# Genomic Divergence in Differentially Adapted Wild and Domesticated Barley

Girma Bedada
*Faculty of Natural Resources and Agricultural Sciences*
*Department of Plant Biology*
*Uppsala*

Doctoral Thesis
Swedish University of Agricultural Sciences
Uppsala 2014

Acta Universitatis agriculturae Sueciae
2014:108

Cover: Circos graph showing different genomic analyses of wild and domesticated barley. Inner to outer circular bar graphs: analysed gene fragments, nucleotide variation from transcriptome data, part of targeted genes, novel *BARE* insertions and the barley chromosomes. The outer scatter graphs show barley gene density (IBGS Consortium *et al*., 2012). Photo - wild & cultivated barley spike (© – G. Bedada)

ISSN 1652-6880
ISBN (print version) 978-91-576-8164-5
ISBN (electronic version) 978-91-576-8165-2

Print: SLU Service/Repro, Uppsala 2014

# Genomic Divergence in Differentially Adapted Wild and Domesticated Barley


## Abstract

Genomic divergence is responsible for plant differential adaptation to diverse and contrasting environments and different biotic stresses. This thesis focuses on the analyses of the adaptive genomic divergence in wild and domesticated barley and the driving evolutionary forces, and to identify genes and genetic variation with signature of adaptive selection.

By applying genome scanning, transcriptome sequencing and customized target-enriched pool sequencing approaches, we found strong adaptive patterns of genomic divergence in wild barley across environmental gradients in Israel, which is about two-thirds of the variation found in samples from the whole species range. Hence, high level of population structure driven by natural selection and neutral evolutionary forces was observed at large and small geographical scales. Strong phenotypic and genomic differentiation was detected between wild barley ecotypes from the desert and Mediterranean environments. The desert ecotype had better water use efficiency and higher leaf relative water content. The majority of the transcripts were non-shared between the ecotypes and hence novel transcripts were identified. The genomic divergence was about 2-fold higher in the desert ecotype and it harbored more deleterious mutations than the Mediterranean ecotype, which is genetically closer to cultivated barley. Novel transcripts from the desert ecotype and genes differentially expressed in another drought-tolerant ecotype showed higher genomic divergence than the average genes. Using the targeted captured pooled sequencing, we identified genes and genetic variation with signature of selection in wild and Ethiopian cultivated barley genotypes. Ethiopian barley had high genomic divergence similar to wild barley, retained large proportion of ancestral variation, and showed low genomic differentiation from the wild ancestor.

Using the targeted sequence capturing method, we were able to detect known *BARE* retroelement insertions and further identify genome-wide novel insertions from pooled sequencing of wild and Ethiopian barley genotypes.



*Keywords: BARE*, drought tolerance, *Hordeum*, Evolution Canyon, genome divergence, population structure, transcriptome, targeted capture, transposable element, wild barley.



*Author's address:* Girma Bedada, SLU, Department of Plant Biology,
P.O. Box 7080, 750 07 Uppsala, Sweden
*E-mail:* Girma.Bedadaa@slu.se

# Dedication

*To my mother Alem Dadi (Alu) and my better-half Hiwot Amenu (Hiwotukiya)*

# Contents

# List of Publications

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

I **Girma Bedada**, Anna Westerbergh, Eviatar Nevo, Abraham Korol and Karl J Schmid (2014). DNA sequence variation of wild barley *Hordeum spontaneum* (L.) across environmental gradients in Israel. *Heredity* 112, 646-655.

II **Girma Bedada**, Anna Westerbergh, Thomas Mueller, Eyal Galkin, Eyal Bdolach, Menachem Moshelion, Eyal Fridman and Karl J Schmid (2014). Transcriptome sequencing of two wild barley (*Hordeum spontaneum* L.) ecotypes differentially adapted to drought stress reveals ecotype-specific transcripts. *BMC Genomics* 2014, 15:995. DOI:10.1186/1471-2164-15-995.

III **Girma Bedada**, Anna Westerbergh, Ivan Barilar and Karl J Schmid. Targeted capture sequencing of selected genes in wild and domesticated barley populations adapted to diverse environments. *Manuscript*.

IV **Girma Bedada**, Anna Westerbergh and Karl J Schmid. TE-Capture: Genome-wide enrichment and pooled sequencing uncover novel *BARE* insertions in diverse wild and domesticated barley. *Manuscript*.

Papers I and II are reproduced with the permission of the publishers.

The contribution of Girma Bedada to the papers included in this thesis was as follows:

I Performed data analysis and wrote the paper with the guidance of supervisors.

II Participated in experiment planning, performed the laboratory experiment, analysed the data and wrote the paper with the guidance of supervisors.

III Designed the experiment, carried out the experimental work, analysed the data and wrote the manuscript with the guidance of supervisors.

IV Designed the experiment, carried out the experimental work, analysed the data and wrote the manuscript with the guidance of supervisors.

# Abbreviations

| | |
|---|---|
| B1K2 | Desert wild barley ecotype from Negev desert of Israel (B1K-2-8) |
| B1K30 | Mediterranean wild barley ecotype from Israel (B1K-30-9) |
| B1K4 | Wild barley ecotype from Ein Prat, Israel (B1K-4-12) |
| B1K | Combined data generated from B1K2 and B1K30 |
| BARE | Barely retroelement |
| CNV | Copy number variation |
| EC | Evolution Canyon |
| Fl-cDNA | Full-length cDNA |
| HC | High-confidence (barley genes) |
| Hs | *Hordeum spontaneum* |
| Hv | *Hordeum vulgare* |
| IBGS | International Barley Genome Sequencing |
| InDel | Insertion and/or deletion |
| LD | Linkage disequilibrium |
| LTR | Long terminal repeat |
| NFS | North-facing slope |
| NGS | Next-generation sequencing |
| Pool-seq | Pool sequencing |
| PUTs | Putative transcripts |
| RBH | Reciprocal (bi-directional) blast hit |
| RNA-Seq | High-throughput transcriptome sequencing |
| RWC | Relative water content |
| SFS | South-facing slope |
| SNV/SNP | Single-nucleotide variation/polymorphism |
| SV | Structural variation |
| TE | Transposable element |
| WGS | Whole genome sequence |
| WUE | Water use efficiency |

# 1 Introduction

Genomic divergence, the variation between the genomes of individuals or populations within the same species, is ranging from small-scale nucleotide variations to large-scale structural variations at gene and chromosomal levels (Marroni *et al.*, 2014; Zmienko *et al.*, 2014). Adaptive genomic divergence is the main factor responsible for differential adaptation of individuals or populations to heterogeneous or contrasting environments with different biotic and abiotic stress pressures. Such divergence is a predominant source of important genes and genetic variants for breeding and development of environmentally adapted and stress tolerant crop plants. So far, using such variations significant achievements have been made in enhancing crop productivity (Godfray *et al.*, 2010; Tester & Langridge, 2010).

The genomic resources in wild crop relatives and landraces have immensely contributed towards enhancing crop productivity, and this untapped resources will be the main source of variation to develop and improve crop plants to meet the global food demand under ever-changing environmental climates (Huang & Han, 2014; Tester & Langridge, 2010). Towards these ends, the followings are pivotal areas of research: (i) systematic collection, characterization and comparison of the genomes of individuals and populations adapted to contrasting environments and under different stresses, (ii) identification and efficient utilization of the responsible genes and genetic variations for differential adaptation in plant breeding, and (iii) dissection of the genetic basis of adaptation and the evolutionary processes driving the adaptive genomic divergence (Huang & Han, 2014; Bevan & Uauy, 2013; Langridge & Fleury, 2011; Morrell *et al.*, 2011).

Both the wild ancestor (*Hordeum spontaneum* C. Koch, the wild barley) and the domesticated barely (*Hordeum vulgare* L.) have been intensively used as important model plants for the genetic and genomics of the Triticeae tribe and for ecological adaptation in efforts to dissect the genetic basis of adaptation. Barley is a source of gene pool for the characterization and

identification of important genes and genetic variation utilized for breeding (Munoz-Amatriain *et al.*, 2014; IBGS Consortium *et al.*, 2012; Nevo, 2006). Moreover, barley, which is the fourth most important cereal crop (FAOSTAT: http://faostat.fao.org/), and its wild ancestor can grow in very diverse environments. The large geographical distribution ranging from the desert to highland climate and its adaptation to diverse ecological habitats with multiple environmental stresses (Figure 1) make wild barley an ideal model plant to explore the genetic basis of adaptation in natural populations under selection and for characterization and identification of important genes and genetic variations.

## 1.1 Barley: botany, ecology and domestication

### 1.1.1 Botanical classification

The genus *Hordeum* belongs to the Triticeae tribe of the grass family (Poaceae or Gramineae) along with wheat and rye. *Hordeum* consists of more than 30 diploid (2n = 14) and polyploid (2n = 28 and 42) species in which *H. vulgare* is the only domesticated species in the genus. However, other species such as *H. spontaneum* and *H. bulbosum* are important genetic resources for breeding (Blattner *et al.*, 2010; Blattner, 2009; Linde-Laursen *et al.*, 2008). Wild barley is a highly self-fertile species (Abdel-Ghani *et al.*, 2004; Brown *et al.*, 1978) and fully interfertile with cultivated barley, whereas *H. bulbosum* is a self-incompatible and an obligate outcrossing perennial species (Lundqvist, 1962) but can be crossed with domesticated barley.

### 1.1.2 Ecological distribution

*H. spontaneum* is a plant with an extraordinary ecological distribution and adaptation. It can grow in all extreme environments such as in the desert, on saline and poor soils and mountainous places. The Fertile Crescent, region with cold rainy winter and dry summer, is the main wild barley distribution center (Zohary *et al.*, 2012). It covers parts of Israel, Lebanon, Jordan, Syria, South Turkey, Iraqi Kurdistan, and South-West Iran (Figure 1). The wild barley distribution extends further over the Mediterranean shore (of Egypt, Libya, Algeria and Morocco), North-East Iran, Central Asia, Turkmenia and Tibet. The hook-like structure (arrowhead shape) formed in matured and degenerated lateral spikelets (Figure 2D) can easily attach to animal coats and hence facilitate seed dispersal (Sakuma *et al.*, 2011).

