2: Overview of the field experiment and exemplary images demonstrating phenotypic traits; (A) and (B): Overview of the field and phenotypic variation among accession; (C): Bolting (BBCH51) and (D) flowering (BBCH60) stage; Glomerulate (E) and amarantiform (F) panicle shapes; red (G) and green (H) stem color; red (I) and green (J) flower/inflorescence; Growth type 1 (K) and type 5 (L); (M): Plant height and maturity variation between two accessions.

Supplementary Fig. 3: SNP density heat map across the 18 quinoa chromosomes. Different colors depict SNP density.

Supplementary Fig. 4: Chromosome wide LD decay in genome A (A) and genome B (B). Colors are depicting different chromosomes. (C) Genome-wide average LD decay of the A sub-genome (blue) and B sub-genome (red).

Supplementary Fig. 5: SNP based PCA across all 18 quinoa chromosomes. Red circles are depicting the two clusters of Lowland accessions.

Supplementary Fig. 6 (A) ADMIXTURE ancestry coefficients for K ranging from 3 to 7 and 9. Each vertical bar represents an accession, and color proportions on the bar correspond to the genetic ancestry. (B) Cross-validation error in ADMIXTURE run.

Supplementary Fig. 7: Diversity of populations along chromosomes measured based on 10 kb non-overlapping windows. Nucleotide diversity (π) distribution of 10 kb windows in population Highland (A) and Lowland (B). (C) Nucleotide diversity ratios (π Lowland/ π Highland). (D) Pairwise genome-wide fixation index (FST) between Highland and Lowland. The broken horizontal line represents the top 1% threshold.

Supplementary Fig. 8: Distribution of Tajima's D along chromosomes in Highland (B) and Lowland (D) populations. Density distribution of Tajima's D between populations. Different colors represent the quartiles.

Supplementary Fig. 9: Graphical presentation of correlations between years among 12 traits. Pearson correlation value (R) with P-values are shown. DTB: days to bolting (inflorescence emergence), DTF: days to flowering, DTB to DTF: days between bolting and flowering, DTM; days to maturity, PH: plant height (cm), PL: panicle length (cm), PD: panicle density (cm), NoB: Number of branches, STL: stem lying, Saponin: saponin content as foam height (mm), Seed yield: seed yield per plant (g), TSW: thousand seed weight (g),

Supplementary Fig. 10: Pearson correlations among 12 quinoa traits. Best linear unbiased estimates across two years were used. Below the diagonal, scatter plots are shown with the fitted line in red. Above the diagonal, the Pearson correlation coefficients are shown with significance levels, *** = $p < 0.001$, ** = $p < 0.01$.

Supplementary Fig. 11: PCA of 12 quantitative phenotypes. A: Individual factor map colored according to populations identified from SNP analysis. B: Variables factor map of the PCA.

Supplementary Fig. 12: Manhattan plots from GWAS with data from 2018 (left), 2019 (center), and the mean of both years (right): The blue horizontal line indicates the suggestive threshold $-\log_{10}(8.98E-7)$. The red horizontal line indicates the significant threshold (Bonferroni correction) $-\log_{10}(1.67e-8)$.

Supplementary Fig. 13: Quantile-quantile plots of GWAS in two years, 2018 (left) and 2019 (center), and BLUE (right).

Supplementary Fig. 14: Local Manhattan plots for (A) flower color, (B) saponin content, and (C) mildew infection. Candidate genes are shown in the color legend. LD heat maps are placed at the Bottom. The colors of the heat map represent the pairwise correlation between individual SNPs.

Supplementary Fig. 15: PCA of 4 quantitative traits (DTF, DTM, PH, and PL). A: Individual factor map, B: variables factor map of the PCA, C: distribution of the first three principal components which were used for GWAS analysis.

Supplementary Fig. 16: GWAS analysis of principal components, PC1 (A), PC2 (B), PC3 (C): Manhattan plots (left), and quantile-quantile plots (right): The blue horizontal line in the

Chapter 2

Manhattan plots indicates the suggestive threshold $-\log_{10}(8.98E-7)$. The red horizontal line indicates the significance threshold (Bonferroni correction) $-\log_{10}(1.67e-8)$.

Supplementary Fig. 17: Haplotypes of two genes, CqPP2C and CqRING controlling seed size in quinoa. Geographic origin of the accessions and haplotype networks are displayed below the gene structure.

2.7.2 Supplementary tables

Supplementary Table 1: Accessions from the quinoa diversity panel and results from re-sequencing

Supplementary Table 2: Summary of high-quality SNPs identified in quinoa accessions

Supplementary Table 3: Variance components analysis of 12 quantitative traits

Supplementary Table 4: Summary of marker trait associations (MTA)

Supplementary Table 5: Candidate genes linked to SNP with significant trait associations

Supplementary Table 6: Summary of MTA associated with DTF, DTM, PD and PH identified on chromosome Cq2A

Supplementary Table 7: Candidate genes located within the 50kb flanking regions of significantly associated SNPs from the multivariate GWAS analysis

2.8 References

- Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome research* 19 (9):1655-1664
- Bates D, Mächler M, Bolker B, Walker S (2015) Fitting Linear Mixed-Effects Models Using lme4. *J Stat Soft* 67:48. doi:10.18637/jss.v067.i01
- Bean A, Sunnadaniya R, Akhavan N, Campbell A, Brown M, Lloyd A (2018) Gain-of-function mutations in beet *DODA2* identify key residues for betalain pigment evolution. *New Phytologist* 219 (1):287-296
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30 (15):2114-2120
- Bonifacio A, Gomez-Pando L, Rojas W (2013) Quinoa breeding and modern variety development. State of the Art Report on Quinoa Around the World
- Chakravarty T, Sopory S (1998) Blue light stimulation of cell proliferation and glyoxalase I activity in callus cultures of *Amaranthus paniculatus*. *Plant science* 132 (1):63-69
- Choi Y-J, Danielsen S, Lübeck M, Hong S-B, Delhey R, Shin H-D (2010) Morphological and molecular characterization of the causal agent of downy mildew on quinoa (*Chenopodium quinoa*). *Mycopathologia* 169 (5):403-412

Chapter 2

- Christensen SA, Pratt DB, Pratt C, Nelson PT, Stevens MR, Jellen EN, Coleman CE, Fairbanks DJ, Bonifacio A, Maughan PJ (2007) Assessment of genetic diversity in the USDA and CIP-FAO international nursery collections of quinoa (*Chenopodium quinoa* Willd.) using microsatellite markers. *Plant Genetic Resources: Characterisation and Utilisation* 5:82-95. doi:10.1017/S1479262107672293
- Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w¹¹¹⁸; iso-2; iso-3. *Fly* 6 (2):80-92
- Coles N, Coleman C, Christensen S, Jellen E, Stevens M, Bonifacio A, Rojas-Beltran J, Fairbanks D, Maughan P (2005) Development and use of an expressed sequenced tag library in quinoa (*Chenopodium quinoa* Willd.) for the discovery of single nucleotide polymorphisms. *Plant Science* 168 (2):439-447
- Colque-Little CX, Abondano MC, Lund OS, Amby DB, Piepho H-P, Andreasen C, Schmöckel SM, Schmid KJ (2020) Genetic variation for tolerance to the downy mildew pathogen *Peronospora variabilis* in genetic resources of quinoa (*Chenopodium quinoa*). bioRxiv
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Group GPA (2011) The variant call format and VCFtools. *Bioinformatics* 27 (15):2156-2158. doi:10.1093/bioinformatics/btr330
- Devanathan S, Erban A, Perez-Torres Jr R, Kopka J, Makaroff CA (2014) *Arabidopsis thaliana* glyoxalase 2-1 is required during abiotic stress but is not essential under normal plant growth. *PLoS One* 9 (4):e95971
- Emrani N, Hasler M, Patiranage DSR, Maldonado N, Rey E, Jung C (2020) An efficient method to produce segregating populations in quinoa (*Chenopodium quinoa* Willd.).
- Gandarillas H (1979) Botánica. In: Tapia ME (ed) Quinoa y kañiwa: cultivos Andinos. CIID, Bogotá,
- Gandarillas H, Alandia S, Cardozo A, Mujica A (1979) Quinoa y Kaniwa cultivos Andinos. Instituto Interamericano de Ciencias Agrícolas, Bogotá, Colombia
- Gomez-Pando L (2015) Quinoa breeding. *Quinoa: Improvement and Sustainable Production*:87-108
- González JA, Eisa S, Hussin S, Prado FE (2015) Quinoa: an Incan crop to face global changes in agriculture. *Quinoa: Improvement and sustainable production*:1-18
- Hatlestad GJ, Sunnadeniya RM, Akhavan Na, Gonzalez A, Goldman IL, McGrath JM, Lloyd AM (2012) The beet *R* locus encodes a new cytochrome P450 required for red betalain production. *Nature Genetics* 44:816-820. doi:10.1038/ng.2297
- Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS (2017) UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* 35 (2):518-522

Chapter 2

- Huang X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature genetics* 42 (11):961
- Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23 (14):1801-1806
- James LEA (2009) Quinoa (*Chenopodium quinoa* Willd.): composition, chemistry, nutritional, and functional properties. *Advances in food and nutrition research* 58:1-31
- Jarvis DE, Ho YS, Lightfoot DJ, Schmöckel SM, Li B, Borm TJ, Ohyanagi H, Mineta K, Mitchell CT, Saber N (2017) The genome of *Chenopodium quinoa*. *Nature* 542 (7641):307
- Jia G, Huang X, Zhi H, Zhao Y, Zhao Q, Li W, Chai Y, Yang L, Liu K, Lu H (2013) A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). *Nature genetics* 45 (8):957-961
- Kalyaanamoorthy S, Minh BQ, Wong TK, von Haeseler A, Jermin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature methods* 14 (6):587
- Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, Sabatti C, Eskin E (2010) Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics* 42:348-354. doi:10.1038/ng.548
- Kolano B, Siwinska D, Pando LG, Szymanowska-Pulka J, Maluszynska J (2012) Genome size variation in *Chenopodium quinoa* (Chenopodiaceae). *Plant systematics and evolution* 298 (1):251-255
- Lê S, Josse J, Husson F (2008) FactoMineR: an R package for multivariate analysis. *Journal of statistical software* 25 (1):1-18
- Letunic I, Bork P (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research* 44 (W1):W242-W245. doi:10.1093/nar/gkw290
- Li H (2011) Tabix: fast retrieval of sequence features from generic TAB-delimited files. *Bioinformatics* 27 (5):718-719
- Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26 (5):589-595
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25 (16):2078-2079
- Li M-X, Yeung JM, Cherny SS, Sham PC (2012) Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Human genetics* 131 (5):747-756

Chapter 2

- Li N, Xu R, Li Y (2019) Molecular networks of seed size control in plants. *Annual review of plant biology* 70:435-463
- Li T, Yang X, Yu Y, Si X, Zhai X, Zhang H, Dong W, Gao C, Xu C (2018) Domestication of wild tomato is accelerated by genome editing. *Nature biotechnology* 36 (12):1160-1163
- Lu X, Xiong Q, Cheng T, Li Q-T, Liu X-L, Bi Y-D, Li W, Zhang W-K, Ma B, Lai Y-C (2017) A *PP2C-1* allele underlying a quantitative trait locus enhances soybean 100-seed weight. *Molecular Plant* 10 (5):670-684
- Mason S, Stevens M, Jellen E, Bonifacio A, Fairbanks D, Coleman C, McCarty R, Rasmussen A, Maughan P (2005) Development and use of microsatellite markers for germplasm characterization in quinoa (*Chenopodium quinoa* Willd.). *Crop Science* 45 (4):1618-1630
- Mather KA, Caicedo AL, Polato NR, Olsen KM, McCouch S, Purugganan MD (2007) The extent of linkage disequilibrium in rice (*Oryza sativa* L.). *Genetics* 177 (4):2223-2232
- Maughan PJ, Chaney L, Lightfoot DJ, Cox BJ, Tester M, Jellen EN, Jarvis DE (2019) Mitochondrial and chloroplast genomes provide insights into the evolutionary origins of quinoa (*Chenopodium quinoa* Willd.). *Scientific Reports* 9:185. doi:10.1038/s41598-018-36693-6
- Maughan PJ, Smith SM, Rojas-Beltrán JA, Elzinga D, Raney JA, Jellen EN, Bonifacio A, Udall JA, Fairbanks DJ (2012) Single Nucleotide Polymorphism Identification, Characterization, and Linkage Mapping in Quinoa. *The Plant Genome Journal* 5:114. doi:10.3835/plantgenome2012.06.0011
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* 20 (9):1297-1303
- Milner SG, Jost M, Taketa S, Mazón ER, Himmelbach A, Oppermann M, Weise S, Knüpfper H, Basterrechea M, König P, Schüler D, Sharma R, Pasam RK, Rutten T, Guo G, Xu D, Zhang J, Herren G, Müller T, Krattinger SG, Keller B, Jiang Y, González MY, Zhao Y, Habekuß A, Färber S, Ordon F, Lange M, Börner A, Graner A, Reif JC, Scholz U, Mascher M, Stein N (2019) Genebank genomics highlights the diversity of a global barley collection. *Nature Genetics* 51:319-326. doi:10.1038/s41588-018-0266-x
- Murphy KM, Matanguihan JB, Fuentes FF, Gómez-Pando LR, Jellen EN, Maughan PJ, Jarvis DE (2018) Quinoa breeding and genomics. *Plant Breeding Reviews* 42:257-320
- Palomino G, Hernández LT, de la Cruz Torres E (2008) Nuclear genome size and chromosome analysis in *Chenopodium quinoa* and *C. berlandieri* subsp. *nuttalliae*. *Euphytica* 164 (1):221
- Peterson A, Jacobsen SE, Bonifacio A, Murphy K (2015) A crossing method for Quinoa. *Sustainability (Switzerland)* 7:3230-3243. doi:10.3390/su7033230

Chapter 2

- Rodríguez LA, Isla MT (2009) Comparative analysis of genetic and morphologic diversity among quinoa accessions (*Chenopodium quinoa* Willd.) of the South of Chile and highland accessions. *Journal of Plant Breeding and Crop Science* 1 (5):210-216
- Rojas W, Pinto M, Alanoca C, Gómez Pando L, León-Lobos P, Alercia A, Diulgheroff S, Padulosi S, Bazile D (2015) Quinoa genetic resources and ex situ conservation.
- Ruas PM, Bonifacio A, Ruas CF, Fairbanks DJ, Andersen WR (1999) Genetic relationship among 19 accessions of six species of *Chenopodium* L., by Random Amplified Polymorphic DNA fragments (RAPD). *Euphytica* 105 (1):25-32
- Ruden DM, Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Lu X (2012) Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, SnpSift. *Frontiers in genetics* 3:35
- Ruiz KB, Biondi S, Oses R, Acuña-Rodríguez IS, Antognoni F, Martínez-Mosqueira EA, Coulibaly A, Canahua-Murillo A, Pinto M, Zurita-Silva A (2014) Quinoa biodiversity and sustainability for food security under climate change. A review. *Agronomy for sustainable development* 34 (2):349-359
- Silvestri V, Gil F (2000) Alogamia en quinua. *Revista de la Facultad de Ciencias Agrarias* 32 (1)
- Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW (2013) Pleiotropy in complex traits: challenges and strategies. *Nature Reviews Genetics* 14 (7):483-495
- Stetter MG, Gates DJ, Mei W, Ross-Ibarra J (2017) How to make a domesticate. *Current Biology* 27 (17):R896-R900
- Turner SD (2014) qqman: an R package for visualizing GWAS results using QQ and manhattan plots. *Biorxiv*:005165
- Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J (2013) From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current protocols in bioinformatics* 43 (1):11.10. 11-11.10. 33
- Varshney RK, Saxena RK, Upadhyaya HD, Khan AW, Yu Y, Kim C, Rathore A, Kim D, Kim J, An S (2017a) Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. *Nature Genetics* 49 (7):1082
- Varshney RK, Shi C, Thudi M, Mariac C, Wallace J, Qi P, Zhang H, Zhao Y, Wang X, Rathore A (2017b) Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. *Nature biotechnology* 35 (10):969-976
- Varshney RK, Thudi M, Roorkiwal M, He W, Upadhyaya HD, Yang W, Bajaj P, Cubry P, Rathore A, Jian J, Doddamani D, Khan AW, Garg V, Chitikineni A, Xu D, Gaur PM, Singh NP, Chaturvedi SK, Nadigatla GVPR, Krishnamurthy L, Dixit GP, Fikre A, Kimurto PK, Sreeman SM, Bharadwaj C, Tripathi S, Wang J, Lee SH, Edwards D, Polavarapu KKB, Penmetsa RV, Crossa J, Nguyen HT, Siddique KHM, Colmer TD, Sutton T, von Wettberg E, Vigouroux Y, Xu X, Liu X (2019) Resequencing of 429

Chapter 2

chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. *Nature Genetics* 51:857-864. doi:10.1038/s41588-019-0401-3

Vega-Gálvez A, Miranda M, Vergara J, Uribe E, Puente L, Martínez EA (2010) Nutrition facts and functional potential of quinoa (*Chenopodium quinoa* willd.), an ancient Andean grain: a review. *Journal of the Science of Food and Agriculture* 90 (15):2541-2547

Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, Wu Z, Li M, Zheng T, Fuentes RR, Zhang F (2018) Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 557 (7703):43-49

Wilson HD (1988) Allozyme variation and morphological relationships of *Chenopodium hircinum* (s.l.). *Syst Bot* 13. doi:10.2307/2419100

Wu D, Liang Z, Yan T, Xu Y, Xuan L, Tang J, Zhou G, Lohwasser U, Hua S, Wang H (2019) Whole-genome resequencing of a worldwide collection of rapeseed accessions reveals the genetic basis of ecotype divergence. *Molecular plant* 12 (1):30-43

Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, Dong Y, Gutenkunst RN, Fang L, Huang L (2012) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nature biotechnology* 30 (1):105-111

Zhang C, Dong S-S, Xu J-Y, He W-M, Yang T-L (2018) PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35 (10):1786-1788. doi:10.1093/bioinformatics/bty875

Zhang R, Zheng F, Wei S, Zhang S, Li G, Cao P, Zhao S (2019) Evolution of disease defense genes and their regulators in plants. *International Journal of Molecular Sciences* 20 (2):335

Zhang T, Gu M, Liu Y, Lv Y, Zhou L, Lu H, Liang S, Bao H, Zhao H (2017) Development of novel InDel markers and genetic diversity in *Chenopodium quinoa* through whole-genome re-sequencing. *BMC Genomics* 18 (1):685. doi:10.1186/s12864-017-4093-8

Zhao G, Lian Q, Zhang Z, Fu Q, He Y, Ma S, Ruggieri V, Monforte AJ, Wang P, Julca I (2019) A comprehensive genome variation map of melon identifies multiple domestication events and loci influencing agronomic traits. *Nature Genetics* 51 (11):1607-1615

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28 (24):3326-3328. doi:10.1093/bioinformatics/bts606

Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, Li W, Yu Y, Shu L, Zhao Y, Ma Y (2015) Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nature biotechnology* 33 (4):408-414

3 Haplotype variations of major flowering time genes in quinoa unveil their role in the adaptation to different environmental conditions

3.1 Summary

Response to photoperiod is of major importance in crop production. It defines the adaptation of plants to local environments. Quinoa is a short day plant which had been domesticated in the Andean regions. We wanted to understand the adaptation to long-day conditions by studying orthologs of two major flowering time regulators of *Arabidopsis*, *FLOWERING LOCUS T (FT)* and *CONSTANS (CO)* in quinoa accessions with contrasting photoperiod response. By searching the quinoa reference genome sequence, we identified 24 *FT* and six *CO* homologs. *CqFT* genes displayed remarkably different expression patterns between long-day and short-day conditions, whereas the influence of the photoperiod on *CqCOL* expressions was moderate. Cultivation of 276 quinoa accessions under short-day and long-day conditions revealed great differences in photoperiod sensitivity. After sequencing their genomes, we identified large sequence variations in 12 flowering time genes. We found non-random distribution of haplotypes across accessions from different geographical origins, highlighting the role of *CqFT* and *CqCOL* genes in the adaptation to different day-length conditions. We identified five haplotypes causing early flowering under long days. This study provides assets for quinoa breeding because superior haplotypes can be assembled in a predictive breeding approach to produce well-adapted early flowering lines under long-day photoperiods.

3.2 Introduction

Quinoa (*Chenopodium quinoa* Willd.) is a traditional Andean crop that was domesticated around 5000-7000 years ago. The production and consumption of quinoa have rapidly expanded in recent years due to its broad adaptability and nutritional value (Murphy and Matanguihan 2015). The crop is widely adaptable to different agroecological conditions due to its ecological plasticity and diversity (González et al. 2015). Quinoa germplasm possesses a broad genetic diversity, which can benefit its adaptation and cultivation in Northern Europe. Understanding the underlying genetics of flowering is essential for the fast improvement of the crop.

The onset of flowering is preceded by a major transition of the shoot apical meristem resulting in floral organ primordia formation. This transition results from signal cascades involving transcription factors and epigenetic modification, which is well understood in the model plant *Arabidopsis thaliana*. *FLOWERING LOCUS T (FT)* is the major floral activator integrating signals from upstream regulators, which are indirectly responding to photoperiod, temperature, plant age, and hormones (Blümel et al. 2015). *FT* is expressed in leaves, and the protein encoded is transferred to the shoot apical meristem to initiate flowering in combination with other genes. In *Arabidopsis*, under long-day (LD) conditions *FT* expression peaks in the morning and around dusk (Song et al. 2018). Photoperiod is an essential factor that affects flowering in plants, and most of the genes involved are highly conserved among crops (Eshed and Lippman 2019). *CONSTANS (CO)* is a key factor in photoperiodic sensing, and it directly governs the expression of *FT* under LD conditions. *CO* is controlled by circadian clock genes, and expression peaks at the end of the day. *CO* encodes for two adjacent B-BOX (BBX) and CCT (*CO*, *CO-like*, *TOC*) domains. The *CO* protein is stable during the light, it is degraded during the dark. *CO* is not involved flowering under short-day (SD) A recent study showed that flowering was promoted under SD conditions in *co* mutants, indicating its role in delaying flowering under SDs (Luccioni et al. 2019).

Quinoa belongs to the plant family Amaranthaceae together with sugar beet (*Beta vulgaris*), spinach, amaranth, *Chenopodium album*, and *C. rubrum*. Sugar beet is an LD species but needs vernalization followed by LDs for flowering. It has two *FT* homologs, *BvFT1* and *BvFT2*, that

Chapter 3

act antagonistically. Before vernalization, *BvFT1* is active and suppresses floral transition. After exposure to a long period of cold temperatures, *BvFT1* expression decreases, and *BvFT2* becomes transcriptionally active to induce flowering (Pin et al. 2010). Sugar beet is lacking a true *CO* ortholog, which encodes both CCT and BBX domains. Two *CONSTANS-LIKE (COL)* genes have been investigated for their involvement in floral induction in sugar beet. *BvCOL1* complemented flowering in an *Arabidopsis co* mutant, but it did not show the typical *CO* expression pattern (Chia et al. 2008), nor was there any genetic evidence for its function as a flowering time gene. However, two other genes have been proven to be upstream regulators of *BvFT1* and *BvFT2*. *BOLTING TIME CONTROL 1 (BTC1)* is a *PSEUDO RESPONSE REGULATOR (PRR)* gene with a CCT domain but lacking the BBX domains (Pin et al. 2012), whereas *BvBBX19* carries the BBX but is lacking the CCT domain (Dally et al. 2014). Both proteins form a dimer, thus acquiring a CO function to regulate flowering (Dally et al. 2018).

Flowering time regulation has also been studied in the SD species *C. rubrum*. This wild relative of beet and quinoa has two *FT* homologs; *CrFTL1* and *CrFTL2* (Cháb et al. 2008), but only *CrFTL1* was found to be essential to initiate flowering (Drabešová et al. 2014b). In the same study, *CrCOL1* and *CrCOL2* genes were investigated. Their expression patterns were similar to beet *COL* genes, but they did not complement *Arabidopsis co* mutants. A recent transcriptome study in two other close relatives, *C. ficifolium* and *C. suecicum*, showed that the expression of *FT* like (*CfFTL1* and *CsFTL1*) genes is correlated with flowering since they showed characteristic diurnal *FT* expression patterns. However, under LDs, their expression was negligible (Štorchová et al. 2019), indicating *C. ficifolium* flowers under LD without upregulation of *FT*-like genes.

Quinoa is a short day species owing to its origin. Nevertheless, it has been shown that quinoa can flower under varying photoperiods depending on the genotype. Therefore, quinoa has been regarded as a facultative SD crop (Bertero et al. 1999b). The day-length response varies vastly between accessions from different geographical regions (Bertero 2003b). In a study with quinoa germplasm from Bolivia, the latest accession flowered 85 days after the earliest accession (Rojas 2003). Strong photoperiod sensitivity is the major hurdle for quinoa cultivation under LD conditions in temperate regions such as Northern Europe and Southern Chile. So far, flowering time research in quinoa has been in its infancy. Christiansen et al. (2010) investigated the phenological development of two quinoa accessions with contrasting days to maturity under different photoperiod regimes (SD, LD, and reciprocal transfer). They speculated that day-length sensitivity in quinoa is governed by a few genes only, but the mode of inheritance and position of genes in the genome could not be identified. Recently, a reference genome sequence has been published, which laid the foundation to investigate flowering time genes on a genome-wide scale. Five *FT* homologs were identified as putative floral inducers. Sequence similarity to *BvFT* suggested their putative function as floral activators/repressors (Jarvis et al. 2017). A study was performed to uncover the evolutionary history of *FT*-like genes in quinoa and identified 14 *FTL* genes belonging to seven *FTL* families. This study proposed a unified gene nomenclature of *FT* genes in the Amaranthaceae family species to avoid confusion when comparisons are made between species. Interestingly members of the *FTL3* gene family specific for Amaranthaceae species, were also identified in quinoa, and their potential role in the adaptation to different environments was hypothesized (Štorchová, 2020). Another genome-wide survey uncovered 611 sequences with some homology to *Arabidopsis* flowering time genes and 459 quinoa specific flowering genes. This study was based on a bioinformatics approach that employed BLAST search, orthologue, and collinearity identification (Golicz 2019). However, genetic or functional evidence surmounting their role as flowering time regulators is still lacking.

This study aimed to understand the molecular reasons of day-length sensitivity in quinoa. The general hypothesis was that *FT* and *COL* homologs play a similar role in quinoa as in sugar beet. However, we expected divergent functions of at least one *FT* ortholog because beet has a strong requirement for vernalization. Selected quinoa genotypes with contrasting day-length requirements were grown under controlled conditions, and the transcriptional activities of six *FT* and six *COL* homologs were analyzed at different stages of development. We found strong day-length responses of *CqFT* genes, whereas the influence of the photoperiod on *CqCOL* expressions was moderate. We further analyzed haplotypes of 12 genes based on SNPs identified from a diversity panel of 303 quinoa accessions. For some genes, haplotypes were non randomly distributed among accession with different geographical origins. This indicates the function of these genes as major factors for the adaptation to varying photoperiods. We identified accessions carrying specific haplotypes, which do not respond to photoperiod, whereas others respond to photoperiod. We then compared the geographical distribution of haplotypes. We found that individual haplotypes are more frequent in the southern latitude compared to other haplotypes. Our results shed new light on the regulation of photoperiod response and vernalization requirement in related species with contrasting life cycle regimes. The data will be important for quinoa breeding because haplotypes can be assembled in a predictive breeding approach to produce early flowering lines well adapted to northern latitudes.

3.3 Results

3.3.1 Flowering time variation under different photoperiodic conditions

Among 303 accessions from our quinoa diversity panel, 276 accessions were selected for the growth experiment (E1) under LD and SD conditions. We observed a broader variation for DTF under LDs (CV=19.66%) compared to SDs (CV=12.4%). Under LD conditions, DTF ranged from 38 to 123 days (mean: 60.02 ± 10.4), whereas under SDs, it ranged from 35 to 72 days (mean: 50 ± 5.2 days). We identified all three types of photoperiodic response ('short day', 'long day', and 'day neutral') among the 276 accessions. We identified 67.39% of the accessions to be 'short day' accessions, which flowered significantly earlier under SD conditions. We also found that 16.67% and 7.6% of the accessions are 'day neutral' and 'long day' accessions, respectively. Then, we compared DTF under long- and short- days with the geographical origin of the accessions. Notably, we found that accessions from southern Chile are photoperiod insensitive, which indicates that day-length sensitivity was lost when cultivation was extended towards lower latitudes (Table 4 and Supplementary Fig. 18).

Table 4: Summary statistics of photoperiod response of 276 quinoa accessions. Ten plants from each accession were grown in a growth chamber with 16h, and 8h light and DTF was determined. Estimation graphics with 5000 bootstraps at 99% confidence interval was used to compare DTF of each accession under SDs and LDs.

Photoperiodic sensitivity	No. of accessions	Percentage %
Short day responsive ^a	186	67.39
Day neutral	46	16.67
Long day responsive ^b	21	7.61
Not assigned	23	8.33

^a flowering earlier under SD

^b flowering earlier under LD

3.3.2 Expression analysis of putative *FT* and *COL* orthologs

We reasoned that putative *COL* and *FT* orthologs function in quinoa in the same way as in their close relatives. Searching the quinoa genome for *FT* homologs, we identified 24 *PEBP* (*phosphatidyl ethanolamine-binding proteins*) family genes, five of which had been identified

Chapter 3

before as putative *FT* orthologs (Jarvis et al. 2017). Protein identity between 24 *PEBP* genes and BvFT2 ranged from 40% to 92% (Supplementary Table 9 and Supplementary Fig. 21). We generated phylogenetic trees to analyze the phylogenetic relationship of the candidate genes. We identified 5, 2, 4, and 13 genes belonging to *TFL1*-like, *BFT*-like, *MFT*-like, and *FT*-like clades, respectively. Like in *C. rubrum*, the *FT*-like clade could be split into three subclades, which were belonging to *FTL1*, *FTL2*, and *FTL3*. Seventeen *PEBP* genes were encoding the complete protein with the PEBP domain (Supplementary Fig. 21A). We observed discrepancies between reference annotation genome version 1 and improved version 2, resulting from differences between genome assemblies. Unfortunately, genes have been given different names by Jarvis et al. (2017) and Štorchová (2020). Therefore, we provide a side-by-side comparison of all genes from both publications in Table S2. We identified six *COL* genes closely resembling *AtCOL2*, *AtCOL4*, and *AtCOL5*. Quinoa *COL* proteins showed 40% to 89% protein identity to BvCOL1 and 37% to 55% identity to CO (Supplementary Table 10 and Supplementary Fig. 24). All genes encode two B-boxes and CCT domains. A phylogenetic tree was constructed with sequences from CO and *COL* genes of quinoa and four other species. *CqCOL* genes were distributed across three clades, demonstrating their resemblance to beet and *Chenopodium* spp. *COL* genes (Supplementary Fig. 21B).

Then, we chose six *FTL* genes and six *COL* genes for expression analysis, which showed the highest protein similarity to BvFT2 and BvCOL, respectively (Supplementary Table 11). We questioned whether their transcriptional activities would respond to different photoperiods and growth stages. We also expected different expressions between ‘early’ and ‘late’ genotypes. Therefore, we grew four quinoa accessions with strikingly different flowering times under LDs and SDs (E2). We observed more pronounced vegetative growth under LDs than under SDs. Moreover, we observed that red color pigmentation in the stem is more prominent under LDs than under SDs (Supplementary Fig. 19).

We also phenotyped those four accessions for DTB and DTF (E2). The ‘early’ accessions Titicaca and PI-614886 started bolting 23 and 27 DAS under SDs and LDs, respectively, indicating that both accessions are not responding to photoperiod. In contrast, the ‘late’ accessions were bolting differently. The accession PI-587173 bolted nine days earlier under SDs, whereas accession CHEN-109 bolted five days earlier under LDs. As expected, accessions Titicaca and PI-614886 flowered earlier than PI-587173 and CHEN-109 under both conditions. Accessions Titicaca and PI-614886 did not respond to different day-length, whereas PI-614886 and PI-587173 flowered earlier under SDs (3 and 19 days, respectively) (Supplementary Fig. 20). We decided to use the ‘day-neutral’ accession Titicaca (early) and the ‘short-day’ accession PI-587173 (late) for expression profiling (E2), due to their contrasting flowering time under different photoperiodic regimes, and due to their contrasting responses to photoperiod.

Chapter 3

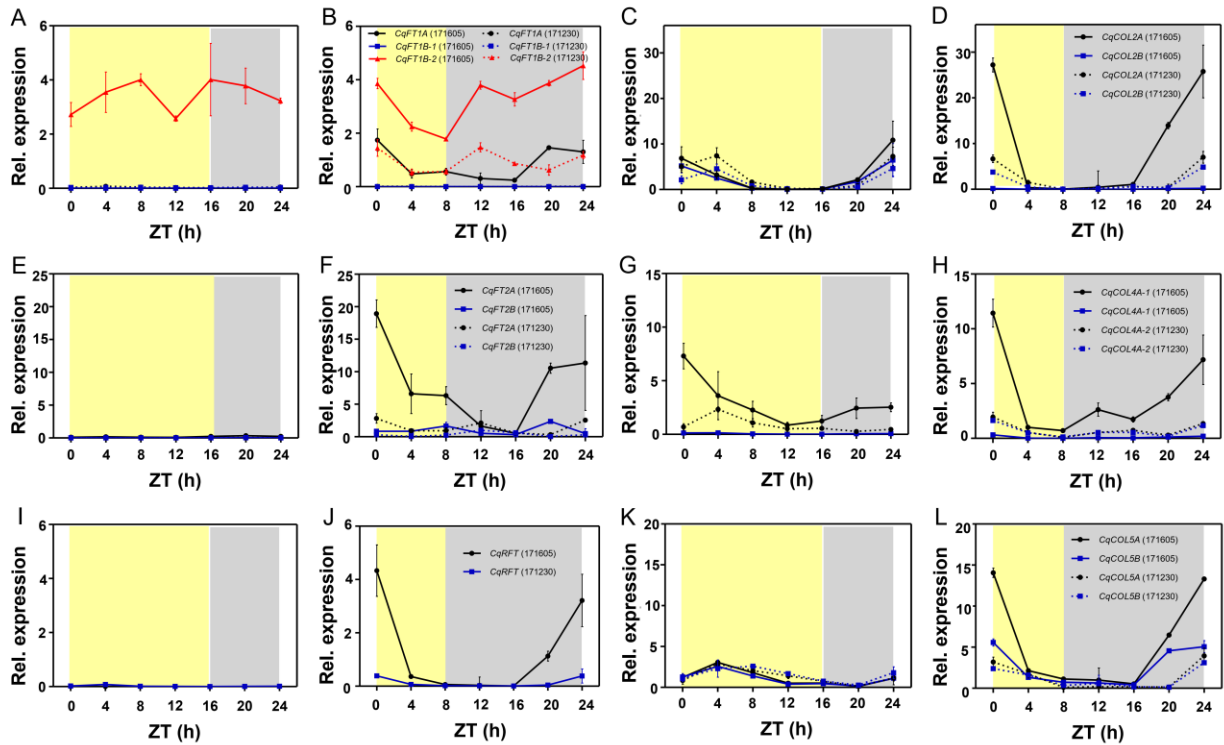


Figure 7: Expression profiles of 12 *FT*-like and *COL* genes in an early (Titicaca) and a late quinoa accession (PI-587173) grown under LDs (A, C, E, G, I, K) and SDs (B, D, F, H, J, L). RNA was isolated at the bolting stage. ZT, zeitgeber time. Continuous lines: PI-587173; Broken lines: Titicaca; Error bars, \pm SEM. Data are normalized against the geometric mean of *CqPTB* and *CqIDH-A*.

In *Arabidopsis* and sugar beet, *FT* and *COL* genes are diurnally expressed. We reasoned that their putative orthologs from quinoa are also diurnally regulated under different photoperiods. We collected leaves from two accessions (E2) at the bolting stage. We observed significantly different expression patterns of all six *CqFT*-like genes investigated between LDs and SDs (Figure 7) in both accessions. The transcriptional activities of all *CqFT* genes were negligible under LDs except for *CqFT1B-2*, which showed stable expression under LDs (accession PI-587173). All *CqFT* genes except *CqFT1B-1* were diurnally regulated under SDs, although at varying levels of expression. *CqFT2A*, *CqFT1A*, and *CqRFT* showed similar expression patterns under SDs; those genes showed the highest expression in the morning and gradually reduced expression during the day. The lowest expression was observed at 16ZT. In general, expression levels of *CqFT* genes in the early accession Titicaca were relatively low compared to the late accession under both LDs and SDs. Diurnal expression of *CqFT2B* under SDs in accession PI-587173 showed late evening peak similar to floral activators (Pin et al. 2010; Hayama et al. 2007; Andres and Coupland 2012), but the expression level was lower than *CqFT2A* (Figure 7).

Chapter 3

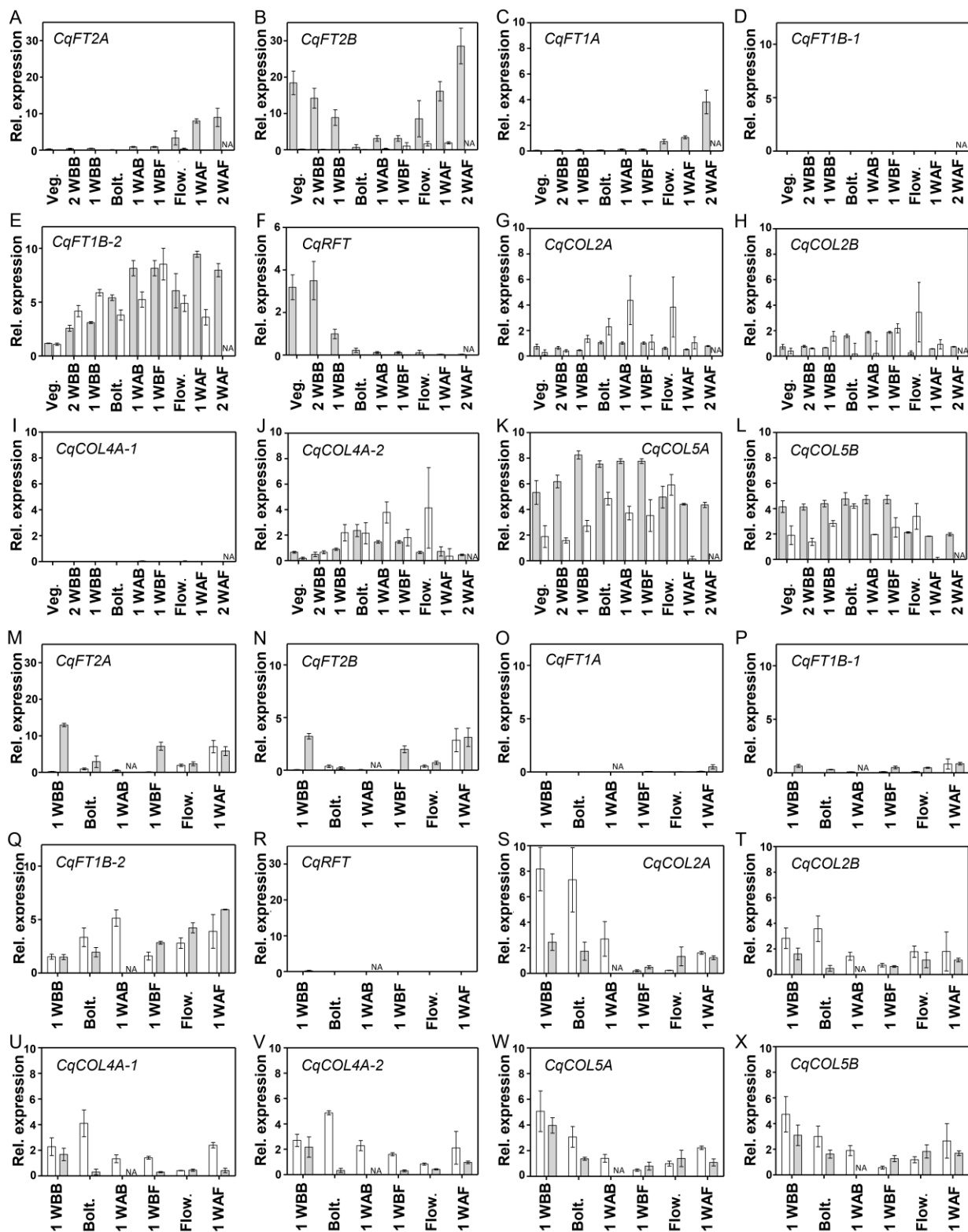


Figure 8: Expression profiles at different stages of development of six *FT*-like and six *COL* genes in the late accession PI-587173 (A-L) and in the early accession Titicaca (M-X). Plants were grown in the climate chamber under LDs (open boxes) and SDs (filled boxes). ZT: zeitgeber time, Error bars: \pm SEM, Data were normalized against the geometric mean of *CqPTB* and *CqIDH-A*. (Veg.: Vegetative, WBB: Week/s before bolting, Bolt.: Bolting, WAB: Week/s after bolting, WBF: Week/s before flowering, Flow.: Flowering, WAF: Week/s after flowering)

Then, we analyzed the expression pattern of six *CqCOL* genes. *CqCOL4A-2* (LD and SD) and *CqCOL2B* (SD) showed very low expression, while the other four genes showed diurnal expression under both conditions (Figure 7). Generally, the relative expression of *CqCOL* genes was higher under SDs. The expression levels in the early accession (Titicaca) were lower compared to the late accession (PI-587173). In Titicaca, all *CqCOL* genes showed a peak after 4h of light under LDs. The *CqCOL5* gene showed different expression patterns between SDs and LDs, but expression patterns were not different between accessions (Figure 7). Taken together, SD expression patterns of *CqCOL* genes were similar to what has been reported for sugar beet *BvCOL1* (Chia et al. 2008) and Arabidopsis *COL1* and *COL2* (Ledger et al. 2001).

As a next step, we analyzed the temporal expression of the candidate genes taking leaf samples at ZT4 over 11 weeks, starting from the second week after sowing (E2). We found strikingly different expression profiles between genes in response to day-length. Under LDs, we could not detect any *CqFT2A* transcripts in the late accession; however, in the early accession, both paralogs were expressed at the flowering stage and one week after flowering (Figure 8). In contrast, under SDs, the expression of *CqFT2B* was higher at the vegetative stage and decreased until bolting. *CqFT2A* expression could be detected only at the flowering stage and at later developmental stages under SDs in the late accession. It was expressed one week before bolting in the early accession. In the early accession, both *CqFT2* genes showed similar expression patterns under SDs. Also, *CqFT1A* expression was similar to *CqFT2A* expression under SDs. Moreover, we did not detect *CqFT1B-1* expression under SD and LD in the late accession, whereas we observed a weak expression in the early accession. *CqFT1B-2* expression showed similar expression patterns under both conditions in both accessions, indicating that it is not responding to photoperiod. The *CqRFT* was, on the other hand, influenced by the photoperiod, and under SDs, its expression gradually dropped with the developmental stage. In contrast, the *CqRFT* expression of the early accession was very low or undetectable under both photoperiods (Figure 8).