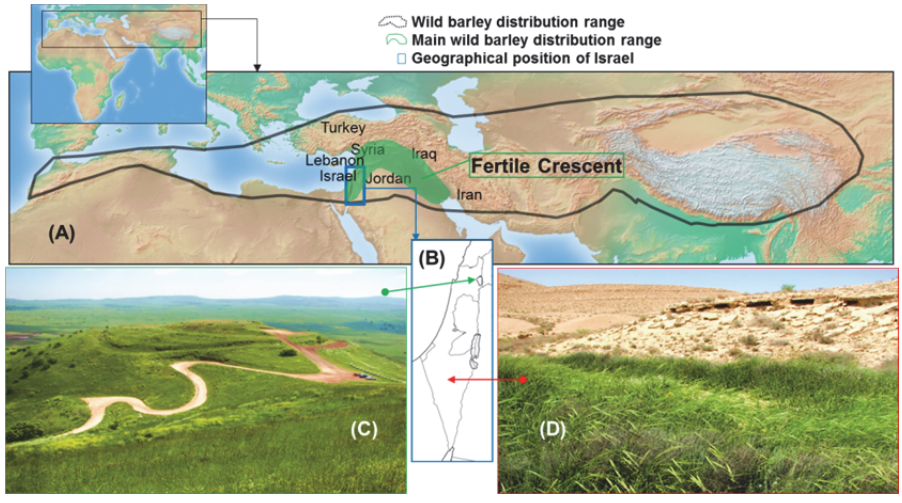


*Figure 1.* Geographical distribution of wild barley. (A) The wider wild barley distribution area with different ecological habitats (black broken line) ranging from the west Mediterranean shore of Morocco to Central Asia and Tajikistan, with the main distribution center in the Fertile Crescent (in green). (B) Map showing geographical areas of Israel, one of the wild barley centres of distribution and where the wild barley accessions used for this thesis were collected. Picture showing wild barley growing (C) in the northern part and (D) in the Negev desert in the southern part of Israel.

### 1.1.3 Domestication and diversification

Barley and wheat are the first domesticated cereals in the world. Cultivated barley was domesticated from its wild ancestor *H. spontaneum* at the early development of agriculture over 10,000 years ago in the Fertile Crescent (Zohary *et al.*, 2012; Pourkheirandish & Komatsuda, 2007). This is based on measurements of the radioactive $^{14}C$ isotope concentration in the remains of barley grain. To date, the wild ancestor is growing in its natural habitats in the Mediterranean area and in South-West Asia.

The available genetic evidence showed that barley has undergone a second domestication in the east of the Fertile Crescent, which served as a source of diversity in barley from Central Asia to the Far East (Morrell & Clegg, 2007; Saisho & Purugganan, 2007). The latest report based on high-throughput datasets indicated, however, that barley has a polyphyletic origin, with further domestication in Tibet (Dai *et al.*, 2014). The polyphyletic origin of barley domestication and diversification events is demonstrated in Figure 2.

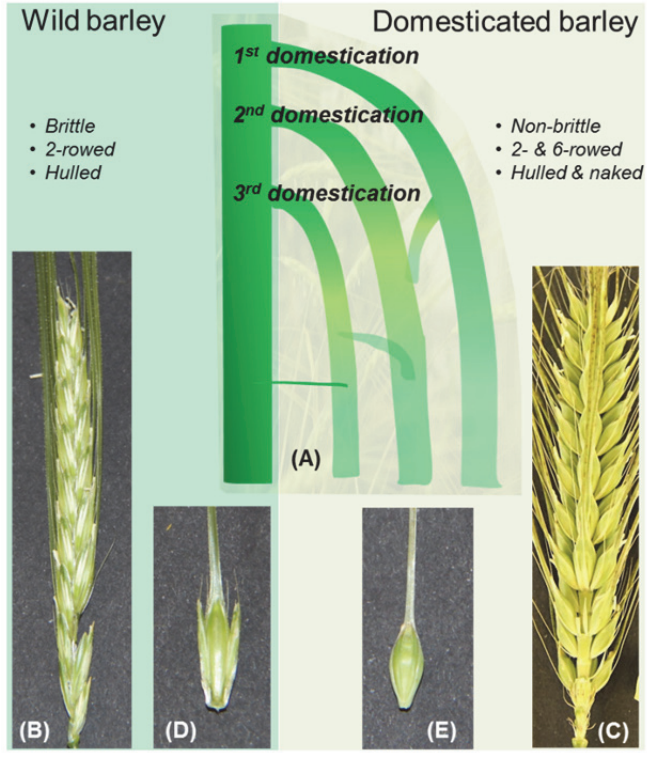


*Figure 2.* Domestication and diversification events of barley. Wild barley traits and pictures (B and D) are on the left panel (green shaded) and domesticated barley traits and pictures (C & E) on the right panel (light green shaded). (A) Diagram depicting the 1st, 2nd and 3rd domestication events in barley. (B) Two-rowed wild barley spike and (C) six-rowed domesticated barley spike. (D) Arrow-like spikelet of wild barley and (E) spikelet of domesticated barley.

Plant domestication involves selection of phenotypic traits that distinguish the cultivated plant from its wild ancestor, which is known as the 'domestication syndrome' (reviewed in Doebley *et al.*, 2006; Salamini *et al.*, 2002). Some traits have been further selected after the domestication events (post-domestication selection) during the expansion and adaptation of domesticated crop plants to different environmental climates, which is referred to as the diversification event (Meyer & Purugganan, 2013). Domestication and diversification related traits are controlled by single or multiple genes and affected by different types and levels of nucleotide and structural variations (Meyer & Purugganan, 2013; Olsen & Wendel, 2013). The three main domestication and diversification-related genes and traits that differentiate the wild and domesticated barley are described in Table 1.

Table 1. *Domestication and diversification traits, the responsible genes and variations in barley. Hs; H. spontaneum (wild barley), Hv; H. vulgare (domesticated barley), SNP; single-nucleotide polymorphism, InDel; insertion and/or deletion and SV; structural variation.*

| Trait | Gene | Phenotype | Mutation | Reference |
|---|---|---|---|---|
| Non-brittle rachis | *Btr1* & *Btr2* | Hs – brittle (*Btr1Btr2*)<br>Hv – nonbrittle (*Btr1btr2* or *btr1Btr2*) | | (Komatsuda *et al.*, 2004) |
| Row type | *VRS1* – HD-ZIP I (homeodomain-leucin zipper I-class) | Hs – 2-rowed (*Vrs1*)<br>Hv – 2- & 6-rowed (*vrs1*) | SNP, InDel and SV | (Komatsuda *et al.*, 2007) |
| Kernel type | *Nud* – ERF (Ethylene response factor) | Hs – covered (*Nud*)<br>Hv – covered (*Nud*) and naked (*nud*) | SV | (Taketa *et al.*, 2008) |

## 1.2 The barley genome

Barley (*H. spontaneum* and *H. vulgare*) is a diploid grass with seven chromosomes (2n = 14) and a large haploid genome of 5.1 gigabases (Gb) (IBGS Consortium *et al.*, 2012). Barley has the 3rd largest cereal genome after diploid bread wheat (17 Gb) and rye (8 Gb) (Bolger *et al.*, 2014), which makes its genome size 2.2x of the maize (2.3 Gb), 7.3x of the sorghum (0.7 Gb) and 13.1x of the rice (0.389 Gb) genomes (Bolger *et al.*, 2014; Bevan & Uauy, 2013). Through the International Barley Genome Sequencing (IBGS) Consortium, the draft barley genome was sequenced and released in May, 2012 using whole-genome shotgun (WGS), full-length complementary DNAs (fl-cDNA) and RNA sequencing data generated by Sanger and next-generation sequencing (NGS) approaches (IBGS Consortium *et al.*, 2012).

The IBGS Consortium estimated the barley genes to be 30,400. So far 26,159 (86%) of them have been identified as 'high-confidence' (HC) genes with homology support from other plant genomes from the total of 79,379 predicted transcript clusters. The rest 53,220 transcripts were categorized as 'low-confidence' (LC) genes without homology and gene family clustering. Based on RNA sequencing data (RNA-Seq) obtained from eight developmental stages, 72-84% of HC genes are expressed in more than one tissue or developmental stage, and 36-55% of them are differentially regulated among samples (IBGS Consortium *et al.*, 2012). Moreover, 73% of intron-containing HC genes showed alternative splicing in which majority of them are unique to the sample.

Genetically barley is a very diverse plant. The genome-wide comparisons of four barley cultivars and one wild barley accession against a reference cultivar 'Morex' uncovered over 15 million single-nucleotide variants (SNVs) in which

up to 350,000 SNVs are associated with exons (IBGS Consortium *et al.*, 2012). The genome survey revealed the presence of low genomic variation at centromeric and peri-centromeric regions of all chromosomes, particularly in cultivated barley, due to low recombination in these regions. Nonetheless, there is an intact genomic diversity in wild barley throughout the genome, which can serve as a source of genetic variation (IBGS Consortium *et al.*, 2012). Structural variations (SVs) due to copy number variations (CNVs) such as deletions, insertions and duplications of over 50 bp are also more prevalent in the barley genome (Munoz-Amatriain *et al.*, 2013). Higher CNVs across all chromosomes were found in the wild than in the cultivated barley.

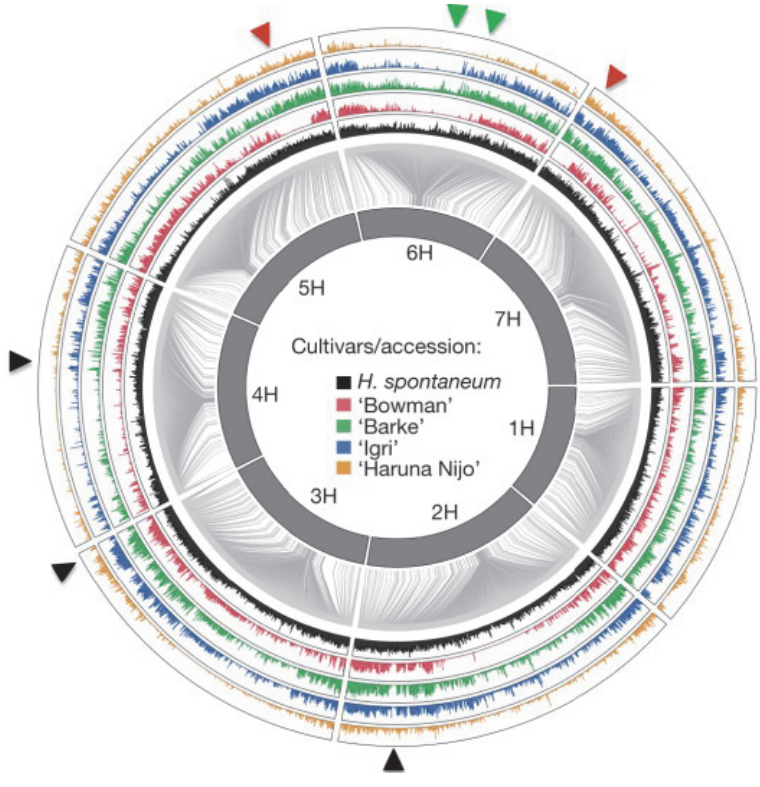


*Figure 3.* Distribution patterns of SNV in barley. Genome-wide frequency distribution of SNV per 50 kb in wild (inner black circular histograms) and cultivated barley (four external circular histograms) on all chromosomes (inner grey bars). The arrowheads show regions with deviated SNV frequency for the respective accession (adapted from IBGS Consortium *et al.*, 2012).

## 1.3 Adaptive genomic divergence in barley

Throughout their wider geographical distribution, both wild and domesticated barley exposed and adapted to multiple environmental factors where drought, salinity and high temperature are the main abiotic stresses. Wild barley successfully adapted to such highly diverse environments that differentiate over short to long geographical distances (Bedada *et al.*, 2014b, Paper I; Russell *et al.*, 2014; Hubner *et al.*, 2013; Hubner *et al.*, 2012; Fitzgerald *et al.*, 2011; Hubner *et al.*, 2009; Yang *et al.*, 2009).

As sessile organisms, plants have developed three different mechanisms to adapt to drought stress (reviewed in Juenger, 2013; Blum, 2011; Verslues & Juenger, 2011; Barnabas *et al.*, 2008). (1) Drought escaping – by undergoing early flowering and maturity, plants can escape the grain filling growth stage before the onset of seasonal drought. (2) Drought avoidance – reducing or avoiding dehydration and maintaining high water status despite exposure to water-deficit using different mechanisms such as stomatal closure to maintain turgor pressure. (3) Drought tolerance – tolerating dehydration and undergoing functional growth and development under low water status by accumulation of protective proteins such as late embryogenesis abundant, dehydrins and chaperons.

Wild barley adaptation to drought stress likely involves combinations of strategies. Water-use efficiency (WUE) is one of the physiological responses associated with drought stress response and it describes the association between carbon fixation to biomass (photosynthesis rate) and water loss (transpiration rate), and is expressed as their ratio. It indicates plant efficiency in biomass gain through photosynthesis (carbon assimilation) while minimizing water loss through transpiration and hence a commonly used parameter to evaluate plant adaptation potential to drought stress or water limited environments (Eppel *et al.*, 2013; Suprunova *et al.*, 2007). Analysis based on WUE of differentially adapted wild barley genotypes to drought stress led to the discovery of the barley dehydration-responsive *Hsdr4* gene (Suprunova *et al.*, 2007). The differential adaptation patterns to diverse environments thus make wild barley an ideal plant for analysis and identification of adaptive genes and genetic variants.

## 1.4 The molecular bases of genome divergence and evolution

The natural genomic divergence within and among populations and closely related species (for instance wild and domesticated barley) are responsible for differential patterns of plant adaptation to heterogeneous environments, and is an important genetic resource for crop plant improvement (Henry & Nevo,

2014; Munoz-Amatriain *et al.*, 2014; Morrell *et al.*, 2011; Alonso-Blanco *et al.*, 2009). These divergences arise from naturally occurring genetic changes or spontaneous mutations that are preserved by natural and artificial selections, and other neutral evolutionary processes.

Mutations arise in the genome cover genetic changes at: (i) the nucleotide scale – single-nucleotide variations (SNVs) and insertions and/or deletions (InDels) (Figure 4G and H); (ii) the gene scale – structural variations (SVs), which include copy number variations (CNVs) and present and/or absent variations (PAVs) (Figure 4A-F); and (iii) the chromosomal scale – such as large deletions and translocations (Marroni *et al.*, 2014; Rensing, 2014; Zmienko *et al.*, 2014; Alkan *et al.*, 2011; Innan & Kondrashov, 2010). The molecular mechanisms responsible for the creation of nucleotides and SVs and their functional impacts on the genotypic and phenotypic differentiations within barley and other plants are described in the following sections.

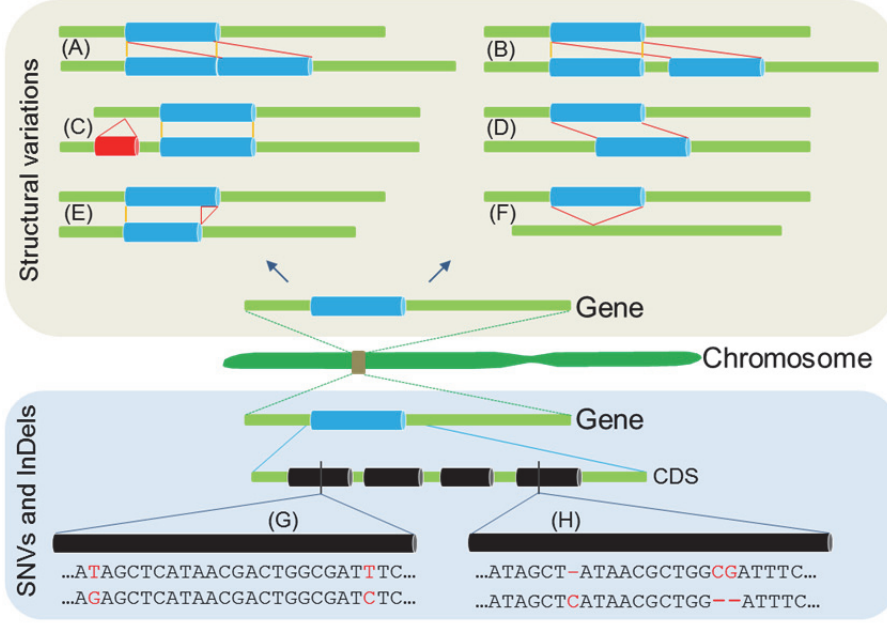


*Figure 4.* Genomic variations and their possible effects. Different possible types of structural variations affecting a gene are indicated in A-F; and SNVs and InDels affecting one to few nucleotides are indicated in G-H. Possible types of SVs: (A) tandem gene duplication, (B) interspersed gene duplication, (C) insertion of TE at regulatory region, (D) translocation of gene, (E) partial gene deletion and (F) complete gene deletion. Nucleotide variations in exonic regions: (G) SNVs and (H) insertion and deletion (InDels).

### 1.4.1 Nucleotide variations

*Single-nucleotide variation (SNV)*: SNV, single nucleotide polymorphism (SNP), is a substitution of a single base pair occurs during the DNA duplication process and/or by other external factors such as chemical substances and UV radiation. SNV occurs in coding regions of the genome can cause a silent/synonymous mutation (sSNP or sSNV), or non-silent/non-synonymous mutation (nsSNP or nsSNV) – a mutation that results in amino acid change. Phenotype causing SNVs can alter the existing gene structures through either frame shift mutation or alternative splicing and thereby change the function of the gene.

*Small insertion/deletion (InDel)*: InDel of one or more nucleotides (usually under 50 bp), which arises due to an error during the DNA duplication process, contributes to plant genomic and phenotypic divergence. InDel mutations in the coding regions of the genome can affect the protein coding reading frame in different ways. InDels of multiple of three nucleotides in the coding regions affect the length of the protein sequence without affecting the reading frame of the original protein. InDels involving one or two nucleotides, however, disturb the reading frame and cause frame shift mutation, which can further lead to the creation of a new gene structure and function that can potentially cause genomic and phenotypic divergence.

*Functional impact of SNV and InDel*: SNV and InDel mutations can cause genomic and phenotypic divergence within and among wild and cultivated plants (reviewed in Meyer & Purugganan, 2013; Olsen & Wendel, 2013; Alonso-Blanco *et al.*, 2009). In barley, SNVs and InDels have affected several domestication and diversification genes and genes controlling agronomically important traits. A single nsSNP at the coding region of *uzu (BRI1)* (Chono *et al.*, 2003) and the intronic region of *sdw1* (Jia *et al.*, 2009) genes cause dwarfed barley plants. Similarly, a single sSNP at the exonic regions of cleistogamous *Cly1* gene, a region targeted by microRNA (miR172), results in cleistogamous flower – a flower that sheds its pollen before opening (Nair *et al.*, 2010). SNPs and InDels at the exonic regions of *VRS1* gene have led to the creation of six-rowed barley (Komatsuda *et al.*, 2007). Nucleotide variations at the coding region of *Ppd-H1* (Turner *et al.*, 2005) and *Vrn-H3* (Yan *et al.*, 2006) genes have caused late-flowering barley phenotypes. Differential adaptation among winter and spring barley types are also due to a single nsSNP at the *Antirrhinum Centroradialis HvCen* gene (Comadran *et al.*, 2012).

The aforementioned and other similar results therefore clearly demonstrate the significant contributions of SNVs and InDels in creating genomic and

phenotypic differentiations among and within wild and domesticated barley. Further analysis and characterization of the diverse cultivated and wild barley gene pools through different genomic approaches is therefore a vital strategy to uncover more beneficial variations.

### 1.4.2 Structural variations

Structural variations (SVs) are another major source of genomic and phenotypic divergence in crop plants (reviewed in Marroni *et al.*, 2014; Rensing, 2014; Zmienko *et al.*, 2014). SVs were initially considered as insertions, deletions, inversions, duplications and translocations of DNA segments over 1 kb, but now redefined as genomic rearrangements covering over 50 bp DNA sequence (Munoz-Amatriain *et al.*, 2013; Alkan *et al.*, 2011). SVs can be categorized as: (i) CNVs – duplications, deletions and insertions of sequences that lead to the occurrence of different sequence copy number among individual genomes; and (ii) PAVs – the presence of sequences in one but complete absent in another individual genome within a species (Marroni *et al.*, 2014; Saxena *et al.*, 2014; Olsen & Wendel, 2013).

Different mechanisms are responsible for the creation of SVs. This includes nonallelic homologous recombination, nonhomologous end joining and transposable element (TE) dynamics (reviewed in Bickhart & Liu, 2014; Chen *et al.*, 2013; Long *et al.*, 2013; Kaessmann *et al.*, 2009; Conrad & Hurles, 2007). Nonallelic homologous recombination is TE-mediated large genome rearrangements in which nonallelic homologous recombination (unequal crossing-over) occurs among: (i) direct repeats leading to deletions and duplications, (ii) inverted repeats causing inversions, and (iii) repeats on different chromosomes resulting in translocations. Nonhomologous end joining mechanism involves ligation of the ends of two double stranded breaks in the DNA sequence.

Gene duplication contributes to genome complexity and phenotype diversity through creation of new genes, and gene structures and functions (Chen *et al.*, 2013; Long *et al.*, 2013; Conant & Wolfe, 2008). Large proportion of duplicated genes are erased due to the accumulation of deleterious mutations, while the remaining few proportion can be retained and become nonfunctional (pseudogenized or silenced), acquire a novel function (neofunctionalization), or divide the original function (subfunctionalization) (Rensing, 2014; Chen *et al.*, 2013; Long *et al.*, 2013; Carretero-Paulet & Fares, 2012; Rutter *et al.*, 2012; Kaessmann, 2010; Conant & Wolfe, 2008).