CqCOL expression patterns were drastically different between the accessions. In the early accession, *CqCOL* genes were higher expressed at the early stages of development and gradually decreased during later developmental stages, showing similar expression patterns like the sugar beet floral repressor *BvFT1* (Pin et al. 2010). *CqCOL2A*, *CqCOL2B*, and *CqCOL4A-1* showed similar expression patterns in the late accession, while *CqCOL4A-2* was expressed at a very low rate. In the late accession, the *CqCOL5A* and *CqCOL5B* genes showed opposite expression patterns compared to the *CqFT2B* gene under SDs (Figure 8). This may indicate that the *CqCOL5A* and *CqCOL5B* are upstream repressors of *CqFT2B*.

3.3.3 Sequence variation in *FT* and *COL* orthologs are correspond with the geographical origin and flowering time

We analyzed the haplotypes of *CqFT* genes and *CqCOL* genes based on high confidence SNPs (HC-SNPs) identified from a diversity panel of 303 quinoa accessions. We excluded individuals with heterozygous SNPs and missing SNPs in the corresponding gene during haplotype identification. We identified 2 to 6 haplotypes across all genes investigated (49 haplotypes in total) (Table 5).

Then, we studied the association between sequence variation in 12 putative flowering time regulators and flowering time under SDs and LDs. Moreover, we analyzed whether sequence variations altered the function of the predicted proteins. *CqFT1B-1* displayed the highest sequence variation (5 haplotypes) among *CqFT* genes, whereas *CqCOL2A* showed the highest

sequence variation (6 haplotypes) among *CqCOL* genes. In contrast, seven genes did not display any SNPs causing amino acid (aa) substitutions (*CqFT1B-2*, *CqFT2A*, *CqFT2B*, *CqRFT*, *CqCOL4A-1*, *CqCOL5A*, *CqCOL5B*), indicating high protein conservation. The highest number of aa substitutions was identified in the *CqCOL2A* gene. Both *CqCOL5* genes were highly conserved, and SNPs were only found in the promoter regions (Figure 9 and Supplementary Fig. 22).

Table 5: Classification of haplotypes based on their photoperiod sensitivity (*P*-values obtained from LD and SD mean comparison). For haplotype identification, high-quality SNP data were used from 303 accessions of a worldwide core collection. Haplotypes were identified based on HC-SNP. Accessions with missing or heterozygous SNPs within the respective gene were excluded from the analysis.

Gene	No. of Accessions	Haplotype		
		Insensitive (<i>P</i> >0.05)	Moderately sensitive (0.001< <i>P</i> <0.05)	Highly sensitive (<i>P</i> <0.001)
<i>CqFT1A</i>	197	<i>CqFT1A_c</i>	<i>CqFT1A_a</i>	<i>CqFT1A_b</i>
		<i>CqFT1A_d</i>		
<i>CqFT1B-1</i>	168		<i>CqFT1B-1_b</i>	<i>CqFT1B-1_a</i>
			<i>CqFT1B-1_c</i>	
			<i>CqFT1B-1_d</i>	
			<i>CqFT1B-1_e</i>	
<i>CqFT1B-2</i>	216	<i>CqFT1B-2_b</i>	<i>CqFT1B-2_a</i>	<i>CqFT1B-2_c</i>
				<i>CqFT1B-2_d</i>
<i>CqFT2A</i>	183	<i>CqFT2A_e</i>	<i>CqFT2A_b</i>	<i>CqFT2A_a</i>
		<i>CqFT2A_d</i>		<i>CqFT2A_c</i>
<i>CqFT2B</i>	185	<i>CqFT2B_e</i>		<i>CqFT2B_a</i>
		<i>CqFT2B_f</i>		<i>CqFT2B_b</i>
				<i>CqFT2B_c</i>
				<i>CqFT2B_d</i>
<i>CqRFT</i>	213		<i>CqRFT_b</i>	<i>CqRFT_a</i>
				<i>CqRFT_c</i>
				<i>CqRFT_d</i>
<i>CqCOL2A</i>	142	<i>CqCOL2A_e</i>	<i>CqCOL2A_c</i>	<i>CqCOL2A_a</i>
		<i>CqCOL2A_f</i>	<i>CqCOL2A_d</i>	<i>CqCOL2A_b</i>
<i>CqCOL2B</i>	145		<i>CqCOL2B_b</i>	<i>CqCOL2B_a</i>
				<i>CqCOL2B_c</i>
<i>CqCOL4A-1</i>	223	<i>CqCOL4A-1_b</i>		<i>CqCOL4A-1_a</i>
<i>CqCOL4A-2</i>	222		<i>CqCOL4A-2_b</i>	<i>CqCOL4A-2_a</i>
				<i>CqCOL4A-2_c</i>
<i>CqCOL5A</i>	176	<i>CqCOL5A_b</i>	<i>CqCOL5A_c</i>	<i>CqCOL5A_a</i>
		<i>CqCOL5A_d</i>		
<i>CqCOL5B</i>	212		<i>CqCOL5B_c</i>	<i>CqCOL5B_a</i>
				<i>CqCOL5B_b</i>

We then analyzed haplotype variation related to day length response. Seven haplotypes (*CqFT1B-1_b*, *CqFT1B-1_e*, *CqFT2A_a*, *CqRFT_d*, *CqCOL4A-2_c*, *CqCOL5A_a*, and *CqCOL5A_c*) resulted in earlier flowering under SDs, whereas accessions with five haplotypes (*CqFT1A_a*, *CqFT2B_c*, *CqCOL2A_c*, *CqCOL2B_b*, and *CqCOL5B_c*) flowered earlier under LDs. We identified 12 haplotypes of six genes (*CqFT1A*, *CqFT1B-2*, *CqFT2A*, *CqFT2B*, *CqCOL2A*, *CqCOL4A-1*, and *CqCOL5A*) not significantly associated with photoperiod response. Besides, we identified 14

Chapter 3

haplotypes with moderate photoperiod sensitivity, whereas the remaining haplotypes were highly sensitive to photoperiod (Table 5 and Supplementary Fig. 22).

We hypothesize that the adaptation to LD conditions at Southern (e.g., Chile) and Northern latitudes was due to mutations within *CqFT* and *CqCOL* genes. Therefore, we grouped the haplotypes according to their abundance on geographical distribution. We found six genes (*CqFT1A*, *CqFT2B*, *CqCOL2A*, *CqCOL2B*, *CqCOL5A*, and *CqCOL5B*) with strong associations between geographical origin and haplotype (Figure 9 and Supplementary Fig. 22). For example, Southern Chilean accessions predominantly carry haplotype *CqFT1A_a*, whereas most of the Peruvian and Highland accessions have haplotype *CqFT1A_b* (Figure 9). Also, haplotype *CqFT2B_c* (Supplementary Fig. 22) and *CqCOL5B_c* (Figure 9) were mostly present in Southern Chilean accessions. These alleles should be preferentially used in European and North American breeding programs.

Haplotypes of six genes (*CqFT1B-1*, *CqFT1B-2*, *CqFT2A*, *CqRFT*, *CqCOL4A-1*, and *CqCOL4A-2*) were equally dispersed across all geographical regions (Supplementary Fig. 22). We did not compare the geographical origin and haplotypes of the accession obtained from USDA gene banks and accessions grown in Europe because the exact origin of those accessions was not clear.

Chapter 3

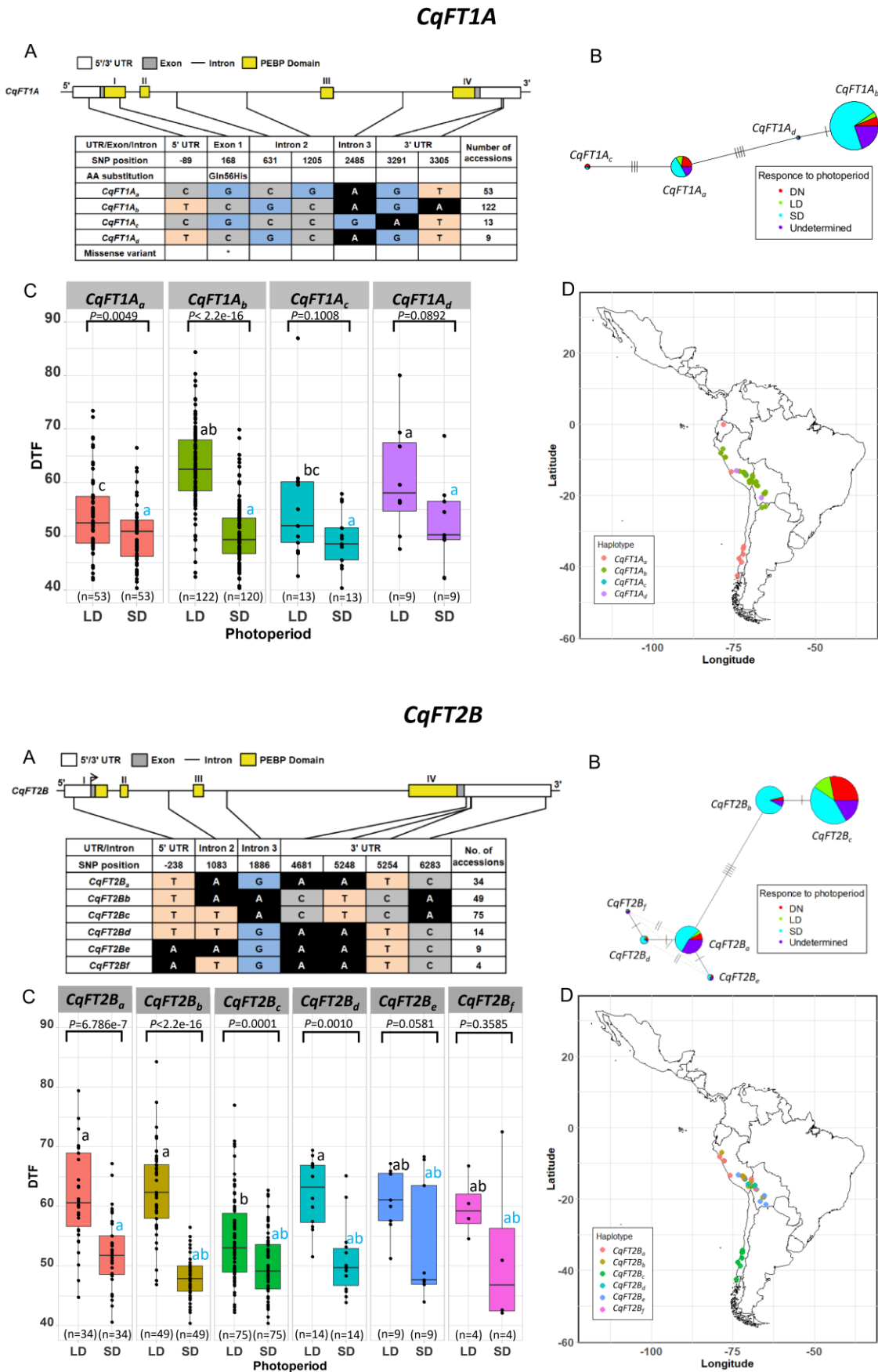
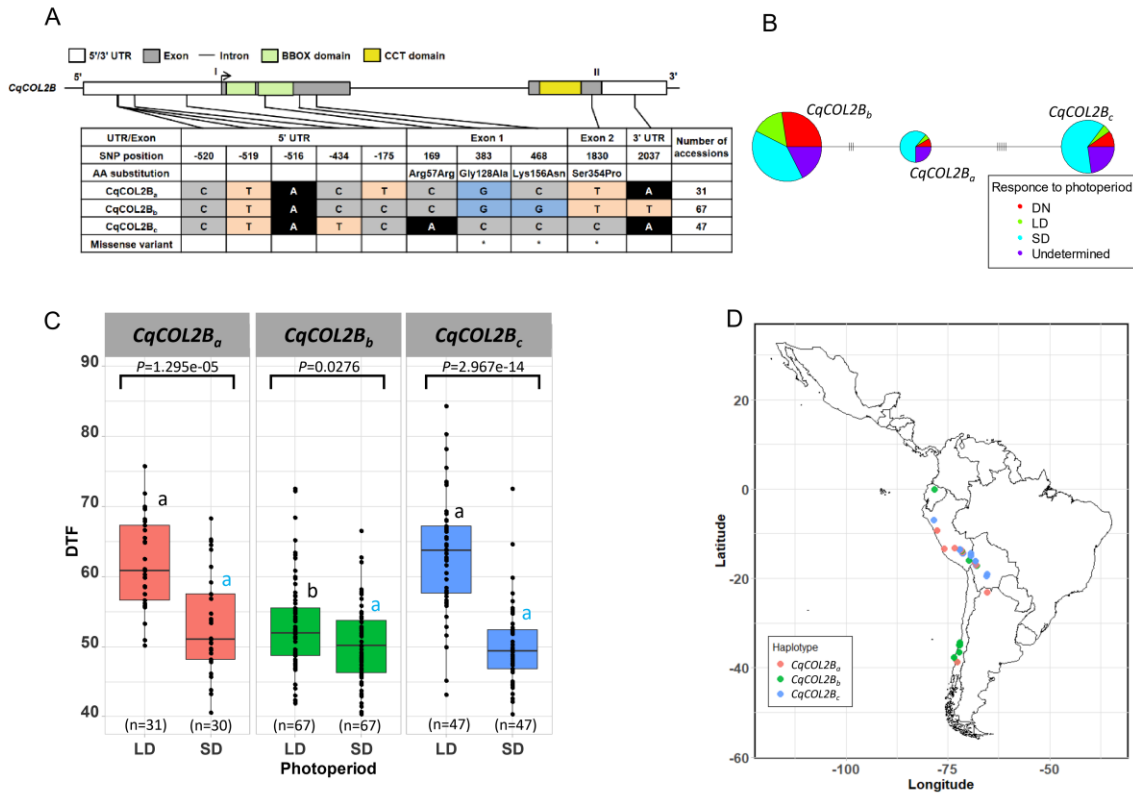


Figure 9: (continued on next page, legend follows).

Chapter 3

CqCOL2B



CqCOL5B

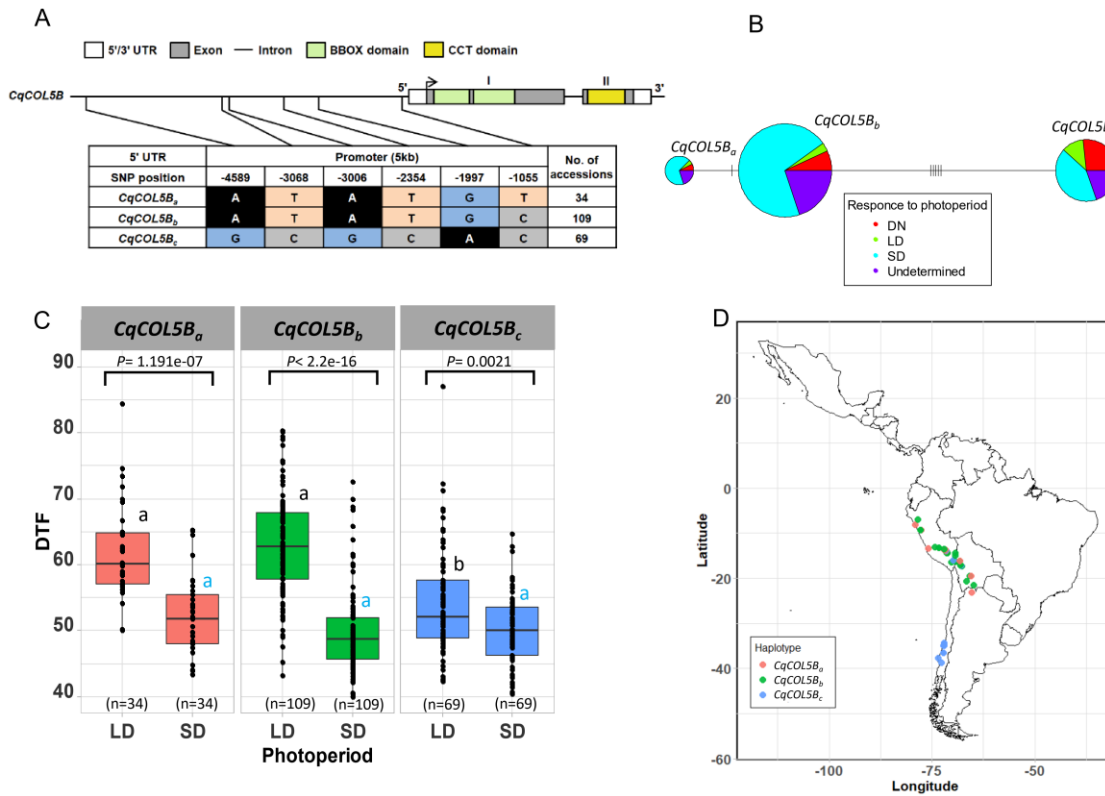


Figure 9: (continued on next page, legend follows).

Figure 9: Phenotypic effects of haplotype variations in four different flowering time genes (*CqFT1A*, *CqFT2B*, *CqCOL2B*, and *CqCOL5B*). The number of accessions analyzed is given in Table 2. (A) Gene structure and nucleotide polymorphisms. Different colors indicate polymorphic nucleotides. SNP positions are given relative to the start codon. The asterisks depict missense variants. (B) Haplotype networks. Circle size illustrates the frequency of the corresponding haplotype. The colors of the pie chart depict different classes of photoperiod response. SD: early flowering under short days, LD: early flowering under long days, DN: day-neutral, no difference between short and long days. Crosslines between two circles are representing the number of nucleotide polymorphisms between haplotypes. (C) Haplotype depending flowering time (days to flowering, DTF) as determined in the climate chamber under SD and LD growth conditions. The *P*-values of mean comparisons between SD and LD within haplotypes are written above the boxplots. An ANOVA was performed to determine significance among haplotypes and grouping of pairwise multiple comparison was obtained using Tukey's post hoc multiple comparison test. Student's t-test was performed to compare DTF means within haplotypes. The black and blue letters indicate significant DTF differences among haplotypes under LD and SD photoperiod, respectively. 'n' indicates the number of accessions. (D) Geographical origin of quinoa accessions with their respective haplotypes. Only accessions are shown where passport data are available.

3.4 Discussion

This study provides the first comprehensive analysis of major flowering time regulators in quinoa. The photoperiod effect on flowering time in quinoa has been reported as early as 1949 (Fuller 1949). But in the meantime, neither genetic nor molecular mechanism of flowering time regulation in quinoa have been discovered. We hypothesized that sequences with high similarity to Arabidopsis *FT*- and *COL*- genes are major flowering time regulators in quinoa and that haplotype variation is associated with flowering time variation and geographical distribution. We showed that under LD conditions, both *CqFT1* and *CqFT2* genes were expressed at an extremely low level. In contrast, the *CqCOL* genes were expressed under LD and SD conditions. However, under LD conditions, expression levels were lower compared to SDs. We could uncover sequence variation in those flowering time genes and their association with flowering under different photoperiodic regimes and geographical distribution. Associations were found between haplotype variation, day-length responsiveness, and geographical distribution, adding to the understanding of adaptation to varying environmental conditions.

Regulation of flowering is controlled by a complex network that involves multiple pathways with hundreds of genes (Blümel et al. 2015; Bouché et al. 2015). The *FT* gene plays a major role in flowering regulation in plant species (Wickland and Hanzawa 2015). TFL is a closely related protein of *FT* with only 39 nonconservative amino acid differences. In contrast to *FT*, *TFL* strongly suppresses flowering (Ho and Weigel 2014b). During domestication, flowering time and plant architecture were modified due to mutations in *FT/TFL* (florigen/anti-florigen) (Eshed and Lippman 2019). Quinoa, as a polyploid species with a short evolutionary history, carries multiple copies of *FT* homologs. Therefore, we focused on genes that are highly related to proven *FT* orthologs from the Amaranthaceae family (Pin et al. 2010; Drabešová et al. 2014a; Štorchová et al. 2019). We selected five *FT* homologs that had already been found in the reference genome sequence (Jarvis et al. 2017). In addition, we selected an *RFT* homolog gene because, in the short day crop rice, *RFT* acts as a floral activator under LDs (Komiya et al. 2009). We expected a regulatory pathway as in other related species where *FT* is under the transcriptional control of *CO* or *COL* genes.

The *CO* protein activates *FT* transcription by binding to a unique *cis*-element in the *FT* promoter (Tiwari et al. 2010). Therefore, we decided to analyze *CO* in our study as it is an important factor in understanding the photoperiodic flowering pathway. Two B-box domains and a conserved C-terminal (CCT) domain are characteristic of Arabidopsis *CO* and related pseudo-response regulator (*PRR*) genes (Griffiths et al. 2003). All *CqCOL* genes we analyzed in this study were encoding two B-boxes and CCT domains.

Chapter 3

Due to gene redundancy in polyploids, it is challenging to identify functional orthologues. During evolution, gene function changes due to the neo- or sub-functionalization of paralogs. Moreover, gene silencing and gene loss are reducing the complexity of paralogs over time. To get a clue about the function of our candidate genes, we measured their transcriptional activities. In allopolyploid species, subgenome-specific gene expression has been described (Bird et al. 2018). For instance, in shoots of tetraploid rapeseed (*Brassica napus*) the transcriptional activity of C genome genes was higher as compared to their A genome paralogs. Contrasting expression patterns were found in the roots of the same species (Chalhoub et al. 2014). Likewise, we observed expression pattern bias toward one subgenome (e.g., *CqFT2A* and *CqFT2B*). All the genes in our study were expressed at least at one developmental stage, which may indicate they are not pseudogenes. Nevertheless, there are genes (*CqFT1B-1*, *CqFT1B-2*) that are unlikely to function as flowering time regulators because their transcriptional profiles are in stark contrast to their Arabidopsis orthologs.

To which extent are *FT* gene functions conserved across species of the plant family Amaranthaceae? Since quinoa and beet belong to the same family, we hypothesized that *CqFT* genes are functional orthologs of the *B. vulgaris* *FT* genes (Pin et al. 2010). Under SDs, three *CqFT* genes were highly expressed before the onset of flowering, which is similar to the floral repressor *BvFT1*. But only *CqFT2B* showed a diurnal expression pattern typical for *FT* orthologs. Moreover, this gene was also highly active before bolting. Therefore, we reason that *CqFT2B* is a floral activator under SDs. Our results suggest that, this gene is the only functional ortholog of sugar beet florigen gene *BvFT2*. The discrepancy between beet and quinoa *FT* gene expression can be explained by the different life-history regimes of both species. In contrast to quinoa, sugar beet needs a long period of vernalization for the transition from the vegetative to the generative phase, which starts with shoot elongation ('bolting'). In quinoa, the *FT* gene's function as a floral repressor would be highly unlikely because quinoa has no vernalization requirement to flower. Therefore, it is likely that quinoa has other floral repressors such as *TFL* homologs because the balance between floral repressors and activators is important for the adaptation of crops to different environments (Eshed & Lippman, 2019). *CqFT2A* may also have alternative functions apart from flowering because it is highly expressed at very early and late developmental stages. It has been reported from other plants that FT proteins are involved in many developmental processes such as fruit setting, vegetative growth, tuberization, stomatal development, and spikelet development (Blümel et al. 2015; Bi et al. 2019). Evidently, *FT* genes in quinoa underwent neo-functionalization and gene silencing during evolution typical for polyploids (Jiao et al. 2011).

Downregulation of *CqFT* genes under LD condition may be the reason for late flowering under LD like in soybean. Soybean is a SD plant where flowering is controlled by two *FT* homologs (*GmFT2a* and *GmFT5a*). Under LDs, both *GmFT2a* and *GmFT5a* genes are downregulated, and as a result, flowering is significantly delayed (Kong et al. 2010). Interestingly, under LDs, quinoa produces flowers without any *CqFT2* expression. Similarly, in *C. ficifolium* under LDs, the expression of *FT* homologs was negligible, although plants flowered regularly (Štorchová et al. 2019). It is tempting to speculate that flowering in quinoa under LDs is controlled by other regulatory pathways. The Arabidopsis *ft/tsf* double mutant can flower without florigen (Yamaguchi et al. 2005). Flowering of these double mutants is age-dependent, which is controlled by the microRNA156 and SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE (miRNA156/SPL) module (Hyun et al., 2019). Depending on day-length, rice produces two different florigens. In LDs, flowering is initiated by the activation of *RFT1*, a homolog of *Hd3a* (Tsuji et al. 2011b). However, quinoa *CqRFT* is unlikely to be a functional ortholog of *BvFT2* because it is expressed only in late accessions under SDs.

Chapter 3

We questioned whether there is a *CO* ortholog in quinoa. Arabidopsis *CO* shows an evening peak in its diurnal expression (Andres and Coupland 2012), which contrasts with all quinoa *COL* genes. *CqCOL* expression under SDs was higher than LDs, similar to sugar beet *COL1* (Chia et al. 2008). Also, *CqCOL* expressions in SDs are more similar to Arabidopsis *COL1* and *COL2* expression patterns. Therefore, these genes cannot be considered *CO* orthologues. Although Arabidopsis *COL1* and *COL2* are not involved in flowering regulation (Ledger et al. 2001), *COL* genes have been identified as floral activators in other species. In legumes, *COL* genes show similar expression patterns to quinoa *COL* genes. In soybean (*Glycine max*) under LDs, *GmCOL1a* and *GmCOL1b* are repressors of florigen encoding genes *GmFT2a* and *GmFT5a*, whereas under SDs they are involved in the upregulation of florigen genes (Cao et al. 2015). In rice, the *CO* homolog *Hdl* acts as the SD floral activator, whereas under LDs *Hdl* turns into a repressor (Lee and An 2015). Interestingly, *CqCOL* and *Hdl* display highly similar expression patterns under SDs. In conclusion, we reason that like in sugar beet, quinoa is lacking a true *CO* orthologue (Dally et al. 2014). Furthermore, evidence has been given for the function of *COL* genes as upstream regulators of *CqFT* genes.

We have found substantial haplotype variation among flowering time genes in quinoa. We found varying degrees of photoperiod sensitivity depending on *FT*- and *COL* haplotypes. Allelic diversity in major flowering time genes are tightly linked with ecotype differentiation in crops, such as, *Bna.A10.FLC* (Schiessl et al. 2017) and *BnaA3.FRI* (Yi et al. 2018) in rapeseed and *Ppd-D1* (Guo et al. 2010) in wheat. Overall, haplotype variations of *CqFT* genes are higher compared to *BvFT* genes (Höft et al. 2018), which is expected due to the short breeding history of quinoa. Remarkably, we did not identify non-functional haplotypes indicating the functional significance of *FT*-like and *COL* genes in quinoa adaptation. Interestingly, four genes (*CqFT1A*, *CqCOL2B*, *CqCOL4A-1*, and *CqCOL5B*) were associated with flowering time variation under LDs, but not under SDs, indicating a putative effect of these genes on flowering time. This might also be related to their different expression patterns under different photoperiods, which supports the assumption that quinoa uses different regulatory pathways to regulate flowering time under LDs and SDs. Notably, five alleles (*CqFT1A_a*, *CqFT2B_c*, *CqCOL2A_c*, *CqCOL2B_b*, and *CqCOL5B_c*) promoted early flowering under LD conditions. Accessions carrying those haplotypes are abundant in southern Chile with LDs during summer. However, differences in flowering time under LD and SD conditions of low abundant haplotypes should be taken with care due to small sample sizes (e.g., *CqFT2B_f*). Nevertheless, the early flowering alleles found in this study are important for quinoa breeding in northern latitudes.

How can haplotype variation uncovered in this study be used for breeding locally adapted quinoa varieties? Whenever crops have been moved from their center of origin to other geographical regions, photoperiod sensitivity has been essential for local adaptation (Meyer and Purugganan 2013). Genetic regulation of photoperiodic sensitivity has been described in several crops. In soybean, allele combinations of three major maturity loci *E1*, *E3*, and *E4* were important for local adaptation (Lu et al. 2017a). Likewise, *Vgt1* and *ZmCCT* haplotype variation in maize (Huang et al. 2017) and *StCDF1* haplotype variation in potato (Gutaker et al. 2019) were major reasons for the worldwide distribution of these important crops. Consequently, breeders can fine-tune photoperiodic sensitivity and thus adjust flowering time by combining different alleles. Selecting quinoa genotypes that are early flowering under LDs is essential for cultivation in North America and Europe. Our study identified haplotypes that are causing photoperiod insensitivity associated with early flowering under LDs. We propose a haplotype-based breeding strategy by pyramiding favorable alleles (e.g., *CqFT1A_a*, *CqFT2B_c*, *CqCOL2A_c*, *CqCOL2B_b*, and *CqCOL5B_c*) to breed quinoa varieties which are early maturing under LDs and are thus suitable for cultivation in Northern latitudes.

3.5 Experimental procedures

3.5.1 Plant materials and growth conditions

We assembled a diversity panel of 303 quinoa accessions representing different geographical regions of the world and different breeding intensities from landraces without any breeding history to inbred lines. Seeds from 276 accessions (Supplementary Table 8) were sown in 35x multi-tray pots, and 10 plants/accession were grown in the climate chamber under SD (8h light and 16h dark) and LD (16h light and 8h dark cycles) photoperiodic conditions at 20°C. Halogen lamps with an intensity of 300 $\mu\text{mol m}^{-2} \text{s}^{-1}$ were used in both chambers. We randomized the multi-tray pots weekly to minimize spatial effects. We recorded flowering when the first flower opened, and days to flowering (DTF) was measured from the sowing date (experiment 1, E1).

To analyze gene expression (experiment 2, E2), we selected two early- (Titicaca, seed code 171230 and PI-614886, seed code 170867) and two late- (PI-587173 and CHEN-109) flowering accessions from different geographical regions. Seeds were sown in 9cm pots, and 15 plants/accession were grown in climate chambers under SD and LD. DTF and days to bolting (DTB) were recorded. For RNA isolations and expression analysis, we collected samples for two different experiments: diurnal expression and temporal expression. We performed diurnal sampling at two different growth stages, stage 0 (vegetative phase) and stage 7 (bolting; pyramid shape of the inflorescence) (Jacobsen and Stølen (1993)). To measure temporal expression, we took leaf samples from plants at nine different time points: two weeks after sowing (vegetative), two weeks before bolting (2WBB), one week before bolting (1WBB), bolting (B), one week after bolting (1WAB), one week before flowering (1WBF), flowering (F), one week after flowering (1WAF) and two weeks after flowering (2WAF). All samples were immediately frozen in liquid nitrogen.

3.5.2 *In-silico* identification of quinoa *FT*-like genes and *CONSTANS* like genes

We used the BvFT2 (Pin et al. 2010) protein sequence as a query for a BLAST search against an improved version of the quinoa reference genome (Jarvis et al. 2017). Moreover, we checked the Conserved Domain Database (CDD) in NCBI to identify conserved domains present in the proteins. For *CONSTANS*-like gene identification, we used the AtCO protein sequence as a query, and then we followed the same procedure as explained above.

3.5.3 Phylogenetic tree construction

We used the protein sequence obtained from candidate genes for multiple sequence alignment. As outliers for phylogenetic tree construction of *FT*-like genes, we used AtFT, CrFTL1, CrFTL2, BvFT1, BvFT2, Hd3a, and RFT protein sequences. For *CONSTANS*-like protein alignment, we used AtCOL, BvCOL, CrCOL, and HD1 protein sequences. Then, we generated a phylogenetic tree using the CLC-main workbench 8.1.3 based on the Neighbor-Joining algorithm with 1000 bootstraps. We obtained a protein pairwise identity percentage matrix from the Clustal Omega (Madeira et al. 2019).

3.5.4 RNA isolation and gene expression analysis

Total RNA was isolated using the NucleoSpin RNA isolation kit (Macherey-Nagel) and treated with DNaseI (Life Technologies GmbH, ThermoFisher Scientific Darmstadt, Germany). One microgram (1 μg) of total RNA was taken for cDNA synthesis using the First Strand cDNA Synthesis Kit (Life Technologies GmbH, ThermoFisher Scientific Darmstadt, Germany). We diluted the cDNA by 20-fold (1:20) and used 2 μl of diluted cDNA as a template for RT-qPCR with three biological replicates per sampling point and three technical replicates. We performed RT-qPCR with SYBR qPCR Supermix w/ROX (Life Technologies GmbH, ThermoFisher Scientific Darmstadt, Germany), using a CFX96 Real-Time System (Bio-Rad Laboratories

GmbH, Germany). We used paralog specific primers to conduct the RT-qPCR (Supplementary Table 11). We analyzed amplification curves using Bio-Rad CFX Manager 3.1 and calculated the relative expression of each gene using the ΔCt method. As reference genes, we used the geometric mean of *CqIDH-A* and *CqPTB* genes. These genes turned out to be best suited as reference genes because they were stably expressed during different growth stages and different time points of the day (Maldonado-Taipe et al. 2020).

3.5.5 DNA sequencing and haplotype analysis

We used whole-genome re-sequences from 303 quinoa accessions from the diversity panel. We mapped Illumina short-read sequences against an improved version of the quinoa reference genome (Jarvis et al. 2017). SNP variants were identified on a population scale using the “—emitRefConfidence” function of the HaplotypeCaller algorithm on each sample, followed by joint genotyping on all samples using GenotypeGVCFs from the Genome Analysis Toolkit (McKenna et al. 2010) (data not shown). The resulting VCF file was further filtered based on minor allele frequency (MAF) ≥ 0.05 , minimum mean depth (min-meanDP) ≥ 5.0 , and maximum-missing (max-missing) ≤ 0.5 using VCFtools (Danecek et al. 2011) and obtained a high confident SNP data set (HC-SNPs). Then, for each gene, we created a bed file containing chromosome name, start position, and end position. We added $-/+ 20\text{kb}$ regions to the start position and end positions to obtain the 20kb flanking regions of genes. For each gene, we created a VCF file using VCFtools (Danecek et al. 2011). Then, we converted VCF files into Flapjack format. After that, we used the Flapjack software (Milne et al. 2010) to sort accessions according to similarity. Subsequently, we manually checked the sequence of each accession to assemble them into haplotypes. We excluded accessions with missing and heterozygous SNPs. For haplotype calling, we first analyzed the transcribed regions. If SNPs were absent in the transcribed region, we further analyzed the regions 5kb upstream of the transcription start. We inferred haplotype networks using haplotNet function from the R package “pegas” (Paradis 2010).

3.5.6 Statistical analysis

We compared DTF within haplotypes using “Welch’s” two-sample t-test ($P < 0.05$). Based on the P -values obtained from the t-test, we classified haplotypes into three groups (1) insensitive ($P > 0.05$), (2) moderately sensitive ($0.001 < P < 0.05$), and (3) highly sensitive ($P < 0.001$). We performed an analysis of variance to compare the DTF of haplotype groups within a photoperiod treatment (SD or LD). The population criteria (Highland and Lowland) obtained from Patiranage et al. (2020) were included in the model as a cofactor to control confounding effects from the origin accessions. Haplotypes were then grouped based on a post hoc comparison with the Tukey test at $P < 0.05$ level. All statistical analyses were performed in R software version 3.6.1 (RCore 2019). We used the recently proposed estimation graphics (J. Ho, Tumkaya, Aryal, Choi, & Claridge-Chang, 2019) in combination with Student’s t-test to compare SD and LD flowering time data for each accession. We classified an accession as ‘short day’ or ‘long day’ if flowering dates were significantly different between both light regimes. ‘day neutral’ accessions did not show any significant difference between LDs and SDs

3.6 Acknowledgments

We thank Prof. David Jarvis for providing access to the improved version of the quinoa reference genome and Prof. George Coupland for patiently read through the manuscript and provide valuable suggestions to improve it. Also, we thank Monika Bruisch, Brigitte Neidhardt-Olf, and Verena Kowalewski, for technical assistance and the Institute of Clinical Molecular Biology (IKMB, Kiel University) for Sanger sequencing. This work was financially supported

by the Competitive Research Grant (Grant No. OSR-2016-CRG5-2966-02) of the King Abdullah University of Science and Technology, Saudi Arabia.

Author Contributions

D.S.R.P, E.A, and N.M planned the experiments, produced and analyzed the data, and involved in manuscript writing. E.R, N.E, M.T, and C.J examined the data and were involved in manuscript writing.

Conflict of interest

The authors declare no conflict of interests.

3.7 Supporting Information

Supplementary Fig. 18: Flowering time variation of 276 accessions grown under short (8h) (B) and long day (16h) (A) conditions. C and D, geographic coordinates of 123 quinoa accessions. Color gradient corresponds to flowering time which was determined under long days (C) and under short days (D).

Supplementary Fig. 19: Comparison between growth under short day (SD) and long day conditions (LD). Plants were grown in 9cm pots in controlled growth chamber conditions. A: Three weeks after sowing B: Seven weeks after sowing. C: Different pigmentation under LDs (left) and SDs (right) in two accessions sensitive to photoperiod; CHEN-109 and PI-587173.

Supplementary Fig. 20: Phenological development and yield components of four accessions grown in the growth chamber under short day (SD) and long day (LD) conditions. Error bars: SEM

Supplementary Fig. 21: Phylogenetic relationships between PEBP family genes (A) and COL genes (B) in quinoa.

Supplementary Fig. 22: Phenotypic effects of haplotype variations in eight different flowering time genes. The number of accessions analyzed is given in Table 2 (A) Gene structure and nucleotide polymorphisms. Polymorphic nucleotides are indicated by different colors. SNP positions are given relative to the start codon. The asterisks depict missense variants. (B) Haplotype networks. Circle size illustrates the frequency of the corresponding haplotype. Colors of the pie chart depict different classes of photoperiod response. (SD) early flowering under short days, (LD) early flowering under long days, (DN) no difference between short- and long day flowering time. Crosslines between two circles are representing the number of nucleotide polymorphisms between haplotypes. (C) Haplotype depending flowering time (days to flowering, DTF) as determined in the climate chamber under SD and LD growth conditions. The black and blue letters indicate significant differences of DTF among haplotypes under LD and SD photoperiod, respectively. The *P*-values of mean comparisons between SD and LD within haplotypes are written above the boxplots. ANOVA with Tukey's method was performed to determine significance among haplotypes, and Student's t-test was performed to compare DTF means within haplotypes. 'n' indicates the number of accessions. (D) Geographical origin of quinoa accessions with their respective haplotypes. Only accessions are shown for which passport data are available.