*Functional impact of SVs*: Unlike ample documented studies on the functional impacts of SVs in different model organisms, little is known about their

impacts in plants. The advent of NGS technologies is, however, uncovering the functional contributions of SVs to plant genomic and phenotypic divergence (Marroni *et al.*, 2014; Saxena *et al.*, 2014; Zmienko *et al.*, 2014; reviews therein). SVs can have profound effects on plant genome structure and complexity, and gene expression and function. These effects include complete duplication or deletion of a gene, deletion and/or duplication of exonic or enhancer region, or insertion of transposable elements in the regulatory or coding region of a gene (Figure 4A-F).

In barley, the genome-wide analysis revealed the presence of high SVs (Munoz-Amatriain *et al.*, 2013; Matsumoto *et al.*, 2011). Recently, it has been revealed that the barley *VRS1* gene responsible for row-types (Komatsuda *et al.*, 2007) is the outcome of a duplication and neofunctionalization process (Sakuma *et al.*, 2013). That means that, *VRS1* is a duplicate of the *HvHox2* gene, which is conserved among cereals. Similarly, duplication of the boron transporter *Bot1* gene coding for the boron efflux transporter causes boron-toxicity tolerance in an African barley landrace from Algeria (Sutton *et al.*, 2007). The tolerant barley genotype has four-times higher number of copies of the *Bot1* gene than the intolerant genotypes. More transcripts provide tolerance by enhancing boron efflux transporter activity and capacity. Insertion of a 1-kb sequence in the upstream of the barley aluminum-activated citrate transporter1 *HvAACT1* gene encoding for the citrate transporter causes aluminum (Al) toxicity tolerance (Fujii *et al.*, 2012). Al-tolerant cultivars have higher expression of the *HvAACT1* gene, which is enhanced by a 1-kb insertion (Fujii *et al.*, 2012). The recent large-scale array-based comparative genome hybridization (CGH) study (Munoz-Amatriain *et al.*, 2013) further shows the prevalence and patterns of SVs in wild and domesticated barley in that 9.5% of the coding sequences represented on the array showed CNVs, 41.8% exon-affecting CNVs are only present in wild barley, and stress and resistance genes such as nucleotide-binding site leucine-rich repeat (NBS-LRR) and resistance *(R)* genes are affected by CNV. The above studies therefore clearly indicate the significant contributions of SVs to phenotypic and genomic divergence within and among wild and cultivated barley adapted to different environments.

## 1.5 Transposable elements dynamics and genomic divergence

Transposable elements (TEs) are DNA sequences that are capable to move around and integrate into new positions in the genome (Wicker *et al.*, 2007). They were initially discovered in maize DNA by Barbara McClintock in 1956 (McClintock, 1956). In the past, they were described as “Junk” DNA or genomic parasite and selfish genes (Doolittle & Sapienza, 1980; Orgel *et al.*,

1980). Nonetheless, now-a-days due to their significant contributions to the evolution and adaptation of organisms, they are considered as key players in reshaping the genome (reviewed in Bennetzen & Wang, 2014; Bonchev & Parisod, 2013; Casacuberta & Gonzalez, 2013; Lisch, 2013; Rebollo *et al.*, 2012).

Based on their transposition mechanisms (i.e., the presence or absence of an RNA transposition intermediate), TEs are generally grouped into two major classes: Class I and Class II elements (Wicker *et al.*, 2007). Class I elements or retrotransposons transpose through a 'copy-and-paste' mechanism (Figure 5A) via a reverse-transcribed RNA intermediate to integrate into a new position in the genome by an integrase enzyme. Class II elements or DNA transposons transpose through a 'cut-and-paste' mechanism using TE encoded transposase enzyme. Class I elements are further classified as long terminal repeat (LTR) elements and non-LTR elements in which they differ in the presence/absence of LTR and their internal structural domains (Figure 5B).

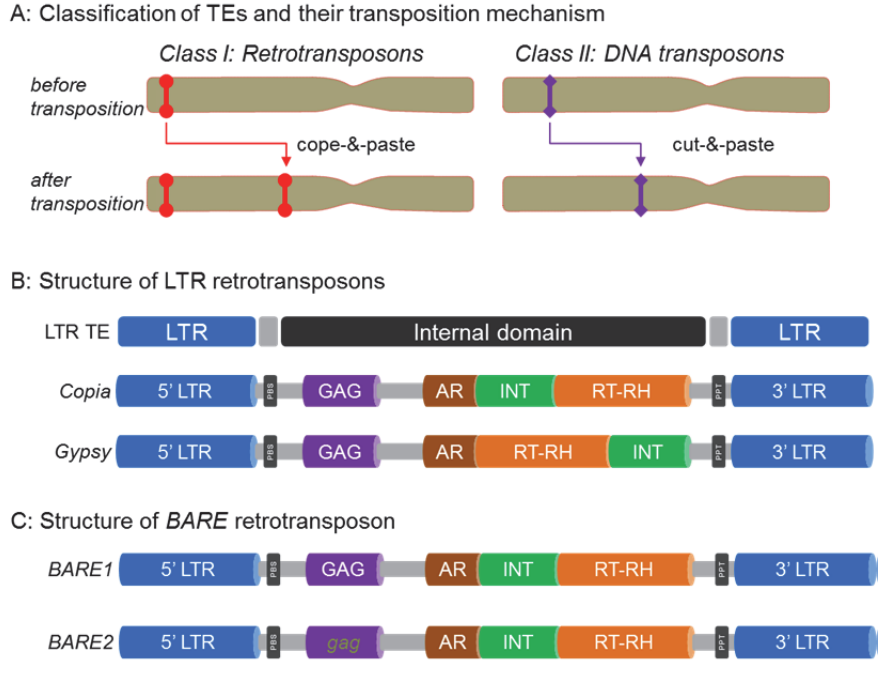


*Figure 5.* Classification of TEs and structure of LTR retrotransposons. (A) The two classes of TEs and their transposition mechanisms. (B) Structure of LTR retrotransposons showing the difference in the arrangement of the internal domains between *Copia* and *Gypsy* superfamilies. (C) Structure of *BARE1* and *BARE2* showing the inactive GAG domain of *BARE2* due to mutation. Figures B and C are based on information in Schulman (2012).

### 1.5.1 Transposable elements in barley

TEs constitute a significant proportion of the plant genome (Vitte *et al.*, 2014), which ranges from 10% in *Arabidopsis thaliana* (Arabidopsis Genome Initiative, 2000) to 85% in maize (Schnable *et al.*, 2009). In barley, TEs constitute about 84% of the genome, with the majority belongs to retrotransposons (IBGS Consortium *et al.*, 2012). Almost all retrotransposons are LTR elements where the *Gypsy* transposons superfamily is the most abundant elements followed by the *Copia* superfamily (Mazaheri *et al.*, 2014; IBGS Consortium *et al.*, 2012; Wicker *et al.*, 2009).

The LTR retroelements particularly occupy the pericentromeric and centromeric regions of the barley chromosomes where the gene density is low, whereas the DNA transposons are abundant in the gene-rich regions (IBGS Consortium *et al.*, 2012). These are the commonly observed chromosomal distribution patterns of TEs in plants (Kejnovsky *et al.*, 2012) in that LTR retroelements occupy the heterochromatic regions – highly condensed, gene-poor and transcriptionally silent regions, whereas the DNA transposons are commonly found in the euchromatic regions – less condensed, gene-rich and transcriptionally accessible regions. The insertion preference and abundance of TEs in the gene-poor pericentromeric and heterochromatic regions with no or low recombination is associated with less deleterious effects of TE insertions at these regions (Kejnovsky *et al.*, 2012).

Barley retroelement 1 (*BARE1*) from *Copia* superfamily (Manninen & Schulman, 1993) is the most abundant type of TEs constituting over 10% of the barley genome, with full-length insert alone constituting 2.9% (Middleton *et al.*, 2012; Wicker *et al.*, 2009; Soleimani *et al.*, 2006; Vicient *et al.*, 1999), followed by *Sabrina* (~8%) form *Gypsy* superfamily. The *BARE* family has three different members: *BARE1* – fully autonomous, *BARE2* – non-autonomous type that depends on *BARE1* and *BARE3* – similar to wheat *WISE*-2 retroelement (Vicient *et al.*, 2005). *BARE1* is the first described *Copia* retrotransposons that is expressed and inherited from generation to generation (Chang *et al.*, 2013; Jaaskelainen *et al.*, 2013; Jaaskelainen *et al.*, 1999). The autonomous *BARE1* life cycle is maintained through its structure composed of LTRs and protein coding internal domains consisting of capsid protein (GAG), aspartic proteinase (AP), integrase (IN), reverse transcriptase and RNase H (RT-RH) (Schulman, 2012; Wicker *et al.*, 2007; Vicient *et al.*, 1999). Unlike *BARE1*, the GAG domain of *BARE2* is transcriptionally inactive due to the deletion of the first codon of the *gag* open reading frame (ORF) through mutation (Tanskanen *et al.*, 2007; Vicient *et al.*, 2005).

The *BARE1* domains are arranged as LTR-GAG-AP-IN-RT-RH-LTR (Figure 5C) and responsible for the transposition process that involves

transcription, translation, packaging, reverse transcription and integration into the genome (Schulman, 2012). The transcription starts with the promoter that resides at 5' LTR and terminates and polyadenylates with a signal provided from 3'LTR. Transcribed RNA is translated into either a separate GAG and pol ORFs or a GAG and polyprotein (Schulman, 2013; Schulman, 2012).

### 1.5.2 Transposable elements drive genomic divergence

TEs dynamics are one of the evolutionary forces that generates genome complexity and variability in plants (Bennetzen & Wang, 2014; Marroni *et al.*, 2014; Mirouze & Vitte, 2014; Buchmann *et al.*, 2013; Lisch, 2013; Slotkin *et al.*, 2012; Morgante *et al.*, 2007). Transposons movements throughout the genome generate different types and levels of structural variations that can lead to genome rearrangement and changes in genome size (Bennetzen & Wang, 2014; Vitte *et al.*, 2014). Moreover, their dynamics can alter gene expression or function by creating novel features, disrupting regulatory (enhancer and promoter) or coding regions of the gene, or through epigenetic mechanisms (reviewed in Marroni *et al.*, 2014; Mirouze & Vitte, 2014; Lisch, 2013; Slotkin *et al.*, 2012).

Little is known about the functional impacts of TEs in Triticeae species. There are, however, several evidences from other plants that TEs dynamics regulate the genes. These regulation can have adaptive, neutral or deleterious impacts on the plant fitness (reviewed in Bennetzen & Wang, 2014; Vitte *et al.*, 2014; Lisch, 2013). The functional impacts of TEs are associated with the regulatory information found in TEs. *BARE1* carries a promoter of abscisic acid (ABA)-responsive element (Suoniemi *et al.*, 1996), which is similar to the promoter of stress-responsive genes. Drought stress induces the expression of *BARE1* GAG protein (Jaaskelainen *et al.*, 2013). An association of *BARE1* diversity with stress has been documented in natural wild barley populations from the 'Evolution canyon' (EC). Population from the drier and rocky South-facing slope (SFS) of the EC1 in Israel has higher *BARE1* copy number than a population from the moist North-facing slope (NFS), indicating the adaptive role of *BARE1* under stress (Kalendar *et al.*, 2000). The promoter region of the barley dehydration-responsive *Hvdr4* gene (Suprunova *et al.*, 2007) responsible for dehydration stress tolerance contains a miniature inverted-repeat transposable element (MITE), a DNA transposon capable to form a hair-pin-like secondary structure. Several nucleotide variations were observed at the MITE insertion region among stress tolerant and sensitive genotypes, which suggested to cause different folding patterns in the tolerant and sensitive genotypes, and hence potentially leading to different patterns of adaptation to drought stress (Suprunova *et al.*, 2007).

## 1.6 Evolutionary processes driving genomic divergence

### *1.6.1* Domestication and diversification

Plant domestication and diversification processes, which involved rigorous conscious and unconscious selection events, led to profound genetic and phenotypic changes (reviewed in Meyer & Purugganan, 2013; Olsen & Wendel, 2013; Meyer *et al.*, 2012; Sakuma *et al.*, 2011; Pourkheirandish & Komatsuda, 2007; Doebley *et al.*, 2006). Domestication and diversification caused genome-wide loss of both neutral and adaptive genetic variations in domesticated barley through genetic bottlenecks.

### *1.6.2* Adaptive selection

Plants adapted to divergent and heterogeneous environmental habitats are facing continuous natural selection pressures that lead to adaptive genomic and phenotypic divergence among individuals and populations (Franks & Hoffmann, 2012; Schoville *et al.*, 2012). Adaptive genomic divergence is caused by natural selection. It primarily acts on loci under divergent selection and its effect further spill over to those tightly linked neutral loci through genetic hitchhiking and thereby affecting their allele frequency (Vitti *et al.*, 2013; Franks & Hoffmann, 2012; Schoville *et al.*, 2012; Hohenlohe *et al.*, 2010; Nosil *et al.*, 2009).

Adaptive genomic divergence can be due to novel mutation (hard selective sweep) or standing genetic variation (soft selective sweep) (reviewed in Hendry, 2013; Messer & Petrov, 2013; Vitti *et al.*, 2013; Franks & Hoffmann, 2012; Schoville *et al.*, 2012; Hohenlohe *et al.*, 2010; Nosil *et al.*, 2009; Barrett & Schluter, 2008). Adaptation from standing genetic variation or soft sweeps occurs when variation already present in a population as neutral or deleterious variant or introduced through gene flow is favored and increased in frequency following natural selection. Soft sweep involves either single or multiple genes (alleles) (Pritchard & Di Rienzo, 2010). Unlike soft sweep, hard selective sweep occurs when a single novel mutation appears in the population is favored and swept to high frequency and thereby causing adaptive genomic divergence. Several studies on both plants and animals are indicating that standing genetic variation is mainly responsible for the adaptive genomic divergence (reviewed in Messer & Petrov, 2013; Schoville *et al.*, 2012; Hohenlohe *et al.*, 2010; Pritchard & Di Rienzo, 2010).

In barley, both soft and hard sweeps have contributed to adaptive genomic divergence. For instance, boron toxicity resistance (Sutton *et al.*, 2007) is due to an increase in the frequency of copy number. Similarly, differential adaptation among winter and spring barley cultivars is due to an increase in the

frequency of the preexisting genetic variant (nsSNP) at the *HvCEN* gene (Comadran *et al.*, 2012). On the other hand, barley tolerance to acidic soil (aluminum toxicity tolerance) is due to the novel insertion of a 1-kb sequence upstream of the aluminum-activated citrate transporter *HvAACT1* gene (Fujii *et al.*, 2012). Adaptive genomic variation can also arise from hybridization among wild and cultivated crop plants (Ellstrand *et al.*, 2013; Hufford *et al.*, 2013; Stapley *et al.*, 2010), which is a soft selection sweep supplied by gene flow. For instance, gene flow among wild and domesticated barley was suggested as a source of adaptive variation observed in the domesticated barley (Russell *et al.*, 2011). Similarly, the introgression of adaptive alleles from the wild relatives of maize, *Zea mays* ssp. *mexicana*, into cultivated maize improved the adaptation of maize to the highland environment (Hufford *et al.*, 2013). Local adaptation in *Arabidopsis thaliana* is also caused by both standing genetic variation (Fournier-Level *et al.*, 2011) and novel mutations (Hancock *et al.*, 2011).

### 1.6.3 Neutral evolutionary processes

Like natural selection, the neutral evolutionary processes (non-selective forces) are not causing adaptive patterns of genomic divergence among individuals and populations, but can influence the adaptive genomic variation by increasing or decreasing the genome-wide level of diversity. These neutral driving forces include gene flow, isolation by dispersal limitation (IBDL) and demographic processes (reviewed in Schoville *et al.*, 2012; Hohenlohe *et al.*, 2010; Suzuki, 2010). Natural selection causes locus-specific adaptive divergence, while the neutral evolutionary processes have genome-wide effects.

Gene flow is a homogenizing evolutionary force that acts uniformly throughout the genome (Aitken & Whitlock, 2013; Orsini *et al.*, 2013; Savolainen *et al.*, 2013; Schoville *et al.*, 2012; Via, 2012; Nosil *et al.*, 2009). Hence, gene flow can either counteract or enhance adaptive genomic divergence based on different factors (Anderson *et al.*, 2010). This includes the environmental habitats of the differentiating populations, the level of gene flow itself, and the strength of adaptive selection. Strong adaptive selection can reduce gene flow among populations adapted to ecologically divergent environments since immigrants from different environments can poorly establish and adapt to the new contrasting environment (Orsini *et al.*, 2013; Nosil *et al.*, 2009; Jump & Peñuelas, 2005). Unless there is strong or equivalent level of natural selection, strong gene flow can also reduce or remove the adaptive divergence. In contrary, gene flow can enhance adaptive divergence by introducing adaptive variation, which is the case of adaptation

from standing genetic variation (Hufford *et al.*, 2013; Schoville *et al.*, 2012; Russell *et al.*, 2011).

Barley is predominantly self-fertile. Gene flow over longer geographical distances is thus mainly through seed dispersal, while both seed and pollen dispersals are responsible for gene flow over shorter distances (Volis *et al.*, 2010). Gene flow among wild barley populations adapted to different environments over both micro- and macro-environmental gradients have been documented in different studies (Bedada *et al.*, 2014b, Paper I; Hubner *et al.*, 2013; Hubner *et al.*, 2012; Russell *et al.*, 2011; Volis *et al.*, 2010; Hubner *et al.*, 2009; Morrell *et al.*, 2003). Volis *et al.* (2010) observed that the level of gene flow within wild barley populations varies in different environmental habitats. Even though several studies are indicating the presence of gene flow, little is known about how gene flow is shaping adaptive genomic divergence in barley. However, the gene flow most likely plays a significant role in shaping the genomic divergence in wild and domesticated barley. Further investigations using genome-wide analysis of systematically collected large number of wild barley accessions from its distribution range and domesticated barley genotypes growing around the collection sites of wild barley is therefore required to dissect how gene flow is shaping the neutral and adaptive genomic divergence over shorter and longer geographical distances with diverse environmental habitats.

Genomic divergence among individuals and populations can also arise due to isolation by dispersal limitation (IBDL), a neutral driving force that can lead to isolation by distance (IBD) pattern of genomic variation (Orsini *et al.*, 2013; Slatkin, 1993). This pattern of genomic divergence can occur when there is no adaptive selection and the gene flow among populations is reduced with increasing geographical distance (Orsini *et al.*, 2013). Nonetheless, when IBD is coupled with strong adaptive selection, both neutral and adaptive genomic variations occur (Orsini *et al.*, 2013; Nosil *et al.*, 2009). IBD pattern of genomic divergence is a commonly observed pattern of variation in wild barley over short to long geographical distances (Bedada *et al.*, 2014b, Paper I; Fang *et al.*, 2014; Russell *et al.*, 2014; Hubner *et al.*, 2013; Hubner *et al.*, 2012; Hubner *et al.*, 2009).

Demographic processes such as change in population size due to bottleneck, expansion, admixture and colonization can also affect patterns of genomic divergence within and among populations. Population bottleneck, which can be caused by different factors such as domestication, can cause genome-wide loss of genetic diversity and hence reduce the adaptive variation, whereas an expanding population can enhance adaptive genomic divergence by sweeping

favored variants to high frequency and fixation (Hohenlohe *et al.*, 2010; Suzuki, 2010).

## 1.7 Approaches for analysis of adaptive genomic divergence

### 1.7.1 Experimental and genomic approaches

*Experimental approaches*

Adaptive genomic variation can be investigated using different experimental methods applied to either natural populations adapted to heterogeneous and contrasting environments or experimental populations derived from crossing of differentially adapted populations (Franks & Hoffmann, 2012; Anderson *et al.*, 2011).

There are four commonly applied experimental approaches (Merila & Hendry, 2014; Franks & Hoffmann, 2012; Anderson *et al.*, 2011) used to infer the genetic basis of adaptive variation. (1) Common-garden experiment – an approach applied by growing populations collected from different environments under common laboratory or field conditions to identify adaptive variations among populations. (2) Reciprocal transplant experiment – this method is implemented by reciprocal transplanting of populations from different climates between environments to investigate the adaptive fitness or fitness advantage of populations at their native and foreign environments. (3) Individuals- or population-based experiment – an approach implemented using individuals or populations collected from contrasting environments to analyze the adaptive genomic divergence. (4) Qualitative trait loci (QTL) mapping approach – this method is applied by generating mapping populations from differentially adapted parental populations or individuals to identify genomic regions associated with divergence or adaptation.

Different approaches have their own merits and disadvantages, but selection of the appropriate method is based on different factors such as the number of individuals or populations included in the experiment and type of applied genomic approach for genotyping of the samples. For adaptive selection analysis of few individuals or populations using high-throughput data, the first three approaches can be implemented.

*Genomic approaches*

Adaptive genomic analysis can be performed using either both phenotypic and genomic data or only genomic data generated from different populations. Genomic data can be generated from few to multiple genomic regions for genome-scan based divergence analysis or from the whole genome to perform NGS based analysis of genomic divergence for identification of signature of

adaptive selection and thereby uncover genes and variations associated with and responsible for adaptation (Vitti *et al.*, 2013; Franks & Hoffmann, 2012; Ekblom & Galindo, 2011; Stapley *et al.*, 2010).