Supplementary Fig. 22: Protein alignment of PEBP proteins of quinoa, sugar beet, *C. rubrum*, rice and Arabidopsis

Supplementary Fig. 23: Protein alignment of COL and CO proteins of quinoa, sugar beet, *C. rubrum*, rice and Arabidopsis

Supplementary Table 8: Plant material used in this study

Supplementary Table 9: List of *PEBP* family genes in quinoa

Supplementary Table 10: *CONSTANS-like* genes in quinoa analyzed in this study

Supplementary Table 11: Primer combinations used for RT-qPCR

Supplementary Table 12: Summary of sampling stages for temporal expression analysis

3.8 References

- Andres F, Coupland G (2012) The genetic basis of flowering responses to seasonal cues. *Nature Reviews Genetics* 13 (9):627-639. doi:10.1038/nrg3291
- Bertero H, King R, Hall A (1999) Photoperiod-sensitive development phases in quinoa (*Chenopodium quinoa* Willd.). *Field Crops Research* 60 (3):231-243
- Bertero HD (2003) Response of developmental processes to temperature and photoperiod in quinoa (*Chenopodium quinoa* Willd.). *Food Reviews International* 19:87-97. doi:10.1081/FRI-120018870
- Bi X, van Esse GW, Mulki MA, Kirschner G, Zhong J, Simon R, von Korff M (2019) CENTRORADIALIS interacts with *FLOWERING LOCUS T*-like genes to control spikelet initiation, floret development and grain number. *Plant Physiology*:pp.01454.02018. doi:10.1104/pp.18.01454
- Bird KA, VanBuren R, Puzey JR, Edger PP (2018) The causes and consequences of subgenome dominance in hybrids and recent polyploids. *New Phytologist* 220:87-93. doi:10.1111/nph.15256
- Blümel M, Dally N, Jung C (2015) Flowering time regulation in crops-what did we learn from Arabidopsis? *Current Opinion in Biotechnology* 32:121-129. doi:10.1016/j.copbio.2014.11.023
- Bouché F, Lobet G, Tocquin P, Périlleux C (2015) FLOR-ID: an interactive database of flowering-time gene networks in *Arabidopsis thaliana*. *Nucleic Acids Research* 44 (D1):D1167-D1171. doi:10.1093/nar/gkv1054
- Cao D, Li Y, Lu S, Wang J, Nan H, Li X, Shi D, Fang C, Zhai H, Yuan X (2015) *GmCOL1a* and *GmCOL1b* function as flowering repressors in soybean under long-day conditions. *Plant and Cell Physiology* 56 (12):2409-2422
- Cháb D, Kolář J, Olson MS, Štorchová H (2008) Two *FLOWERING LOCUS T* (*FT*) homologs in *Chenopodium rubrum* differ in expression patterns. *Planta* 228:929-940. doi:10.1007/s00425-008-0792-3

- Chalhoub B, Denoeud F, Liu S, Parkin IaP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, Correa M, Da Silva C, Just J, Falentin C, Koh CS, Le Clainche I, Bernard M, Bento P, Noel B, Labadie K, Alberti a, Charles M, Arnaud D, Guo H, Daviaud C, Alamery S, Jabbari K, Zhao M, Edger PP, Chelaifa H, Tack D, Lassalle G, Mestiri I, Schnel N, Le Paslier M-C, Fan G, Renault V, Bayer PE, Golicz aa, Manoli S, Lee T-H, Thi VHD, Chalabi S, Hu Q, Fan C, Tollenaere R, Lu Y, Battail C, Shen J, Sidebottom CHD, Canaguier a, Chauveau a, Berard a, Deniot G, Guan M, Liu Z, Sun F, Lim YP, Lyons E, Town CD, Bancroft I, Meng J, Ma J, Pires JC, King GJ, Brunel D, Delourme R, Renard M, Aury J-M, Adams KL, Batley J, Snowdon RJ, Tost J, Edwards D, Zhou Y, Hua W, Sharpe aG, Paterson aH, Guan C, Wincker P (2014) Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* 345:950-953. doi:10.1126/science.1253435
- Chia TYP, Müller a, Jung C, Mutasa-Göttgens ES (2008) Sugar beet contains a large *CONSTANS-LIKE* gene family including a *CO* homologue that is independent of the early-bolting (B) gene locus. *Journal of experimental botany* 59:2735-2748. doi:10.1093/jxb/ern129
- Christiansen JL, Jacobsen S-E, Jørgensen ST (2010) Photoperiodic effect on flowering and seed development in quinoa (*Chenopodium quinoa* Willd.). *Acta Agriculturae Scandinavica, Section B-Soil & Plant Science* 60 (6):539-544
- Dally N, Eckel M, Batschauer A, Höft N, Jung C (2018) Two *CONSTANS-LIKE* genes jointly control flowering time in beet. *Scientific Reports* 8:16120. doi:10.1038/s41598-018-34328-4
- Dally N, Xiao K, Holtgräwe D, Jung C (2014) The B2 flowering time locus of beet encodes a zinc finger transcription factor. *Proceedings of the National Academy of Sciences* 111 (28):10365-10370
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Group GPA (2011) The variant call format and VCFtools. *Bioinformatics* 27 (15):2156-2158. doi:10.1093/bioinformatics/btr330
- Drabešová J, Cháb D, Kolař J, Haškovcová K, Štorchová H (2014) A dark-light transition triggers expression of the floral promoter *CrFTL1* and downregulates *CONSTANS*-like genes in a short-day plant *Chenopodium rubrum*. *Journal of experimental botany* 65 (8):2137-2146. doi:10.1093/jxb/eru073
- Eshed Y, Lippman ZB (2019) Revolutions in agriculture chart a course for targeted breeding of old and new crops. *Science* 366 (6466)
- Fuller HJ (1949) Photoperiodic responses of *Chenopodium quinoa* Willd. and *Amaranthus caudatus* L. *American Journal of Botany*:175-180
- Golicz AA (2019) Analysis of the quinoa genome reveals conservation and divergence of the flowering pathways.
- González JA, Eisa S, Hussin S, Prado FE (2015) Quinoa: an Incan crop to face global changes in agriculture. *Quinoa: Improvement and sustainable production*:1-18

Chapter 3

- Griffiths S, Dunford RP, Coupland G, Laurie DA (2003) The evolution of *CONSTANS*-like gene families in barley, rice, and Arabidopsis. *Plant physiology* 131 (4):1855-1867
- Guo Z, Song Y, Zhou R, Ren Z, Jia J (2010) Discovery, evaluation and distribution of haplotypes of the wheat *Ppd-D1* gene. *New Phytologist* 185 (3):841-851
- Gutaker RM, Weiß CL, Ellis D, Anglin NL, Knapp S, Fernández-Alonso JL, Prat S, Burbano HA (2019) The origins and adaptation of European potatoes reconstructed from historical genomes. *Nature Ecology & Evolution* 3 (7):1093-1101
- Hayama R, Agashe B, Luley E, King R, Coupland G (2007) A Circadian Rhythm Set by Dusk Determines the Expression of *FT* Homologs and the Short-Day Photoperiodic Flowering Response in *Pharbitis*. *The Plant Cell* 19:2988-3000. doi:10.1105/tpc.107.052480
- Ho J, Tumkaya T, Aryal S, Choi H, Claridge-Chang A (2019) Moving beyond P values: data analysis with estimation graphics. *Nature Methods* 16 (7):565-566. doi:10.1038/s41592-019-0470-3
- Ho WWH, Weigel D (2014) Structural features determining flower-promoting activity of Arabidopsis *FLOWERING LOCUS T*. *The Plant cell* 26:552-564. doi:10.1105/tpc.113.115220
- Höft N, Dally N, Hasler M, Jung C (2018) Haplotype Variation of Flowering Time Genes of Sugar Beet and Its Wild Relatives and the Impact on Life Cycle Regimes. *Frontiers in Plant Science* 8. doi:10.3389/fpls.2017.02211
- Huang C, Sun H, Xu D, Chen Q, Liang Y, Wang X, Xu G, Tian J, Wang C, Li D, Wu L, Yang X, Jin W, Doebley JF, Tian F (2017) *ZmCCT9* enhances maize adaptation to higher latitudes *Proceedings of the National Academy of Sciences* 115:E334-E341. doi:10.1073/pnas.1718058115
- Jacobsen S-E, Stølen O (1993) Quinoa - Morphology, phenology and prospects for its production as a new crop in Europe. *European Journal of Agronomy* 2:19-29. doi:10.1016/S1161-0301(14)80148-2
- Jarvis DE, Ho YS, Lightfoot DJ, Schmöckel SM, Li B, Borm TJ, Ohyanagi H, Mineta K, Mitchell CT, Saber N (2017) The genome of *Chenopodium quinoa*. *Nature* 542 (7641):307
- Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS (2011) Ancestral polyploidy in seed plants and angiosperms. *Nature* 473 (7345):97-100
- Komiya R, Yokoi S, Shimamoto K (2009) A gene network for long-day flowering activates *RFT1* encoding a mobile flowering signal in rice. *Development* 136:3443-3450. doi:10.1242/dev.040170
- Kong F, Liu B, Xia Z, Sato S, Kim BM, Watanabe S, Yamada T, Tabata S, Kanazawa A, Harada K (2010) Two coordinately regulated homologs of *FLOWERING LOCUS T* are involved in the control of photoperiodic flowering in soybean. *Plant physiology* 154 (3):1220-1231

Chapter 3

- Ledger S, Strayer C, Ashton F, Kay SA, Putterill J (2001) Analysis of the function of two circadian-regulated *CONSTANS-LIKE* genes. *The Plant Journal* 26 (1):15-22
- Lee Y, An G (2015) Complex regulatory networks of flowering time in rice. *J Rice Res* 3 (2)
- Lu S, Zhao X, Hu Y, Liu S, Nan H, Li X, Fang C, Cao D, Shi X, Kong L (2017) Natural variation at the soybean J locus improves adaptation to the tropics and enhances yield. *Nature genetics* 49 (5):773-779
- Luccioni L, Krzymuski M, Sánchez-Lamas M, Karayekov E, Cerdán PD, Casal JJ (2019) *CONSTANS* delays *Arabidopsis* flowering under short days. *The Plant Journal* 97 (5):923-932. doi:10.1111/tpj.14171
- Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey AR, Potter SC, Finn RD (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic acids research* 47 (W1):W636-W641
- Maldonado-Taípe N, Sarange D, Schmöckel S, Jung C, Emrani N (2020) Validation of suitable genes for normalization of diurnal gene expression studies in *Chenopodium quinoa*. bioRxiv
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* 20 (9):1297-1303
- Meyer RS, Purugganan MD (2013) Evolution of crop species: genetics of domestication and diversification. *Nature reviews Genetics* 14:840-852. doi:10.1038/nrg3605
- Milne I, Shaw P, Stephen G, Bayer M, Cardle L, Thomas WTB, Flavell AJ, Marshall D (2010) Flapjack--graphical genotype visualization. *Bioinformatics (Oxford, England)* 26 (24):3133-3134. doi:10.1093/bioinformatics/btq580
- Murphy KS, Matanguihan J (2015) *Quinoa: Improvement and sustainable production*. John Wiley & Sons,
- Paradis E (2010) pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* 26 (3):419-420
- Pin PA, Benlloch R, Bonnet D, Wremmerth-Weich E, Kraft T, Gielen JJ, Nilsson O (2010) An antagonistic pair of *FT* homologs mediates the control of flowering time in sugar beet. *Science* 330 (6009):1397-1400
- Pin PA, Zhang W, Vogt SHH, Dally N, Büttner B, Schulze-Buxloh G, Jelly NSS, Chia TYPYP, Mutasa-Göttgens ESS, Dohm JCC, Himmelbauer H, Weisshaar B, Kraus J, Gielen JLLJL, Lommel M, Weyens G, Wahl B, Schechert A, Nilsson O, Jung C, Kraft T, Müller AEE (2012) The role of a pseudo-response regulator gene in life cycle adaptation and domestication of beet. *Current biology : CB* 22:1095-1101. doi:10.1016/j.cub.2012.04.007

Chapter 3

- Rojas W (2003) Multivariate Analysis of Genetic Diversity of Bolivian Quinoa Germplasm. *Food Reviews International* 19 (1-2):9-23. doi:10.1081/FRI-120018864
- Schiessl SV, Huettel B, Kuehn D, Reinhardt R, Snowdon RJ (2017) Flowering Time Gene Variation in *Brassica* Species Shows Evolutionary Principles. *Frontiers in Plant Science* 8:1-13. doi:10.3389/fpls.2017.01742
- Song YH, Kubota A, Kwon MS, Covington MF, Lee N, Taagen ER, Laboy Cintrón D, Hwang DY, Akiyama R, Hodge SK, Huang H, Nguyen NH, Nusinow DA, Millar AJ, Shimizu KK, Imaizumi T (2018) Molecular basis of flowering under natural long-day conditions in *Arabidopsis*. *Nature Plants* 4:824-835. doi:10.1038/s41477-018-0253-3
- Štorchová H, Hubáčková H, Abeyawardana OA, Walterová J, Vondráková Z, Eliášová K, Mandák B (2019) *Chenopodium ficifolium* flowers under long days without upregulation of *FLOWERING LOCUS T* (FT) homologs. *Planta* 250 (6):2111-2125
- Tiwari SB, Shen Y, Chang H-C, Hou Y, Harris A, Ma SF, McPartland M, Hymus GJ, Adam L, Marion C, Belachew A, Repetti PP, Reuber TL, Ratcliffe OJ (2010) The flowering time regulator *CONSTANS* is recruited to the *FLOWERING LOCUS T* promoter via a unique cis-element. *New Phytologist* 187 (1):57-66. doi:10.1111/j.1469-8137.2010.03251.x
- Tsuji H, Taoka KI, Shimamoto K (2011) Regulation of flowering in rice: Two florigen genes, a complex gene network, and natural variation. *Current Opinion in Plant Biology* 14:45-52. doi:10.1016/j.pbi.2010.08.016
- Wang J-W, Czech B, Weigel D (2009) miR156-regulated SPL transcription factors define an endogenous flowering pathway in *Arabidopsis thaliana*. *Cell* 138 (4):738-749
- Wickland DP, Hanzawa Y (2015) The *FLOWERING LOCUS T/TERMINAL FLOWER 1* Gene Family: Functional Evolution and Molecular Mechanisms. *Molecular Plant* 8:983-997. doi:10.1016/j.molp.2015.01.007
- Yamaguchi A, Kobayashi Y, Goto K, Abe M, Araki T (2005) *TWIN SISTER OF FT (TSF)* acts as a floral pathway integrator redundantly with FT. *Plant and Cell Physiology* 46 (8):1175-1189
- Yi L, Chen C, Yin S, Li H, Li Z, Wang B, King GJ, Wang J, Liu K (2018) Sequence variation and functional analysis of a *FRIGIDA* orthologue (*BnaA3.FRI*) in *Brassica napus*. *BMC Plant Biology* 18:1-13. doi:10.1186/s12870-018-1253-1

4 Closing discussion

Quinoa is an ancient Andean crop, for which breeding is still in its initial stages. There is an untapped and promising potential in quinoa as a climate-resilient crop. It can be well adapted to the Northern European climate (Jacobsen 2017). Thus, it is essential to identify the underlying genetic mechanism of adaptive traits for the efficient and systematic improvement of quinoa. This study aimed to understand the underlying genetic mechanism of flowering time and other agronomically important traits in quinoa.

The results of my study provide a substantial contribution to quinoa genomics and breeding. This is achieved by presenting an unprecedented large-scale genotypic and phenotypic characterization of a large diversity collection. I analyzed re-sequencing data from a diversity panel and identified two populations, namely Highland and Lowland. These populations showed substantial differences in their patterns of variations. The Lowland population showed considerably slow LD decay and low nucleotide diversity. In comparison to the five sub-populations identified within Highland quinoa, only two sub-populations were identified within the Lowland quinoa. These results suggest that Lowland quinoa has undergone through stronger selection bottleneck. Moreover, Highland quinoa is more diverse, attributed to the geographical diversity of the Andean highlands. I then performed a GWAS analysis for 17 different traits and identified 600 consistent marker-trait associations (MTA). I confirmed previously identified candidate genes for saponin content and flower color. In addition, candidate genes for thousand seed weight, mildew infection, days to flowering, days to maturity, plant height, and panicle length were identified.

I also analyzed the flowering time of 276 accessions under short day and long day photoperiods. Flowering time variation was higher under LD grown plants compared to SD grown plants. The majority of accessions (67.39%) flowered significantly earlier under SDs. I also identified that 16.67% of accessions do not respond to photoperiods, whereas 7.6% of the accessions flowered earlier under LDs. Next, I analyzed the expression profiles of six *CqFT* genes and six *CqCOL* genes under SD and LD photoperiods. I showed that quinoa *FT* genes (*CqFT1* and *CqFT2*) do not act antagonistically, as in sugar beet. Furthermore, I demonstrated the effect of photoperiod on flowering time genes. Further investigation of sequence variations in *CqFT* and *CqCOL* genes revealed the haplotypes associated with the flowering time under SDs and LDs. Five haplotypes showed early flowering under LD conditions, and of these, the most were characteristic to the southern Chilean accessions. I found that the haplotype distributions of six genes are strongly associated with the geographical origin of accessions.

4.1 Sequence variations and expression of quinoa *CO* and *FT* homologs influences flowering time and geographical adaptation

First, I wanted to know whether the functions of the flowering time genes are conserved between quinoa and beet. Comparative genetic analysis has been done based on the expectation of conservation of regulatory pathway genes between species (Thompson et al. 2015); hence it is expected to be true for the flowering time regulation as well. There are numerous genes involved in controlling flowering time variation in plants (Blümel et al. 2015; Bouché et al. 2015). In many plant species such as *Arabidopsis* (Andrés and Coupland 2012), rice (Kojima et al. 2002), soybean (Fan et al. 2014), aspen (Böhlenius et al. 2006), a conserved function of the FT-CO module for flowering control has been revealed. Therefore, I expected *CqFT* is under the transcriptional control of *CqCO* or *CqCOL*, like in other species. I decided to focus on the homologs of two major regulators (*CO* and *FT*). Since quinoa is a polyploid species, multiple gene copies are expected. Nevertheless, not all detected paralogs might be functional, or might have gained new function(s), features which are often found in polyploids (Jiao et al. 2011). Since the sequence conservation correlates with functional conservation (Joshi and Xu 2007), I selected the genes sharing the highest homology with *AtCO* and sugar beet *FTs*.

Closing discussion

The *FT* gene is under the regulation of circadian clock genes, with diurnal expression peaking around dusk in LDs (Andres and Coupland 2012). Sugar beet floral activator, *BvFT2* showed a similar diurnal expression with a typical late evening peak, whereas the floral repressor *BvFT1* showed an expression peak in the morning (Pin et al. 2010). Both Arabidopsis and sugar beet are typical long-day plants, and *FT* and *BvFT2* expression were negligible under SDs (Pin et al. 2010; Andres and Coupland 2012). In contrast, in the short-day species rice, the *FT* homolog *Hd3a* has been shown to peak at dusk under SDs. Whereas under LD, rice *Hd3a* showed no expression (Komiya et al. 2008). I expected *CqFT1* and *CqFT2* genes to act as floral repressors and activators, respectively, based on their homology to *BvFT* genes. Considering expression patterns of floral repressors and activators under LDs and SDs, I conclude that none of the *CqFT* genes act as floral repressors because they did not resemble the expression of *BvFT1*.

In sugar beet as a biennial crop, to avoid flowering before vernalization, the necessity of a floral repressor is obvious (Pin et al. 2010). However, I assume there is no evolutionary advantage for quinoa to have a floral repressor because it is an annual crop that flowers without vernalization. Moreover, my data suggest that *CqFT1* genes might not play a role in quinoa flowering time. Interestingly, I observed that under LDs, all *CqFT* expressions were undetectable except for *CqFT1B-2*. This raises the question: what is the florigen of quinoa under LD conditions? I hypothesize under LDs, quinoa undergoes flowering without activation of *CqFT* genes, and alternative pathways may contribute to flowering. It has been shown that genes in the aging pathway (*SPL*) activate floral meristem identity genes without *FT* activity (Wang et al. 2009). Furthermore, gibberellic acids are directly involved in the activation of floral meristem identity genes (Bao et al. 2020). Nonetheless, functional studies are needed to test this hypothesis. As a starting point, it is worthwhile to produce knockout mutants/and overexpression lines of each of the candidate genes and phenotype flowering under different photoperiods. In this way, the effect of each gene on flowering and photoperiod response is characterized. Efficient transformation methods are essential for creating mutant or overexpression lines, which are not currently available for quinoa.

In Arabidopsis, *CO* expression under LDs coincided with the *FT* expression but with a slightly earlier expression peak before dusk. This indicates the induction of *FT* by *CO*. In the short-day plant rice, *CO* homolog *HDI* represses flowering under LDs, whereas under SD, it induces flowering by activating *Hd3a* (Ishikawa et al. 2011). In contrast, there is no true *CO* in sugar beets, but two *COL* genes are present (Chia et al. 2008). Chia et al. (2008) speculated that *BvCOL* and *AtCO* are not functional homologs because diurnal expression patterns were not similar. Nonetheless, a study using Arabidopsis *co-2* mutants showed that *BvCOL1* complements the late-flowering phenotype (Chia et al. 2008). Compensating for the absence of *CO* in beet, the joint activity of *BTC1* and *BvBBX19* is hypothesized to trigger the expression of *BvFT2* (Dally et al. 2018). Diurnal expression patterns of the *CqCOL* genes were more similar to Arabidopsis (Ledger et al. 2001), beet (Chia et al. 2008), and *C. rubrum* (Drabešová et al. 2014b) *COL* gene homologs. In facultative long-day species, *COL* homologs have different functions under LD and SD (Cao et al. 2015; Tsuji et al. 2011a). Therefore, I wanted to investigate if quinoa *COL* genes also show such expression patterns and functions under different photoperiods. Indeed, I observed that the diurnal expression pattern of *CqCOL5* and *CqCOL2* changes under LD compared to SD. However, the *CqFT* expression patterns were not corresponding to *CqCOL* expressions. Therefore, *CqCOL* does not share the same functions with *AtCO*. There are numerous *PRR* (*PSUEDO RESPONSE REGULATOR*) genes in quinoa. Consequently, it would be challenging to select one of them for further investigations. Nevertheless, as a starting point for a future experiment, I suggest testing the expressions and function of *CqBTC1* genes and *CqBBX19* genes to know if they share common functions with *BvBTC1* and *BvBBX19* genes.

Closing discussion

Two homologs for each *BvBTC1* and *BvBBX19* genes have been identified in quinoa (Golicz et al. 2020; Asare 2020).

Due to the presence of multiple paralogs, the interaction of flowering time genes becomes more complex in polyploidy species. Therefore, the information that could be obtained from RT-qPCR studies are limited and not cost-effective (Nonis et al. 2014). For future studies, I suggest genome-wide transcriptome analysis for quinoa, which has already been done in other polyploid species, to understand flowering time gene expression at different stages and under different environmental conditions (Yu et al. 2016b; Shah et al. 2018; Matar et al. 2020). For quinoa, whole-genome transcriptome analysis of samples obtained from leaves and meristem during floral transition under LD and SD conditions would be a valuable initiative to investigate the flowering time genes further.

Allelic diversity of flowering time genes underlies geographical adaptation of crops (Romero Navarro et al. 2017). Therefore, my next hypothesis was that sequence variation within flowering genes affects the flowering time variation and geographical adaptation in quinoa. To test this hypothesis, I conducted experiments to answer two questions. (1) What is the flowering time response of different quinoa accessions? (2) What are the sequence variations in flowering time candidate genes, and do they correspond with specific geographies? In accordance with previous studies (Bertero et al. 1999a), phenotyping of quinoa accessions under LD and SD showed varying degrees of photoperiod sensitivity. Also, it has been demonstrated that photoperiod sensitivity in quinoa differs between accessions (Bertero 2003a). Different photoperiods also affect flowering and other growth parameters such as plant height, days to maturity, biomass, and chlorophyll content (Adams and Langton 2005). Especially, quinoa accessions are sensitive to photoperiod during seed filling and maturity (Christiansen et al. 2010). Therefore, I suggest phenotyping for multiple traits that are influenced by photoperiod response, such as plant height, days to maturity, and biomass, under different photoperiods to better classify photoperiod sensitivity of accessions. This information is important for quinoa cultivation outside of its center of origin.

Crop adaptation is a result of fine-tuned photoperiod sensitivity and is acquired as a result of sequence variation in flowering genes (Meyer and Purugganan 2013). Therefore, I aimed to identify haplotypes that influence photoperiod sensitivity. I found a strong geographical association of haplotypes from six genes. This reflects that these genes are essential for the geographical distribution of quinoa. I expected to find haplotypes that do not confer photoperiod sensitivity because some quinoa accessions are day-neutral. The varying degree of photoperiod sensitivity was a result of variation haplotype variation in *CqFT* and *CqCOL* genes (see Chapter 3). Interestingly, I found that accessions distributed towards southern lower latitudes accumulated a high number of haplotypes associated with photoperiod insensitivity. Hence, these accessions can be grown under LD conditions. This provides information on superior haplotypes that could be applied for breeding in the northern temperate climates. Also, I demonstrated that there are haplotypes that cause early flowering under LD conditions. These haplotypes are of interest for breeding quinoa in Europe, where early flowering is advantageous. A new question arises after this result: what is the combined effect of superior haplotypes? However, this question remains unanswered, and future experiments are needed to investigate the combined effect of superior haplotypes. For instance, haplotype-assisted crossing parent selection can be performed to pyramid superior haplotypes into a single cultivar (Qian et al. 2017). Also, haplotype-based genomic prediction instead of individual SNPs has been proposed to improve the prediction accuracy (Won et al. 2020; Ballesta et al. 2019; Prakapenka et al. 2020). In this way, the effects of each haplotype are estimated, and haplotype-based breeding values for each accession could be estimated.

4.2 High marker density allows rapid identification of candidate genes for agronomically important traits

Quinoa germplasm shows substantial diversity. However, this diversity is still untapped. Genetic diversity and structure of quinoa germplasm have been reported in previous studies based on different marker technologies (Wilson 1988b; Christensen et al. 2007; Fuentes et al. 2009; Costa Tártara et al. 2012; Rodríguez and Isla 2009; del Castillo et al. 2007; Zhang et al. 2017; Murphy et al. 2018). Although marker densities and the number of markers used in previous studies were limited, all studies reported genetically diverse populations from South American lowlands and highlands. My results were in accordance with the previous studies. With the high-density SNPs, I also aimed to achieve a higher resolution of the subpopulation characterization. Therefore, I analyzed the admixture proportion of these two populations separately. I observed that the number of subpopulations in the Highland population is higher than the Lowland population. These results were expected because of the diversity statistics, where I demonstrated that the Highland was more diverse than the Lowland population. Zhang et al. (2017) analyzed population structure based on 147 markers and reported that $K=4$ is the optimum number of populations in the quinoa accessions they investigated. Although this finding contrasts with my results, a direct comparison is not possible as there are differences in marker resolutions and number of accessions. Recent studies have shown that finer population division could be identified by higher K and high density of SNPs (Milner et al. 2019). Nonetheless, I did not analyze the potential cause for population differentiation. A detailed description of domestication history and population diversion is given in Chapter 2.

Whole genome-wide LD reflects the population history, breeding system, and patterns of geographical subdivisions. Local LD of a genomic region can be used to study the history of selection (Slatkin 2008). Hence, LD decay is used to study the breeding and domestication history of a crop. It is well known that with selection, the size of LD blocks increases, and therefore LD decay slows down (Wang et al. 2020). I observed similar results in the Lowland population, which had substantially slower LD decay. It will be worthwhile to investigate these differences between populations in more detail. Possible causes for these differences are discussed in Chapter 2. Whole population-wide LD decay in quinoa was found to be low compared to other major crop species (see Chapter 2). Comparisons must be made with caution due to the varying number of markers and accessions, different reproduction types, and the selection intensity.

Hybrid breeding has been discussed in quinoa after heterosis was observed for the first time by Ward and Johnson (1993), but it is still far from realization (Emrani et al. 2020). Nevertheless, the presence of a highly genetically diverged population is an advantage for hybrid breeding. Therefore, it is interesting to know if there is high heterosis between identified populations. Selecting parents for hybrid production can be assisted with the phylogenetic tree presented in this study (Chapter 2). To the best of my knowledge, this is the most thorough and informative phylogenetic tree for quinoa which can be used in the selection of genetically distant parents to make crosses for breeding new cultivars. The identified SNPs can be used for low-costs marker development such as KASP. In addition, I provide resources toward core collection building and conservation.

Another aim of my study was to identify candidate genes for flowering time and other agronomically important traits. First, I asked, is the diversity panel suitable for gene identification? To test this, I used traits like saponin content and flower color, that have already been investigated and for which the candidate genes were identified (Jarvis et al. 2017). As a proof of concept, I could locate MTA with these two traits confirming the reliability of my GWAS analysis. Moreover, this indicates that the scale of the diversity panel was sufficient for

Closing discussion

gene identification. Also, the LD extent of the diversity panel provides a single-gene resolution. However, the power of identification of significant associations and candidate genes depends on the trait architecture. I identified stable MTA associations for ten out of 17 traits investigated. There was a tendency of correlation between the heritability of the traits and the number of stable associations identified. However, it was not linear as some highly heritable traits had a low number of MTA. Due to time and resource limitations, I could only use phenotype data from field trials over two years. I suggest performing further field trials in multiple locations to obtain additional phenotype data, especially for low heritable traits. Nonetheless, these MTA can be used to design markers for selection in breeding programs in Europe. Before using MTA as markers, it would be worthwhile to perform field experiments in multiple locations to identify stable MTA (Varshney et al. 2019a) since this study provides consistent MTAs from a single location.

4.3 Perspectives for quinoa breeding by haplotype building

The history of crop breeding is divided into four stages by Wallace et al. (2018). At stage 1 (breeding 1.0), the mass selection is made by the farmers, while in stage 2 (breeding 2.0), advanced experimental designs and statistical methods are used to improve the selection process. At stage 3 (breeding 3.0), genetic and genomic data are used to facilitate the development of varieties. Combining known alleles to obtain optimal combinations is done at stage 4 (breeding 4.0). Most of the modern crops are either at breeding 3.0 or 4.0 (Wallace et al. 2018). In contrast, quinoa as an underutilized crop is still at stages 1.0 and 2.0. With my research, I wanted to accelerate and bring quinoa breeding into more advanced stages. To achieve this, I briefly discuss how to exploit the existing variation in quinoa germplasm. Quinoa germplasm is a “gold mine” for the identification of new alleles. Extensive utilization of genetic variation can be achieved based on an adequate understanding of quinoa genomics.

The general goal of breeding is to incorporate more desirable characters into a single genotype to obtain an ideal genotype (ideotype). This can be achieved by combining superior alleles into a variety. However, prior to this, ideotypes should be defined according to different agro-climatic regions. The first step is to define the target environment and understand its features. In this case, information on biotic and abiotic stresses, soil type, temperature, photoperiod, rainfall, and also market requirements needs to be considered.

We further need to understand the causes and necessity of adaptation of the crop. Breeding objectives could be different based on the local climates. For example, if quinoa is bred for Northern European conditions, a major hurdle is the photoperiod adaptation to achieve maturity at the right time. On the other hand, parts of the world such as North Africa and the Middle East, prioritize drought and salt tolerance; traits that are important for these geographies. However, a high and stable yield is a primary breeding objective for any climate. The stability of the yield depends on other traits such as drought, salt, and disease tolerance. The pre-requisite is to identify candidate genes or QTLs for traits that need to be integrated. Therefore, the general aim of my work was to identify the candidate genes and QTLs that underlie agronomically important traits in quinoa. Nevertheless, I could not identify candidate genes for all the traits I investigated. It is certainly not realistic to uncover the genetic architecture of all traits in a single analysis, as most agronomic traits are controlled by multiple genes (Wallace et al. 2018).

What are the genes and alleles that we should target and integrate into an ideal quinoa variety? López-Marqués et al. (2020) have discussed the potential candidate genes and traits to be targeted in order to improve quinoa. They mainly focused on (1) seed size improvement, (2) reduction of seed shattering, pre-harvest sprouting, and saponin content, (3) optimization of plant height and flowering time, and (4) improvement of heat and mildew tolerance. Based on

Closing discussion

sequence homology, this study has suggested candidate genes to be targeted, but no genetic evidence was given. With the experience of my study, I recommend also focusing on the optimization of the number of branches, days to maturity/photoperiod sensitivity, uniformity, and reduction of lodging. Certainly, with my GWAS study, I identified candidate genes for ten traits (see chapter 3). These genes can be used with novel breeding techniques to improve quinoa cultivars (Ahmar et al. 2020). Therefore, I propose starting with these genes for functional analysis. Also, for conventional breeding, stable MTAs can be used to design molecular markers that can facilitate the selection in breeding programs. In addition, I showed desirable haplotypes for DTF, DTM, PH, PL, TSW, and mildew tolerance. With those haplotypes, I propose a haplotype building strategy, as shown in Figure 10. It is essential to emphasize again that tailoring haplotypes should be specific to local environments. For hybrid breeding, also parents can be genotyped and used in breeding programs. Indeed, investigation of epistasis interactions for haplotype building would be a future fundamental research question since it might also impact hybrid production.

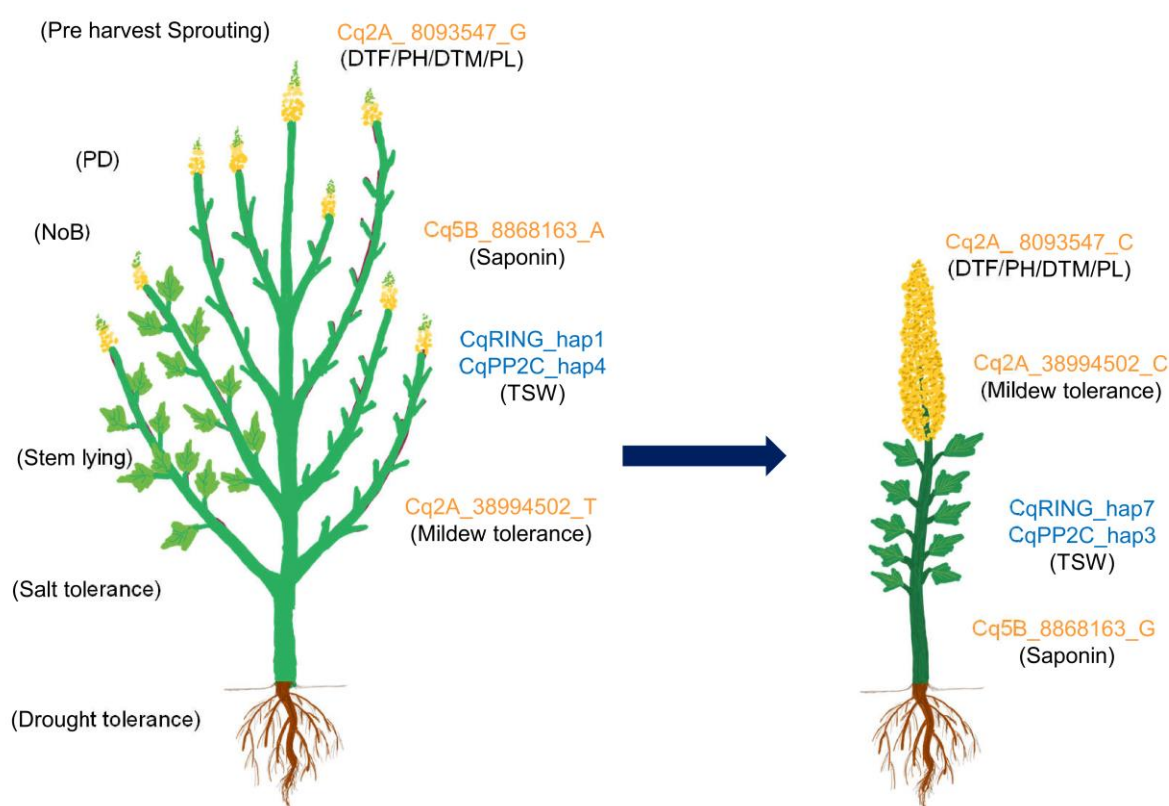


Figure 10: Developing ideotypes by converting partially domesticated quinoa into a fully domesticated crop. The mentioned traits have major priorities for breeding. Blue color is for haplotypes, and dark orange is for marker alleles. A plant with inferior phenotypes is a result of inferior marker alleles and haplotypes. The traits with no MTA found or investigated are shown without haplotypes above the trait (left). A plant with a superior phenotype (ideotype) is obtained by pyramiding superior marker alleles and haplotypes (right).

In Figure 10, I depict a strategy for haplotype building to increase quinoa productivity. The left figure represents a hypothetical scenario of accessions with undesired characters (e.g. accessions: CHEN-110 and D-12081). For example, this accession has undesired branching and height, disease susceptibility, high amount of saponin, small panicles, and small seeds. These characters are influenced by inferior alleles, as shown in the figure. On the right, an example of an ideotype

Closing discussion

obtained by accumulating superior haplotypes is presented. This could be achieved by pyramiding superior haplotypes. Utilization of accession with multiple superior haplotypes in combination with speed breeding/rapid generation cycles would substantially speed up the haplotype pyramiding process (Ghosh et al. 2018). In this model, no interaction of haplotypes is assumed. Also, it should be noted that this model is exclusively based on the associations identified in the long-day photoperiod. If the target environments are short days, corresponding phenotyping and association data need to be considered (Figure 10). Therefore, this ideotype should be customized and fine-tuned according to the demand of the target environment.

None of the accessions will have all undesired alleles or all desired alleles. In practice, breeders have to eliminate inferior alleles while retaining superior alleles. This could be done based on the haplotype-based genomic selection (Haplo-GS), as recently proposed (Abbai et al. 2019). Especially for a small and underutilized crop, this method is cost-effective. Also, the targeted breeding and domestication of crops have been discussed (Li et al. 2018; Lemmon et al. 2018). In this case, the CRISPR-Cas9 system would be a method of choice in order to modify wild alleles into domesticated ones. The availability of efficient transformation methods for quinoa is essential for genome editing, which so far have not been established. This would unlock the potential to achieve variations beyond what is currently available from the quinoa germplasm.

5 Summary

Quinoa (*Chenopodium quinoa* Willd.) is a traditional Andean crop with a long history of cultivation and was domesticated around 5000-7000 years ago. It is broadly adaptable and has high nutritional value; therefore, quinoa production and consumption have rapidly expanded in recent years. Quinoa germplasm possesses a broad genetic diversity which can benefit its adaptation and cultivation in Northern Europe. Nevertheless, it remains untapped due to a lack of implementation of modern breeding tools. Understanding the underlying genetics of flowering and other agronomic traits is essential for the fast improvement of the crop.

This study analyzed genetic and phenotypic variation within the species *Chenopodium quinoa*. It was split into two work packages (WP). The first WP aimed to study haplotype and expression variation in flowering time genes that affect photoperiod sensitivity and quinoa adaptation. In the second work package, I used a diversity panel of 310 quinoa accessions and aimed to analyze population structure, linkage disequilibrium decay, and marker-trait association for flowering time and other agronomically important traits.

I conducted a climate chamber experiment using 276 quinoa accessions to phenotype flowering under long- and short-day photoperiod. I observed days to flowering variation was higher under LD compared to SD. The majority of the accessions showed a ‘short day’ (67.39%) response, where the plants flowered significantly earlier under SD conditions. In contrast, 7.6% accession flowered significantly earlier under LDs. I found that accessions originating from southern Chile were photoperiod insensitive, suggesting loss of day-length sensitivity facilitated the extension of cultivation towards lower latitudes. Next, I studied homologs of major flowering time regulators of Arabidopsis, *FLOWERING LOCUS T (FT)*, and *CONSTANS (CO)* in quinoa. I identified 24 *FT* and six *CO* homologs. Based on the highest homology to sugar beet floral repressor and activator *BvFT1* and *BvFT2*, and Arabidopsis *CO*, I selected six *CqFT* and six *CqCOL* genes, respectively, for expression analysis. Diurnal expression analysis showed that under LD, *CqFT* genes are highly downregulated except *CqFT1B-2*. The expression of *CqCOL* genes was not affected by photoperiods. *CqFT2B* under SD showed a late evening peak, which is a typical diurnal expression for *FT* orthologs. Thus, indicating that *CqFT2B* is a floral activator under SDs. *CqCOL* expressions did not resemble *CO* expression, but they resembled *BvCOL* and *AtCOL* expressions, indicating quinoa has no true *CO* orthologs.

I analyzed the haplotypes of six *CqFT* genes and six *CqCOL* genes based on high confidence SNPs (HC-SNPs) identified from the quinoa diversity panel. I identified 2 to 6 haplotypes across all genes investigated. The highest sequence variation was displayed in the *CqCOL2A* gene with six haplotypes, whereas in seven genes no amino acid substitutions were observed. Interestingly I did not identify any SNP *CqCOL5* genes. Seven haplotypes resulted in early flowering under SD, and five resulted in early flowering under LDs. Also, 12 photoperiod insensitive haplotypes were identified. Haplotypes from six genes displayed a strong association with geography.

To characterize the diversity panel, I performed field trials in two years and phenotyped 17 agronomic characters. I identified substantial phenotypic diversity between accessions. I estimated high heritability for nine traits. Principle component analysis based on phenotypes showed two clusters. Moreover, together with our partners at King Abdullah University of Science and Technology, Saudi Arabia, we performed whole-genome re-sequencing of the diversity panel and identified 2.9 million high-confidence SNPs. In accordance with phenotype data, SNP based principle component analysis also showed highly divergent populations which are corresponding to Highland and Lowland quinoa. I identified $K=8$ as the optimum number of populations using the ADMIXTURE analysis. Moreover, this analysis showed that the Highland population was structured into five groups, while Lowland accessions were split into two.

Summary

Highland and Lowland populations showed a high F_{ST} divergence of 0.36. LD decay of the Highland population (6.5 kb) was substantially fast compared to the Lowland population (49.8 kb). I also observed that the nucleotide diversity is 1.62 fold higher in the Highland, whereas average Tajima's D was negative in the Lowland population. Altogether these results led me to propose three possible scenarios of quinoa domestication and distribution within South America: (1) two independent domestication events with a strong bottleneck on Lowland populations, (2) a single domestication but strong population growth after adaptation of Lowland quinoa or (3) strong adaptive selection after domestication. To understand the history and genetics of domestication, it will be necessary to sequence a large representative set of wild relatives of quinoa.

Next, I performed a genome-wide association analysis for 17 agronomic traits and identified 600 SNP trait associations (MTA) which are consistent across two years. I precisely identified previously reported candidate genes for saponin content and flower color, indicating the reliability of this analysis. I further analyzed candidate genes for thousand seed weight and identified two candidate genes that are reported to be associated with controlling seed size from the Brassinosteroid and the ubiquitin-proteasome pathway. Accessions carrying haplotypes PP2C_hap3 and RING_hap7 displayed larger seeds in both years. I identified a candidate region within a region of 38.99 - 39.03 Mb on chromosome Cq2A associated with downy mildew tolerance. A resistance gene analog with an NBS-LRR (nucleotide-binding site leucine-rich repeat) domain was detected in this region. I found the region 80.50 -81.50 Mb on chromosome Cq2A to be of special interest because it displays stable pleiotropic MTA for DTF, DTM, PH, and PL. The most significant SNP is located within the *CqGLX2-2* gene, which encodes the glyoxalase family enzyme. The allele carrying a cytosine at the most significant SNP resulted in early flowering, maturing, and shorter plants and panicles.

Based on these results, I proposed a haplotype-based breeding strategy by pyramiding superior alleles favorable for northern latitudes. Accessions from the diversity set, which are not well adapted to long-day conditions but with favorable agronomic characters can be used as crossing parents for breeding programs. Then, favorable genotypes could be identified from offspring generations by marker-assisted selection using markers in LD with significant MTAs.

6 Zusammenfassung

Quinoa ist eine alte Kulturart aus der Andenregion und wurde vor ca. 5000-7000 Jahren domestiziert. Sie besitzt eine hohe Anpassungsfähigkeit und einen hohen Ernährungswert. Aus diesem Grund sind Produktion und Verbrauch in den letzten Jahren stark angestiegen. Der Quinoa Genpool besitzt eine hohe genetische Diversität, was einen Vorteil für die Züchtung neuer Sorten darstellt, die an einen Anbau im nördlichen Europa angepasst sind. Dafür ist ein Verständnis der Blühgenetik und anderer agronomischer Eigenschaften unerlässlich.

In dieser Studie wurde die genotypische und phänotypische Variation innerhalb der Art *Chenopodium quinoa* untersucht. Sie war in zwei Arbeitspakete unterteilt. Im ersten wurden Haplotypen und unterschiedliche Genexpression von Blühgenen untersucht, die eine Auswirkung auf die Anpassung an die Tageslänge haben. Im zweiten analysierte ich die Populationsstruktur, den Abfall des Kopplungsungleichgewichts und die Marker-Merkmal-Kopplung für Blühzeitpunkt und andere agronomisch bedeutsame Merkmale in einem Diversitätspanel von 310 Quinoa-Herkünften.

Ich führte ein Klimakammerexperiment mit 276 Quinoa-Herkünften durch, um das Blühverhalten unter Kurz- und Langtagbedingungen zu untersuchen. Dabei beobachtete ich eine höhere Variation unter Langtagbedingungen. Die Mehrzahl der Akzessionen zeigte eine „Kurztagantwort“ (67,4%), d.h., die Pflanzen blühten signifikant früher unter Kurztagsbedingungen. Dagegen blühten nur 7,6% der Herkünfte signifikant früher unter Langtagbedingungen. Akzessionen aus Südchile waren Tageslängen-insensitiv. Die Tageslängenunabhängigkeit erleichterte vermutlich den Anbau in niedrigeren Breitengraden.

Dann studierte ich die homologen Gene der wichtigsten Blühzeitpunktregulatoren, der Arabidopsisgene *FLOWERING LOCUS T (FT)* und *CONSTANS (CO)*, in Quinoa. Ich identifizierte 24 *FT* und sechs *CO* Homologe. Auf der Grundlage der größten Homologie zu den Zuckerrüben Genen *BvFT1* (Blührepressor) und *BvFT2* (Blühaktivator) selektierte ich sechs *CqFT* und sechs *CqCOL* Gene für eine Expressionsanalyse. Die circadiane Expressionsanalyse zeigte unter Langtagbedingungen eine Herunterregulation der *CqFT* Gene mit Ausnahme von *CqFT1B-2*. Die *CqCOL* Expression war tageslängenunabhängig. *CqFT2B* zeigt im Kurztag eine Peak am späten Abend, was für *FT* Orthologe typisch ist und dafür spricht, daß *CqFT2B* ein Blühaktivator im Kurztag ist. Die Expression von *CqCOL* ähnelte nicht der typischen *CO* Expression, jedoch den Expressionen von *BvCOL* und *AtCOL*, was darauf schließen läßt, daß Quinoa kein *CO* Ortholog besitzt.