*(1) Genome-scans*

The genome-scans have been widely used to perform a genomic survey and compare patterns of genetic variation within and among populations and thereby identify candidate loci (outlier loci) associated with adaptation (Schoville *et al.*, 2012; Strasburg *et al.*, 2012; Coop *et al.*, 2010; Nosil *et al.*, 2009; Foll & Gaggiotti, 2008; Nielsen *et al.*, 2005). The outlier loci can be detected by analyzing highly differentiating allele frequency among populations (*Fst*), and allele frequency strongly associated with differentiating environments.

Genome-scan approach has been a method of choice particularly before the NGS technologies were widely accessible due to high costs and other aspects. Hence, it has been applied to scan and analyze patterns of genomic variation in wild and domesticated barley populations and different evolutionary processes driving local adaptation and genomic differentiation (Bedada *et al.*, 2014b, Paper I; Comadran *et al.*, 2012; Hofinger *et al.*, 2011; Russell *et al.*, 2011; Yang *et al.*, 2011; Hubner *et al.*, 2009; Yang *et al.*, 2009; Jilal *et al.*, 2008; Cronin *et al.*, 2007; Baek *et al.*, 2003; Morrell *et al.*, 2003; Volis *et al.*, 2001). Nonetheless, this method has some limitations such as poor resolution to identify gene and genetic variation responsible for adaptation (Strasburg *et al.*, 2012; Narum & Hess, 2011).

*(2) NGS approach*

NGS based analysis is another highly informative and comprehensive approach for genome-wide analysis of genomic divergence (Kiani *et al.*, 2013; Morey *et al.*, 2013). Unlike genome-scans, the NGS approach has high resolution to perform a genome-wide scanning for identification of signature of adaptive divergence among populations and the responsible genes and genetic variations (Morey *et al.*, 2013; Vitti *et al.*, 2013; Stapley *et al.*, 2010). This approach can be implemented using different sequencing techniques, but I will here only focus on the basic and other techniques that we have applied in this thesis work. Other approaches such as reduced sequencing representation methods and genome-wide association studies (GWAS) are not covered.

***Whole genome (re)-sequencing (WGS)***: a comprehensive sequencing approach for the analysis of the genomic divergence among individuals across an entire genome. This method helps to identify all types of variations in both coding

and non-coding regions of the genome (IBGS Consortium *et al.*, 2012) and to discover novel transcripts or genes by resequencing (Lai *et al.*, 2010). WGS can therefore uncover genome-wide adaptive genetic variations. Efficient utilization of WGS data, however, requires a reference genome. Furthermore, resequencing approach is not a method of choice for analysis of large samples or populations, particularly in plants like barley with large repetitive-rich genome, though it can be applied as low-coverage resequencing of pooled samples for medium-sized genome such as rice (He *et al.*, 2011). So far, this method has not been used for adaptive genomic analysis in barley.

***Transcriptome sequencing***: a sequencing of reverse transcribed mRNA (cDNA) from the whole genome is another approach for interrogation of transcriptome divergence within and among populations. RNA-Seq analysis can be done on transcriptome library that is either unnormalized to perform differential gene expression analysis or normalized to analyze many transcripts and thereby uncover novel genes and variants (Ekblom *et al.*, 2012; Ekblom & Galindo, 2011).

Transcriptome-based analysis of adaptive divergence is therefore an informative method particularly to analyze non-model organisms without a reference genome, but it has some limitations (Hirsch *et al.*, 2014; Franks & Hoffmann, 2012; Good, 2011). First, adaptive genomic divergence maybe due to novel or standing genetic variations that is not linked to differential gene expression. RNA-Seq is therefore less informative to identify such adaptive variants. Second, an adaptive divergence maybe associated with tissue-specific and/or developmental stage-specific differentially expressed genes and gene networks. It is therefore unlikely to identify such candidate genes and genetic variations from unrepresentative libraries sampled at different stages or from different tissues.

The RNA-Seq approach has been broadly used to analyze genomic divergence in many cereals (Kiani *et al.*, 2013). In barley, transcriptome sequencing has been used for different studies such as discovery of growth stage and tissue-specific novel transcripts (IBGS Consortium *et al.*, 2012; Thiel *et al.*, 2012) and for identification of differentially expressed genes among drought sensitive and tolerant wild barley ecotypes (Hubner S. *et al*., in preparation). We have also implemented normalized transcriptome sequencing for identification of novel transcripts and SNPs from two differentially adapted wild barley ecotypes under drought stress (Bedada *et al.*, 2014a, Paper II).

***Targeted capture sequencing:*** is a method performed by enriching and sequencing of targeted genomic regions or genes of interest to reduce the

complexity of the genome (Blumenstiel *et al.*, 2010; Gnirke *et al.*, 2009; Hodges *et al.*, 2007) in large individuals or populations. Hence, it is a cost-effective and powerful approach to investigate the adaptive genomic divergence at coding regions of the genome, selected candidate genes or targeted genomic regions in large samples (Andrews & Luikart, 2014; Kiani *et al.*, 2013; Good, 2011),

Targeted or exome sequence capture approach has been successfully applied in different crop plants (as reviwed in Saxena *et al.*, 2014; Kiani *et al.*, 2013). Recently, barley whole exome capture was successfully developed, and the data has been used for phylogenetic-based analysis of genomic divergence within and among wild and domesticated barley (Mascher *et al.*, 2013). Exome or targeted capture is therefore a method of choice for population-based analysis of adaptive genomic divergence. We applied this method to analyze patterns of adaptive divergence at randomly selected and candidate genes (Manuscript III) and to investigate the patterns of *BARE* TE insertions (manuscript IV) in large wild and domesticated barley populations.

### 1.7.2 Bioinformatics techniques

Bioinformatics is a core base for the analysis of high-throughput genomic data and dissection of the genetic basis of divergence within and among populations. Bioinformatic techniques can be applied as a series of pipelines and workflows involving different bioinformatic tools to process and analyze NGS datasets generated from individuals or pool of individuals (Pool-seq). NGS-based analysis of genomic divergence generally involves three main steps (Figure 6).

(1) *NGS data generation, processing and quality control*
The raw NGS data generated from individuals or populations is processed and quality controlled by removing or trimming barcodes (indexes), adapters, primers, poor-quality reads and nucleotides, and further visualized and inspected for the quality parameters. (Guo *et al.*, 2013; Wolf, 2013; Ekblom & Galindo, 2011; Martin & Wang, 2011). This is therefore a critical step that can affect the downstream analysis and the biological conclusions drawn from the data.

(2) *Reads mapping or assembly*
Well-processed and quality controlled raw reads are either *de novo* assembled or mapped onto the reference genome or sequence. Accurate *de novo* or reference-based assembly depends on several factors such as the quality of the reads, the quality of the reference genome or sequences and the alignment tool

and implemented parameters (Wolf, 2013; Lee *et al.*, 2012; Martin & Wang, 2011; Nielsen *et al.*, 2011; Zhang *et al.*, 2011; Kumar & Blaxter, 2010).

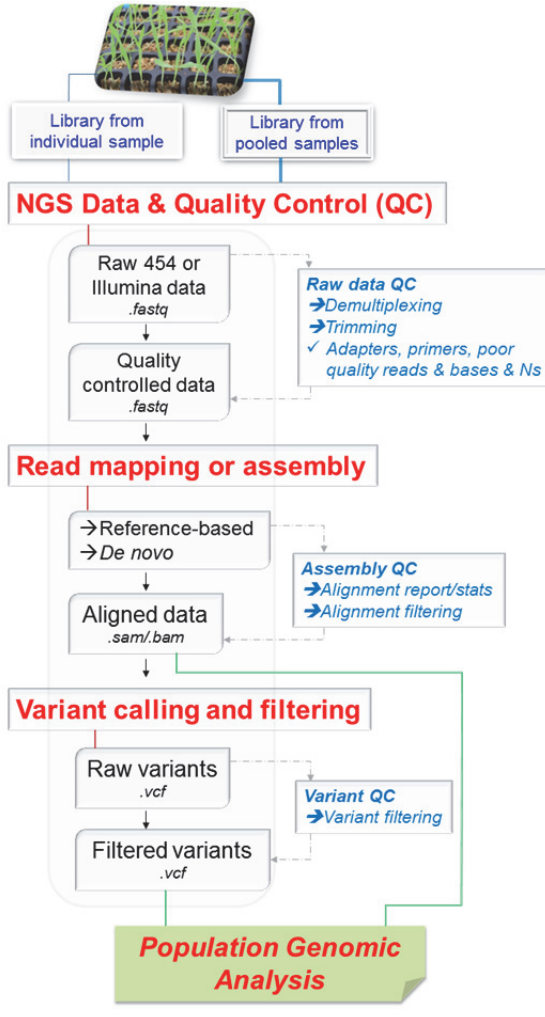


*Figure 6.* Workflow for analysis of NGS data. NGS datasets generated from individuals or pooled samples are processed and quality controlled at different stages for assembly and identification of nucleotide variations that can be used for different population genomic analysis.

(3) *Variant calling and filtering*

Efficient mapping of reads leads to the discovery of high-quality genomic variations. Identification of high-quality variants involves variants calling, filtering and discovery. Variants identification is therefore affected by factors such as quality of the reference sequence, nature and source of sequencing

date, depth of coverage at variant site and alignment and variant calling algorithms (Guo *et al.*, 2013; Kiani *et al.*, 2013; Lee *et al.*, 2012; Martin & Wang, 2011; Nielsen *et al.*, 2011).

There are four different bioinformatic methods that are implemented in different mapping and variant calling programs for the identification of genomic divergence among individuals and populations using NGS data (Marroni *et al.*, 2014; Alkan *et al.*, 2011; Nielsen *et al.*, 2011). These are (1) read depth, (2) split read mapping, (3) pair-end (PE) mapping and (4) sequence assembly.

PE-based analysis of genomic variations is a powerful approach for detection of all types of variants based on the distance among and/or the orientation of PE reads mapped onto the reference sequence. The presence of InDels or SVs leads to the deviation from the expected distance and/or anticipated orientation among PE reads. This approach is also used for the detection of novel TE insertions (Zhuang *et al.*, 2014; Keane *et al.*, 2013; Kofler *et al.*, 2012). This is based on the principle that novel TE insertions cause only one of the PE reads to be mapped to the reference sequence, while the second read maps onto the novel TE insertion.

The read depth approach can identify variations based on the depth of coverage (DOC) or number of reads mapped at a specific genomic region. Split read method is based on the split created in a read during mapping, which leads to mapping of a single read into different locations or mapping of part of a read. The sequence assembly method helps to detect variation among individuals and populations by comparing the reference and *de novo* assembled sequences. This method is particularly suitable for analysis of SVs and identification of novel genes from *de novo* assembly by comparison against high-quality reference genome.

### 1.7.3 Population genomic approaches

Genomic variation can be neutral, adaptive or deleterious (Nielsen, 2005). Neutral variations have no fitness advantage and hence not causing adaptive divergence, i.e., their frequency is not changed under selection. Deleterious mutations, however, reduce plant fitness under natural selection and hence removed from the genome through purifying (negative) or background selection, which leads to conserved genomic regions (Pybus *et al.*, 2009; Nielsen, 2005).

Adaptive (advantageous) variants, however, increase plant fitness under natural selection and hence maintained and increased in frequency – meaning that, they are under positive selection (Figure 7B) (Vitti *et al.*, 2013; Pybus *et al.*, 2009; Nielsen, 2005). Unlike neutral and deleterious variants,

advantageous mutations are therefore causing adaptive genomic divergence among individuals and populations. A balance among positive and negative selections leads to balanced selection, which is the occurrence of multiple genetic variants in a population (Vitti *et al.*, 2013; Schoville *et al.*, 2012; Nielsen, 2005). It is difficult to analyze and associate such pattern of divergence with adaptive selection. However, balanced selection can increase plant fitness through overdominance, which means that there is heterozygote advantage over homozygotes (Pybus *et al.*, 2009).

Under positive selection, the rate at which adaptive variants sweep to higher frequency and become fixed in the population (Figure 7A and B) depends on the strength of selection, the population size and type of variant (novel or standing genetic variation). Strong adaptive variant can sweep to higher frequency in short generation than weaker ones both in large and small populations (Pybus *et al.*, 2009). Novel and standing genetic variants can also vary in sweep rate. Unlike novel variation, standing genetic variations involving multiple alleles can cause adaptive divergence through slight changes in frequencies without reaching fixations.

Identification of adaptive genes and genomic variations is based on the identification of signature of positive selection using different statistical methods (Boitard *et al.*, 2013; Vitti *et al.*, 2013; Franks & Hoffmann, 2012; Strasburg *et al.*, 2012; Hohenlohe *et al.*, 2010; Suzuki, 2010; Excoffier *et al.*, 2009; Nosil *et al.*, 2009; Pybus *et al.*, 2009; Nielsen, 2005). I here only describe three commonly used methods that are relevant for this thesis.

*(1) Allele frequency spectrum-based analysis*

Allele frequency spectrum (AFS) approaches are used to infer signature of positive selection within a population based on the frequency of fitness-enhancing mutations. Positive selection increases the frequency and subsequent fixation of the adaptive variants and nearby linked neutral variants in the hitchhiker genomic regions (reviewed in Vitti *et al.*, 2013; Burke, 2012; Strasburg *et al.*, 2012; Hohenlohe *et al.*, 2010; Suzuki, 2010; Nielsen, 2005). Fixation of an adaptive variant leads to the creation of homogenous genomic region and hence causes low genomic variation within individuals or population around the selected genomic region (Figure 7B). New variations reappear at this homogenous region and cause a surplus of rare low-frequency variants, but do not increase the genomic variation among individuals. AFS methods thus rely on the frequency patterns of fixed and surplus rare variants to infer signature of adaptive selections.

Tajima's D (Tajima, 1989) is one of the commonly used AFS-based statistical tests to detect signal of adaptive divergence. It compares the average

number of pairwise nucleotide differences ($\pi$) between individuals with the total number of segregating polymorphism (S) estimated by Watterson $\theta_W$ (Watterson, 1975) at a given genomic region within a population. When the number of pairwise nucleotide divergences among individuals ($\pi$) is similar with the number of segregating variants (S or $\theta$w), it is assumed that a neutral process or genetic drift is responsible for the observed patterns of variation. A small (negative) Tajima's D, however, arises when there is low nucleotide variation ($\pi$) but an excess of new rare variants (high $\theta$w) exists within a population. Such pattern is associated with adaptive genomic divergence (signature of positive selection) or a non-adaptive demographic process such as population expansion. In contrary, a large Tajima's D arises when there is high nucleotide variation ($\pi$) within a population, which is a sign of balanced selection or presence of population structure (Pybus *et al.*, 2009).

*(2) Population differentiation-based analysis*

Fixation of a beneficial mutation in one population causes adaptive genomic divergence among populations (Figure 7C). The degree of population differentiation is measured by the fixation index – *Fst*, which describes the proportion of genetic variation due to allele frequency differences among populations (Excoffier *et al.*, 2009; Holsinger & Weir, 2009; Nosil *et al.*, 2009; Foll & Gaggiotti, 2008). Similar allele frequencies within each population lead to absence of divergence among populations and give low *Fst*, whereas high difference in adaptive allele frequencies among populations (i.e., a differential allele frequency change in only population) leads to adaptive divergence among populations and give large *Fst* (Vitti *et al.*, 2013; Suzuki, 2010; Holsinger & Weir, 2009): *Fst* = 1 means that the adaptive allele is fully fixed in one population.

*(3) Linkage disequilibrium-based analysis*

Natural selection sweeps the frequency of adaptive allele and linked neutral variants at different genomic loci (Figure 7B). The adaptive and the neighboring linked variants are therefore strongly associated (i.e., in strong linkage disequilibrium – LD), which in turn leads to the creation of different haplotype structures at selective genomic region (Vitti *et al.*, 2013; Suzuki, 2010). That means that adaptive and linked neutral variants that are in strong LD together form frequent and longer haplotypes, while the neutral variants form less frequent haplotypes in non-adaptive individuals or populations. Such haplotype structures persist in the population until the LD breaks down by recombination events. The LD-based approaches for analysis of adaptive

divergence therefore use such patterns of association and haplotype structure to detect signature of adaptive selection within and among populations.

*Limitations with the adaptive divergence identification approaches*

The aforementioned population genomic approaches have both merits and limitations to analyse and identify signature of adaptive genomic divergence (reviewed in Vitti *et al.*, 2013; Strasburg *et al.*, 2012; Hohenlohe *et al.*, 2010; Suzuki, 2010). The limitations are associated with the ability of the approaches to detect hard and soft selective sweeps at different stages of the selection process and the influence of other demographic processes on the analysis.

Adaptation due to standing genetic variation can generate different patterns. For instance, when multiple adaptive alleles slightly change in frequency, diverse genomic regions occur, while complete sweeps of multiple alleles create and nearly homogenous genomic regions. Both AFS- and *Fst*-based approaches have low power to detect adaptive divergence associated with soft sweeps with slight changes in allele frequency. Population bottleneck and purifying selection can also generate a signature similar to selective sweeps, i.e., regions with low genomic diversity, which can affect the AFS-based identification of adaptive selection. The presence of unaccounted population structure can also affect the identification of selective sweeps since the allele frequency distribution with and without population structure are different. Furthermore, approaches for analysis of single and many populations are different. LD-based methods are good in detecting the ongoing or very recent selective sweeps since in old sweeps, the LD might have broken down. The recombination frequency that varies across the genome (Munoz-Amatriain *et al.*, 2013) can also affect LD-based approaches. For instance, LD is low (decay rapidly) in wild barley (Morrell *et al.*, 2005) and hence the LD-based method has low relevance unless used to detect very recent sweeps.

Identification of adaptive genomic diversity within and among populations is therefore challenging even by implementing more than one method. Detection of signature of adaptive selection and further case-by-case and functional analysis of the candidate adaptive genes and genetic variations are therefore a more realistic approach at the moment (Vitte *et al.*, 2014). However, the advancement in both bioinformatics and population genomic approaches in line with high-throughput genomic datasets that can be generated from large individuals and populations will resolve current challenges.

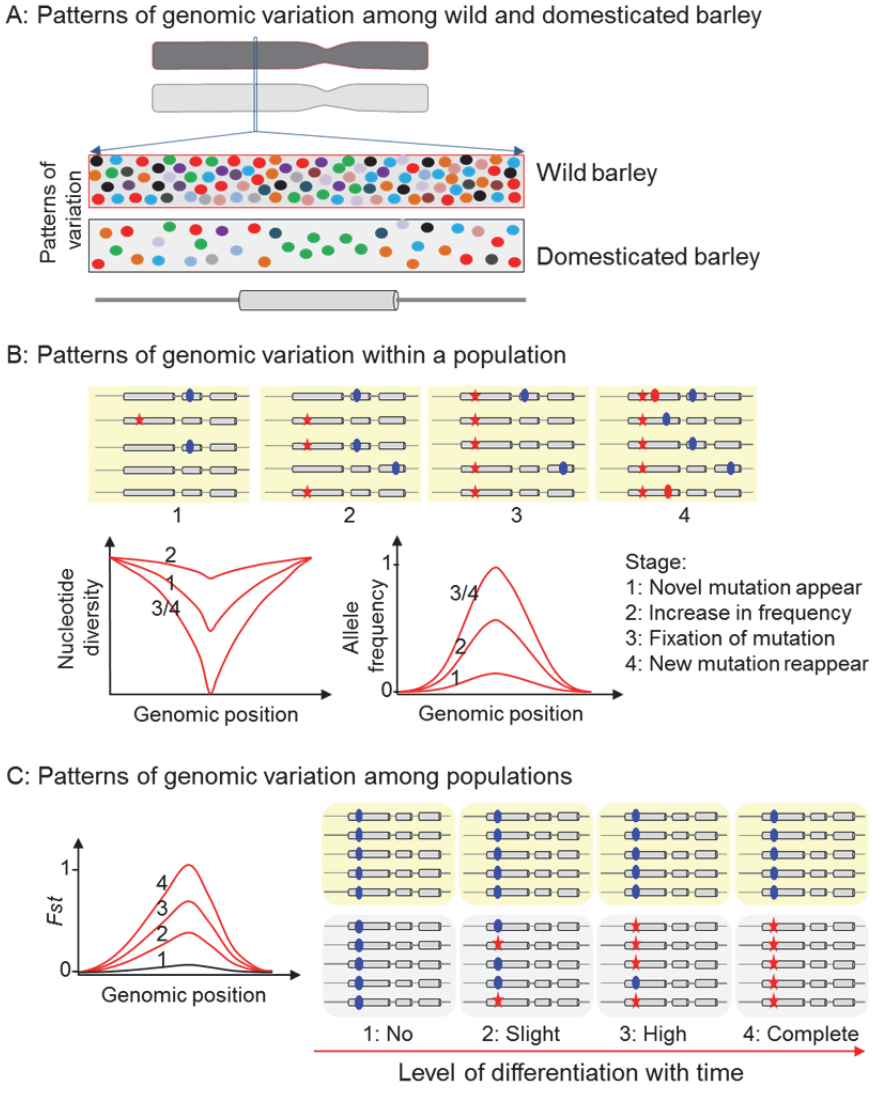


*Figure 7.* Illustrations showing patterns of genomic divergence within and among individuals and populations. (A) Demonstrates the patterns of divergence among wild and cultivated barley due to domestication in which loss of diversity occurs at a gene. (B) Patterns of genomic variation within a population due to an increase in the frequency and fixation of a novel mutation (red star). Nucleotide diversity and allele frequency spectrum can be used to detect such selective sweep. (C) Patterns of genomic variation among populations due to an increase in the frequency of beneficial mutation (red star), which can be due to standing genetic variation. Population differentiation method can be used to detect such patterns of divergence. ● Neutral mutation and ● New mutation.