Ich analysierte die Haplotypen von sechs *CqFT* und sechs *CqCOL* Genen auf der Grundlage von SNPs mit hoher Konfidenz, die im Quinoa-Diversitätspanel identifiziert wurden. Dabei fand ich zwischen 2 und 6 Haplotypen in allen untersuchten Genen. Die höchste Sequenzvariation zeigten die *CqCOL2A* Gene mit 6 Haplotypen, während sieben Genen nicht zu Aminosäuresubstitution führten. Interessanterweise fand ich keine SNP in *CqCOL5*. Sieben Haplotypen führten zu einem früheren Blühzeitpunkt im Kurztag, fünf im Langtag. Es wurden auch 12 tageslängenunabhängige Haplotypen identifiziert. Haplotypen aus sechs Genen zeigten eine starke Assoziation mit der Geographie.

Um das Diversitätspanel weiter zu charakterisieren, führte ich Feldversuche über 2 Jahre durch und analysierte 17 agronomische Merkmale. Ich fand grundlegende phänotypische Unterschiede zwischen den Herkünften. 9 Merkmale zeigten eine hohe Heritabilität. Eine Hauptkomponentenanalyse ergab zwei Cluster. In Zusammenarbeit mit der King Abdullah University of Science and Technology in Saudi-Arabien wurde eine vollständige Genomsequenzierung des Panels durchgeführt und 2,9 Millionen Hoch-Konfidenz-SNP

Zusammenfassung

identifiziert. In Übereinstimmung mit den Phänotypdaten zeigte die SNP basierte Hauptkomponentenanalyse ebenfalls hochdivergente Populationen, die Hochland- und Tieflandquinoa entsprachen. Ich fand mit $K=8$ die optimale Populationszahl in der ADMIXTURE Analyse. Die Hochlandpopulation war darüberhinaus noch in 5 Subpopulationen, die Tieflandpopulation in zwei unterteilt. Hochland- und Tieflandpopulationen zeigten eine hohe F_{ST} Divergenz von 0,36. Der LD Abfall der Hochlandpopulation war wesentlich steiler (6,5kb) als der Abfall der Tieflandpopulation (49,8kb). Die Nukleotid-Diversität war in der Hochlandpopulation ebenfalls um Faktor 1,62 höher, wohingegen der durchschnittliche Tajima D Wert in der Tieflandpopulation negativ war. Aufgrund dieser Daten schlage ich drei mögliche Szenarien der Quinoa-Domestikation und –Verbreitung in Südamerika vor:

- (1) Es gab zwei unabhängige Domestikationsereignisse mit einer starken Engpass bei der Tieflandpopulation
- (2) Es gab nur ein Domestikationsereignis, aber mit starkem Populationswachstum nach Anpassung der Tieflandpopulation, oder
- (3) einer starken adaptiven Selektion nach der Domestikation.

Um die Geschichte und Genetik der Domestikation besser zu verstehen, wird es notwendig sein, eine repräsentative Auswahl von wilden Verwandten von Quinoa zu sequenzieren.

Als nächstes führte ich eine genomweite Assoziationskartierung für 17 agronomische Merkmale durch und identifizierte 600 merkmalsassoziierte SNP (MTA), die über die Anbaujahre stabil waren. Ich konnte bereits früher identifizierte Kandidatengene für die Saponinsynthese und Blütenfarbe genau lokalisieren, was die Verlässlichkeit der Methode demonstrierte. Des weiteren analysierte ich Kandidatengene für Tausendkorngewicht und fand zwei Kandidatengene aus dem Brassinosteroid und Ubiquitin-Proteasom Stoffwechsel, die Samengröße regulieren sollen. Akzessionen mit den Haplotypen PP2C_hap3 und RING_hap7 zeigten größere Samen in beiden Jahren. Auf Chromosom Cq2A fand ich eine Kandidatenregion für Mehltau-Resistenz im Bereich 38.99 - 39.03 Mb. In dieser Region wurde ein Resistenzgen-Analog mit einer NBS-LRR Domäne (Nucleotide-Binding Site Leucine-Rich Repeat) gefunden. Die Region 80.50 -81.50 Mb auf Chromosom Cq2A war besonders interessant, da sie pleiotrope MTAs für DTF, DTM, PH, und PL zeigte. Der bedeutendste SNP lag im *CqGLX2-2* Gen, einem Gen das für ein Enzym der Glyoxalase Familie kodiert. Das Allel, das ein Cytosin enthielt, bewirkte frühe Blüte und Samenreife, kurze Rispen und niedrige Wuchshöhe.

Auf dieser Grundlage schlage ich eine Haplotyp-basierte Züchtungsstrategie vor durch Pyramidisierung von besonders vorteilhaften Allelen zur Anpassung an nördliche Breiten. Dabei können Akzessionen aus dem Diversitätspanel, die nicht so gut an Langtagbedingungen adaptiert sind, aber vorteilhafte agronomischen Merkmale besitzen, zunächst zur Kreuzung genutzt werden. Aus den Nachkommen können dann mit markergestützter Selektion und Markern in LD mit signifikanten MTAs geeignete Genotypen selektiert werden.

Appendix

7 Appendix**7.1 Supplementary tables**

Supplementary Table 1: Accessions from the quinoa diversity panel and results from re-sequencing (*This table is provided with the supplementary data DVD, due to the excessive length*)

Supplementary Table 2: Summary of high-quality SNPs identified in quinoa accessions (*This table is provided with the supplementary data DVD, due to the excessive length*)

Supplementary Table 3: Variance components analysis of 12 quantitative traits (*This table is provided with the supplementary data DVD, due to the excessive length*)

Supplementary Table 4: Summary of marker-trait associations (MTA)

Trait	No. of significant MTA 2018	No. of significant MTA 2019	No. of significant MTA BLUE	No. of consistent MTA
DTB	23	3	9	2
DTB to DTF	4	3	5	0
DTF	18	26	25	13
DTM	640	680	674	387
PH	43	23	54	11
PL	13	13	26	2
NoB	49	0	8	0
PD	91	11	46	2
TKW	316	496	508	168
Yield	7	49	19	0
Saponin	17	36	44	6
STL	13	1	4	0
FC	10	4	10	4
GT	1	27	14	0
SC	3	6	6	3
PSH	27	90	28	2
Mildew	NA	86	NA	NA
Total	1275	1554	1480	600

Supplementary Table 5: Candidate genes linked to SNP with significant trait associations (*This table is provided with the supplementary data DVD, due to excessive length*)

Appendix

Supplementary Table 6: Summary of MTA associated with DTF, DTM, PD, and PH identified on chromosome Cq2A

Chromosome	Position	Marker	DTF	DTM	PH	PL
			<i>P-value</i>			
Cq2A	4242017	02:4242017	3.15E-08	6.85E-10	2.91E-08	4.43E-07
Cq2A	4406829	02:4406829	1.63E-08	8.56E-10	5.91E-08	2.83E-07
Cq2A	8080505	02:8080505	4.52E-07	2.01E-11	3.69E-08	8.20E-08
Cq2A	8080510	02:8080510	5.13E-07	1.83E-11	7.21E-08	7.58E-08
Cq2A	8093547	02:8093547	4.66E-09	3.96E-14	1.42E-08	1.50E-07
Cq2A	8101350	02:8101350	4.08E-07	4.06E-12	9.25E-08	1.46E-07
Cq2A	8116642	02:8116642	3.87E-08	3.43E-09	1.88E-08	7.67E-07

Supplementary Table 7: Candidate genes located within the 50kb flanking regions of significantly associated SNPs from the multivariate GWAS analysis (*This table is provided with the supplementary data DVD, due to excessive length*)**Supplementary Table 8:** Plant material used in this study.

Seed Code	Kiel ID	Accession	Country	Experiment
170867	QP-343	PI-614886	Chile	E2
170876	QP-337	CHEN-109	Peru	E2
171230	QP-346	Titicaca	Denmark	E1/E2
171605	QP-349	PI-587173	Argentina	E2
182087	QP-001	BO-55	Chile	E1
182088	QP-002	Moroccan Yellow	Unknown	E1
182089	QP-003	Bouchane-3	Unknown	E1
182090	QP-004	PI-614889	Chile	E1
182091	QP-005	ICBA-Q5	Unknown	E1
182092	QP-006	PI-614927	Bolivia	E1
182094	QP-010	D-12196	Bolivia	E1
182095	QP-016	Baer-II	Unknown	E1
182096	QP-017	D-12036	Peru	E1
182097	QP-018	Javi	Chile	E1
182098	QP-019	E-DK-4	Unknown	E1
182099	QP-021	Ames-13734	United States	E1
182100	QP-022	D-12420	Bolivia	E1
182101	QP-023	D-12220	Peru	E1
182102	QP-026	Indian Quinoa	Unknown	E1
182103	QP-027	CHEN-5	Unknown	E1
182105	QP-030	PUC-mix-red	Chile	E1
182106	QP-031	PI-614887	Chile	E1
182107	QP-032	Brightest-Brilliant-Rainbow	Unknown	E1
182108	QP-034	CHEN-291	Peru	E1
182109	QP-035	RU-5	United Kingdom	E1
182111	QP-038	NSL-86628	Unknown	E1

Appendix

182112	QP-039	BO-40	Chile	E1
182114	QP-042	Ames-13745	United States	E1
182115	QP-043	Oro-de-Valle	Unknown	E1
182116	QP-044	Ames-13722	United States	E1
182118	QP-046	Ames-13744	United States	E1
182120	QP-048	PI-634917	Chile	E1
182121	QP-049	Ames-13738	United States	E1
182122	QP-050	D-12258	Peru	E1
182123	QP-053	CHEN-389	Peru	E1
182124	QP-055	PUC-mix-green	Chile	E1
182126	QP-059	Ames-13730	United States	E1
182127	QP-060	Ames-13743	Chile	E1
182128	QP-061	BO-58	Chile	E1
182129	QP-062	PI-634924	Chile	E1
182130	QP-063	Ames-13727	United States	E1
182131	QP-065	Vikinga	Denmark	E1
182132	QP-066	PI-614885	Chile	E1
182133	QP-067	PI-584524	Chile	E1
182134	QP-068	BO-16	Chile	E1
182135	QP-069	D-12405	Peru	E1
182136	QP-070	Ames-13760	United States	E1
182137	QP-071	PI-510543	Peru	E1
182138	QP-073	BO-13	Chile	E1
182139	QP-075	CHEN-147	Bolivia	E1
182140	QP-078	D-12184	Peru	E1
182142	QP-083	PI-665283	United States	E1
182143	QP-084	EMBRAPA-Brazil	Unknown	E1
182144	QP-085	D-12320	Peru	E1
182145	QP-086	Nde-09	Chile	E1
182146	QP-087	Ames-13742	United States	E1
182147	QP-089	RU-2	United Kingdom	E1
182148	QP-090	D-12282	Argentina	E1
182149	QP-091	BO-17	Chile	E1
182150	QP-093	0654.Q1.quinoa	Peru	E1
182151	QP-094	CHEN-66	Peru	E1
182152	QP-095	PI-614888	Chile	E1
182153	QP-096	PI-634923	Chile	E1
182154	QP-097	NSL-86649	Unknown	E1
182155	QP-098	BO-46-2	Chile	E1
182157	QP-100	PI-614880	Chile	E1
182158	QP-102	CHEN-415	Peru	E1
182159	QP-103	BO-03	Chile	E1
182160	QP-104	CHEN-218	Peru	E1

Appendix

182161	QP-105	NL-6	Chile	E1
182162	QP-107	BO-32	Chile	E1
182164	QP-109	PI-634925	Chile	E1
182165	QP-111	CHEN-64	Unknown	E1
182166	QP-112	D-12119	Peru	E1
182167	QP-113	Redhead	Unknown	E1
182168	QP-114	Ames-13728	United States	E1
182169	QP-116	Chucapaca	Bolivia	E1
182170	QP-118	Ames-13754	United States	E1
182172	QP-120	Ames-13750	United States	E1
182175	QP-123	Ames-13731	United States	E1
182176	QP-124	Ames-13748	United States	E1
182177	QP-126	BO-30	Chile	E1
182178	QP-127	Bouchane-4	Unknown	E1
182181	QP-133	Ames-13761	United States	E1
182183	QP-135	Ames-13740	United States	E1
182185	QP-137	Ames-13751	United States	E1
182186	QP-138	PI-614929	Bolivia	E1
182187	QP-139	NSL-91567	Unknown	E1
182192	QP-146	D-10005	Peru	E1
182193	QP-147	Ames-13758	United States	E1
182194	QP-148	D-12176	Peru	E1
182195	QP-149	D-12166	Bolivia	E1
182196	QP-150	CHEN-128	Bolivia	E1
182197	QP-152	PI-510549	Peru	E1
182200	QP-160	Ames-13733	United States	E1
182204	QP-165	BO-51	Chile	E1
182206	QP-168	D-12409	Bolivia	E1
182207	QP-169	D-11889	Argentina	E1
182209	QP-173	Ames-13723	United States	E1
182211	QP-175	BO-63	Chile	E1
182212	QP-176	BO-42	Chile	E1
182213	QP-178	D-12229	Peru	E1
182214	QP-179	Titicaca	Denmark	E1
182215	QP-180	Ames-13737	United States	E1
182216	QP-181	BO-11	Chile	E1
182217	QP-182	CHEN-322	Peru	E1
182218	QP-183	PI-614920	Bolivia	E1
182220	QP-185	PI-614882	Chile	E1
182221	QP-186	PI-478408	Bolivia	E1
182222	QP-187	CHEN-174	Bolivia	E1
182223	QP-188	D-11953	Peru	E1
182225	QP-190	D-12431	Bolivia	E1

Appendix

182226	QP-191	Ames-13732	United States	E1
182227	QP-194	CHEN-7	Switzerland	E1
182228	QP-195	D-12123	Peru	E1
182229	QP-197	D-12341	Peru	E1
182231	QP-199	CHEN-411	Peru	E1
182232	QP-202	D-12181	Bolivia	E1
182234	QP-206	CHEN-155	Bolivia	E1
182235	QP-212	CHEN-140	Bolivia	E1
182236	QP-213	BO-52	Chile	E1
182237	QP-215	NSL-92331	Unknown	E1
182238	QP-216	Ames-13725	United States	E1
182239	QP-217	CHEN-344	Peru	E1
182240	QP-218	CHEN-302	Peru	E1
182241	QP-220	PI-634918	Chile	E1
182243	QP-224	D-12279	Unknown	E1
182244	QP-225	Chery vanilla	United States	E1
182245	QP-229	BO-60	Chile	E1
182246	QP-231	Bouchane-2	Unknown	E1
182248	QP-233	ICBA-Q3	Unknown	E1
182250	QP-235	D-9992	Unknown	E1
182251	QP-237	D-12424	Bolivia	E1
182253	QP-241	CHEN-427	Peru	E1
182254	QP-242	CHEN-182	Peru	E1
182255	QP-244	D-11893	Peru	E1
182256	QP-245	CHEN-199	Bolivia	E1
182262	QP-256	D-12060	Peru	E1
182263	QP-258	CHEN-398	Peru	E1
182264	QP-261	CHEN-397	Peru	E1
182265	QP-262	CHEN-417	Peru	E1
182268	QP-269	Ratuqui	Unknown	E1
182269	QP-271	D-12162	Peru	E1
182271	QP-275	PI-614881	Argentina	E1
182272	QP-277	Kamiri	Unknown	E1
182273	QP-278	D-12135	Peru	E1
182274	QP-279	D-12346	Bolivia	E1
182276	QP-283	PI-614919	Bolivia	E1
182277	QP-284	D-9998	Unknown	E1
182281	QP-293	CHEN-390	Peru	E1
182282	QP-294	CHEN-429	Peru	E1
182283	QP-297	D-12082	Peru	E1
182288	QP-308	CHEN-401	Peru	E1
182315	QP-303	G-205	Unknown	E1
182318	QP-246	Salcedo-Inia	Peru	E1

Appendix

182333	QP-232	Bouchane-1	Unknown	E1
182335	QP-040	Ames-13753	United States	E1
182350	QP-248	PI-614914	Bolivia	E1
182355	QP-028	D-12211	Peru	E1
182360	QP-172	PI-634919	Chile	E1
182364	QP-255	CHEN-299	Peru	E1
182375	QP-238	D-12268	Peru	E1
182393	QP-129	D-12387	Peru	E1
182397	QP-228	D-11985	Peru	E1
182398	QP-081	Ames-13726	United States	E1
182402	QP-013	D-12243	Peru	E1
182404	QP-307	CHEN-112	Peru	E1
182408	QP-302	CHEN-412	Peru	E1
182418	QP-119	Ames-13755	United States	E1
182420	QP-110	CHEN-222	Bolivia	E1
182425	QP-270	CHEN-144	Bolivia	E1
182431	QP-184	D-12201	Peru	E1
182443	QP-074	CHEN-382	Peru	E1
182448	QP-064	CHEN-33	Unknown	E1
182454	QP-009	CHEN-298	Peru	E1
182456	QP-011	D-11980	Peru	E1
182462	QP-140	PI-614884	Argentina	E1
182463	QP-041	Ames-13721	United States	E1
182466	QP-153	REAL	Bolivia	E1
182469	QP-037	CHEN-90	Peru	E1
182474	QP-036	Regalona	Chile	E1
182476	QP-125	CHEN-226	Bolivia	E1
182672	QP-012	CHEN-465	Peru	E1
182674	QP-014	CHEN-349	Peru	E1
182675	QP-015	CHEN-420	Peru	E1
182677	QP-020	D-12205	Peru	E1
182679	QP-024	D-12345	Peru	E1
182680	QP-025	CHEN-424	Peru	E1
182683	QP-033	CHEN-470	Peru	E1
182687	QP-051	D-12357	Peru	E1
182688	QP-052	D-12167	Peru	E1
182690	QP-054	D-12208	Bolivia	E1
182691	QP-056	D-12037	Bolivia	E1
182692	QP-057	CHEN-126	Bolivia	E1
182693	QP-058	D-12087	Peru	E1
182696	QP-072	D-12401	Peru	E1
182699	QP-076	CHEN-404	Peru	E1
182700	QP-077	D-12411	Peru	E1

Appendix

182701	QP-079	D-9973	Unknown	E1
182702	QP-080	D-12175	Peru	E1
182704	QP-082	D-12253	Bolivia	E1
182706	QP-088	PI-665277	Bolivia	E1
182707	QP-092	CHEN-109	Peru	E1
182709	QP-101	D-12183	Peru	E1
182711	QP-106	CHEN-65	Peru	E1
182717	QP-117	CHEN-319	Peru	E1
182719	QP-121	CHEN-159	Bolivia	E1
182724	QP-130	CHEN-205	Peru	E1
182725	QP-131	CHEN-256	Peru	E1
182726	QP-132	D-12289	Peru	E1
182727	QP-134	Ollague	Chile	E1
182731	QP-142	PI-596498	Peru	E1
182732	QP-143	CHEN-342	Peru	E1
182733	QP-144	PI-510533	Peru	E1
182742	QP-154	CHEN-91	Peru	E1
182743	QP-155	CHEN-162	Bolivia	E1
182744	QP-156	CHEN-361	Peru	E1
182745	QP-157	CHEN-219	Bolivia	E1
182747	QP-159	CHEN-437	Peru	E1
182748	QP-161	CHEN-71	Bolivia	E1
182749	QP-162	D-11943	Peru	E1
182750	QP-163	D-12416	Peru	E1
182752	QP-166	D-12354	Peru	E1
182753	QP-167	D-11965	Peru	E1
182755	QP-170	CHEN-220	Bolivia	E1
182756	QP-171	CHEN-224	Peru	E1
182757	QP-174	CHEN-225	Peru	E1
182758	QP-177	D-12397	Peru	E1
182765	QP-192	CHEN-406	Peru	E1
182766	QP-193	D-12361	Peru	E1
182772	QP-200	D-11936	Peru	E1
182773	QP-201	03-21-072RM	Peru	E1
182775	QP-203	CHEN-354	Peru	E1
182777	QP-205	CHEN-430	Peru	E1
182779	QP-207	D-12120	Peru	E1
182780	QP-208	CHEN-89	Peru	E1
182781	QP-209	CHEN-391	Peru	E1
182782	QP-210	PI-510545	Peru	E1
182783	QP-211	CHEN-450	Peru	E1
182785	QP-214	D-12393	Peru	E1
182788	QP-219	CHEN-416	Peru	E1

Appendix

182790	QP-222	D-12350	Peru	E1
182791	QP-223	PI-478418	Bolivia	E1
182792	QP-226	PI-674266	Ecuador	E1
182793	QP-227	D-12353	Peru	E1
182795	QP-230	Kancolla	Unknown	E1
182797	QP-236	CHEN-297	Peru	E1
182801	QP-240	D-12177	Peru	E1
182804	QP-243	D-12251	Peru	E1
182808	QP-247	D-11986	Peru	E1
182812	QP-251	D-11973	Peru	E1
182813	QP-252	D-12256	Peru	E1
182814	QP-253	D-12114	Peru	E1
182817	QP-257	CICA	Peru	E1
182819	QP-259	D-10003	Peru	E1
182824	QP-264	D-12074	Peru	E1
182827	QP-268	D-12363	Peru	E1
182831	QP-273	CHEN-413	Peru	E1
182832	QP-274	CHEN-384	Peru	E1
182834	QP-276	Pasankalla	Unknown	E1
182838	QP-280	D-11935	Peru	E1
182839	QP-281	D-9977	Unknown	E1
182840	QP-282	CHEN-339	Peru	E1
182843	QP-286	CHEN-383	Peru	E1
182844	QP-287	D-12027	Peru	E1
182845	QP-288	CHEN-151	Bolivia	E1
182847	QP-290	D-12413	Bolivia	E1
182848	QP-291	D-12182	Peru	E1
182849	QP-292	D-12179	Peru	E1
182852	QP-295	CHEN-212	Peru	E1
182853	QP-296	CHEN-358	Peru	E1
182856	QP-299	D-11918	Peru	E1
182857	QP-300	CHEN-433	Peru	E1
182867	QP-311	D-12021	Peru	E1
182869	QP-313	D-12172	Peru	E1
182872	QP-316	CHEN-142	Bolivia	E1
182881	QP-326	D-12395	Bolivia	E1
182883	QP-328	D-12222	Peru	E1
182885	QP-330	CHEN-425	Peru	E1

Appendix

Supplementary Table 9: List of PEBP family genes in quinoa.

Gene	Gene ID	Chromosome	Start	End	Strand	Gene length (bp)	Protein length	Protein similarity %			Ref. genome annotation
								BvFT2	CrFT1	AtFT	
<i>CqFT2B</i>	CQ019992	Cq4B	3497140	3503985	+	6845	177	92	92.57	73.14	Similar to HD3A: Protein HEADING DATE 3A (Oryza sativa subsp. japonica OX%3D39947)
<i>CqFT2-like</i>	CQ049229	Cq1A	43527958	43528515	+	557	88	89.66	91.95	68.97	Similar to FT: Protein FLOWERING LOCUS T (Arabidopsis thaliana OX%3D3702)
<i>CqFT1B-1</i>	CQ007311	Cq7B	3153968	3157869	-	3901	181	86.29	83.43	74.29	Similar to HD3A: Protein HEADING DATE 3A (Oryza sativa subsp. japonica OX%3D39947)
<i>CqFT1A</i>	CQ052447	Cq9A	51593659	51597564	+	3905	181	85.71	82.86	74.86	Similar to FT: Protein FLOWERING LOCUS T (Arabidopsis thaliana OX%3D3702)
<i>CqFT2A</i>	CQ035263	Cq4A	3164713	3168721	+	4007	198	81.71	83.43	66.86	Similar to FT: Protein FLOWERING LOCUS T (Arabidopsis thaliana OX%3D3702)
<i>CqTFSA</i>	CQ026333	Cq6A	14561770	14566485	-	4715	176	67.25	65.5	62.79	Similar to TSF: Protein TWIN SISTER of FT (Arabidopsis thaliana OX%3D3702)
<i>CqTSFB</i>	CQ000928	Cq6B	15486422	15488654	+	2232	167	64.81	63.58	60.12	Similar to TSF: Protein TWIN SISTER of FT (Arabidopsis thaliana OX%3D3702)

Appendix

<i>CqHD3AA</i>	CQ052448	Cq9A	51606934	51610237	-	3303	132	64.65	61.62	57.58	Similar to HD3A: Protein HEADING DATE 3A (Oryza sativa subsp. japonica OX%3D39947)
<i>CqRFTA</i>	CQ026350	Cq6A	15150517	15151242	-	725	104	63.37	62.38	55.45	Similar to RFT1: Protein RICE FLOWERING LOCUS T 1 (Oryza sativa subsp. japonica OX%3D39947)
<i>CqFT1B-2</i>	CQ007309	Cq7B	3146721	3150386	-	3664	205	63.01	63.01	54.34	
<i>CqHD3AB</i>	CQ000922	Cq6B	14735351	14740133	+	4782	178	62.64	61.49	57.47	Similar to HD3A: Protein HEADING DATE 3A (Oryza sativa subsp. japonica OX%3D39947)
<i>CqTSFB-2</i>	CQ000924	Cq6B	15080357	15084855	+	4498	178	61.49	60.34	56.32	Similar to TSF: Protein TWIN SISTER of FT (Arabidopsis thaliana OX%3D3702)
<i>CqCET4B</i>	CQ009691	Cq7B	69996734	70000056	-	3322	174	58.48	58.48	56.98	Similar to CET4: CEN-like protein 4 (Nicotiana tabacum OX%3D4097)
<i>CqCET2B</i>	CQ032861	Cq9B	4979813	4982251	+	2438	169	57.74	58.33	57.14	Similar to CET2: CEN-like protein 2 (Nicotiana tabacum OX%3D4097)
<i>CqCEN2A-2</i>	CQ040951	Cq7A	897708	900025	-	2317	172	57.65	56.47	56.73	Similar to CET2: CEN-like protein 2 (Nicotiana tabacum OX%3D4097)
<i>CqCET2A-1</i>	CQ040950	Cq7A	877362	879679	-	2317	172	57.65	56.47	56.73	Similar to CET2: CEN-like protein 2 (Nicotiana tabacum OX%3D4097)
<i>CqCET4A</i>	CQ043380	Cq7A	51972849	51975945	-	3096	174	57.31	57.89	55.81	Similar to CET4: CEN-like protein 4 (Nicotiana tabacum OX%3D4097)

Appendix

<i>CqMFTA-1</i>	CQ030683	Cq5A	37714977	37718215	-	3238	174	57.14	54.76	48.52	Similar to MFT: Protein MOTHER of FT and TFL1 (Arabidopsis thaliana OX%3D3702)
<i>CqCET1B</i>	CQ016755	Cq3B	2947577	2949524	-	1947	175	56.73	56.73	56.98	Similar to CET1: CEN-like protein 1 (Nicotiana tabacum OX%3D4097)
<i>CqCET1A</i>	CQ044067	Cq3A	2681971	2684097	-	2126	175	56.73	56.73	56.98	Similar to CET1: CEN-like protein 1 (Nicotiana tabacum OX%3D4097)
<i>CqMFTA-2</i>	CQ038400	Cq8A	4458609	4461250	-	2641	180	51.46	50.88	46.2	Similar to MFT: Protein MOTHER of FT and TFL1 (Arabidopsis thaliana OX%3D3702)
<i>CqMFTB-1</i>	CQ010508	Cq7B	4157584	4160499	-	2914	179	50.88	50.88	46.78	Similar to MFT: Protein MOTHER of FT and TFL1 (Arabidopsis thaliana OX%3D3702)
<i>CqMFTB-2</i>	CQ010508.5	Cq7B	4154966	4156139	-	1172	179	44.77	44.19	40.46	Similar to MFT: Protein MOTHER of FT and TFL1 (Arabidopsis thaliana OX%3D3702)
<i>CqRFTB</i>	CQ000927	Cq6B	15382595	15390216	+	7621	111	40	41	35	Similar to RFT1: Protein RICE FLOWERING LOCUS T 1 (Oryza sativa subsp. japonica OX%3D39947)

Appendix

Supplementary Table 10: CONSTANS-like genes in quinoa analyzed in this study

Gene	Gene ID	Chromosome	Start	End	Strand	Gene length (bp)	Protein length	Protein similarity %			Ref. genome annotation
								BvCOL1	CrCOL1	CO	
<i>CqCOL5A</i>	CQ040185	Cq8A	51827124	51829397	+	2273	389	40.71	41.94	37.34	Similar to COL5: Zinc finger protein CONSTANS-LIKE 5 (Arabidopsis thaliana OX%3D3702)
<i>CqCOL5B</i>	CQ013023	Cq8B	74652550	74654898	-	2348	389	41.59	42.81	37.93	Similar to COL5: Zinc finger protein CONSTANS-LIKE 5 (Arabidopsis thaliana OX%3D3702)
<i>CqCOL4A-2</i>	CQ052851	Cq2A	1028973	1031191	+	2218	381	50.64	51.13	43.31	Similar to COL4: Zinc finger protein CONSTANS-LIKE 4 (Arabidopsis thaliana OX%3D3702)
<i>CqCOL4A-1</i>	CQ037524	Cq4A	54461001	54462646	+	1645	387	51.44	52.24	43.17	Similar to COL4: Zinc finger protein CONSTANS-LIKE 4 (Arabidopsis thaliana OX%3D3702)
<i>CqCOL2A</i>	CQ055009	Cq2A	46542059	46545047	+	2988	367	89.34	94.25	54.87	Similar to COL2: Zinc finger protein CONSTANS-LIKE 2 (Arabidopsis thaliana OX%3D3702)
<i>CqCOL2B</i>	CQ015386	Cq2B	63445581	63448600	+	3019	367	89.89	94.79	55.15	Similar to COL2: Zinc finger protein CONSTANS-LIKE 2 (Arabidopsis thaliana OX%3D3702)

Appendix

Supplementary Table 11: Description of primer combinations used for RT-qPCR, amplification product size, and annealing temperature

Gene abbreviation	Gene ID	Primer ID	Primer sequence 5'-3' (bp)	Amplification product size (bp)	Annealing temperature (°C)
<i>CqFT1A</i>	CQ052447	NEQ017 + NMQ014	F:GGATCCTTTCAACAGGTCTGTAACTC R:CTAGGGTCGAGGTTGGAGGCG	119	59
<i>CqFT1B-1</i>	CQ007311	NEQ001 + NMQ005	F:GGATCCTTTTAATAGGTCCGTAAATC R:GTAATGGTGGATCCTGATGC	159	54
<i>CqFT1B-2</i>	CQ007309	NMQ008 + NMQ009	F:GCCCAAGAAACCCGTCGC R:GCCAAGAGGTTGTTCCGTATG	99	56
<i>CqFT2-B</i>	CQ019992	NMQ023 + NMQ024	F:GATATTCCTGGGTCCACAGGTGC R:CGACAACCTCGGAAGGCAAACCTG	124	58
<i>CqFT2-A</i>	CQ035263	NMQ017 + NMQ018	F:GCCAAGAGATAGTTTGCTATGAGAGC R:CACAACCTAGAGAATTTGCTGAAGTC	146	56
<i>CqRFT1</i>	CQ026350	DSQ303 + DSQ304	F:TTATCA ACGCTTGTGAGCTTAGACCC R:GATTGCCAGGACTTGGAGCG	135	60
<i>CqCOL2A</i>	CQ055009	DSQ314 + DSQ325	F:GTGACACGTGCCGTTTCAGCA R:CAA CTG ATT GGC AGC GTG G	101	61
<i>CqCOL2B</i>	CQ015386	DSQ326 + DSQ316	F:GTGACACGTGCCGTTTCAGCG R:ATGCCAACTGATTGGCAGCGTGA	105	62
<i>CqCOL4A-2</i>	CQ052851	DSQ323 + DSQ328	F:GCAAAGCTGACGCCGCACATA R:GGGGAAAGAGGGACACGATC	97	60
<i>CqCOL4A-1</i>	CQ037524	DSQ321 + DSQ327	F:GTAAAGCTGACGCCGCACACC R:GGGAAAGAGGGACACGGTC	96	60
<i>CqCOL5A</i>	CQ040185	DSQ317 + DSQ318	F:GTTGCTGTAATCCCGTTTCTCGAC R:GACTCGGTATAACCGTCGCG	103	60

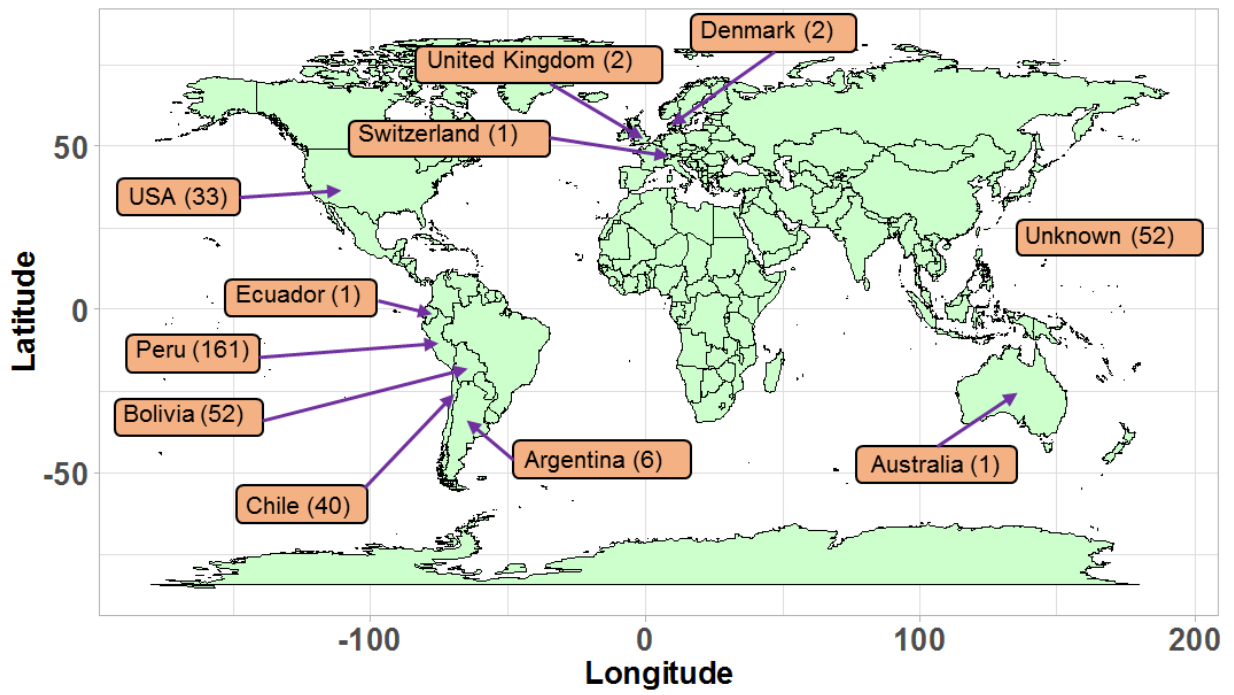
Appendix

<i>CqCOL5B</i>	CQ013023	DSQ319 + DSQ320	F: GTTGCTGTTATCCCGTTTCTTGAA R: TATCACGGCTAGCGTCGAGG	97	60
<i>CqIDH-A</i>	CQ044820	NMQ057+NMQ058	F: GGACGGACTATTGAAGCTGAAGC R: GCATGGACACGTGGGCTTGC	125	60
<i>CqPTB</i>	CQ013990	NMQ070+NMQ071	F: GCAGTGCAACAGGCTCCAAG R: GGCACCTGCACCCTCATG	143	57

Supplementary Table 12: Summary of sampling stages for temporal expression analysis

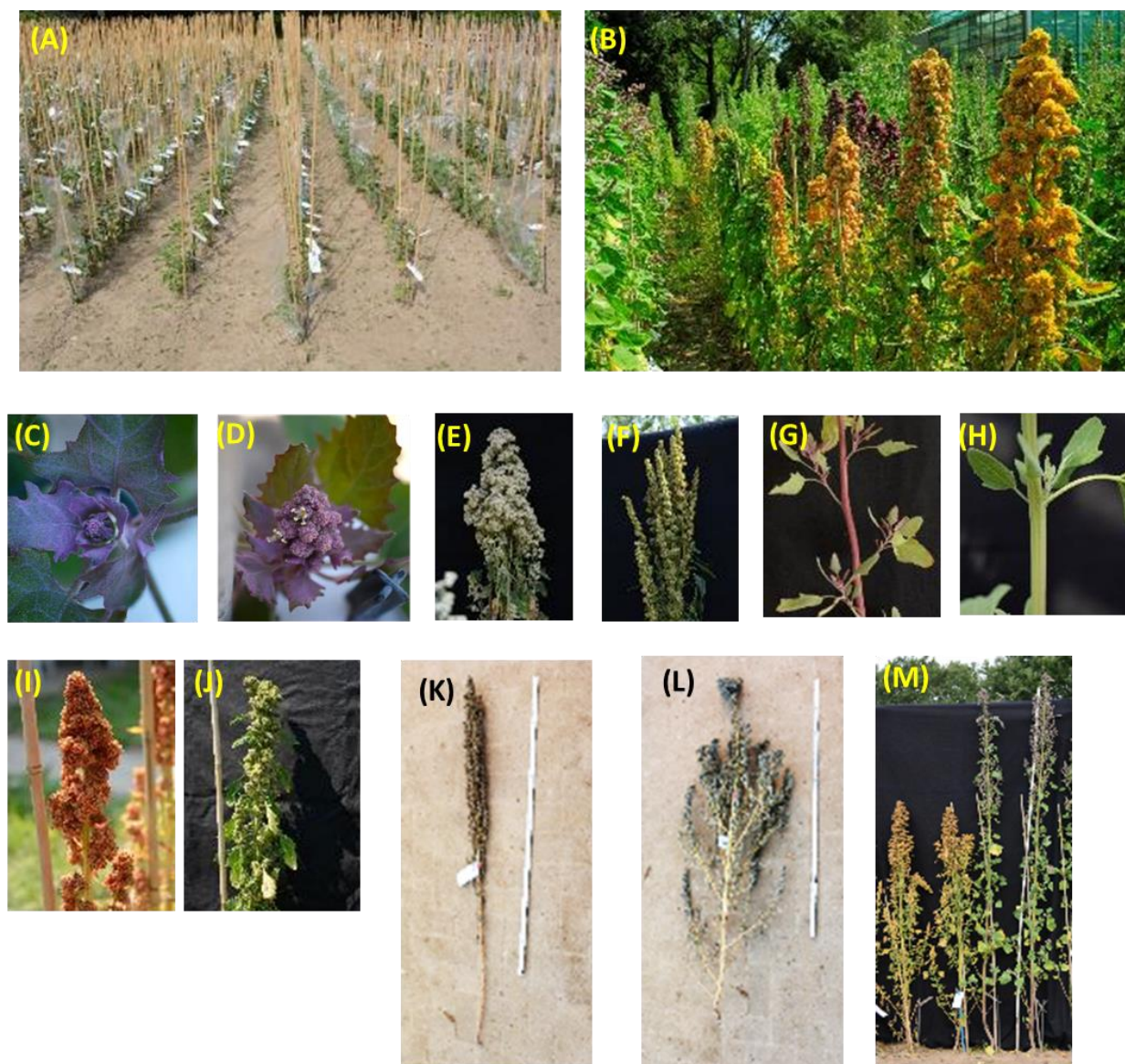
Sampling stages	171230		170605	
	SDs	LDs	SDs	LDs
Vegetative (Two weeks after sowing)	na	na	✓	✓
Two weeks before bolting	na	na	✓	✓
One week before bolting	✓	✓	✓	✓
Bolting	✓	✓	✓	✓
One week after bolting	na	✓	✓	✓
One week before flowering	✓	✓	✓	✓
Flowering	✓	✓	✓	✓
One week after flowering	✓	✓	✓	✓
Two weeks after flowering	na	na	✓	na

7.2 Supplementary figures



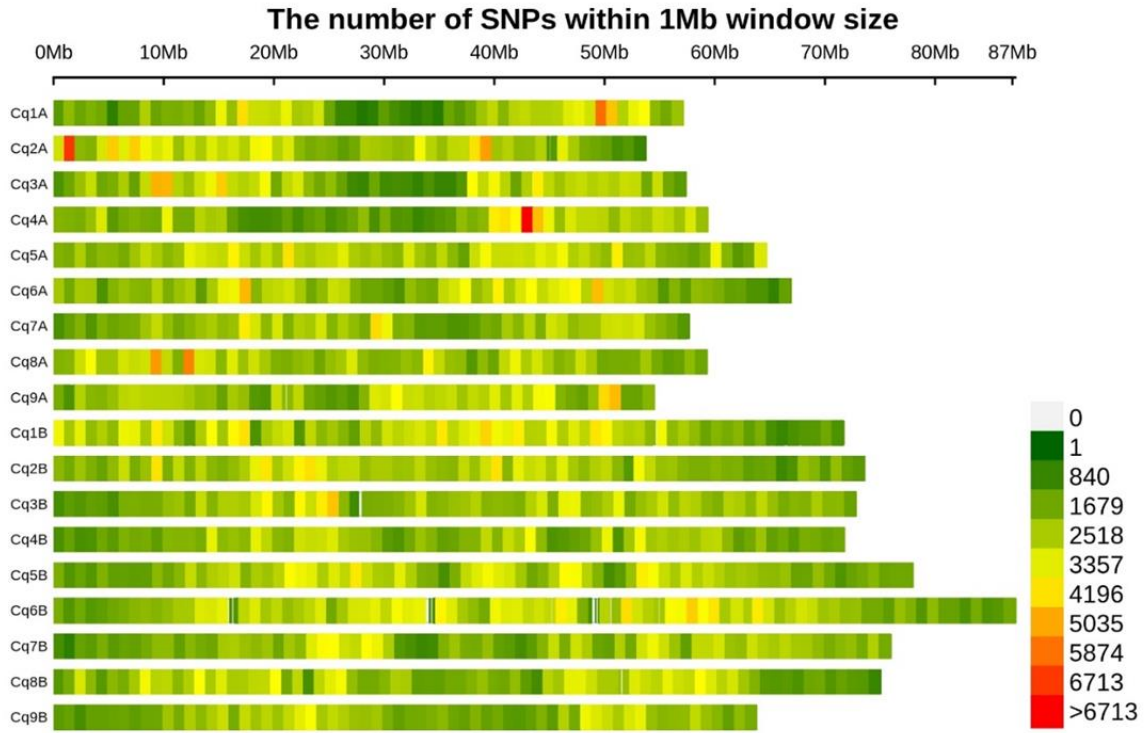
Supplementary Fig. 1: Geographical origin of the accessions forming the quinoa diversity panel.

Appendix



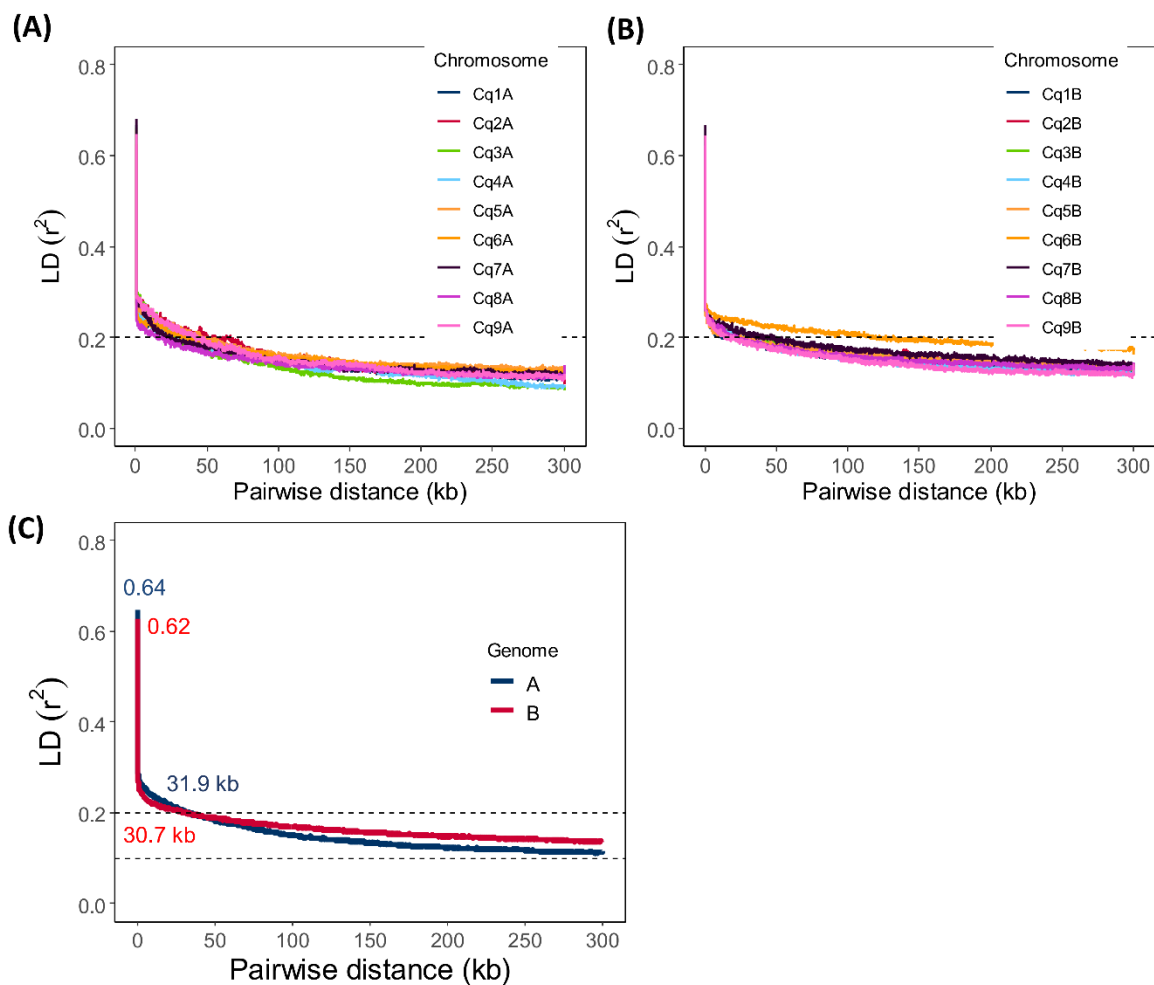
Supplementary Fig. 2: Overview of the field experiment and example images of phenotypic traits. A and B: Overview of the field and phenotypic variation among accession. C: Bolting (BBCH51) and flowering (BBCH60) (D) stage. Panicle shapes: glomerulate (E) and amarantiform (F). Stem color red (G) and green (H). Flower/inflorescence color red (I) and green (J). Growth type 1 (K) and type 5 (L). M: Plant height and maturity variation between the two accessions.

Appendix



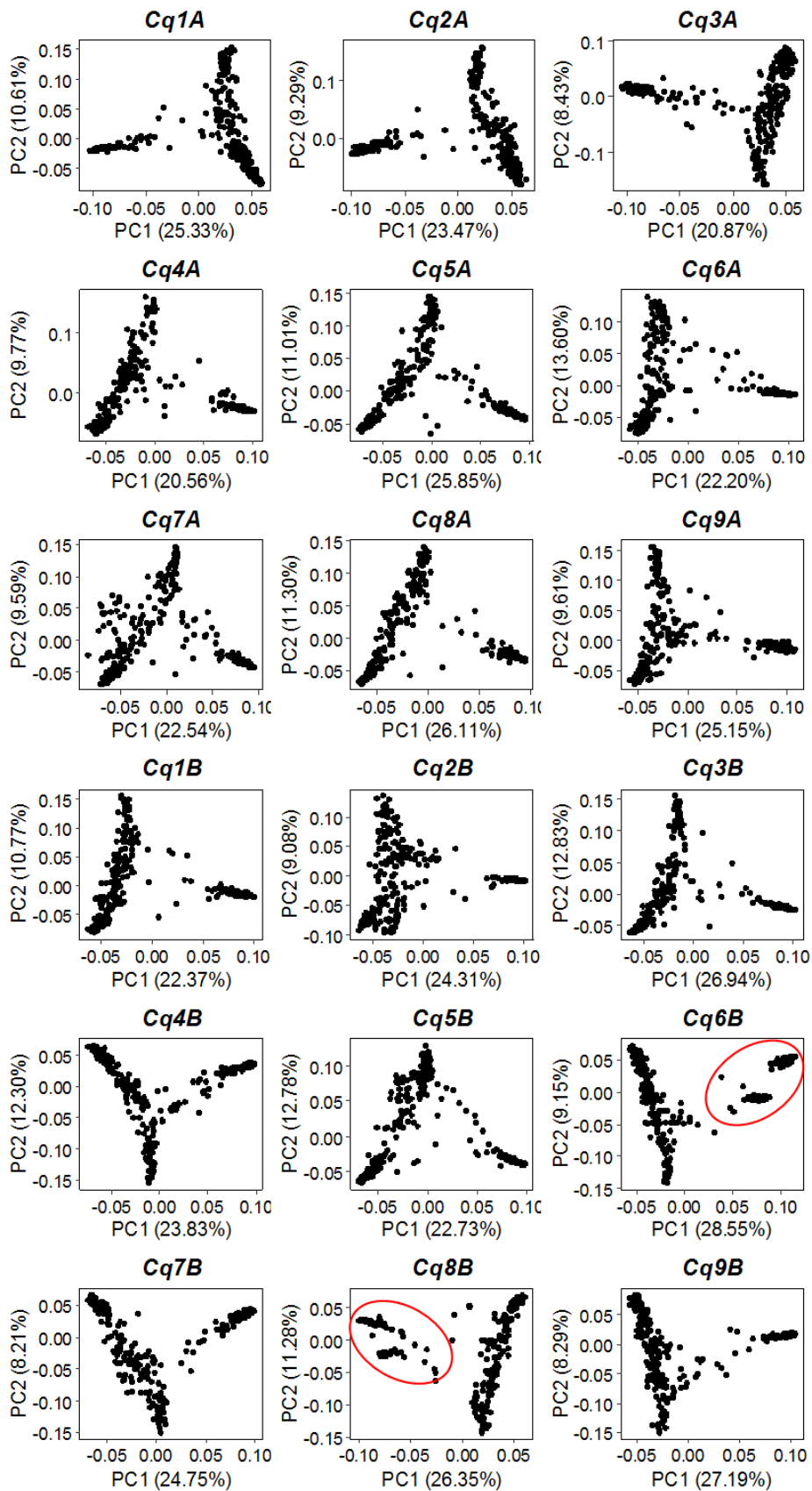
Supplementary Fig. 3: SNP density heat map across the 18 quinoa chromosomes. Different colors depict SNP density.

Appendix



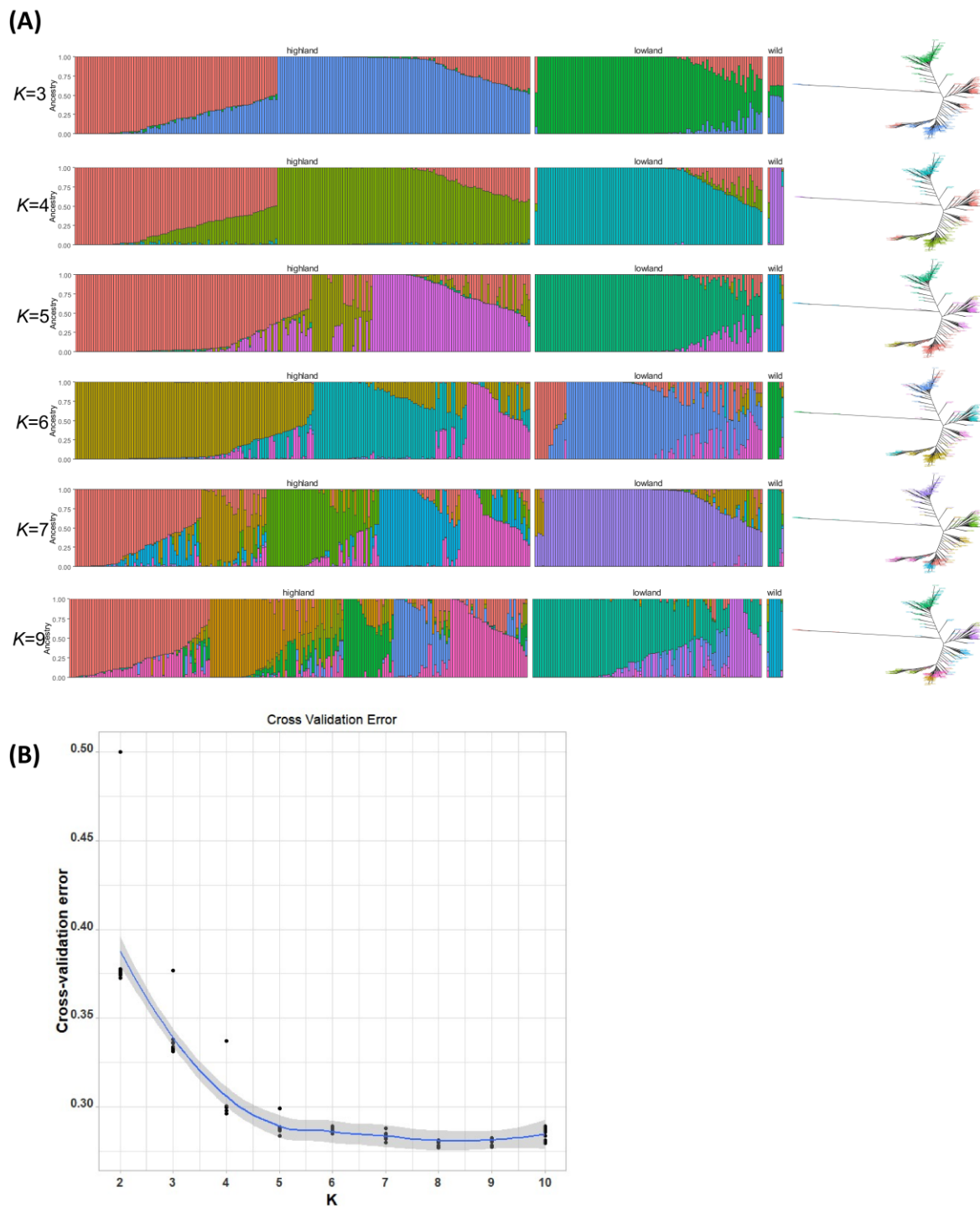
Supplementary Fig. 4: Chromosome wide LD decay in genome A (A) and genome B (B). Colors are depicting different chromosomes. (C) Genome-wide average LD decay of the A sub-genome (blue) and B sub-genome (red).

Appendix



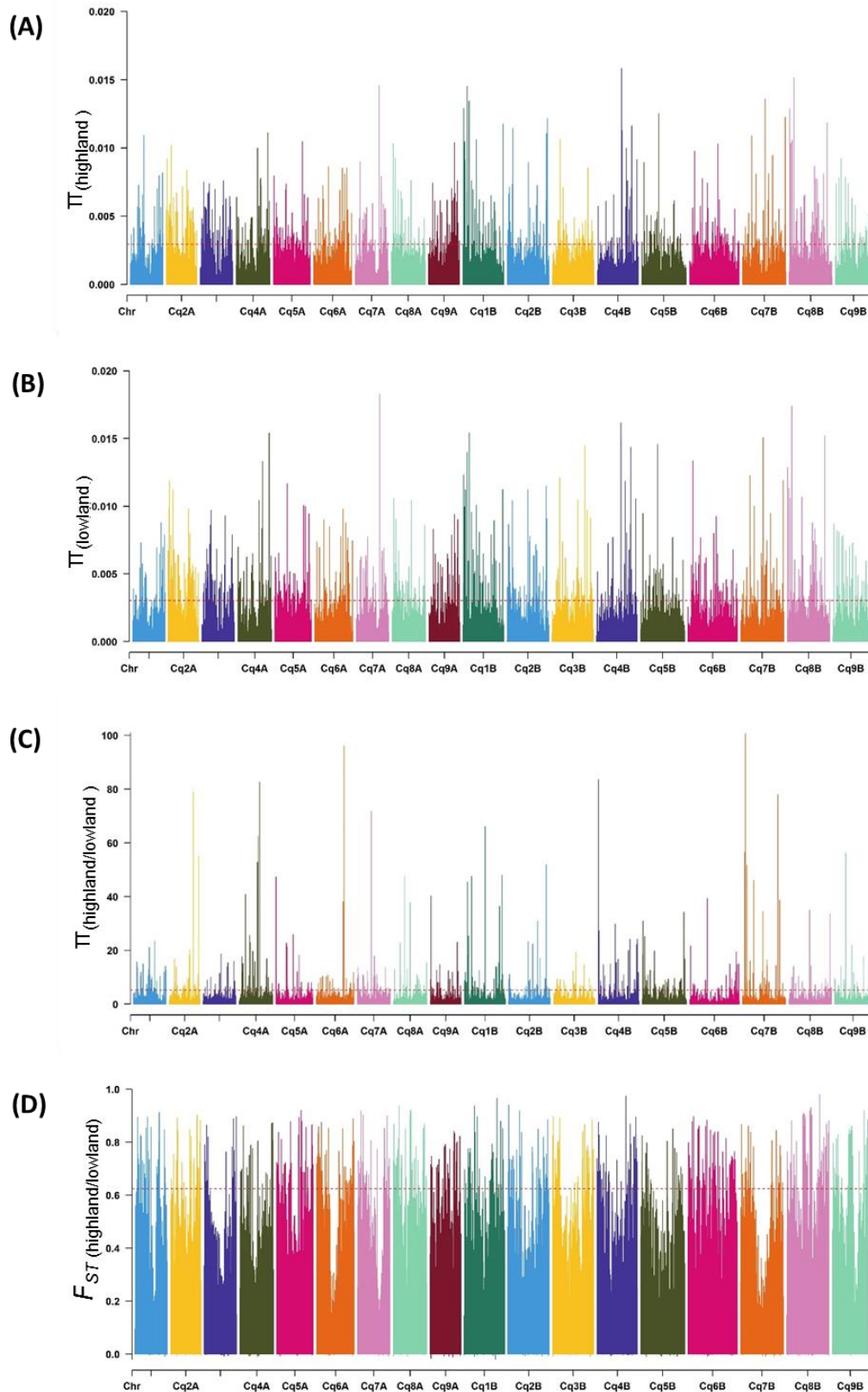
Supplementary Fig. 5: SNP based PCA across all 18 quinoa chromosomes. Red circles are depicting the two clusters of Lowland accessions.

Appendix



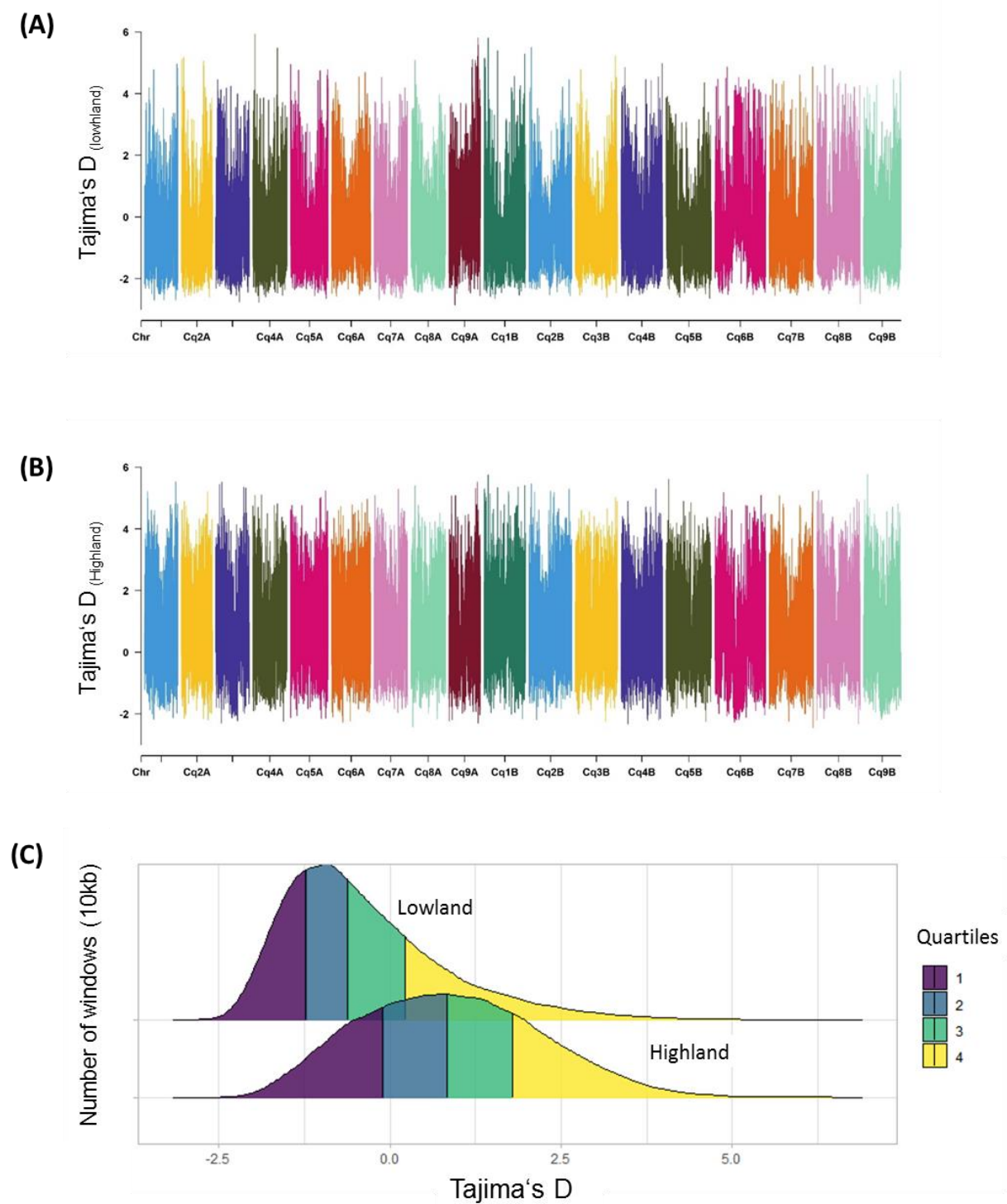
Supplementary Fig. 6: (A) ADMIXTURE ancestry coefficients for K ranging from 3 to 7 and 9. Each vertical bar represents an accession, and color proportions on the bar correspond to the genetic ancestry. (B) Cross-validation error in ADMIXTURE run.

Appendix



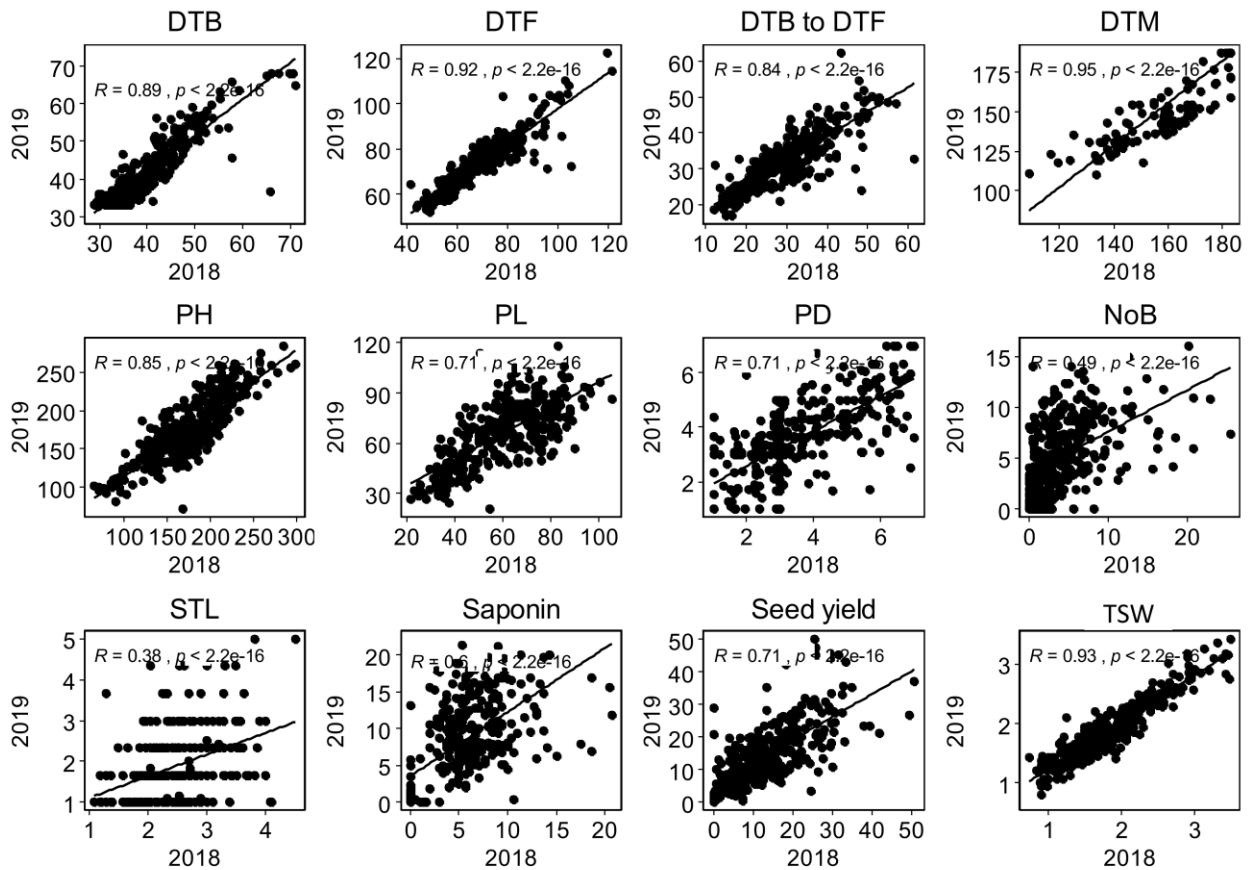
Supplementary Fig. 7: Diversity of populations along chromosomes measured based on 10 kb non-overlapping windows. Nucleotide diversity (π) distribution of 10 kb windows in population Highland (A) and Lowland (B). (C) Nucleotide diversity ratios (π Lowland/ π Highland). (D) Pairwise genome-wide fixation index (F_{ST}) between Highland and Lowland. The broken horizontal line represents the top 1% threshold.

Appendix



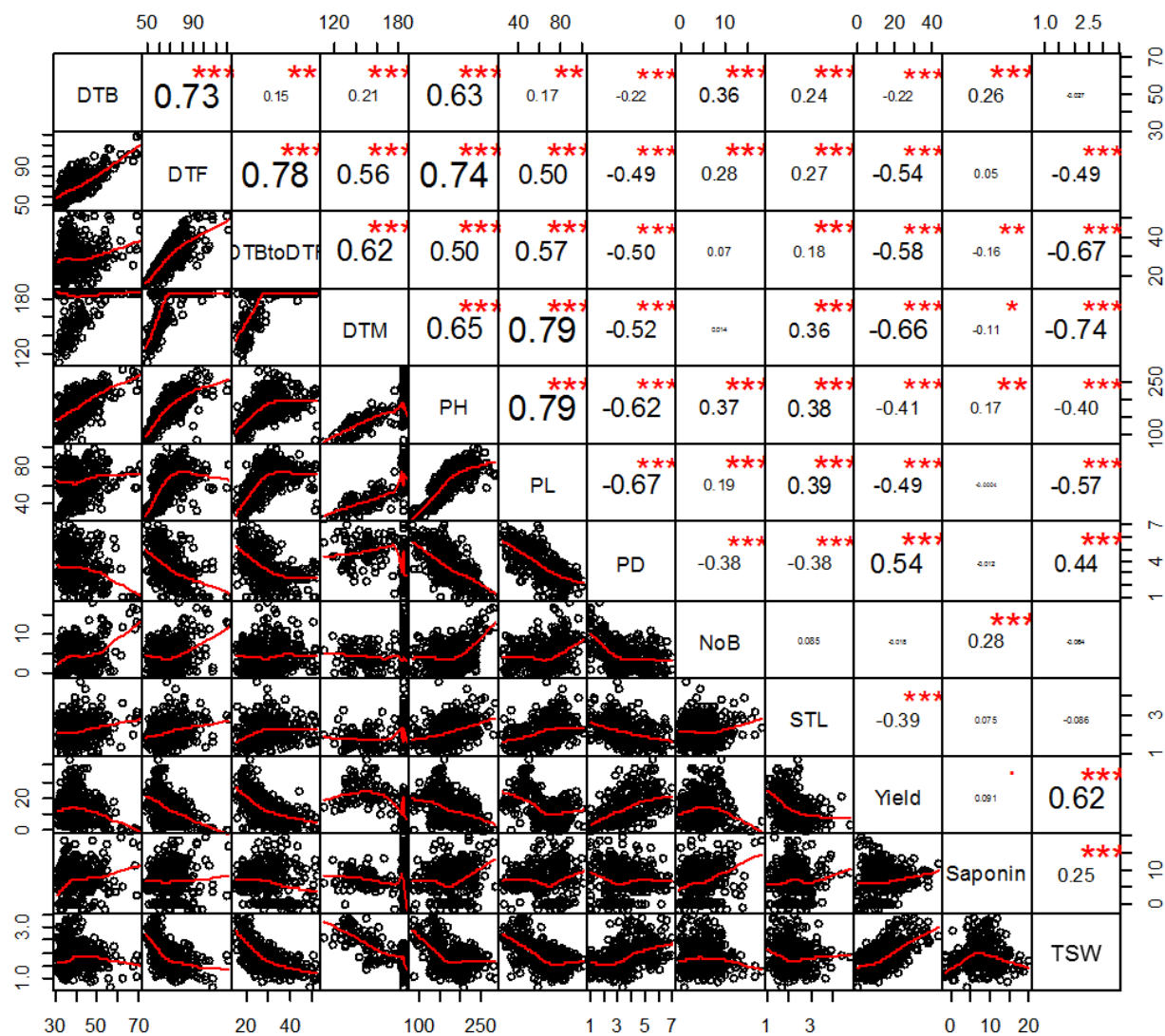
Supplementary Fig. 8: Distribution of Tajima's D along chromosomes in Lowland (A) and Highland (B) populations. Density distribution of Tajima's D between populations (C). Different colors represent the quartiles.

Appendix



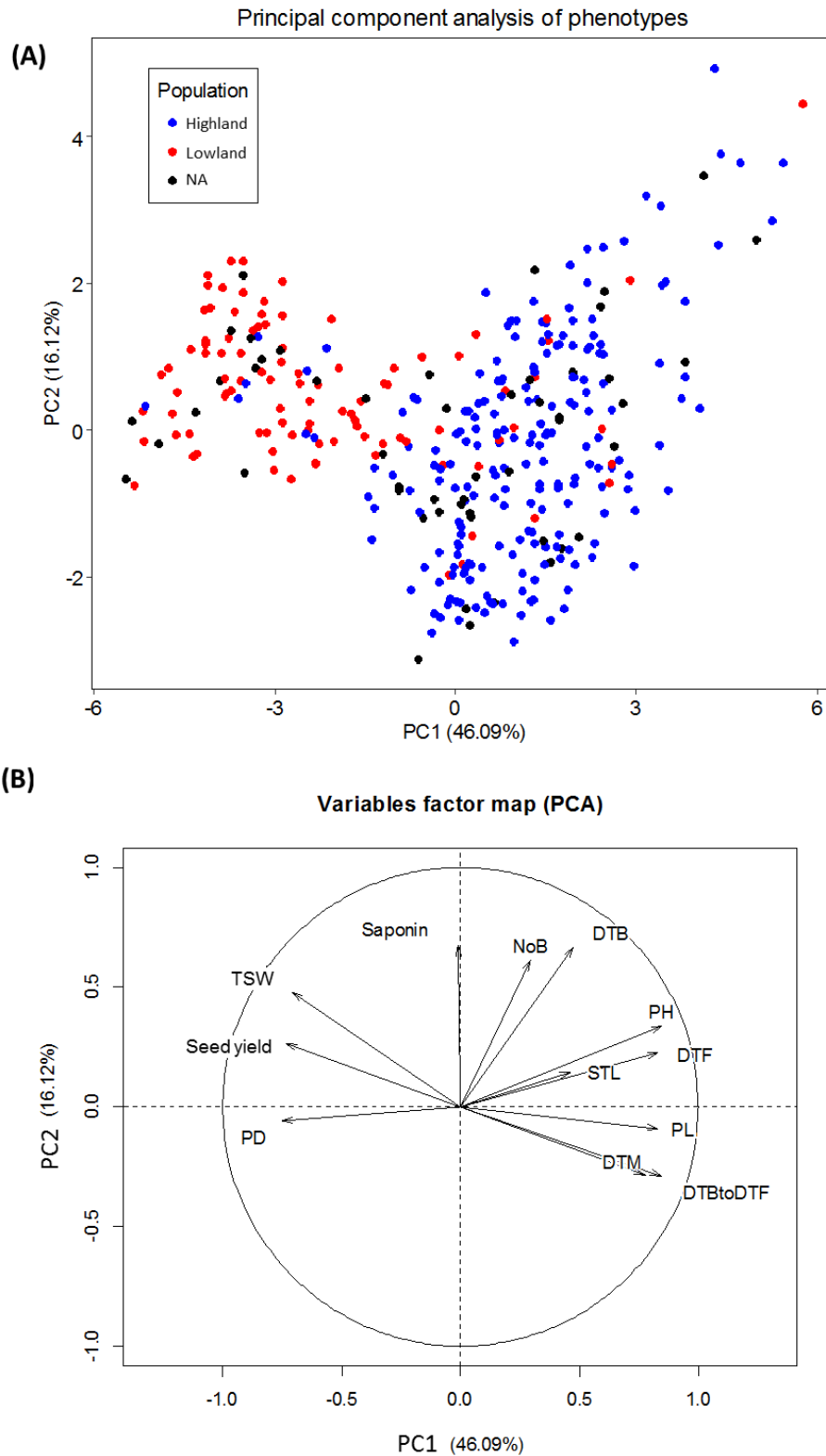
Supplementary Fig. 9: Graphical presentation of correlations between years among 12 traits. Pearson correlation value (R) with P-values are shown. DTB: days to bolting (inflorescence emergence), DTF: days to flowering, DTB to DTF: days between bolting and flowering, DTM; days to maturity, PH: plant height (cm), PL: panicle length (cm), PD: panicle density (cm), NoB: Number of branches, STL: stem lying, Saponin: saponin content as foam height (mm), Seed yield: seed yield per plant (g), TSW: thousand kernel weight (g).

Appendix



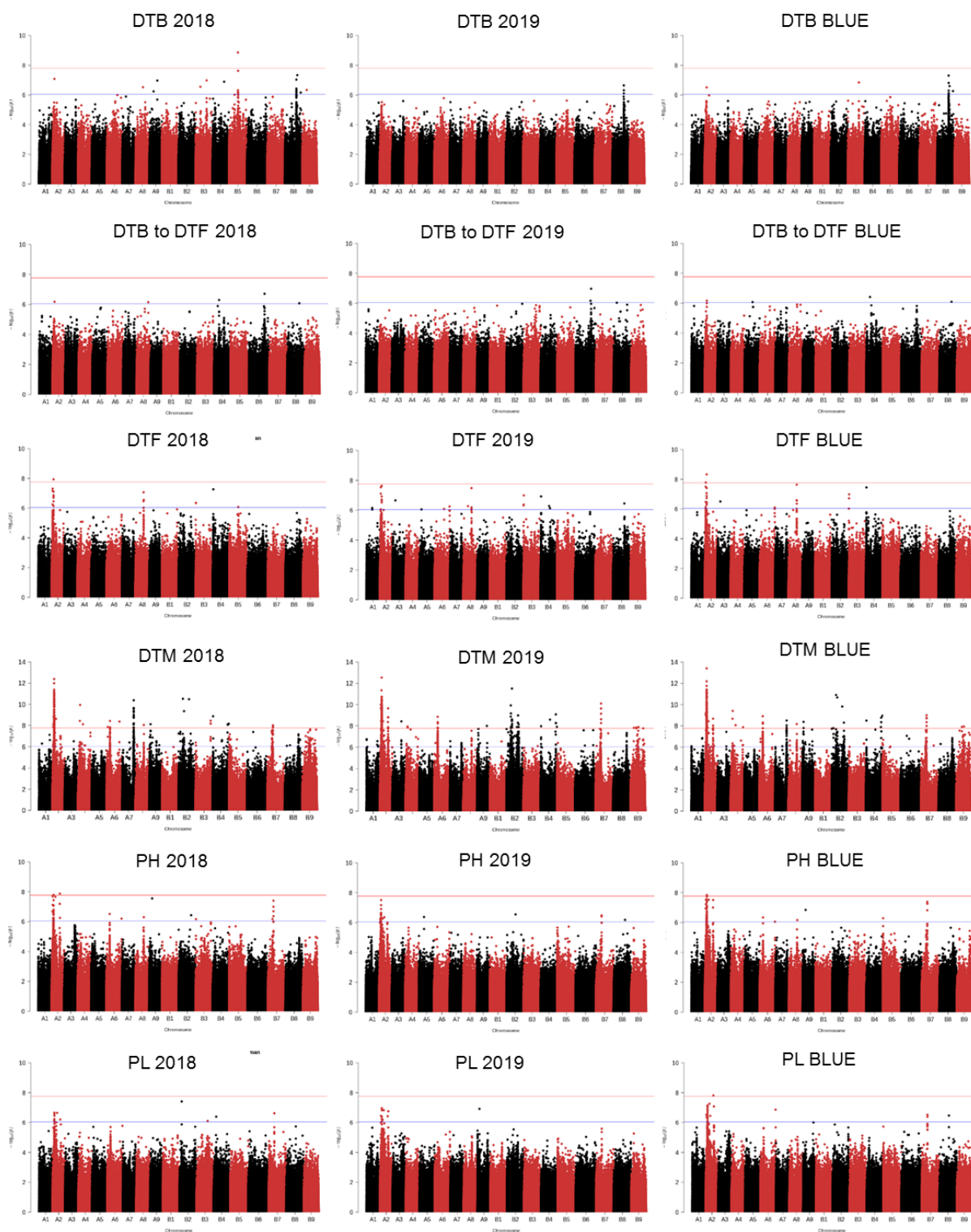
Supplementary Fig. 10: Pearson correlations among 12 quinoa traits. Best linear unbiased estimates across two years were used. Below the diagonal, scatter plots are shown with the fitted line in red. Above the diagonal, the Pearson correlation coefficients are shown with significance levels, *** = $P < 0.001$, ** = $P < 0.01$.

Appendix

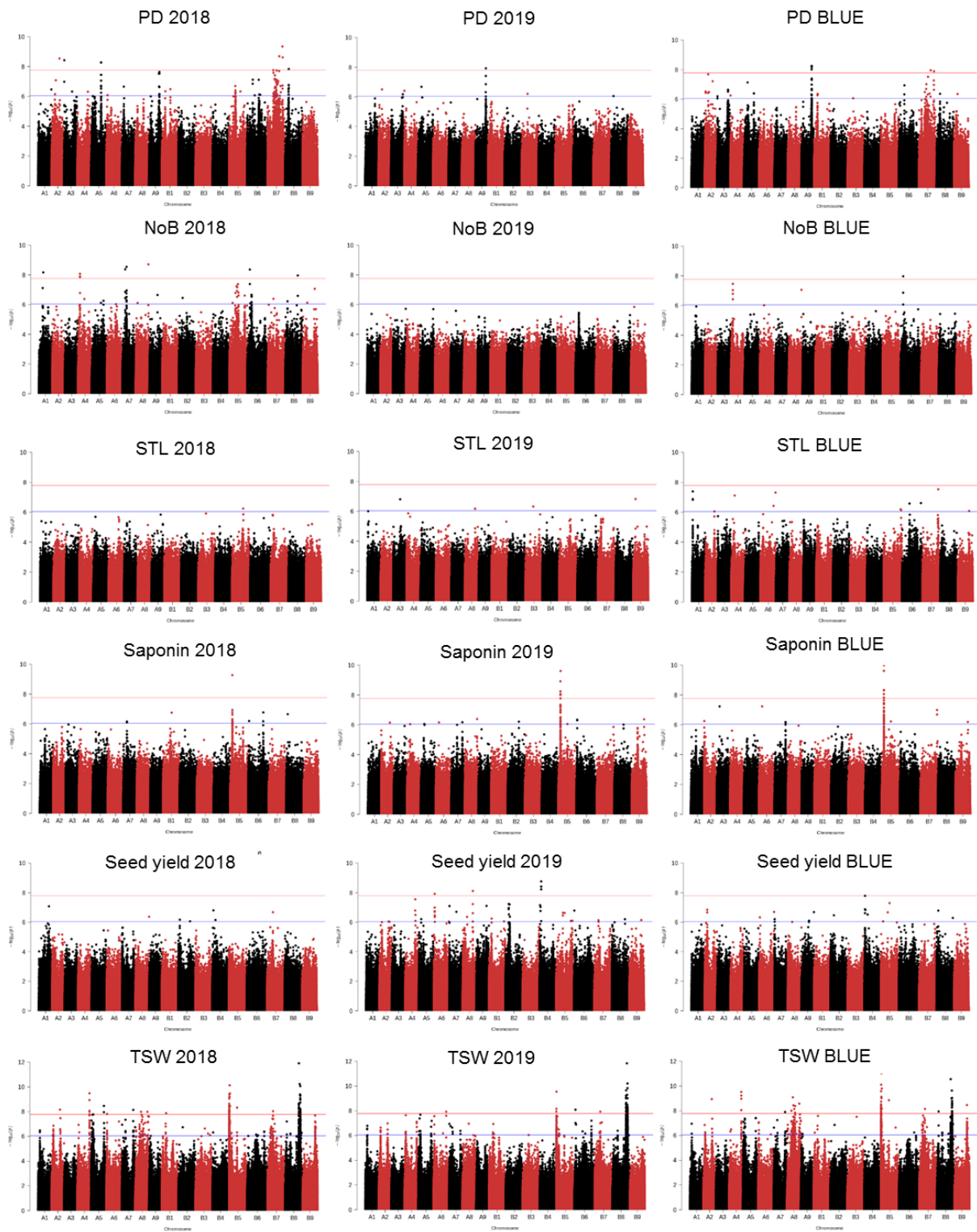


Supplementary Fig. 11: PCA of 12 quantitative phenotypes. A: Individual factor map colored according to populations identified from SNP analysis. B: Variables factor map of the PCA.

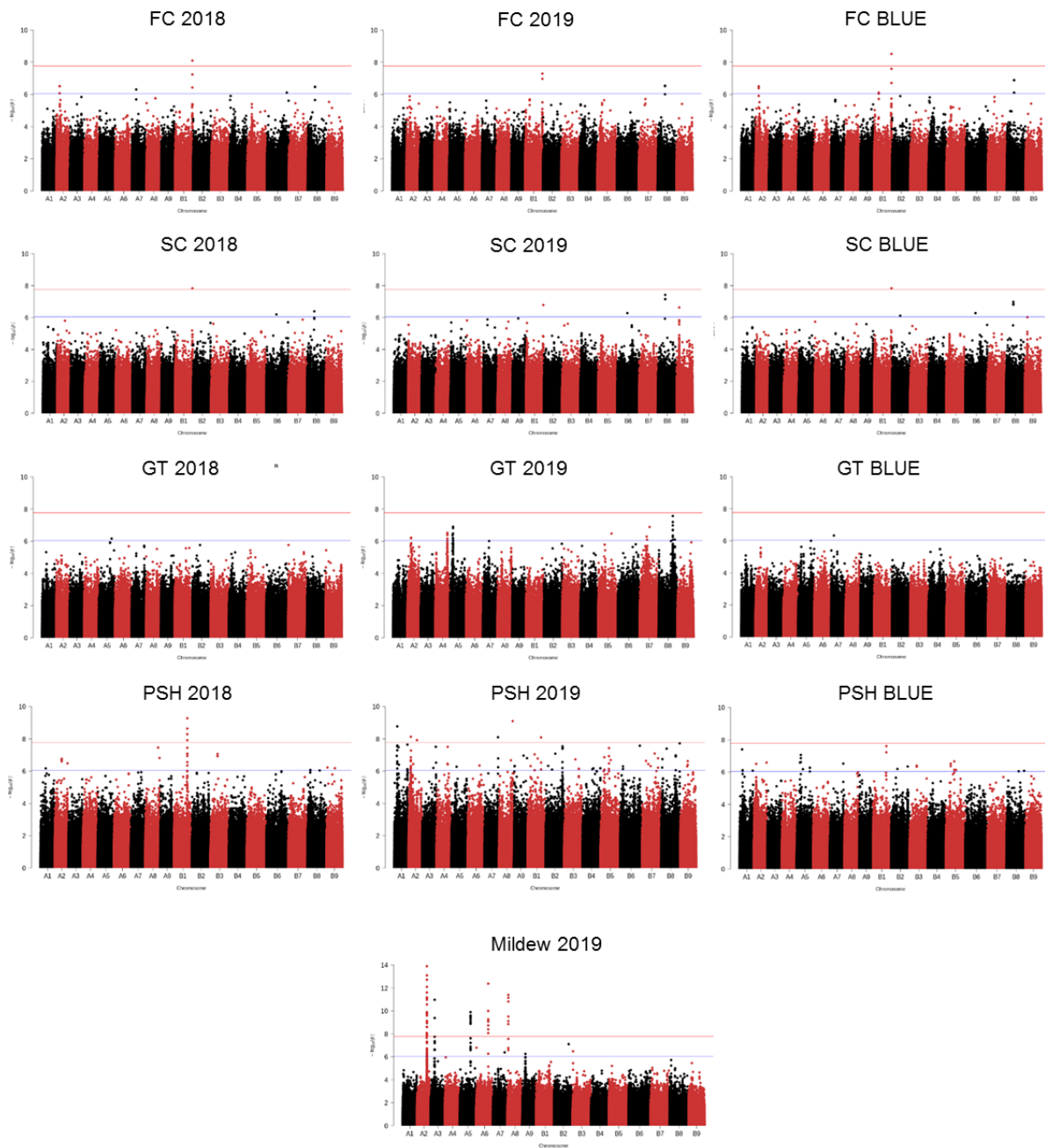
Appendix

Supplementary Fig. 12: *cont.*

Appendix

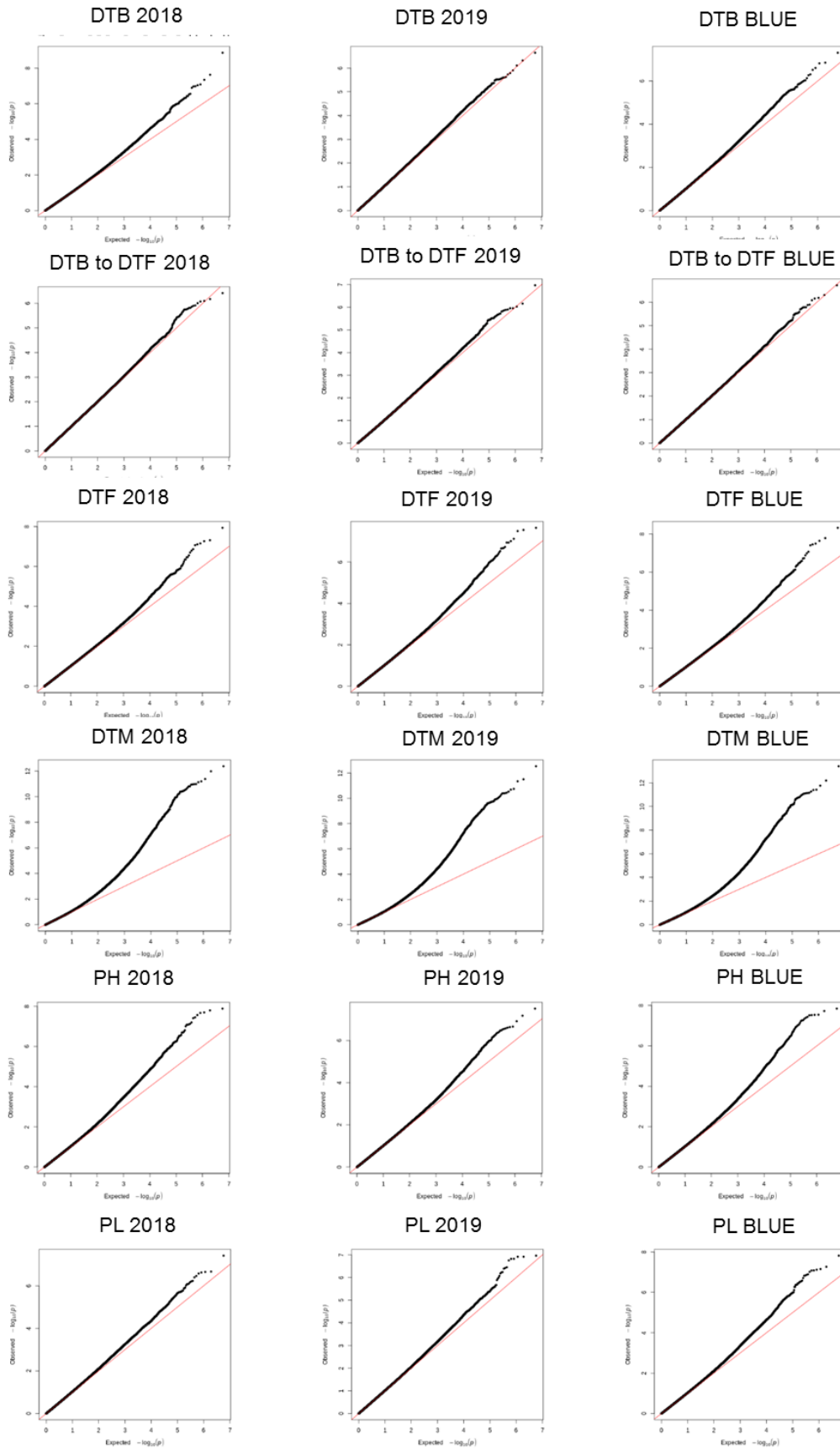
Supplementary Fig. 12: *cont.*

Appendix



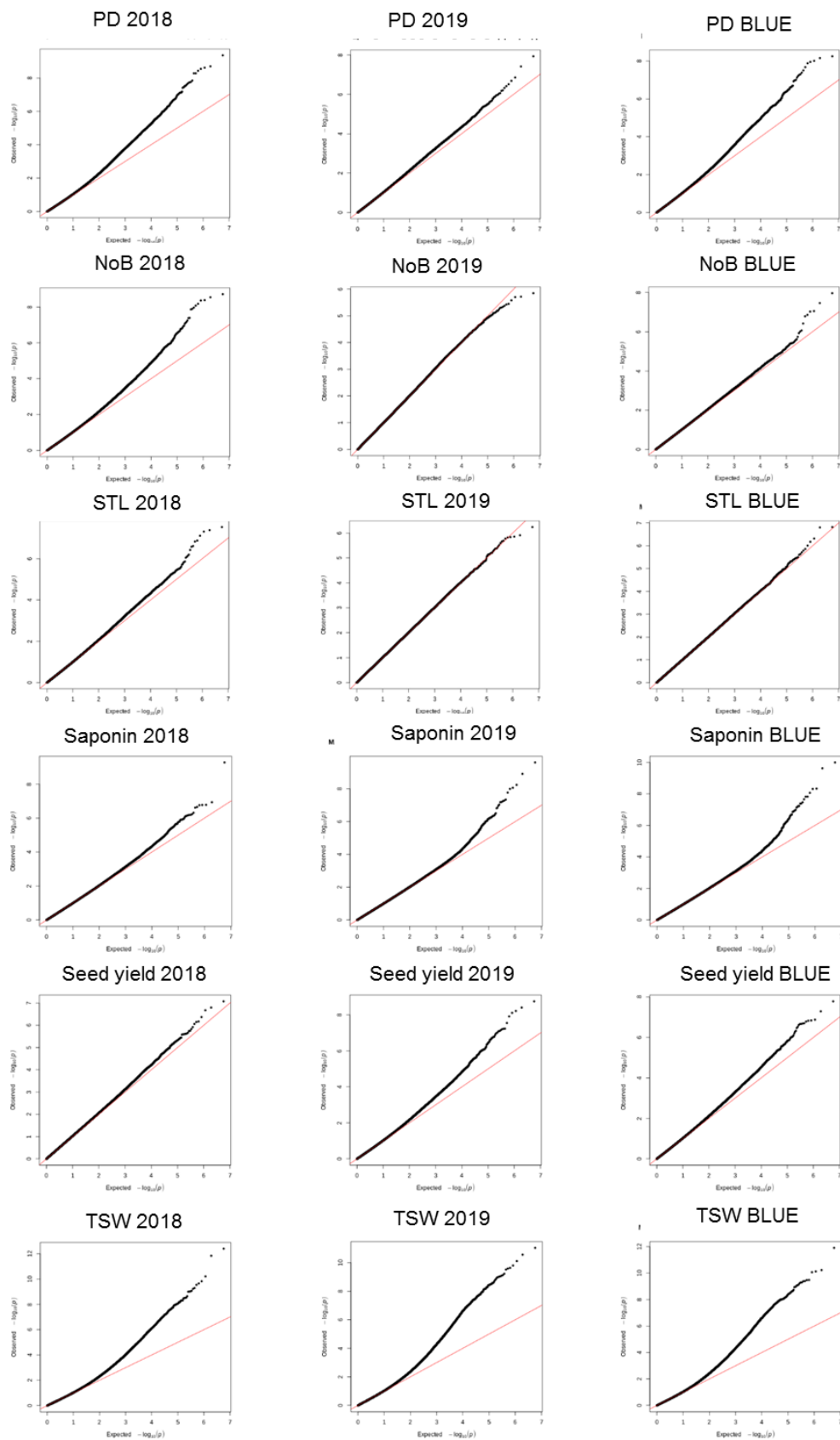
Supplementary Fig. 12: Manhattan plots from GWAS with data from 2018 (left), 2019 (center), and the mean of both years (right): The blue horizontal line indicates the suggestive threshold $-\log_{10}(8.98E-7)$. The red horizontal line indicates the significant threshold (Bonferroni correction) $-\log_{10}(1.67e-8)$.

Appendix

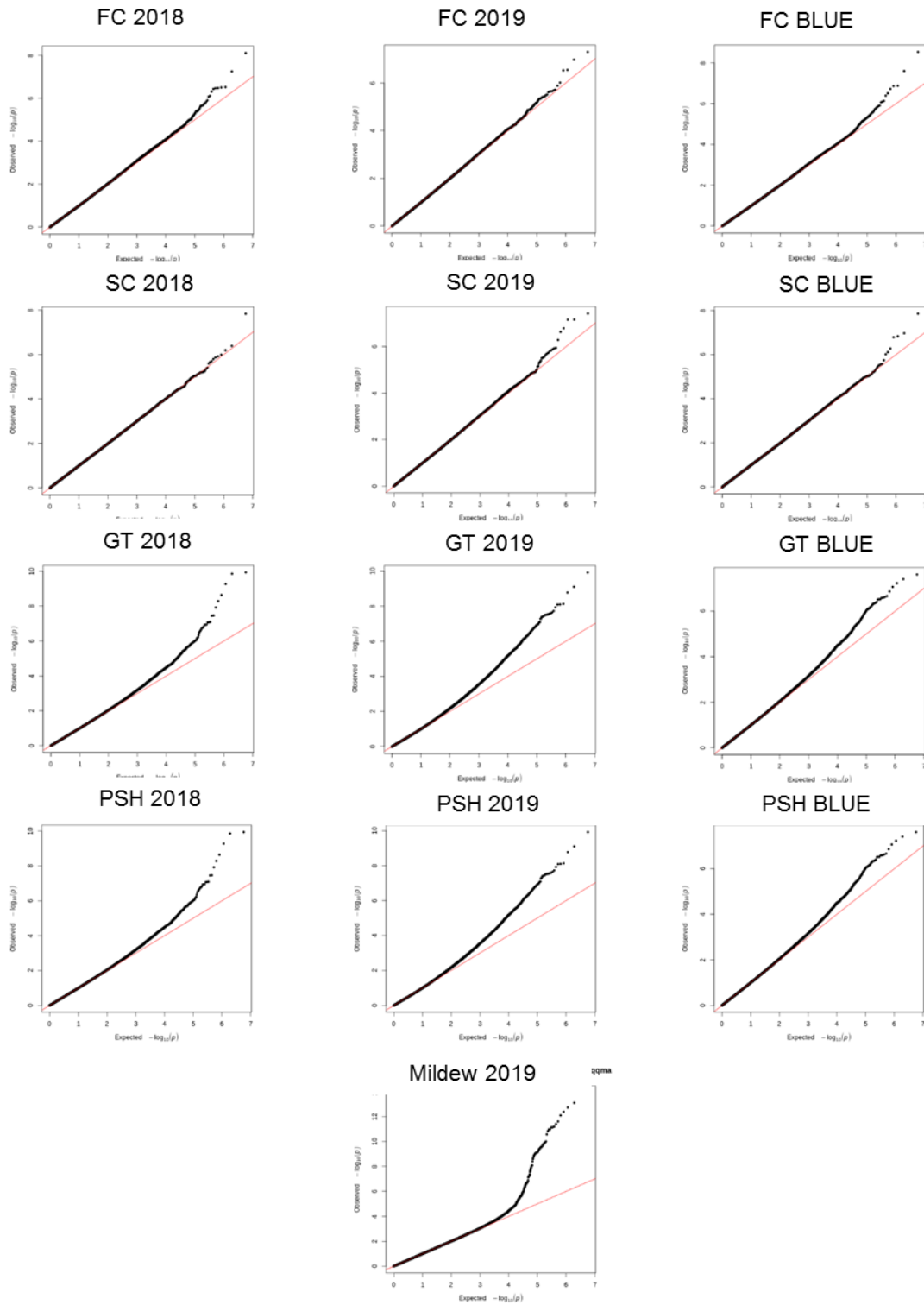


Supplementary Fig. 13: cont.

Appendix

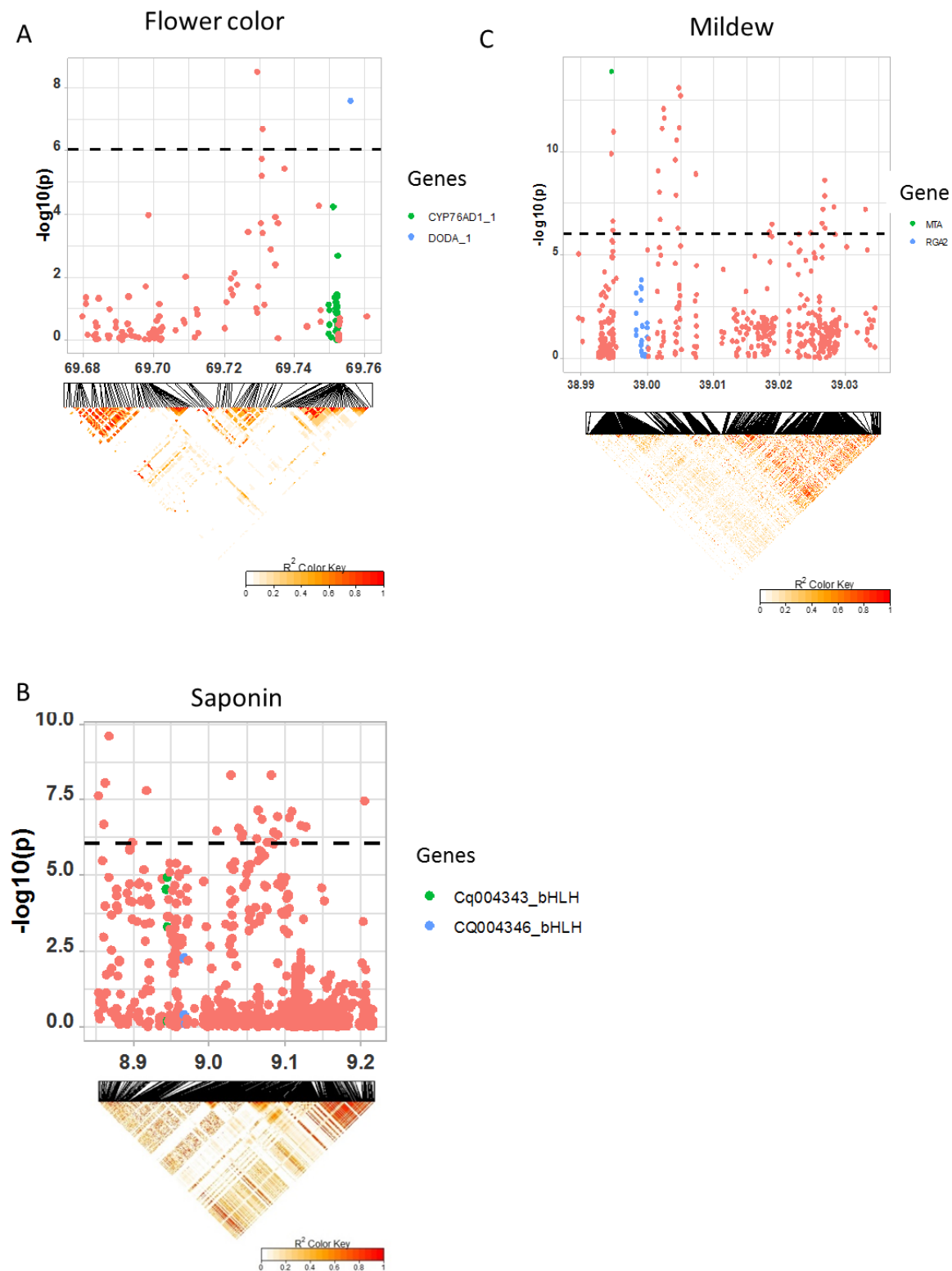
Supplementary Fig. 13: *cont.*

Appendix

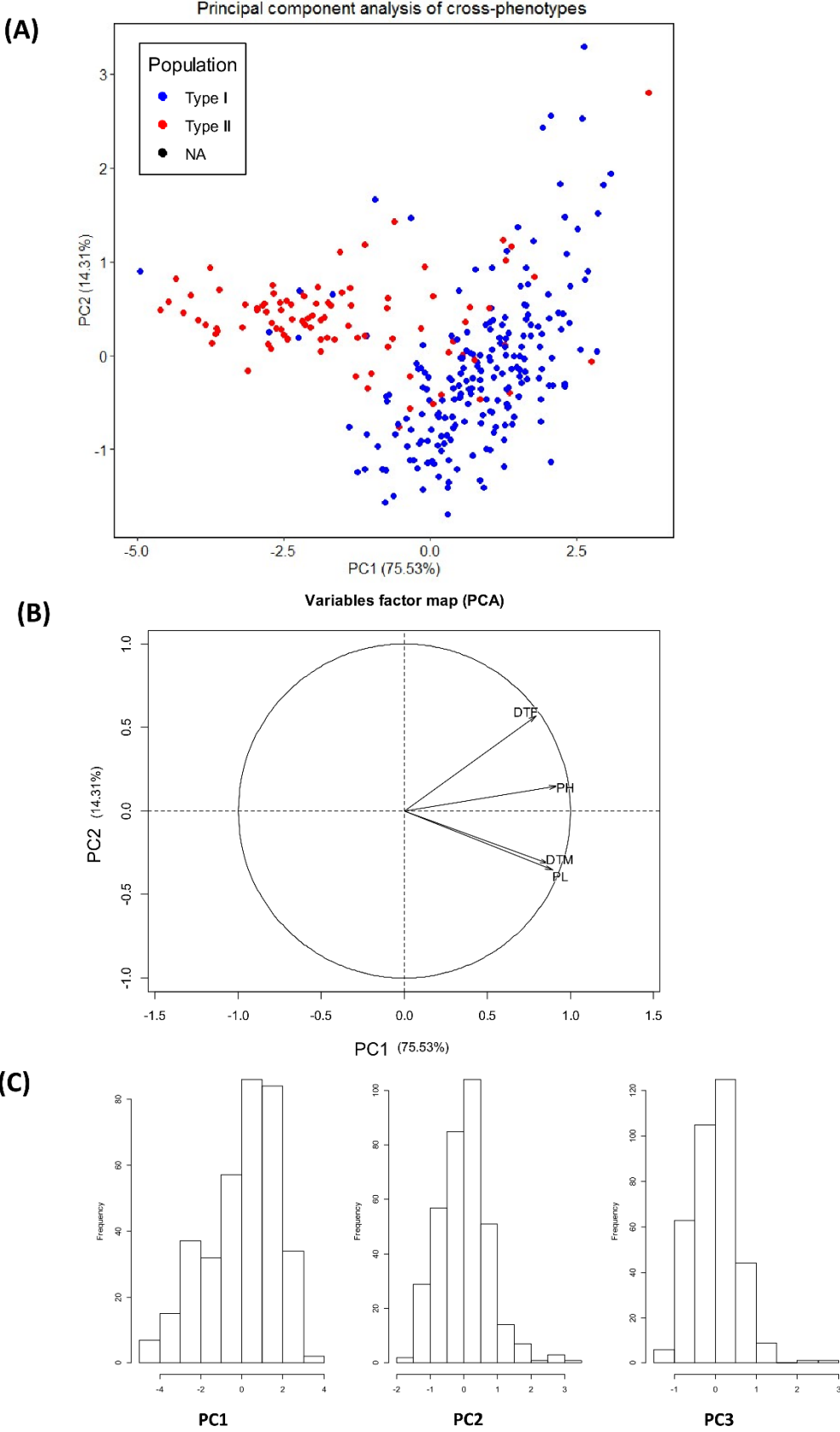


Supplementary Fig. 13: Quantile-quantile plots of GWAS in two years, 2018 (left) and 2019 (center), and BLUE (right).

Appendix

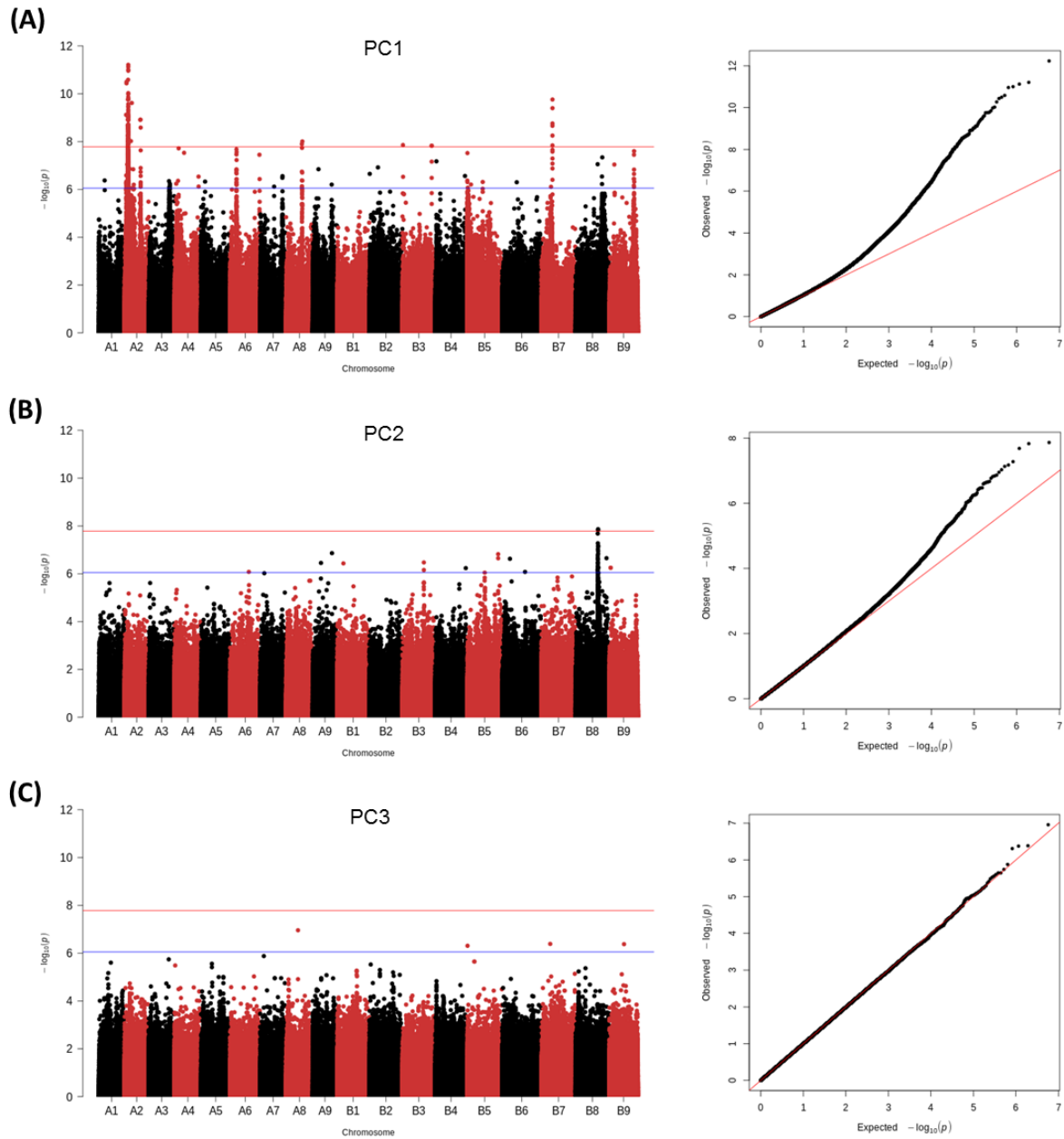


Supplementary Fig. 14: Local Manhattan plots for (A) flower color, (B) saponin content, and (C) mildew infection. Candidate genes are shown in the color legend. LD heat maps are placed at the bottom. The colors of the heat map represent the pairwise correlation between individual SNPs.



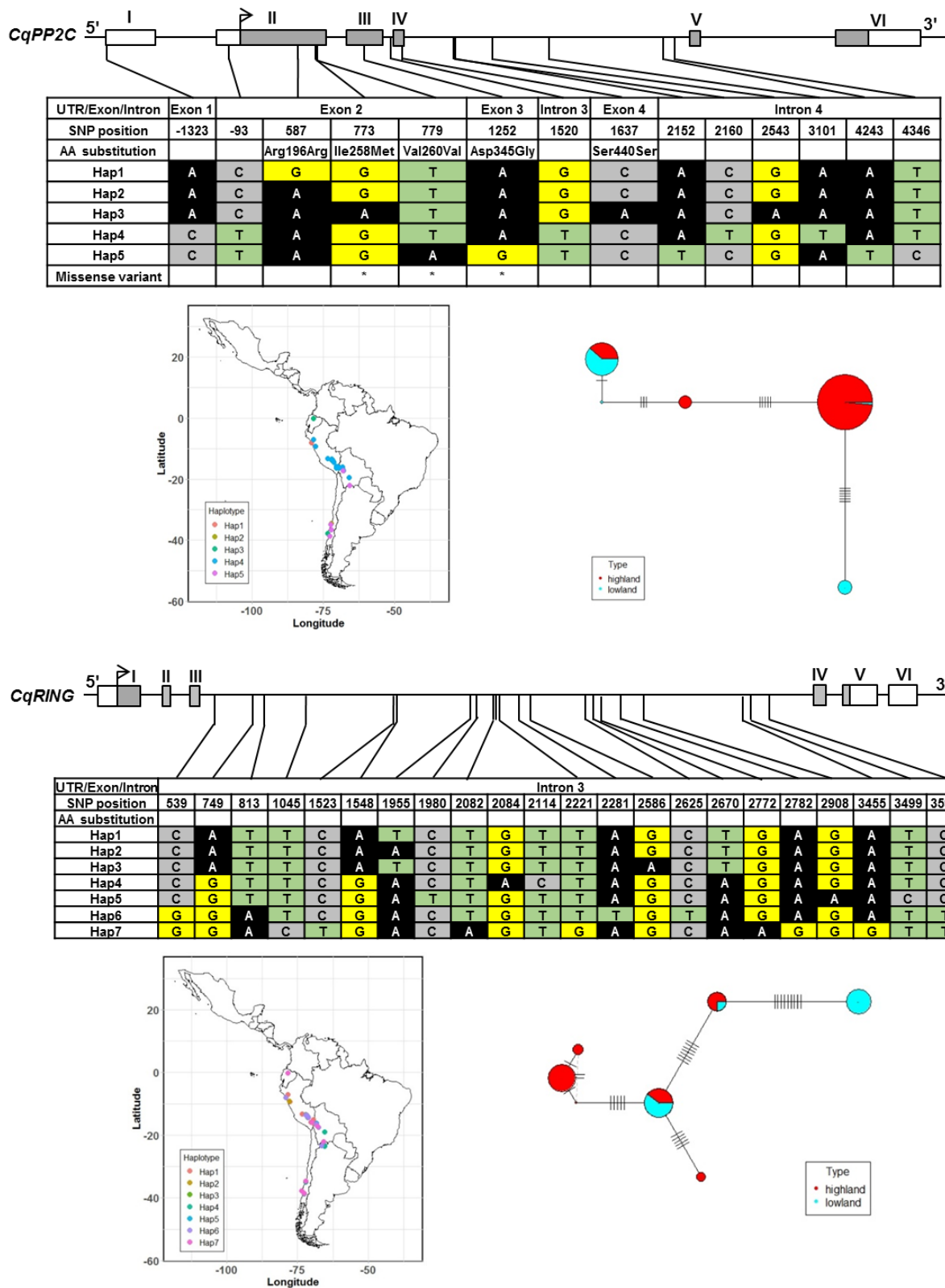
Supplementary Fig. 15: PCA of four quantitative traits (DTF, DTM, PH, and PL). A: Individual factor map, B: variables factor map of the PCA, C: distribution of the first three principal components which were used for GWAS analysis.

Appendix



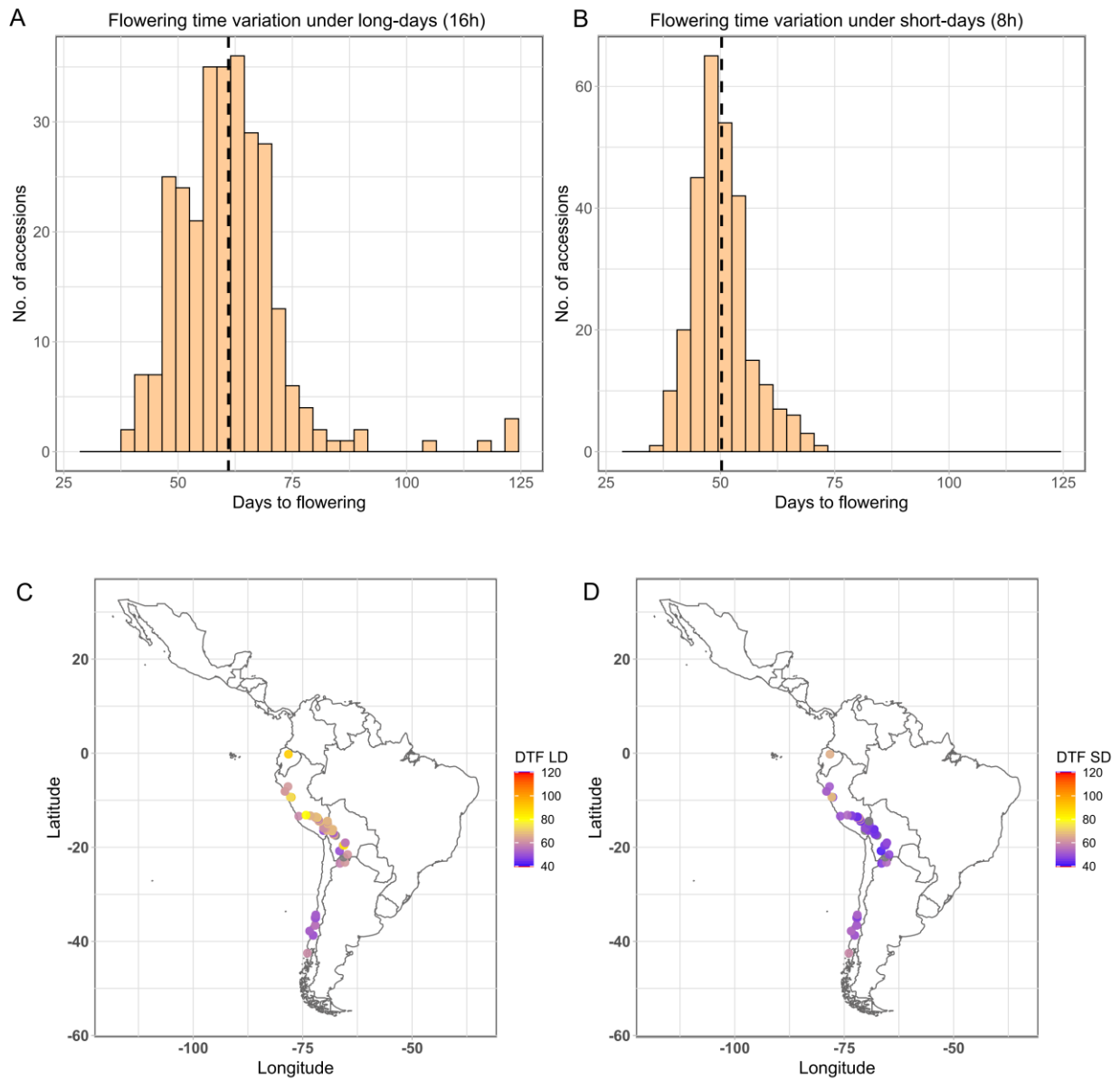
Supplementary Fig. 16: GWAS analysis of principal components, PC1 (A), PC2 (B), PC3 (C): Manhattan plots (left), and quantile-quantile plots (right): The blue horizontal line in the Manhattan plots indicates the suggestive threshold $-\log_{10}(8.98E-7)$. The red horizontal line indicates the significance threshold (Bonferroni correction) $-\log_{10}(1.67E-8)$.

Appendix



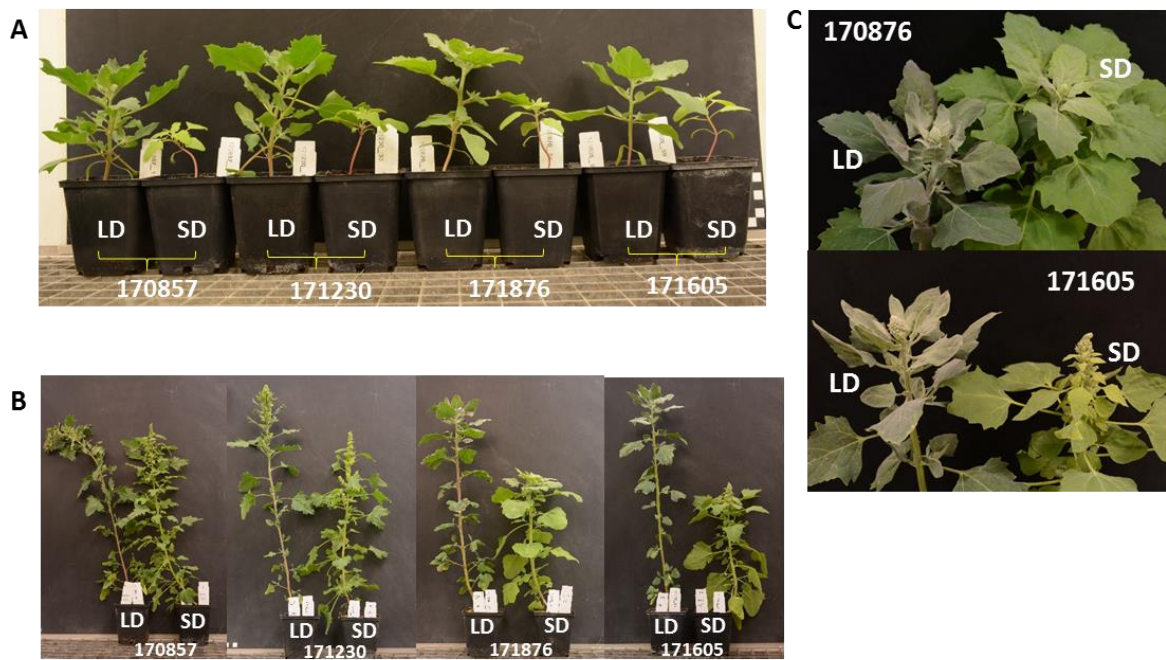
Supplementary Fig. 17: Haplotypes of two genes, *CqPP2C* and *CqRING* controlling seed size in quinoa. Geographic origin of the accessions and haplotype networks are displayed below the gene structure.

Appendix



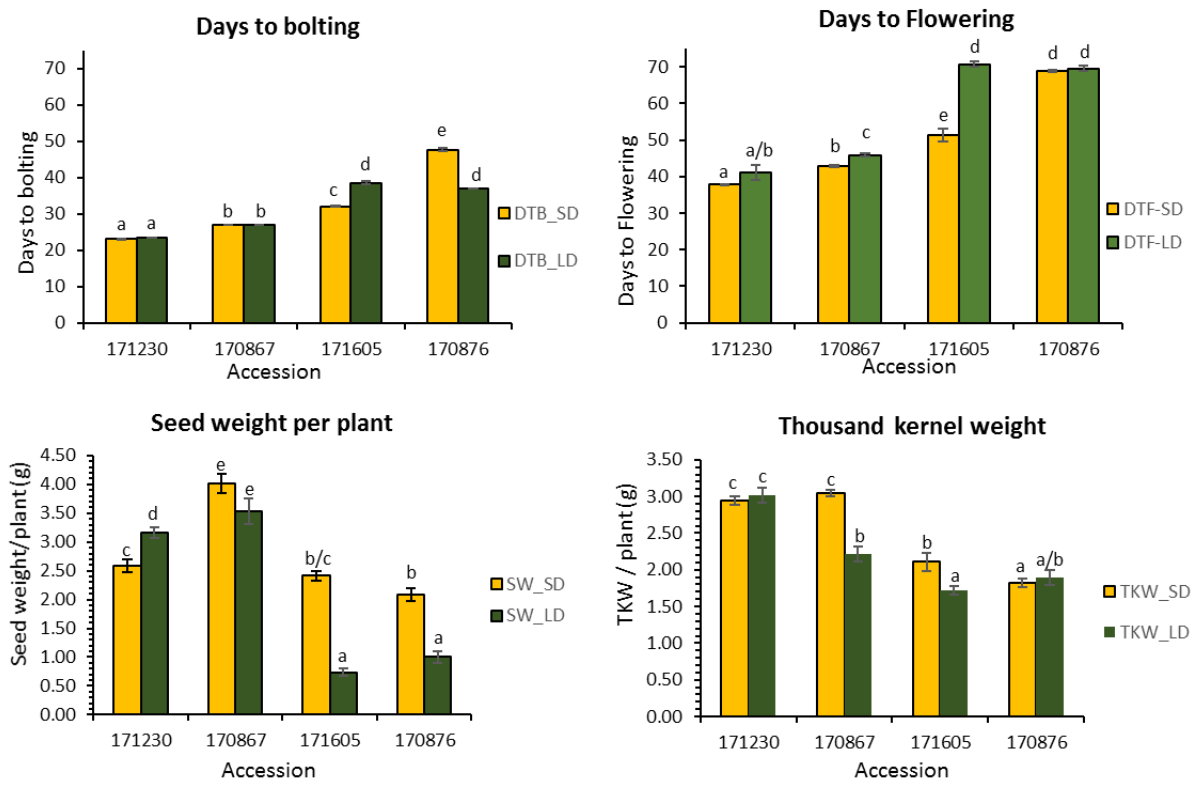
Supplementary Fig. 18: Flowering time variation of 276 accessions grown under short (8h) (B) and long day (16h) (A) conditions. C and D, geographic coordinates of 123 quinoa accessions. Color gradient corresponds to flowering time which was determined under long days (C) and under short days (D).

Appendix

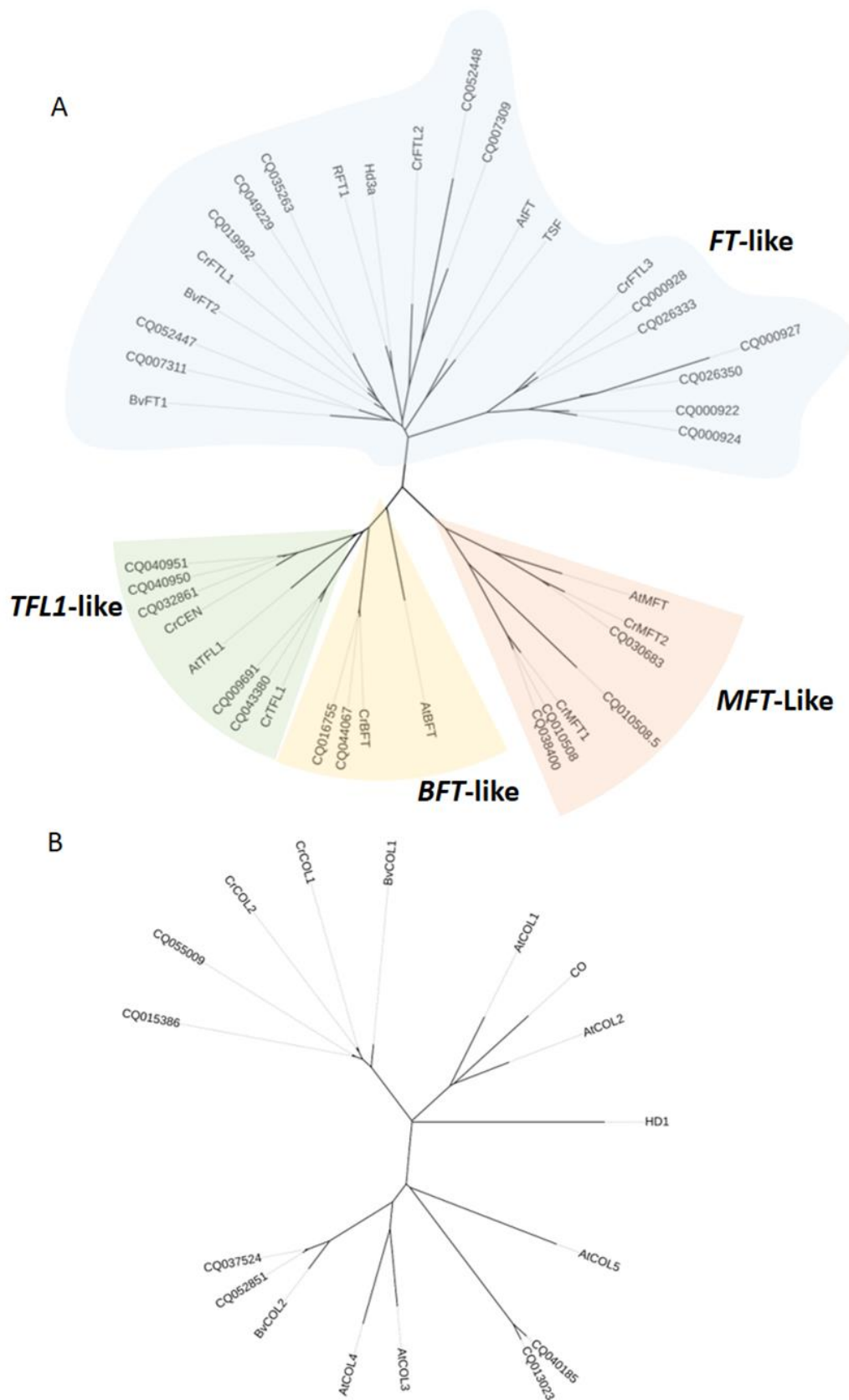


Supplementary Fig. 19: Comparison between growth under short day (SD) and long day conditions (LD). Plants were grown in 9cm pots in controlled growth chamber conditions. A: Three weeks after sowing B: Seven weeks after sowing. C: Different pigmentation under LDs (left) and SDs (right) in two accessions sensitive to photoperiod; CHEN-109 and PI-587173.

Appendix

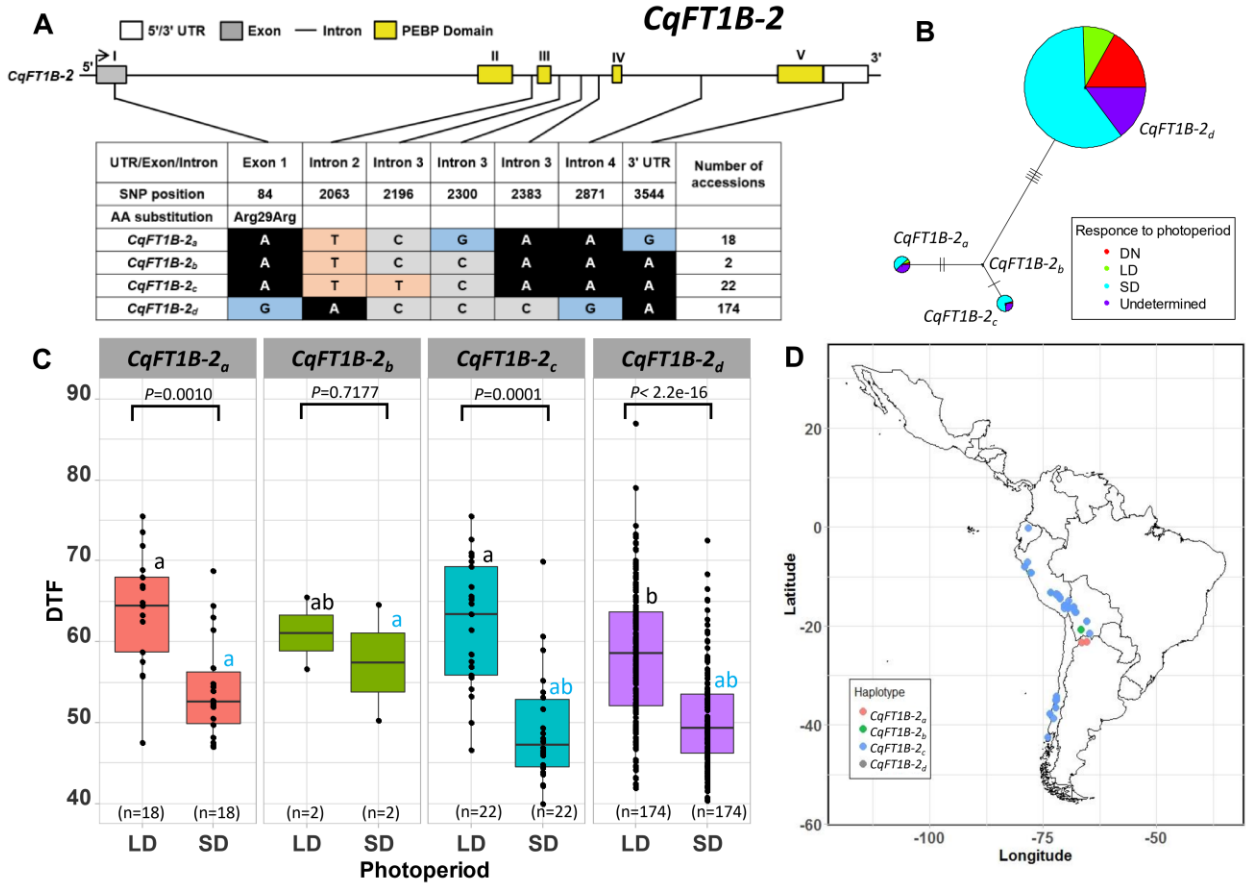
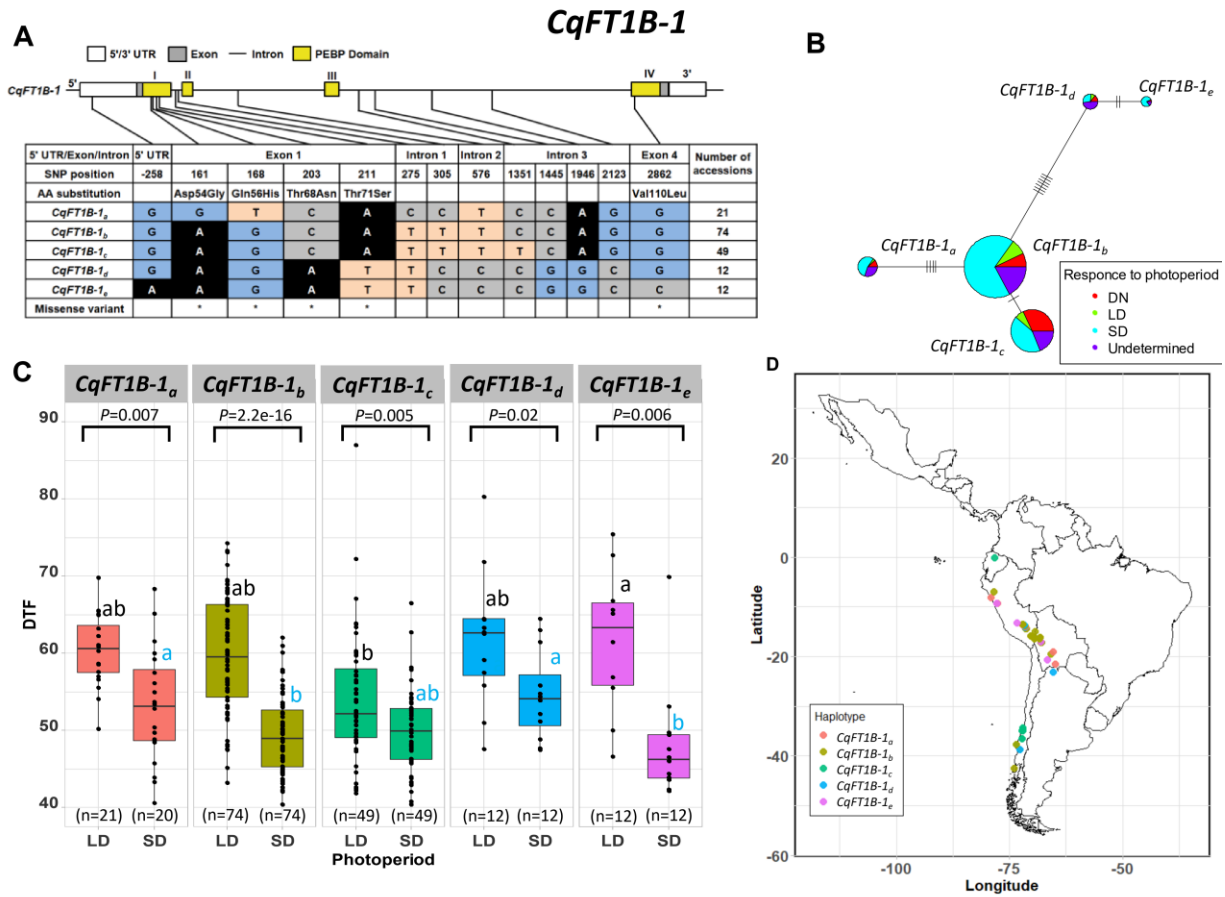


Supplementary Fig. 20: Phenological development and yield components of four accessions grown in the growth chamber under short day (SD) and long day (LD) conditions. Error bars: SEM



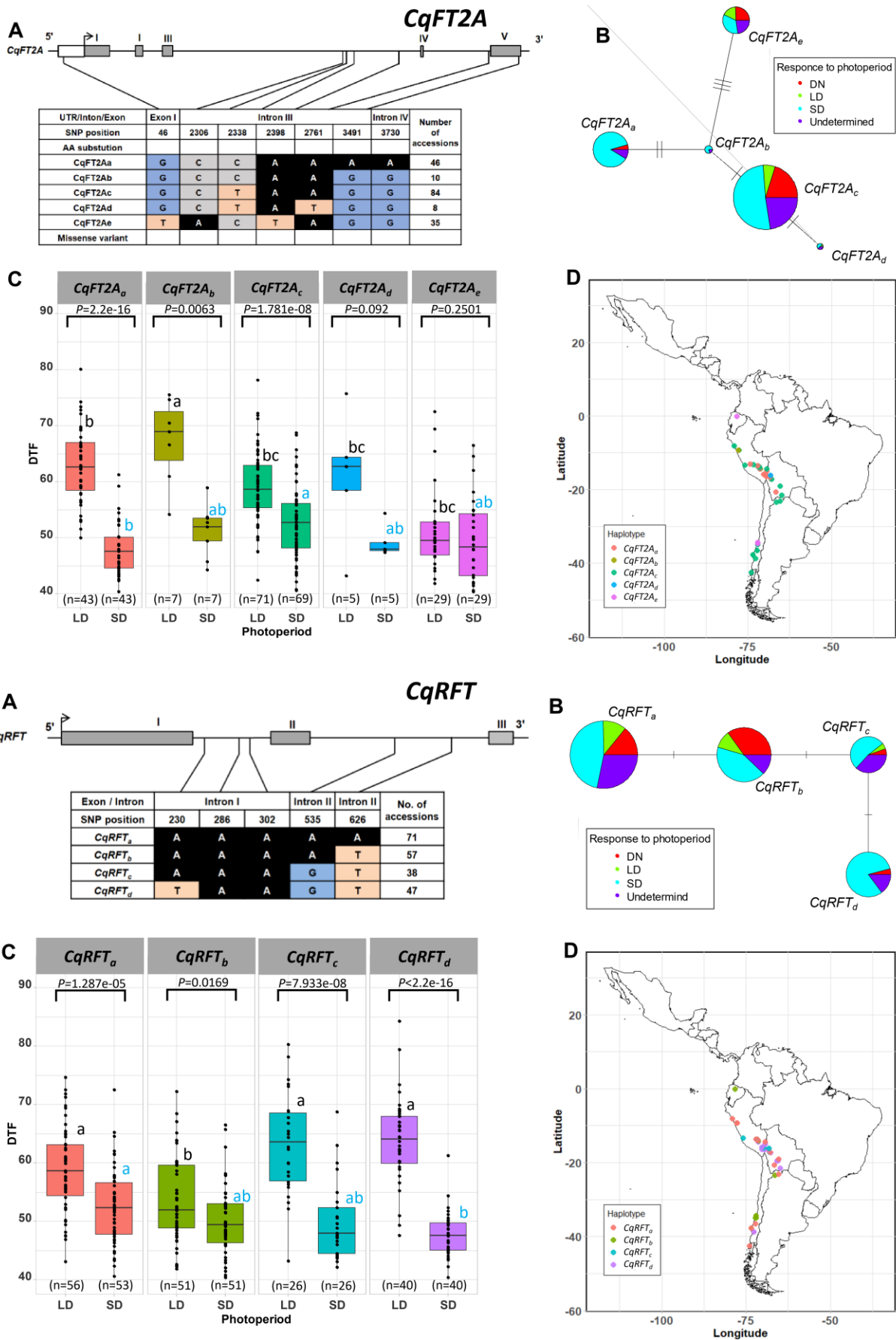
Supplementary Fig. 21: Phylogenetic relationships between PEBP family genes (A) and COL genes (B) in quinoa.

Appendix



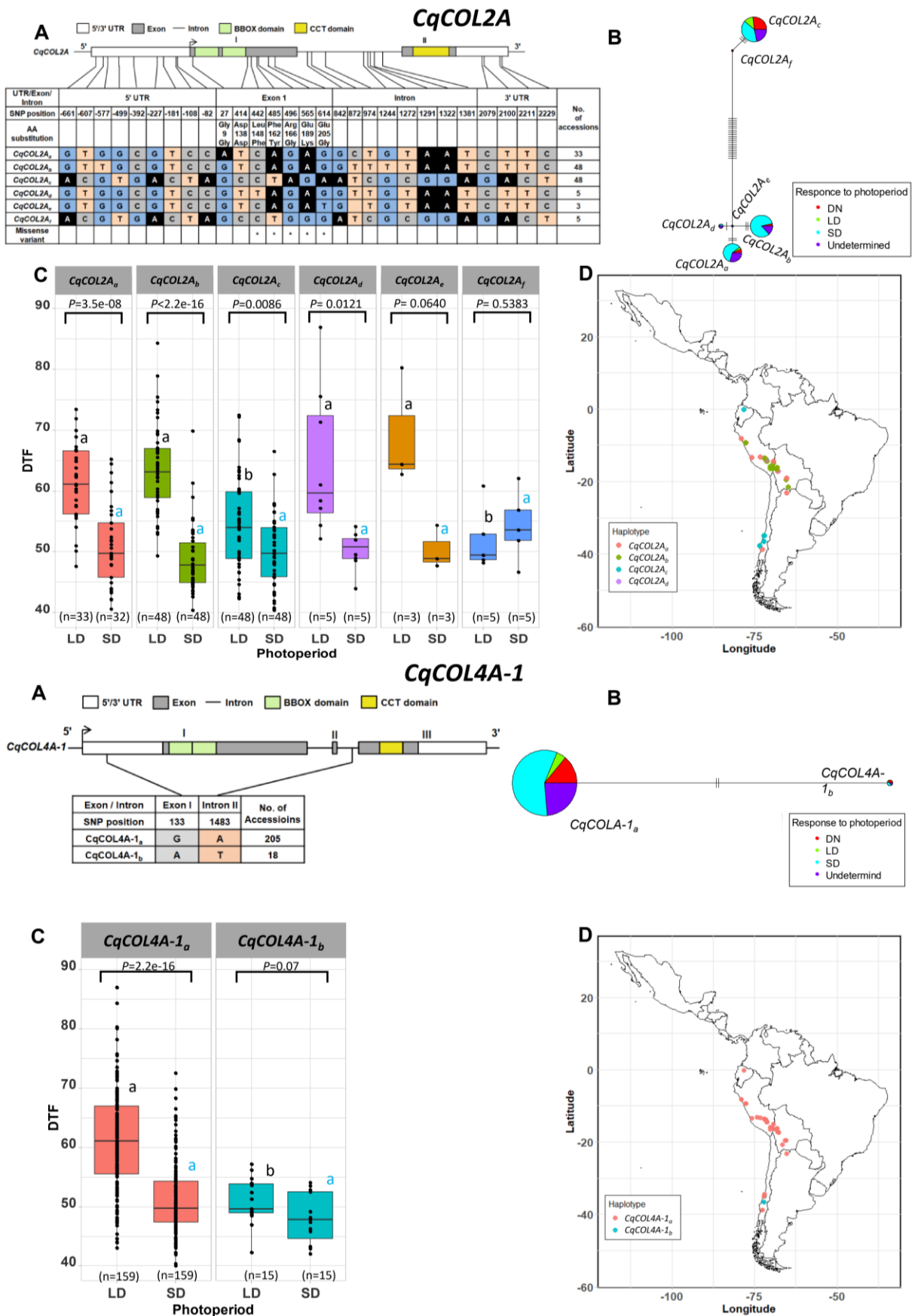
Supplementary Fig. 22: continued on next page and legend follows

Appendix



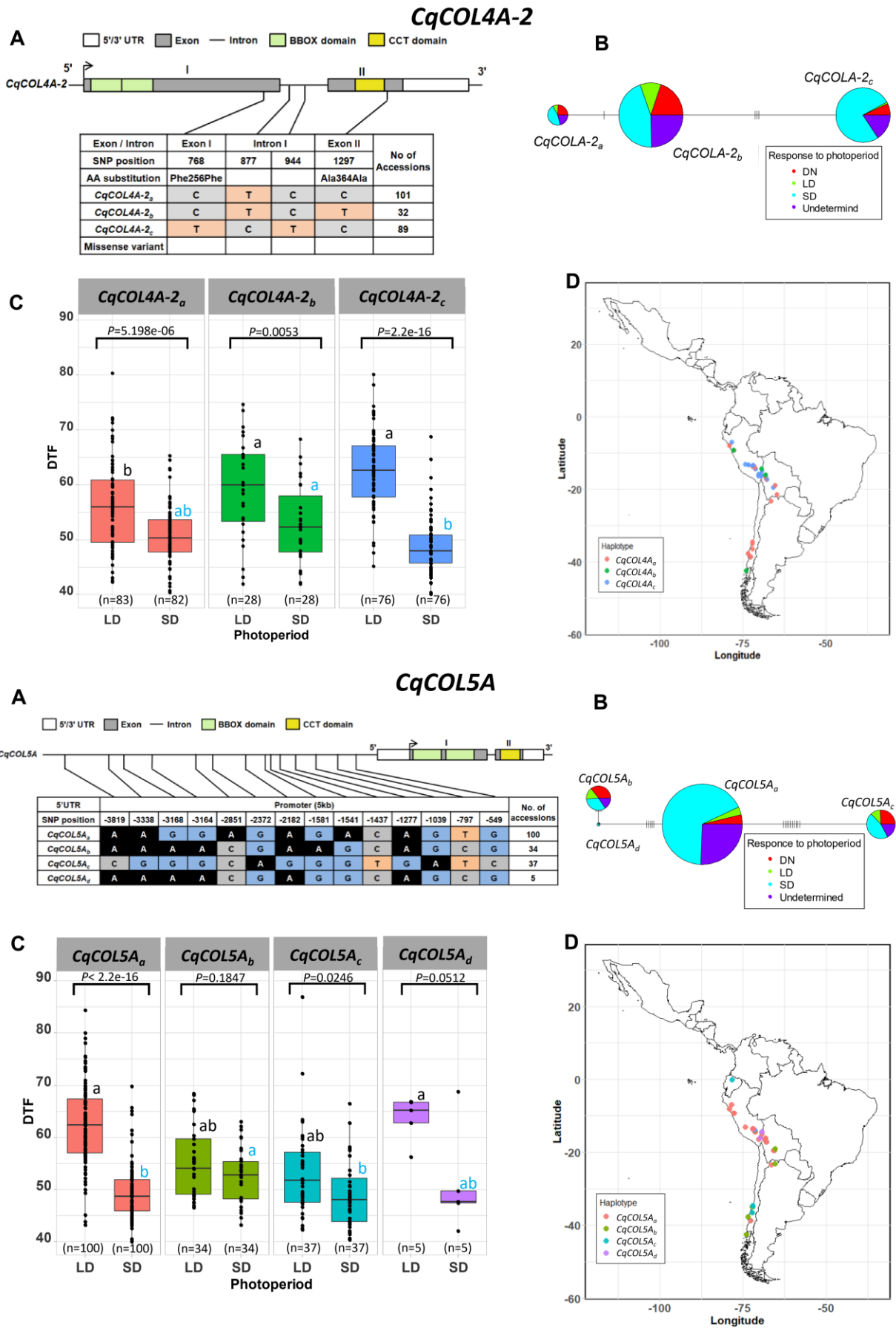
Supplementary Fig. 22: continued on next page and legend follows

Appendix



Supplementary Fig. 22: continued on next page and legend follows

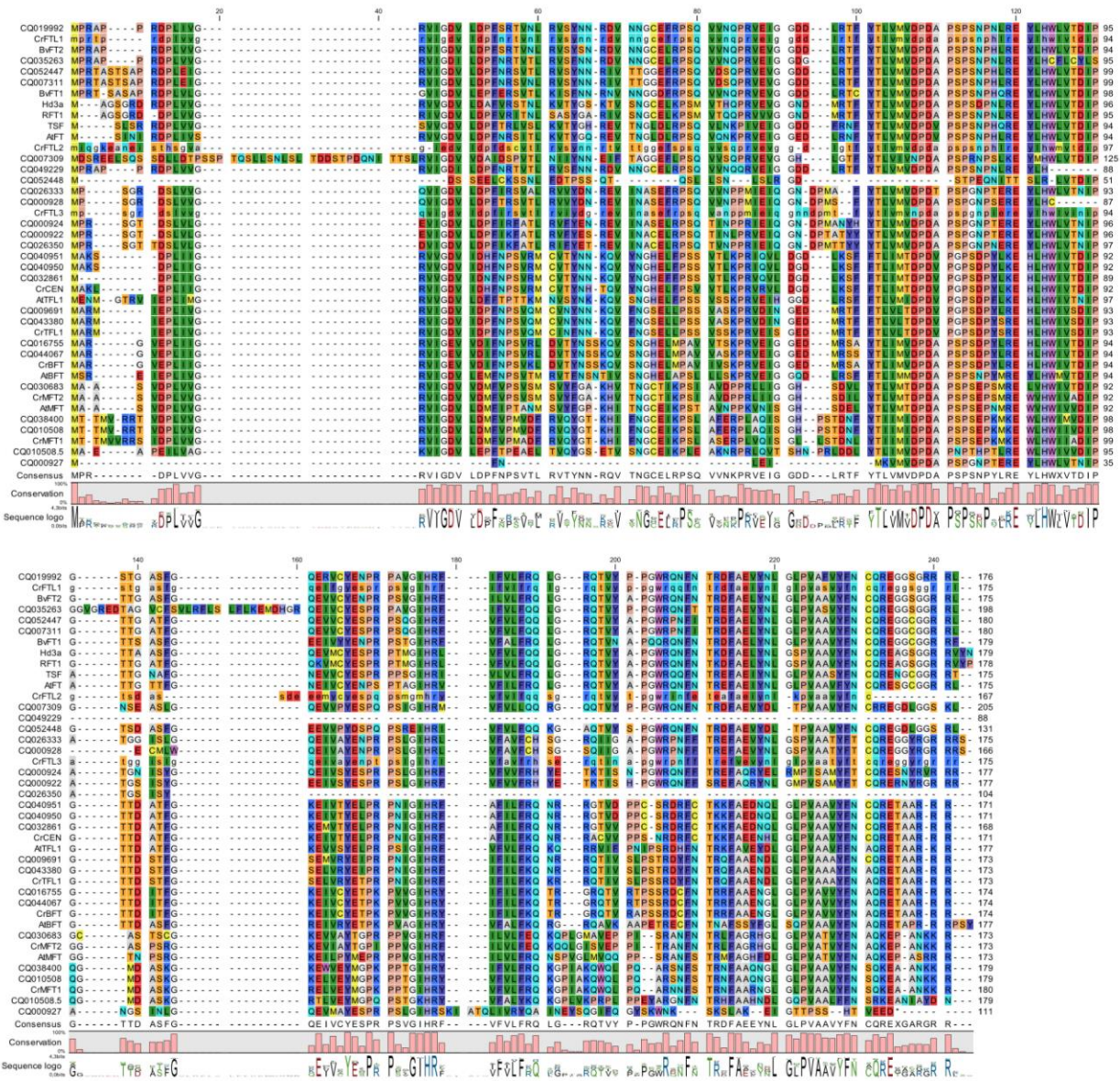
Appendix



Supplementary Fig. 22: continued on next page and legend follows

Supplementary Fig. 22: Phenotypic effects of haplotype variations in eight different flowering time genes. The number of accessions analyzed is given in Table 2. (A) Gene structure and nucleotide polymorphisms. Polymorphic nucleotides are indicated by different colors. SNP positions are given relative to the start codon. The asterisks depict missense variants. (B) Haplotype networks. Circle size illustrates the frequency of the corresponding haplotype. Colors of the pie chart depict different classes of photoperiod response. SD: early flowering under short days, LD: early flowering under long days, DN: no difference between short- and long day flowering time. Crosslines between two circles are representing the number of nucleotide polymorphisms between haplotypes. (C) Haplotype depending flowering time (days to flowering, DTF) as determined in the climate chamber under SD and LD growth conditions. The black and blue letters indicate significant differences of DTF among haplotypes under LD and SD photoperiod, respectively. The P-values of mean comparisons between SD and LD within haplotypes are written above the boxplots. ANOVA with Tukey's method was performed to determine significance among haplotypes, and Student's t-test was performed to compare DTF means within haplotypes. 'n' indicates the number of accessions. (D) Geographical origin of quinoa accessions with their respective haplotypes. Only accessions are shown for which passport data are available.

Appendix



Supplementary Fig. 23: Protein alignment of PEBP proteins of quinoa, sugar beet, *C. rubrum*, rice and Arabidopsis.

8 References

- Abbai R, Singh VK, Nachimuthu VV, Sinha P, Selvaraj R, Vipparla AK, Singh AK, Singh UM, Varshney RK, Kumar A (2019) Haplotype analysis of key genes governing grain yield and quality traits across 3K RG panel reveals scope for the development of tailor-made rice with enhanced genetic gains. *Plant biotechnology journal* 17 (8):1612-1622
- Adams S, Langton F (2005) Photoperiod and plant growth: a review. *The Journal of Horticultural Science and Biotechnology* 80 (1):2-10
- Ahmar S, Gill RA, Jung K-H, Faheem A, Qasim MU, Mubeen M, Zhou W (2020) Conventional and Molecular Techniques from Simple Breeding to Speed Breeding in Crop Plants: Recent Advances and Future Outlook. *International Journal of Molecular Sciences* 21 (7):2590
- Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome research* 19 (9):1655-1664
- Alqudah AM, Sallam A, Baenziger PS, Börner A (2020) GWAS: Fast-forwarding gene identification and characterization in temperate Cereals: lessons from Barley—A review. *Journal of Advanced Research* 22:119-135
- Anderson JT, Willis JH, Mitchell-Olds T (2011) Evolutionary genetics of plant adaptation. *Trends in Genetics* 27 (7):258-266
- Andres F, Coupland G (2012) The genetic basis of flowering responses to seasonal cues. *Nat Rev Genet* 13 (9):627-639. doi:10.1038/nrg3291
- Andrés F, Coupland G (2012) The genetic basis of flowering responses to seasonal cues. *Nature Reviews Genetics* 13 (9):627-639
- Asare E (2020) Identification of allelic variations associated with flowering time within homologs of flowering time genes in quinoa (*Chenopodium quinoa* Willd.). Unpublished Master's Thesis. Christian-Albrechts-University of Kiel, Kiel, Germany,
- Balasubramanian S, Sureshkumar S, Lempe J, Weigel D (2006) Potent induction of *Arabidopsis thaliana* flowering by elevated growth temperature. *PLoS Genet* 2 (7):e106
- Ballesta P, Maldonado C, Pérez-Rodríguez P, Mora F (2019) SNP and haplotype-based genomic selection of quantitative traits in *Eucalyptus globulus*. *Plants* 8 (9):331
- Bao S, Hua C, Shen L, Yu H (2020) New insights into gibberellin signaling in regulating flowering in *Arabidopsis*. *Journal of Integrative Plant Biology* 62 (1):118-131
- Bates D, Mächler M, Bolker B, Walker S (2015) Fitting Linear Mixed-Effects Models Using lme4. *J Stat Soft* 67:48. doi:10.18637/jss.v067.i01
- Bazile D, Jacobsen S-E, Verniau A (2016a) The global expansion of quinoa: trends and limits. *Frontiers in Plant Science* 7:622
- Bazile D, Pulvento C, Verniau A, Al-Nusairi MS, Ba D, Breidy J, Hassan L, Mohammed MI, Mambetov O, Otambekova M, Sepahvand NA, Shams A, Souici D, Miri K, Padulosi S (2016b) Worldwide Evaluations of Quinoa: Preliminary Results from Post International

References

- Year of Quinoa FAO Projects in Nine Countries. *Frontiers in Plant Science* 7 (850). doi:10.3389/fpls.2016.00850
- Bean A, Sunnadeniya R, Akhavan N, Campbell A, Brown M, Lloyd A (2018) Gain-of-function mutations in beet *DODA2* identify key residues for betalain pigment evolution. *New Phytologist* 219 (1):287-296
- Bertero H (2003a) Response of developmental processes to temperature and photoperiod in quinoa (*Chenopodium quinoa* Willd.). *Food Reviews International* 19 (1-2):87-97
- Bertero H, De la Vega A, Correa G, Jacobsen S, Mujica A (2004) Genotype and genotype-by-environment interaction effects for grain yield and grain size of quinoa (*Chenopodium quinoa* Willd.) as revealed by pattern analysis of international multi-environment trials. *Field Crops Research* 89 (2-3):299-318
- Bertero H, King R, Hall A (1999a) Modelling photoperiod and temperature responses of flowering in quinoa (*Chenopodium quinoa* Willd.). *Field crops research* 63 (1):19-34
- Bertero H, King R, Hall A (1999b) Photoperiod-sensitive development phases in quinoa (*Chenopodium quinoa* Willd.). *Field Crops Research* 60 (3):231-243
- Bertero HD (2003b) Response of developmental processes to temperature and photoperiod in quinoa (*Chenopodium quinoa* Willd.). *Food Reviews International* 19:87-97. doi:10.1081/FRI-120018870
- Bhargava A, Fuentes F (2010) Mutational dynamics of microsatellites. *Molecular biotechnology* 44 (3):250-266
- Bhargava A, Shukla S, Ohri D (2006) Karyotypic studies on some cultivated and wild species of *Chenopodium* (Chenopodiaceae). *Genetic Resources and Crop Evolution* 53 (7):1309-1320
- Bhargava A, Srivastava S (2013) *Quinoa: Botany, production and uses*. CABI,
- Bi X, van Esse GW, Mulki MA, Kirschner G, Zhong J, Simon R, von Korff M (2019) *CENTRORADIALIS* interacts with *FLOWERING LOCUS T*-like genes to control spikelet initiation, floret development and grain number. *Plant Physiology*:pp.01454.02018. doi:10.1104/pp.18.01454
- Bird KA, VanBuren R, Puzey JR, Edger PP (2018) The causes and consequences of subgenome dominance in hybrids and recent polyploids. *New Phytologist* 220:87-93. doi:10.1111/nph.15256
- Blümel M, Dally N, Jung C (2015) Flowering time regulation in crops-what did we learn from *Arabidopsis*? *Current Opinion in Biotechnology* 32:121-129. doi:10.1016/j.copbio.2014.11.023
- Böhlenius H, Huang T, Charbonnel-Campaa L, Brunner AM, Jansson S, Strauss SH, Nilsson O (2006) *CO/FT* regulatory module controls timing of flowering and seasonal growth cessation in trees. *Science* 312 (5776):1040-1043
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30 (15):2114-2120

References

- Bonifacio A (1995) Interspecific and intergeneric hybridization in chenopod species. Brigham Young University. Department of Botany and Range Sciences.,
- Bonifacio A (2003) *Chenopodium* sp.: genetic resources, ethnobotany, and geographic distribution. *Food Reviews International* 19 (1-2):1-7
- Bonifacio A, Gomez-Pando L, Rojas W (2013) Quinoa breeding and modern variety development. State of the Art Report on Quinoa Around the World
- Bouché F, Lobet G, Tocquin P, Périlleux C (2015) FLOR-ID: an interactive database of flowering-time gene networks in *Arabidopsis thaliana*. *Nucleic Acids Research* 44 (D1):D1167-D1171. doi:10.1093/nar/gkv1054
- Bradley D, Ratcliffe O, Vincent C, Carpenter R, Coen E (1997) Inflorescence commitment and architecture in *Arabidopsis*. *Science* 275 (5296):80-83
- Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, Goodman MM, Harjes C, Guill K, Kroon DE, Larsson S, Lepak NK, Li H, Mitchell SE, Pressoir G, Peiffer Ja, Rosas MO, Rocheford TR, Romay MC, Romero S, Salvo S, Sanchez Villeda H, da Silva HS, Sun Q, Tian F, Upadyayula N, Ware D, Yates H, Yu J, Zhang Z, Kresovich S, McMullen MD (2009) The genetic architecture of maize flowering time. *Science (New York, NY)* 325:714-718. doi:10.1126/science.1174276
- Cao D, Li Y, Lu S, Wang J, Nan H, Li X, Shi D, Fang C, Zhai H, Yuan X (2015) *GmCOL1a* and *GmCOL1b* function as flowering repressors in soybean under long-day conditions. *Plant and Cell Physiology* 56 (12):2409-2422
- Cháb D, Kolář J, Olson MS, Štorchová H (2008) Two *FLOWERING LOCUS T (FT)* homologs in *Chenopodium rubrum* differ in expression patterns. *Planta* 228:929-940. doi:10.1007/s00425-008-0792-3
- Chakravarty T, Sopory S (1998) Blue light stimulation of cell proliferation and glyoxalase I activity in callus cultures of *Amaranthus paniculatus*. *Plant science* 132 (1):63-69
- Chalhoub B, Denoeud F, Liu S, Parkin IaP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, Correa M, Da Silva C, Just J, Falentin C, Koh CS, Le Clainche I, Bernard M, Bento P, Noel B, Labadie K, Alberti a, Charles M, Arnaud D, Guo H, Daviaud C, Alamery S, Jabbari K, Zhao M, Edger PP, Chelaifa H, Tack D, Lassalle G, Mestiri I, Schnel N, Le Paslier M-C, Fan G, Renault V, Bayer PE, Golicz aa, Manoli S, Lee T-H, Thi VHD, Chalabi S, Hu Q, Fan C, Tollenaere R, Lu Y, Battail C, Shen J, Sidebottom CHD, Canaguier a, Chauveau a, Berard a, Deniot G, Guan M, Liu Z, Sun F, Lim YP, Lyons E, Town CD, Bancroft I, Meng J, Ma J, Pires JC, King GJ, Brunel D, Delourme R, Renard M, Aury J-M, Adams KL, Batley J, Snowdon RJ, Tost J, Edwards D, Zhou Y, Hua W, Sharpe aG, Paterson aH, Guan C, Wincker P (2014) Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* 345:950-953. doi:10.1126/science.1253435
- Chia TYP, Müller a, Jung C, Mutasa-Göttgens ES (2008) Sugar beet contains a large *CONSTANS-LIKE* gene family including a *CO* homologue that is independent of the early-bolting (*B*) gene locus. *Journal of experimental botany* 59:2735-2748. doi:10.1093/jxb/ern129

References

- Choi Y-J, Danielsen S, Lübeck M, Hong S-B, Delhey R, Shin H-D (2010) Morphological and molecular characterization of the causal agent of downy mildew on quinoa (*Chenopodium quinoa*). *Mycopathologia* 169 (5):403-412
- Christensen SA, Pratt DB, Pratt C, Nelson PT, Stevens MR, Jellen EN, Coleman CE, Fairbanks DJ, Bonifacio A, Maughan PJ (2007) Assessment of genetic diversity in the USDA and CIP-FAO international nursery collections of quinoa (*Chenopodium quinoa* Willd.) using microsatellite markers. *Plant Genetic Resources: Characterisation and Utilisation* 5:82-95. doi:10.1017/S1479262107672293
- Christiansen JL, Jacobsen S-E, Jørgensen ST (2010) Photoperiodic effect on flowering and seed development in quinoa (*Chenopodium quinoa* Willd.). *Acta Agriculturae Scandinavica, Section B-Soil & Plant Science* 60 (6):539-544
- Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w¹¹¹⁸; iso-2; iso-3. *Fly* 6 (2):80-92
- Coles N, Coleman C, Christensen S, Jellen E, Stevens M, Bonifacio A, Rojas-Beltran J, Fairbanks D, Maughan P (2005) Development and use of an expressed sequenced tag library in quinoa (*Chenopodium quinoa* Willd.) for the discovery of single nucleotide polymorphisms. *Plant Science* 168 (2):439-447
- Colque-Little CX, Abondano MC, Lund OS, Amby DB, Piepho H-P, Andreasen C, Schmöckel SM, Schmid KJ (2020) Genetic variation for tolerance to the downy mildew pathogen *Peronospora variabilis* in genetic resources of quinoa (*Chenopodium quinoa*). *bioRxiv*
- Costa Tártara SM, Manifesto MM, Bramardi SJ, Bertero HD (2012) Genetic structure in cultivated quinoa (*Chenopodium quinoa* Willd.), a reflection of landscape structure in Northwest Argentina. *Conservation Genetics* 13:1027-1038. doi:10.1007/s10592-012-0350-1
- Dally N, Eckel M, Batschauer A, Höft N, Jung C (2018) Two *CONSTANS-LIKE* genes jointly control flowering time in beet. *Scientific reports* 8 (1):1-10
- Dally N, Xiao K, Holtgräwe D, Jung C (2014) The *B2* flowering time locus of beet encodes a zinc finger transcription factor. *Proceedings of the National Academy of Sciences* 111 (28):10365-10370
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Group GPA (2011) The variant call format and VCFtools. *Bioinformatics* 27 (15):2156-2158. doi:10.1093/bioinformatics/btr330
- del Castillo C, Winkel T, Mahy G, Bizoux J-P (2007) Genetic structure of quinoa (*Chenopodium quinoa* Willd.) from the Bolivian altiplano as revealed by RAPD markers. *Genetic Resources and Crop Evolution* 54 (4):897-905
- Devanathan S, Erban A, Perez-Torres Jr R, Kopka J, Makaroff CA (2014) *Arabidopsis thaliana* glyoxalase 2-1 is required during abiotic stress but is not essential under normal plant growth. *PLoS One* 9 (4):e95971

References

- Drabešová J, Černá L, Mašterová H, Koloušková P, Potocký M, Štorchová H (2016) The evolution of the *FT/TFL1* genes in *Amaranthaceae* and their expression patterns in the course of vegetative growth and flowering in *Chenopodium rubrum*. *G3: Genes, Genomes, Genetics* 6 (10):3065-3076
- Drabešová J, Cháb D, Kolař J, Haškovcová K, Štorchová H (2014a) A dark-light transition triggers expression of the floral promoter *CrFTL1* and downregulates *CONSTANS*-like genes in a short-day plant *Chenopodium rubrum*. *J Exp Bot* 65 (8):2137-2146. doi:10.1093/jxb/eru073
- Drabešová J, Cháb D, Kolař J, Haškovcová K, Štorchová H (2014b) A dark–light transition triggers expression of the floral promoter *CrFTL1* and downregulates *CONSTANS*-like genes in a short-day plant *Chenopodium rubrum*. *Journal of experimental botany* 65 (8):2137-2146
- Emrani N, Hasler M, Patiranage DSR, Maldonado N, Rey E, Jung C (2020) An efficient method to produce segregating populations in quinoa (*Chenopodium quinoa* Willd.).
- Eshed Y, Lippman ZB (2019) Revolutions in agriculture chart a course for targeted breeding of old and new crops. *Science* 366 (6466)
- Fairbanks D, Burgener K, Robison L, Andersen W, Ballon E (1990) Electrophoretic characterization of quinoa seed proteins. *Plant Breeding* 104 (3):190-195
- Fairbanks D, Waldrigues A, Ruas C, Ruas P, Maughan P, Robison L, Andersen W, Riede C, Pauley C, Caetano L (1993) Efficient characterization of biological diversity using field DNA extraction and random amplified polymorphic DNA markers. *Revista Brasileira de Genética* 16:11-11
- Fan C, Hu R, Zhang X, Wang X, Zhang W, Zhang Q, Ma J, Fu Y-F (2014) Conserved CO-FT regulons contribute to the photoperiod flowering control in soybean. *BMC Plant Biology* 14 (1):1-14
- Fuentes F, Bazile D, Bhargava A, Martinez EA (2012) Implications of farmers' seed exchanges for on-farm conservation of quinoa, as revealed by its genetic diversity in Chile. *Journal of Agricultural Science-London* 150 (6):702
- Fuentes F, Martinez E, Hinrichsen P, Jellen E, Maughan P (2009) Assessment of genetic diversity patterns in Chilean quinoa (*Chenopodium quinoa* Willd.) germplasm using multiplex fluorescent microsatellite markers. *Conservation Genetics* 10 (2):369-377
- Fuentes F, Zurita-Silva A (2013) 11 Molecular Studies.
- Fuller HJ (1949) Photoperiodic responses of *Chenopodium quinoa* Willd. and *Amaranthus caudatus* L. *American Journal of Botany*:175-180
- Galvao VC, Horrer D, Kuttner F, Schmid M (2012) Spatial control of flowering by DELLA proteins in *Arabidopsis thaliana*. *Development*. doi:10.1242/dev.080879
- Galwey N (1992) The potential of quinoa as a multi-purpose crop for agricultural diversification: a review. *Industrial crops and products* 1 (2-4):101-106

References

- Gandarillas H (1969) Esterilidad genética y citoplásmica de la quinua (*Chenopodium quinoa*). Turrialba 19 (3):429-430
- Gandarillas H (1979) Botánica. In: Tapia ME (ed) Quinoa y kañiwa: cultivos Andinos. CIID, Bogotá,
- Gandarillas H, Alandía S, Cardozo A, Mujica A (1979) Quinoa y Kaniwa cultivos Andinos. Instituto Interamericano de Ciencias Agrícolas, Bogotá, Colombia
- Gandarillas H, Luizaga J (1967) Número de cromosomas de *Chenopodium quinoa* Willd en radículas y raicillas. Turrialba 17 (3):275-+
- Ghosh S, Watson A, Gonzalez-Navarro OE, Ramirez-Gonzalez RH, Yanes L, Mendoza-Suárez M, Simmonds J, Wells R, Rayner T, Green P (2018) Speed breeding in growth chambers and glasshouses for crop breeding and model plant research. Nature protocols 13 (12):2944-2963
- Golicz AA (2019) Analysis of the quinoa genome reveals conservation and divergence of the flowering pathways.
- Golicz AA, Steinfort U, Arya H, Singh MB, Bhalla PL (2020) Analysis of the quinoa genome reveals conservation and divergence of the flowering pathways. Functional & integrative genomics 20 (2):245-258
- Gómez-Pando LR, Aguilar-Castellanos E, Ibañez-Tremolada M Quinoa (*Chenopodium quinoa* Willd.) Breeding. In, 2019.
- Gomez-Pando LR, Eguiluz-de la Barra A (2013) Developing genetic variability of quinoa (*Chenopodium quinoa* Willd.) with gamma radiation for use in breeding programs.
- Gomez-Pando L (2015) Quinoa breeding. Quinoa: Improvement and Sustainable Production:87-108
- González JA, Eisa S, Hussin S, Prado FE (2015) Quinoa: an Incan crop to face global changes in agriculture. Quinoa: Improvement and sustainable production:1-18
- Griffiths S, Dunford RP, Coupland G, Laurie DA (2003) The evolution of *CONSTANS*-like gene families in barley, rice, and Arabidopsis. Plant physiology 131 (4):1855-1867
- Guo A-Y, Zhu Q-H, Gu X, Ge S, Yang J, Luo J (2008) Genome-wide identification and evolutionary analysis of the plant specific SBP-box transcription factor family. Gene 418 (1-2):1-8
- Guo Z, Song Y, Zhou R, Ren Z, Jia J (2010) Discovery, evaluation and distribution of haplotypes of the wheat *Ppd-D1* gene. New Phytologist 185 (3):841-851
- Gupta PK, Kulwal PL, Jaiswal V (2019) Association mapping in plants in the post-GWAS genomics era. In: Advances in genetics, vol 104. Elsevier, pp 75-154
- Gutaker RM, Weiß CL, Ellis D, Anglin NL, Knapp S, Fernández-Alonso JL, Prat S, Burbano HA (2019) The origins and adaptation of European potatoes reconstructed from historical genomes. Nature Ecology & Evolution 3 (7):1093-1101

References

- Hatlestad GJ, Sunnadeniya RM, Akhavan Na, Gonzalez A, Goldman IL, McGrath JM, Lloyd AM (2012) The beet *R* locus encodes a new cytochrome P450 required for red betalain production. *Nature Genetics* 44:816-820. doi:10.1038/ng.2297
- Hayama R, Agashe B, Luley E, King R, Coupland G (2007) A Circadian Rhythm Set by Dusk Determines the Expression of *FT* Homologs and the Short-Day Photoperiodic Flowering Response in *Pharbitis*. *The Plant Cell* 19:2988-3000. doi:10.1105/tpc.107.052480
- Hinojosa L, Matanguihan JB, Murphy KM (2019) Effect of high temperature on pollen morphology, plant growth and seed yield in quinoa (*Chenopodium quinoa* Willd.). *Journal of Agronomy and Crop Science* 205 (1):33-45
- Ho J, Tumkaya T, Aryal S, Choi H, Claridge-Chang A (2019) Moving beyond P values: data analysis with estimation graphics. *Nature Methods* 16 (7):565-566. doi:10.1038/s41592-019-0470-3
- Ho WWH, Weigel D (2014a) Structural features determining flower-promoting activity of *Arabidopsis* FLOWERING LOCUS T. *The Plant Cell* 26 (2):552-564
- Ho WWH, Weigel D (2014b) Structural features determining flower-promoting activity of *Arabidopsis* FLOWERING LOCUS T. *The Plant cell* 26:552-564. doi:10.1105/tpc.113.115220
- Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS (2017) UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* 35 (2):518-522
- Höft N, Dally N, Hasler M, Jung C (2018) Haplotype Variation of Flowering Time Genes of Sugar Beet and Its Wild Relatives and the Impact on Life Cycle Regimes. *Frontiers in Plant Science* 8. doi:10.3389/fpls.2017.02211
- Huang C, Sun H, Xu D, Chen Q, Liang Y, Wang X, Xu G, Tian J, Wang C, Li D, Wu L, Yang X, Jin W, Doebley JF, Tian F (2017) *ZmCCT9* enhances maize adaptation to higher latitudes *Proceedings of the National Academy of Sciences* 115:E334-E341. doi:10.1073/pnas.1718058115
- Huang X, Han B (2014) Natural variations and genome-wide association studies in crop plants. *Annual review of plant biology* 65:531-551. doi:10.1146/annurev-arplant-050213-035715
- Huang X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature genetics* 42 (11):961
- Hyun Y, Richter R, Coupland G (2017) Competence to flower: Age-controlled sensitivity to environmental cues. *Plant Physiology* 173:36-46. doi:10.1104/pp.16.01523
- Ishikawa R, Aoki M, Kurotani K-i, Yokoi S, Shinomura T, Takano M, Shimamoto K (2011) Phytochrome B regulates *Heading date 1* (*Hd1*)-mediated expression of rice florigen *Hd3a* and critical day length in rice. *Molecular Genetics and Genomics* 285 (6):461-470
- Izquierdo J, Mujica A, Marathe J, Jacobsen S-E (2003) Horizontal, technical cooperation in research on quinoa (*Chenopodium quinoa* Willd.). *Food Reviews International* 19 (1-2):25-29

References

- Jacobsen S-E, Christiansen JL, Rasmussen J (2010) Weed harrowing and inter-row hoeing in organic grown quinoa (*Chenopodium quinoa* Willd.). *Outlook on Agriculture* 39 (3):223-227
- Jacobsen S-E, Stølen O (1993) Quinoa - Morphology, phenology and prospects for its production as a new crop in Europe. *European Journal of Agronomy* 2:19-29. doi:10.1016/S1161-0301(14)80148-2
- Jacobsen S, Bendevis M (2013) Adaptation and scope for quinoa in northern latitudes of Europe. State of the art report on quinoa around the world in:436-446
- Jacobsen SE (2017) The scope for adaptation of quinoa in Northern Latitudes of Europe. *Journal of Agronomy and Crop Science* 203:603-613. doi:10.1111/jac.12228
- Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23 (14):1801-1806
- James LEA (2009) Quinoa (*Chenopodium quinoa* Willd.): composition, chemistry, nutritional, and functional properties. *Advances in food and nutrition research* 58:1-31
- Jarvis D (2020) EMS mutagenesis in quinoa: Demonstrating the utility of genomic resources International Quinoa Research Symposium.
- Jarvis D, Kopp O, Jellen E, Mallory M, Pattee J, Bonifacio A, Coleman C, Stevens M, Fairbanks D, Maughan P (2008) Simple sequence repeat marker development and genetic mapping in quinoa (*Chenopodium quinoa* Willd.). *Journal of Genetics* 87 (1):39-51
- Jarvis DE, Ho YS, Lightfoot DJ, Schmöckel SM, Li B, Borm TJ, Ohyanagi H, Mineta K, Michell CT, Saber N (2017) The genome of *Chenopodium quinoa*. *Nature* 542 (7641):307
- Jia G, Huang X, Zhi H, Zhao Y, Zhao Q, Li W, Chai Y, Yang L, Liu K, Lu H (2013) A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). *Nature genetics* 45 (8):957-961
- Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS (2011) Ancestral polyploidy in seed plants and angiosperms. *Nature* 473 (7345):97-100
- Joshi T, Xu D (2007) Quantitative assessment of relationship between sequence similarity and function similarity. *BMC genomics* 8 (1):222
- Kaakinen M, Mägi R, Fischer K, Heikkinen J, Järvelin M-R, Morris AP, Prokopenko I (2017) MARV: a tool for genome-wide multi-phenotype analysis of rare variants. *BMC bioinformatics* 18 (1):1-8
- Kalyaanamoorthy S, Minh BQ, Wong TK, von Haeseler A, Jermini LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature methods* 14 (6):587
- Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, Sabatti C, Eskin E (2010) Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics* 42:348-354. doi:10.1038/ng.548

References

- Kojima S, Takahashi Y, Kobayashi Y, Monna L, Sasaki T, Araki T, Yano M (2002) *Hd3a*, a rice ortholog of the *Arabidopsis FT* gene, promotes transition to flowering downstream of *Hdl* under short-day conditions. *Plant and cell physiology* 43 (10):1096-1105
- Kolano B, Siwinska D, Pando LG, Szymanowska-Pulka J, Maluszynska J (2012) Genome size variation in *Chenopodium quinoa* (Chenopodiaceae). *Plant systematics and evolution* 298 (1):251-255
- Komiya R, Ikegami A, Tamaki S, Yokoi S, Shimamoto K (2008) *Hd3a* and *RFT1* are essential for flowering in rice. *Development* 135 (4):767-774
- Komiya R, Yokoi S, Shimamoto K (2009) A gene network for long-day flowering activates *RFT1* encoding a mobile flowering signal in rice. *Development* 136:3443-3450. doi:10.1242/dev.040170
- Kong F, Liu B, Xia Z, Sato S, Kim BM, Watanabe S, Yamada T, Tabata S, Kanazawa A, Harada K (2010) Two coordinately regulated homologs of *FLOWERING LOCUS T* are involved in the control of photoperiodic flowering in soybean. *Plant physiology* 154 (3):1220-1231
- Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. *Plant methods* 9 (1):1-9
- Kumar SV, Lucyshyn D, Jaeger KE, Alós E, Alvey E, Harberd NP, Wigge PA (2012) Transcription factor PIF4 controls the thermosensory activation of flowering. *Nature* 484 (7393):242
- Lê S, Josse J, Husson F (2008) FactoMineR: an R package for multivariate analysis. *Journal of statistical software* 25 (1):1-18
- Ledger S, Strayer C, Ashton F, Kay SA, Putterill J (2001) Analysis of the function of two circadian-regulated *CONSTANS-LIKE* genes. *The Plant Journal* 26 (1):15-22
- Lee Y, An G (2015) Complex regulatory networks of flowering time in rice. *J Rice Res* 3 (2)
- Lemmon ZH, Reem NT, Dalrymple J, Soyk S, Swartwood KE, Rodriguez-Leal D, Van Eck J, Lippman ZB (2018) Rapid improvement of domestication traits in an orphan crop by genome editing. *Nature plants* 4 (10):766-770
- Letunic I, Bork P (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research* 44 (W1):W242-W245. doi:10.1093/nar/gkw290
- Lexander K Present knowledge of sugar-beet bolting mechanisms. In: Proceedings 43rd Winter Congress of the Institut International de Recherches Betteravieres, Brussels., 1980. pp 245-258
- Li H (2011) Tabix: fast retrieval of sequence features from generic TAB-delimited files. *Bioinformatics* 27 (5):718-719
- Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26 (5):589-595

References

- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25 (16):2078-2079
- Li N, Xu R, Li Y (2019) Molecular networks of seed size control in plants. *Annual review of plant biology* 70:435-463
- Li T, Yang X, Yu Y, Si X, Zhai X, Zhang H, Dong W, Gao C, Xu C (2018) Domestication of wild tomato is accelerated by genome editing. *Nature biotechnology* 36 (12):1160-1163
- Lindhout P, Danial D (2006) Participatory genomics in quinoa. *Taylor Biotechnol* 2:31-50
- Listgarten J, Lippert C, Heckerman D (2013) FaST-LMM-Select for addressing confounding from spatial structure and rare variants. *Nature genetics* 45 (5):470-471
- Liu HJ, Yan J (2019) Crop genome-wide association study: a harvest of biological relevance. *The Plant Journal* 97 (1):8-18
- López-Marqués RL, Nørrevang AF, Ache P, Moog M, Visintainer D, Wendt T, Østerberg JT, Dockter C, Jørgensen ME, Salvador AT (2020) Prospects for the accelerated improvement of the resilient crop quinoa. *Journal of Experimental Botany* 71 (18):5333-5347
- Lu S, Zhao X, Hu Y, Liu S, Nan H, Li X, Fang C, Cao D, Shi X, Kong L (2017a) Natural variation at the soybean *J* locus improves adaptation to the tropics and enhances yield. *Nature genetics* 49 (5):773-779
- Lu X, Xiong Q, Cheng T, Li Q-T, Liu X-L, Bi Y-D, Li W, Zhang W-K, Ma B, Lai Y-C (2017b) A *PP2C-I* allele underlying a quantitative trait locus enhances soybean 100-seed weight. *Molecular Plant* 10 (5):670-684
- Luccioni L, Krzymuski M, Sánchez-Lamas M, Karayekov E, Cerdán PD, Casal JJ (2019) CONSTANS delays Arabidopsis flowering under short days. *The Plant Journal* 97 (5):923-932. doi:10.1111/tpj.14171
- Mackay I, Powell W (2007) Methods for linkage disequilibrium mapping in crops. *Trends in Plant Science* 12:57-63. doi:10.1016/j.tplants.2006.12.001
- Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey AR, Potter SC, Finn RD (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic acids research* 47 (W1):W636-W641
- Maldonado-Taípe N, Sarange D, Schmöckel S, Jung C, Emrani N (2020) Validation of suitable genes for normalization of diurnal gene expression studies in *Chenopodium quinoa*. bioRxiv
- Mason S, Stevens M, Jellen E, Bonifacio A, Fairbanks D, Coleman C, McCarty R, Rasmussen A, Maughan P (2005) Development and use of microsatellite markers for germplasm characterization in quinoa (*Chenopodium quinoa* Willd.). *Crop Science* 45 (4):1618-1630
- Matar S, Kumar A, Holtgräwe D, Weisshaar B, Melzer S (2020) The transition to flowering in winter rapeseed during vernalization. *Authorea Preprints*

References

- Mather KA, Caicedo AL, Polato NR, Olsen KM, McCouch S, Purugganan MD (2007) The extent of linkage disequilibrium in rice (*Oryza sativa* L.). *Genetics* 177 (4):2223-2232
- Mathew B, Léon J, Sannemann W, Sillanpää MJ (2018) Detection of epistasis for flowering time using Bayesian multilocus estimation in a barley MAGIC population. *Genetics* 208 (2):525-536
- Maughan PJ, Bonifacio A, Jellen E, Stevens M, Coleman C, Ricks M, Mason S, Jarvis D, Gardunia B, Fairbanks D (2004) A genetic linkage map of quinoa (*Chenopodium quinoa*) based on AFLP, RAPD, and SSR markers. *Theoretical and Applied Genetics* 109 (6):1188-1195
- Maughan PJ, Chaney L, Lightfoot DJ, Cox BJ, Tester M, Jellen EN, Jarvis DE (2019) Mitochondrial and chloroplast genomes provide insights into the evolutionary origins of quinoa (*Chenopodium quinoa* Willd.). *Scientific Reports* 9:185. doi:10.1038/s41598-018-36693-6
- Maughan PJ, Smith SM, Rojas-Beltrán JA, Elzinga D, Raney JA, Jellen EN, Bonifacio A, Udall JA, Fairbanks DJ (2012) Single Nucleotide Polymorphism Identification, Characterization, and Linkage Mapping in Quinoa. *The Plant Genome Journal* 5:114. doi:10.3835/plantgenome2012.06.0011
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* 20 (9):1297-1303
- Meyer RS, Purugganan MD (2013) Evolution of crop species: genetics of domestication and diversification. *Nature reviews Genetics* 14:840-852. doi:10.1038/nrg3605
- Michaels SD, Amasino RM (1999) *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. *The Plant Cell* 11 (5):949-956
- Milne I, Shaw P, Stephen G, Bayer M, Cardle L, Thomas WTB, Flavell AJ, Marshall D (2010) Flapjack--graphical genotype visualization. *Bioinformatics (Oxford, England)* 26 (24):3133-3134. doi:10.1093/bioinformatics/btq580
- Milner SG, Jost M, Taketa S, Mazón ER, Himmelbach A, Oppermann M, Weise S, Knüpfner H, Basterrechea M, König P (2019) Genebank genomics highlights the diversity of a global barley collection. *Nature genetics* 51 (2):319-326
- Murase K, Hirano Y, Sun T-p, Hakoshima T (2008) Gibberellin-induced DELLA recognition by the gibberellin receptor *GID1*. *Nature* 456 (7221):459-463
- Murphy KM, Matanguihan JB, Fuentes FF, Gómez-Pando LR, Jellen EN, Maughan PJ, Jarvis DE (2018) Quinoa Breeding and Genomics. *Plant Breeding Reviews* 42:257-320
- Murphy KS, Matanguihan J (2015) Quinoa: Improvement and sustainable production. John Wiley & Sons,
- Navarro C, Abelenda Ja, Cruz-Oró E, Cuéllar Ca, Tamaki S, Silva J, Shimamoto K, Prat S (2011) Control of flowering and storage organ formation in potato by *FLOWERING LOCUS T*. *Nature* 478:119-122. doi:10.1038/nature10431

References

- Nonis A, De Nardi B, Nonis A (2014) Choosing between RT-qPCR and RNA-seq: a back-of-the-envelope estimate towards the definition of the break-even-point. *Analytical and bioanalytical chemistry* 406 (15):3533-3536
- Ohshima S, Murata M, Sakamoto W, Ogura Y, Motoyoshi F (1997) Cloning and molecular analysis of the Arabidopsis gene *Terminal Flower 1*. *Molecular and General Genetics MGG* 254 (2):186-194
- Palomino G, Hernández LT, de la Cruz Torres E (2008) Nuclear genome size and chromosome analysis in *Chenopodium quinoa* and *C. berlandieri* subsp. *nuttalliae*. *Euphytica* 164 (1):221
- Paradis E (2010) pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* 26 (3):419-420
- Peterson A, Jacobsen SE, Bonifacio A, Murphy K (2015) A crossing method for Quinoa. *Sustainability (Switzerland)* 7:3230-3243. doi:10.3390/su7033230
- Pin PA, Benlloch R, Bonnet D, Wremmerth-Weich E, Kraft T, Gielen JJ, Nilsson O (2010) An antagonistic pair of *FT* homologs mediates the control of flowering time in sugar beet. *Science* 330 (6009):1397-1400
- Pin PA, Zhang W, Vogt SHH, Dally N, Büttner B, Schulze-Buxloh G, Jelly NSS, Chia TYPYP, Mutasa-Göttgens ESS, Dohm JCC, Himmelbauer H, Weisshaar B, Kraus J, Gielen JJJL, Lommel M, Weyens G, Wahl B, Schechert A, Nilsson O, Jung C, Kraft T, Müller AEE (2012) The role of a pseudo-response regulator gene in life cycle adaptation and domestication of beet. *Current biology : CB* 22:1095-1101. doi:10.1016/j.cub.2012.04.007
- Prakapenka D, Wang C, Liang Z, Bian C, Tan C, Da Y (2020) GVCHAP: A Computing Pipeline for Genomic Prediction and Variance Component Estimation Using Haplotypes and SNP Markers. *Frontiers in Genetics* 11:282
- Qian L, Hickey LT, Stahl A, Werner CR, Hayes B, Snowdon RJ, Voss-Fels KP (2017) Exploring and harnessing haplotype diversity to improve yield stability in crops. *Frontiers in plant science* 8:1534
- Rea J (1969) Biología floral de la quinua (*Chenopodium quinoa*). *Publ. IICA N* 19 (1): 91-96. Turrialba, Costra Rica
- Risi J, Galwey N, Wickens G, Haq N, Day P (1989) New crops for food and industry.
- Risi JC, Galwey NW (1984) The *Chenopodium* grains of the Andes: Inca crops for modern agriculture. *Adv Appl Biol* 10
- Roden LC, Song H-R, Jackson S, Morris K, Carre IA (2002) Floral responses to photoperiod are correlated with the timing of rhythmic expression relative to dawn and dusk in *Arabidopsis*. *Proceedings of the National Academy of Sciences* 99 (20):13313-13318
- Rodríguez LA, Isla MT (2009) Comparative analysis of genetic and morphologic diversity among quinoa accessions (*Chenopodium quinoa* Willd.) of the South of Chile and highland accessions. *Journal of Plant Breeding and Crop Science* 1 (5):210-216

References

- Rojas W (2003) Multivariate Analysis of Genetic Diversity of Bolivian Quinoa Germplasm. *Food Reviews International* 19 (1-2):9-23. doi:10.1081/FRI-120018864
- Rojas W, Pinto M, Alanoca C, Gómez Pando L, León-Lobos P, Alercia A, Diulgheroff S, Padulosi S, Bazile D (2015) Quinoa genetic resources and ex situ conservation.
- Romero Navarro JA, Willcox M, Burgueño J, Romay C, Swarts K, Trachsel S, Preciado E, Terron A, Delgado HV, Vidal V, Ortega A, Banda AE, Montiel NOG, Ortiz-Monasterio I, Vicente FS, Espinoza AG, Atlin G, Wenzl P, Hearne S, Buckler ES (2017) A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nature Genetics* 49:476-480. doi:10.1038/ng.3784
- Ruas PM, Bonifacio A, Ruas CF, Fairbanks DJ, Andersen WR (1999) Genetic relationship among 19 accessions of six species of *Chenopodium* L., by Random Amplified Polymorphic DNA fragments (RAPD). *Euphytica* 105 (1):25-32
- Ruden DM, Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Lu X (2012) Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, SnpSift. *Frontiers in genetics* 3:35
- Ruiz KB, Biondi S, Oses R, Acuña-Rodríguez IS, Antognoni F, Martínez-Mosqueira EA, Coulibaly A, Canahua-Murillo A, Pinto M, Zurita-Silva A (2014) Quinoa biodiversity and sustainability for food security under climate change. A review. *Agronomy for sustainable development* 34 (2):349-359
- Schiessl SV, Huettel B, Kuehn D, Reinhardt R, Snowdon RJ (2017) Flowering Time Gene Variation in *Brassica* Species Shows Evolutionary Principles. *Frontiers in Plant Science* 8:1-13. doi:10.3389/fpls.2017.01742
- Shah S, Weinholdt C, Jedrusik N, Molina C, Zou J, Große I, Schiessl S, Jung C, Emrani N (2018) Whole-transcriptome analysis reveals genetic factors underlying flowering time regulation in rapeseed (*Brassica napus* L.). *Plant, cell & environment* 41 (8):1935-1947
- Sheldon CC, Burn JE, Perez PP, Metzger J, Edwards JA, Peacock WJ, Dennis ES (1999) The *FLF* MADS box gene: a repressor of flowering in Arabidopsis regulated by vernalization and methylation. *The Plant Cell* 11 (3):445-458
- Shindo C, Aranzana MJ, Lister C, Baxter C, Nicholls C, Nordborg M, Dean C (2005) Role of *FRIGIDA* and *FLOWERING LOCUS C* in determining variation in flowering time of Arabidopsis. *Plant physiology* 138 (2):1163-1173
- Silvestri V, Gil F (2000) Alogamia en quinua. *Revista de la Facultad de Ciencias Agrarias* 32 (1)
- Simmonds NW (1971) The breeding system of *Chenopodium quinoa*. I *Heredity* 27
- Slatkin M (2008) Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nat Rev Genet* 9 (6):477-485
- Sloat LL, Davis SJ, Gerber JS, Moore FC, Ray DK, West PC, Mueller ND (2020) Climate adaptation by crop migration. *Nature communications* 11 (1):1-9
- Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW (2013) Pleiotropy in complex traits: challenges and strategies. *Nat Rev Genet* 14 (7):483-495

References

- Song YH, Kubota A, Kwon MS, Covington MF, Lee N, Taagen ER, Laboy Cintrón D, Hwang DY, Akiyama R, Hodge SK, Huang H, Nguyen NH, Nusinow DA, Millar AJ, Shimizu KK, Imaizumi T (2018) Molecular basis of flowering under natural long-day conditions in *Arabidopsis*. *Nature Plants* 4:824-835. doi:10.1038/s41477-018-0253-3
- Song YH, Shim JS, Kinmonth-Schultz HA, Imaizumi T (2015) Photoperiodic flowering: time measurement mechanisms in leaves. *Annual review of plant biology* 66:441-464
- Stetter MG, Gates DJ, Mei W, Ross-Ibarra J (2017) How to make a domesticate. *Current Biology* 27 (17):R896-R900
- Štorchová H, Hubáčková H, Abeyawardana OA, Walterová J, Vondráková Z, Eliášová K, Mandák B (2019) *Chenopodium ficifolium* flowers under long days without upregulation of *FLOWERING LOCUS T (FT)* homologs. *Planta* 250 (6):2111-2125
- Teotia S, Tang G (2015) To bloom or not to bloom: role of microRNAs in plant flowering. *Molecular plant* 8 (3):359-377
- Thompson D, Regev A, Roy S (2015) Comparative analysis of gene regulatory networks: from network reconstruction to evolution. *Annual review of cell and developmental biology* 31:399-428
- Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001) *Dwarf8* polymorphisms associate with variation in flowering time. *Nature genetics* 28 (3):286-289
- Tiwari SB, Shen Y, Chang H-C, Hou Y, Harris A, Ma SF, McPartland M, Hymus GJ, Adam L, Marion C, Belachew A, Repetti PP, Reuber TL, Ratcliffe OJ (2010) The flowering time regulator *CONSTANS* is recruited to the *FLOWERING LOCUS T* promoter via a unique cis-element. *New Phytologist* 187 (1):57-66. doi:10.1111/j.1469-8137.2010.03251.x
- Tsuji H, Taoka K, Shimamoto K (2011a) Regulation of flowering in rice: two florigen genes, a complex gene network, and natural variation. *Current Opinion in Plant Biology* 14 (1):45-52. doi:10.1016/j.pbi.2010.08.016
- Tsuji H, Taoka KI, Shimamoto K (2011b) Regulation of flowering in rice: Two florigen genes, a complex gene network, and natural variation. *Current Opinion in Plant Biology* 14:45-52. doi:10.1016/j.pbi.2010.08.016
- Turner SD (2014) qqman: an R package for visualizing GWAS results using QQ and manhattan plots. *Biorxiv*:005165
- Uotila P Chromosome counts on the *Chenopodium album* aggregate in Finland and NE Sweden. In: *Annales Botanici Fennici*, 1972. JSTOR, pp 29-32
- Valverde F, Mouradov A, Soppe W, Ravenscroft D, Samach A, Coupland G (2004) Photoreceptor regulation of *CONSTANS* protein in photoperiodic flowering. *Science* 303 (5660):1003-1006
- Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J (2013) From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current protocols in bioinformatics* 43 (1):11.10. 11-11.10. 33

References

- Varshney RK, Saxena RK, Upadhyaya HD, Khan AW, Yu Y, Kim C, Rathore A, Kim D, Kim J, An S (2017) Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. *Nature Genetics* 49 (7):1082
- Varshney RK, Thudi M, Roorkiwal M, He W, Upadhyaya HD, Yang W, Bajaj P, Cubry P, Rathore A, Jian J (2019a) Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. *Nature genetics* 51 (5):857-864
- Varshney RK, Thudi M, Roorkiwal M, He W, Upadhyaya HD, Yang W, Bajaj P, Cubry P, Rathore A, Jian J, Doddamani D, Khan AW, Garg V, Chitikineni A, Xu D, Gaur PM, Singh NP, Chaturvedi SK, Nadigatla GVPR, Krishnamurthy L, Dixit GP, Fikre A, Kimurto PK, Sreeman SM, Bharadwaj C, Tripathi S, Wang J, Lee SH, Edwards D, Polavarapu KKB, Penmetsa RV, Crossa J, Nguyen HT, Siddique KHM, Colmer TD, Sutton T, von Wettberg E, Vigouroux Y, Xu X, Liu X (2019b) Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. *Nature Genetics* 51:857-864. doi:10.1038/s41588-019-0401-3
- Vega-Gálvez A, Miranda M, Vergara J, Uribe E, Puente L, Martínez EA (2010) Nutrition facts and functional potential of quinoa (*Chenopodium quinoa* willd.), an ancient Andean grain: a review. *Journal of the Science of Food and Agriculture* 90 (15):2541-2547
- Wallace JG, Rodgers-Melnick E, Buckler ES (2018) On the road to breeding 4.0: unraveling the good, the bad, and the boring of crop quantitative genomics. *Annual review of genetics*
- Wang B, Lin Z, Li X, Zhao Y, Zhao B, Wu G, Ma X, Wang H, Xie Y, Li Q (2020) Genome-wide selection and genetic improvement during modern maize breeding. *Nature Genetics*:1-7
- Wang J-W, Czech B, Weigel D (2009) miR156-regulated SPL transcription factors define an endogenous flowering pathway in *Arabidopsis thaliana*. *Cell* 138 (4):738-749
- Wang S, Tsuchiya T, Wilson H (1993) Chromosome studies in several species of *Chenopodium* from North and South America. *Journal of Genetics and Breeding* 47:163-163
- Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, Wu Z, Li M, Zheng T, Fuentes RR, Zhang F (2018) Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 557 (7703):43-49
- Ward SM, Johnson DL (1993) Cytoplasmic male sterility in quinoa. *Euphytica* 66:217-223. doi:10.1007/BF00025306
- Wickland DP, Hanzawa Y (2015) The *FLOWERING LOCUS T/TERMINAL FLOWER 1* Gene Family: Functional Evolution and Molecular Mechanisms. *Molecular Plant* 8:983-997. doi:10.1016/j.molp.2015.01.007
- Wilson HD (1988a) Allozyme variation and morphological relationships of *Chenopodium hircinum* (s.l.). *Syst Bot* 13. doi:10.2307/2419100
- Wilson HD (1988b) Quinoa biosystematics I: Domesticated populations. *Economic Botany* 42 (4):461-477

References

- Wilson HD (1988c) Quinoa biosystematics II: free-living populations. *Economic Botany* 42 (4):478-494
- Wilson HD (1990) Quinoa and relatives (*Chenopodium* sect. *Chenopodium* subsect. *Celluloid*). *Economic Botany* 44 (3):92
- Won S, Park J-E, Son J-H, Lee S-H, Park BH, Park M, Park W-C, Chai H-H, Kim H, Lee J (2020) Genomic Prediction Accuracy Using Haplotypes Defined by Size and Hierarchical Clustering Based on Linkage Disequilibrium. *Frontiers in Genetics* 11:134
- Wu D, Liang Z, Yan T, Xu Y, Xuan L, Tang J, Zhou G, Lohwasser U, Hua S, Wang H (2019) Whole-genome resequencing of a worldwide collection of rapeseed accessions reveals the genetic basis of ecotype divergence. *Molecular plant* 12 (1):30-43
- Wu Z, Fang X, Zhu D, Dean C (2020) Autonomous Pathway: *FLOWERING LOCUS C* Repression through an Antisense-Mediated Chromatin-Silencing Mechanism. *Plant Physiology* 182 (1):27-37
- Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, Dong Y, Gutenkunst RN, Fang L, Huang L (2012) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nature biotechnology* 30 (1):105-111
- Yamaguchi A, Kobayashi Y, Goto K, Abe M, Araki T (2005) *TWIN SISTER OF FT (TSF)* acts as a floral pathway integrator redundantly with *FT*. *Plant and Cell Physiology* 46 (8):1175-1189
- Yamaguchi S (2008) Gibberellin metabolism and its regulation. *Annu Rev Plant Biol* 59:225-251
- Yi L, Chen C, Yin S, Li H, Li Z, Wang B, King GJ, Wang J, Liu K (2018) Sequence variation and functional analysis of a *FRIGIDA* orthologue (*BnaA3.FRI*) in *Brassica napus*. *BMC Plant Biology* 18:1-13. doi:10.1186/s12870-018-1253-1
- Yu D, Hu Y, Wang H, Pan J, Li Y, Lou D (2016a) The *DELLA-CONSTANS* transcription factor cascade integrates gibberellic acid and photoperiod signaling to regulate flowering. *Plant Physiology*
- Yu K, Wang X, Chen F, Chen S, Peng Q, Li H, Zhang W, Hu M, Chu P, Zhang J (2016b) Genome-wide transcriptomic analysis uncovers the molecular basis underlying early flowering and apetalous characteristic in *Brassica napus* L. *Scientific reports* 6:30576
- Zhang C, Dong S-S, Xu J-Y, He W-M, Yang T-L (2018) PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35 (10):1786-1788. doi:10.1093/bioinformatics/bty875
- Zhang R, Zheng F, Wei S, Zhang S, Li G, Cao P, Zhao S (2019) Evolution of disease defense genes and their regulators in plants. *International Journal of Molecular Sciences* 20 (2):335
- Zhang T, Gu M, Liu Y, Lv Y, Zhou L, Lu H, Liang S, Bao H, Zhao H (2017) Development of novel InDel markers and genetic diversity in *Chenopodium quinoa* through whole-genome re-sequencing. *BMC Genomics* 18 (1):685. doi:10.1186/s12864-017-4093-8

References

- Zhao G, Lian Q, Zhang Z, Fu Q, He Y, Ma S, Ruggieri V, Monforte AJ, Wang P, Julca I (2019) A comprehensive genome variation map of melon identifies multiple domestication events and loci influencing agronomic traits. *Nature Genetics* 51 (11):1607-1615
- Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28 (24):3326-3328. doi:10.1093/bioinformatics/bts606
- Zhou C-M, Zhang T-Q, Wang X, Yu S, Lian H, Tang H, Feng Z-Y, Zozomova-Lihová J, Wang J-W (2013) Molecular basis of age-dependent vernalization in *Cardamine flexuosa*. *Science* 340 (6136):1097-1100
- Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, Li W, Yu Y, Shu L, Zhao Y, Ma Y (2015) Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nature biotechnology* 33 (4):408-414
- Zou C, Chen A, Xiao L, Muller HM, Ache P, Haberer G, Zhang M, Jia W, Deng P, Huang R, Lang D, Li F, Zhan D, Wu X, Zhang H, Bohm J, Liu R, Shabala S, Hedrich R, Zhu JK, Zhang H (2017) A high-quality genome assembly of quinoa provides insights into the molecular basis of salt bladder-based salinity tolerance and the exceptional nutritional value. *Cell Research* 27:1327-1340. doi:10.1038/cr.2017.124

9 Supplementary data on CD/DVD

The following supplemental data are available on a CD and can be distributed upon request (contact: Prof. Dr. Christian Jung, c.jung@plantbreeding.uni-kiel.de).

File name	Content	Format
Chapter 2 Supplementary Tables	Supplementary Table 1, 2, 3, 5 and 7	.xlsx
Reference sequences	All genes from investigated in the Chapter 2	.clc annotates sequences
Seed-codes	All seed codes of plant materials used in this study	.xlsx
Phenotype data	All phenotype data used Chapter 2 and 3 of this study	.xlsx
Re-sequencing data analysis README	Absolute path to sequence folders of data and its contents all info will be available upon publication of Chapter 2 on Github.com	.txt
GWAS analysis README	Information and Codes of GWAS and downstream analysis and all info will be available upon publication of Chapter 2 on Github.com	.txt

10 Curriculum Vitae and Publications

10.1.1 Curriculum vitae

Personal Data

Name: Dilan Sarange Rajapaksha Patiranage

Date of birth: 27th September 1985

Place of birth: Walasmulla/Sri Lanka

Nationality: Sri Lankan

Education

01/2018 - present PhD student at the Plant Breeding institute Kiel, University of Kiel.
Research title: “Unraveling genetic mechanisms of flowering time control and other agronomically important traits in quinoa (*Chenopodium quinoa*)”
Supervisors: Dr. Nazgol Emrani, Prof. Dr. Christian Jung

03/2016 - 12/2017 PhD student at the Plant Breeding institute Kiel, University of Kiel.
Research title: “Molecular and quantitative genetics of storage root formation in sugar beet”. Supervisors: Dr. Siegbert Melzer, Prof. Dr. Christian Jung

04/2015 - 10/2015 Master’s thesis at the Plant Breeding Institute, University of Kiel
Research title: “Phenotyping and genotyping a sugar beet x leaf beet RIL population.”
Supervisors: Dr. Siegbert Melzer, Prof. Dr. Christian Jung

10/2013 - 10/2015 Master of Science in AgriGenomics, University of Kiel

08/2007 - 10/2011 Bachelor of Science in Agriculture Technology and Management,
Faculty of Agriculture, University of Peradeniya, Sri Lanka
Research title: “Effect of different concentrations of nitrates on nodulation of common bean (*Phaseolus vulgaris* L.)”. Supervisors: Prof. Dr. Renuka Karunagoda, Prof. Dr. V. Ariya Sumanasinghe

10.1.2 Publications

Articles

Patirange D.S.R., Asare E, Maldonado N., Rey E, Emrani N., Tester M., and Jung C (2020). Haplotype variations of major flowering time genes in quinoa unveil their role in the adaptation to different environmental conditions. Under review.

Patirange D.S.R., Rey E, Emrani N., Wellman G., Schmöckel S., Schmid K., Tester M., and Jung C (2020). Genome-wide association study in the pseudo cereal quinoa reveals selection patterns typical for crops with a short breeding history, manuscript in preparation.

Curriculum Vitae and Publications

Emrani N, Hasler M, **Patiranage D.S.R**, Maldonado N, Rey E, Jung C An efficient method to produce segregating populations in quinoa (*Chenopodium quinoa*). Plant Breeding 2020; 00:1-11. doi:10.1111/pbr.12873

Karunarathna N.L., **Patirange D.S.R**, Harloff H.-J., Sashidhar N., Jung C., Genomic background selection to reduce mutation load using molecular markers and rapid cycle breeding rapeseed (*Brassica napus*), manuscript in preparation

Maldonado-Taibe N, **Sarange D**, Schmöckel S, Jung C, Emrani N (2020) Validation of suitable genes for normalization of diurnal gene expression studies in *Chenopodium quinoa*. bioRxiv

Oral and poster presentations

Patirange D.S.R, Emrani N., Asare E., Böndel B K., Wellman G., Rey E., Schmöckel S., Schmid K., Tester M., Jung C. (2020). Genome-wide association mapping for flowering time in quinoa. International Symposium of the Society for Plant Breeding (GPZ). Digital breeding. 11-13 February 2020, Tulln, Austria. Oral presentation

Patirange D.S.R., Emrani N., and Jung C., (2019). Unraveling genetic mechanisms of flowering time control in quinoa. KLS Retreat, 21-22 November 2019, Schleswig, Germany. Oral presentation

Patirange D.S.R., Emrani N., and Jung C., (2019). Genome-wide association mapping for flowering time in quinoa. Kiel Plant Center Mini-Symposium, 7 November 2019, Kiel, Germany. Oral presentation

Patirange D.S.R., Maldonado N., Emrani N., and Jung C., (2019). Unraveling genetic mechanisms of flowering time control in quinoa. Genome Research for Plant Breeding GPZ Conference, 28-29 March 2019, Schleswig, Germany. Poster presentation

Patirange D.S.R, Emrani N., and Jung C., (2018). Unraveling genetic mechanisms of flowering time control in quinoa. III. Symposium: Genetic Variation of Flowering Time Genes and Applications for Crop Improvement. 14-16 March 2018, Kiel, Germany. Poster presentation

Patirange D.S.R, Melzer S., Capistrano-Goßmann.G., Jung C., (2018). Unraveling quantitative trait loci for storage root formation in sugar beet (*Beta vulgaris*): German plant breeding conference 2018, 28.02-02.03.2018 Wernigerode, Germany. Poster presentation

Patiranage D.S.R, Melzer S., Capistrano-Goßmann.G., Jung C., (2017). Unraveling quantitative trait loci for storage root formation in sugar beet (*Beta vulgaris*) Botanikertagung, 17-21 Sep 2017 Kiel, Germany. Poster presentation

11 Publications and declaration of own contribution

Chapters 2 was prepared as manuscripts for publication. Chapters 3 of this thesis has been submitted for publication. The respective declarations of co-authorship are presented on the following page.

C A U	Christian-Albrechts-Universität zu Kiel	Agrar- und Ernährungs- Wissenschaftliche Fakultät
------------------	---	--

Declaration of co-authorship

1. Doctoral candidate
Name: Dilan Sarange Rajapaksha Patiranage


2. This co-author declaration applies to the following article:
Haplotype variations of major flowering time genes in quinoa unveil their role in the adaptation to different environmental conditions

The extent of the doctoral candidate's contribution to the article is assessed on the following scale:

- A. Has contributed to the work (0-33%)
- B. Has made a substantial contribution (34-66%)
- C. Did the majority of the work independently (67-100%)

3. Declaration on the individual phases of the scientific work (A,B,C)	Extent
Concept: Formulation of the basic scientific problem based on theoretical questions which require clarification, including a summary of the general questions which, it is assumed, will be answerable via analyses or concrete experiments/investigations	B
Planning: Planning of experiments/analyses and formulation of investigative methodology, including choice of method and independent methodological development, in such a way that the scientific questions asked can be expected to be answered	B
Execution: Involvement in the analysis or the concrete experiments/investigation	C
Manuscript preparation: Presentation, interpretation and discussion of the results obtained in article form	C

4. Signature of all co-authors		
Date	Name	Signature
05.11.2020	Edward Asare	
06/11/20	Nathaly Maldonado-Taibe	
05/11/20	Elodie Rey	
05.11.2020	Nazgol Emrani	
04/11/20	Mark Tester	
6.11.2020	Christian Jung	

5. Signature of doctoral candidate		
Date	Name	Signature
06/11/20	Dilan S.R. Patiranage	

C	A	U	Christian-Albrechts-Universität zu Kiel	Agrar- und Ernährungs- wissenschaftliche Fakultät
----------	----------	----------	---	--

Declaration of co-authorship

1. Doctoral candidate

Name: Dilan Sarange Rajapaksha Patiranage

2. This co-author declaration applies to the following article:

Genome-wide association study in the pseudocereal quinoa reveals selection pattern typical for crops with a short breeding history

The extent of the doctoral candidate's contribution to the article is assessed on the following scale:

- A. Has contributed to the work (0-33%)
- B. Has made a substantial contribution (34-66%)
- C. Did the majority of the work independently (67-100%)

3. Declaration on the individual phases of the scientific work (A,B,C)	Extent
Concept: Formulation of the basic scientific problem based on theoretical questions which require clarification, including a summary of the general questions which, it is assumed, will be answerable via analyses or concrete experiments/investigations	B
Planning: Planning of experiments/analyses and formulation of investigative methodology, including choice of method and independent methodological development, in such a way that the scientific questions asked can be expected to be answered	C
Execution: Involvement in the analysis or the concrete experiments/investigation	C
Manuscript preparation: Presentation, interpretation and discussion of the results obtained in article form	C

4. Signature of all co-authors

Date	Name	Signature
05/11/20	Elodie Rey	
05.11.2020	Nazgol Emrani	
5-11-2020	Gordon Wellman	
05.11.20	Karl Schmid	
05.11.2020	Sandra M. Schmöckel	
04/11/20	Mark Tester	
6.11.2020	Christian Jung	

5. Signature of doctoral candidate

Date	Name	Signature
06/11/20	Dilan S.R. Patiranage	

12 Acknowledgments

First, I would like to thank Prof. Dr. Christian Jung for giving me the opportunity to conduct this exciting work at the Plant Breeding Institute. I am grateful for your support, guidance, and providing a well-organized working environment. Your supervision, critical comments, and advice encouraged me to give my best and shaped my research work and thesis.

Then, I would like to express my sincere gratitude to Dr. Nazgol Emrani for believing in me and take the risk to supervise me. This work would have been impossible without your continuous support, guidance, and advice. Thank you very much Nozi, for being patient and sharing knowledge and ideas. My vocabulary is not enough to thank you for the great support all the time.

I am grateful to Moni and Brigitte for helping me at the greenhouse and the field for gigantic phenotyping and harvesting operations; without your support, it was impossible even to think.

My genuine thanks go to Prof. Dr. Mark Tester, Prof. Dr. Karl Schmid, Jun.-Prof. Dr. Sandra S. Schmokel, Dr. Elodie Ray, Dr. Gordon Wellman, and Gabriele Fiene for the excellent collaboration. It was a great pleasure to conduct research with all of you.

I would like express my gratitude to Prof. Dr. Mario Hasler for supporting me with statistical modeling and designing.

I thank Hans for the friendly discussions about research and life during my Ph.D. and for giving me time whenever I need help. I am appreciative of Siegbert for helping me at the beginning of my Ph.D. work. Also, I am grateful to Birgit, Verena, Kerstin, Hilke, Ines, Meike and Bettina for your technical assistance and kindness. I thank Gina, Nadine Dally, Nadine Höft, Smit, and Janina for having a nice time during your time at the institute. Big thanks also go to Antje and other secretaries for your help in dealing with all the complicated documents. I thank Nathaly and Edward for doing your master's thesis as a part of my research, and working with you was a pleasure.

I am extremely thankful to my "Kiel Family" Avneesh, Srijan, Sarah, Tahmina, Yixin, Amar, and Ishani, for giving a great environment. Having you all around kept me moving forward. All the fruitful discussions about research, life, and society helped me think out of the box.

My special thanks go to my lovely wife, Nirosha, for being my whole life. I always feel stronger as an individual and also as a researcher when you are with me. You complete me in every possible way, so it's incredible to have you in my life.

I want to express my gratitude to my parents, sister, and relatives for all your moral support and encouragement. You were always the best source of motivation. I am grateful to my German family Bernd and Wolli, for everything. I could not have come to Germany without your support, and I am here today because of you.

Last but not least, I appreciate the financial support by the Competitive Research Grant (Grant No. OSR-2016-CRG5-466 2966-02) of the King Abdullah University of Science and Technology, Saudi Arabia, and the Sanger sequencing service of the Institute of Clinical Molecular Biology Kiel. Also, I cannot forget German taxpayers who enabled free and quality education for students of the World.