# 2 Aims of the study

The general aims of this thesis were to study the adaptive genomic divergence in wild and domesticated barley individuals and populations across environmental gradients and to identify candidate genes and genetic variation linked to adaptive selection. Towards these ends, we focused on wild barley from Israel and domesticated barley from Ethiopia for several reasons. Israel is part of the Fertile Crescent and geographically a country with strong small- and large-scale ecological gradients with contrasting environments. The large-scale gradient covers the North-South axis, with a humid environment in the northern part and the Negev desert in the southern part. Several small-scale highly differentiating environmental gradients are found across the North-South axis with East-West orientated valleys. Ethiopia is also a country with very diverse ecogeographical environments and a highly distinctive and diverse barley gene pool.

We therefore analyzed wild and domesticated barley genotypes from diverse and contrasting environments to specifically address the following research objectives:

- Understand patterns of genomic divergence in wild barley populations across macro- and micro-environmental gradients in Israel.
- Analyze phenotypic and transcriptome divergence between differentially adapted and drought stressed wild barley ecotypes to identify novel genes and genetic variation.
- Design custom array for targeted-enrichment of selected genes for identification of candidate genes and genetic variation with signature of selection in wild and domesticated barley populations.
- Design an array for genome-wide targeted capturing and detection of known and novel *BARE* insertions in the wild and domesticated barley populations.

# 3 Results and Discussion

## 3.1 Patterns of genomic divergence in wild barley (I, II and III)

The geographical regions where wild barley has adapted are highly differentiating over short- and long-geographical scales for different ecological (temperature, altitude and precipitation) and edaphic factors. We therefore investigated how the spatial scale, the strongly differentiating environmental gradients and neutral evolutionary processes are shaping the patterns of genomic divergence in wild barley across Israel.

### 3.1.1 Adaptive patterns of genomic divergence

To analyze the patterns of genomic variation in wild barley populations across macro- and micro-environmental gradients across Israel, we first performed a genomic survey by sequencing 34 genomic fragments representing single-copy genes in 54 wild barley accessions. We further performed transcriptome sequencing of two differentially adapted wild barley ecotypes from the Negev desert (B1K2) and the Mediterranean moist environment (B1K30). These two accessions were from the large wild barley ecotype collections (Barley1K) (Hubner *et al.*, 2009). The experiments were conducted when large population-based analysis of genomic divergence using high-throughput sequencing approach was not feasible cost-wise. Further, we implemented the customized targeted sequence capture approach to analyze the patterns of divergence in stress-related and other important genes, novel transcripts identified from transcriptome sequencing of differentially adapted ecotypes, and randomly selected single-copy genes.

The genome scanning, transcriptome analyses and targeted sequencing studies revealed the presence of high genomic variation in wild barley from Israel, with an average nucleotide variation $\pi$ of $4.18 \times 10^{-3}$ across 34 gene fragments (Figure 8A)(Bedada *et al.*, 2014b), and a SNP density of 4.4 SNPs/kb based on transcriptome data (Bedada *et al.*, 2014a) and 4.7 SNPs/kb

at targeted genes. Likewise, the genomic variation in 30 wild barley accessions collected from the micro-environmental gradient at EC1 Nahal Oren was high ($\pi = 3.6x10^{-3}$), which was over 85% of the variation across Israel. The variation at the hot and drier SFS ($\pi = 2.2x10^{-3}$) was 1.8-fold lower than the variation at the humid NFS ($3.9x10^{-3}$, Figure 8A). The results indicate the presence of high genomic divergence in wild barley from a smaller geographical region (Israel), which is about two-third of the variation observed ($\pi = 6.8x10^{-3}$) in wild barley distribution range (Morrell & Clegg, 2007).

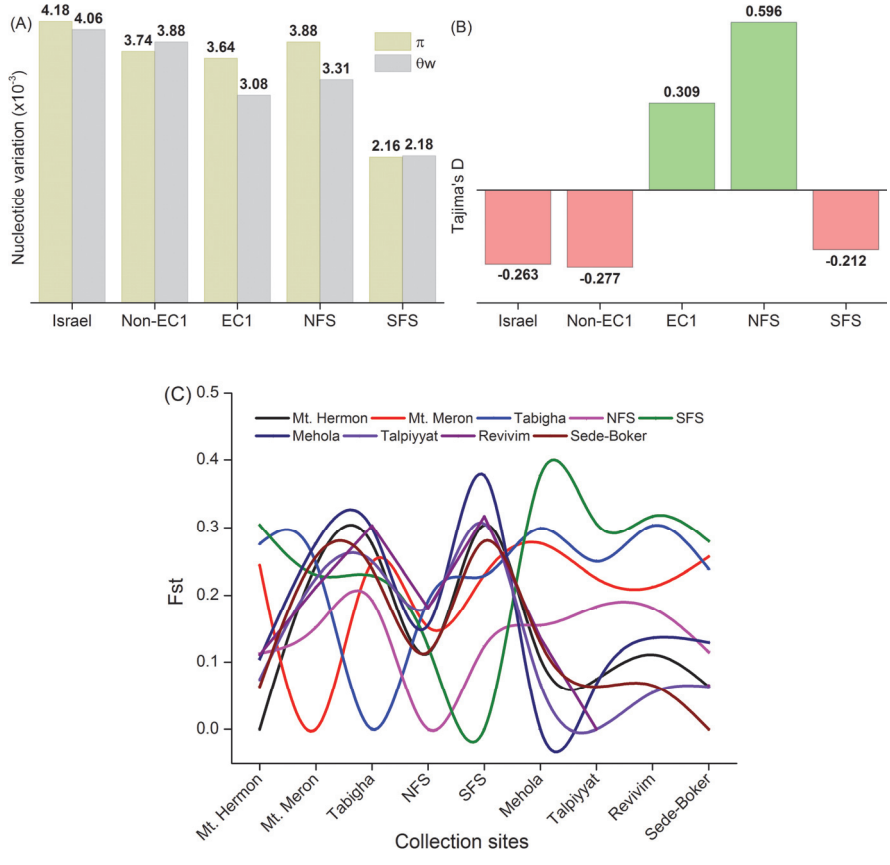


*Figure 8.* Patterns of genomic variation in wild barley across Israel. Average (A) nucleotide variation, (B) Tajima's D and (C) *Fst* across 34 gene fragments in wild barley from macro- and micro-environments across Israel. Grouping of wild barley accessions: 'Israel' – across the country, 'Non-EC1' – across the country except from EC1, 'EC1' – Evolution Canyon 1, 'NFS' – North-facing slope at EC1 and 'SFS' – South-facing slope at EC1.

The patterns of genomic variation across genes were highly variable and deviated from the neutral model of evolution and hence indicating signature of

natural selection. This is because of an overall average negative Tajima's D value (Figure 8B) and 12 gene fragments significantly deviating from neutrality (Bedada *et al.*, 2014b). The pairwise *Fst* analysis further supported the presence of significant genetic differentiations among wild barley populations at several loci (Figure 8C) (Bedada *et al.*, 2014b). Furthermore, we found strong transcriptome divergence between two differentially adapted wild barley ecotypes. Almost half of the transcripts from each ecotype were not shared between the ecotypes, the SNP density of the desert ecotype B1K2 was almost by two-fold higher than that of the Mediterranean ecotype B1K30, and the ratio of nsSNPs to sSNPs was higher in the desert than the Mediterranean ecotype. High SNP density and more deleterious mutations in the desert ecotype B1K2 most likely attributed to the accumulation of both adaptive and neutral variations that can have deleterious effects. That means that it is an adaptive selection likely involving relaxed purified selection, a pattern recently observed in wild and domesticated tomato (Koenig *et al.*, 2013).

Genes associated with adaptation have also different patterns of genomic variation and differentiation. This is because the level of genomic variation in genes differentially expressed in drought-tolerant wild barley ecotype and novel genes from the desert ecotype B1K2 was 1.9- and 1.4-fold higher than the variation in average barley gene (Figure 9), respectively (Manuscript III). This indicates that the level of adaptive genomic variation is positively correlated with the level of differential gene expression, meaning, adaptive genes are highly variable and differentially expressed. Positive correlation among level of gene expression and genomic variation has been documented in *Arabidopsis* (Kliebenstein *et al.*, 2006) and *Drosophila* (Lawniczak *et al.*, 2008). A recent study on tomato (Koenig *et al.*, 2013) further revealed the presence of correlation among selection pressure and level of gene expression in which stress-related and environmental responsive genes showed shift-in expression pattern. Our results therefore support the presence of positive or adaptive selection that most likely shaped the observed patterns of genomic divergence among wild barley populations or ecotypes adapted to diverse environments. Further in-depth analysis similar to the recent study on wild and domesticated tomato by Koenig *et al.* (2013) is therefore important to investigate how natural and artificial selections are shaping the patterns of sequence and expression divergences of different types of genes such as domestication and diversification- as well as stress-related genes in wild and domesticated barley adapted to different environments. Such analyses help to dissect the genetic bases of adaptation in barley and thereby to identify genes and genetic variations related with adaptation for further introgression into barley breeding populations.

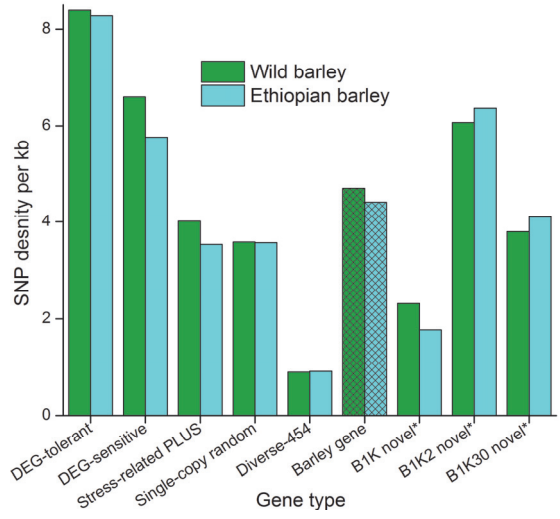


*Figure 9.* Summary of SNP density in different barley genes. The average SNP per kb was generated from pooled sequencing datasets. The pattern at 'barley gene' shows the average value for all annotated barley genes obtained from ENSEMBL database. *Shows the average SNP density at novel transcripts without transcripts with density > 50 SNPs/kb.

### 3.1.2 Adaptive and neutral patterns of population clustering

We used the haplotype data extracted from 34 gene fragments to infer the population structure in wild barley across macro- and micro-environmental gradients. Across the large geographical scale, we detected 3 to 8 clusters with two different programs (STRUCTURE and discriminate analysis of principal components – DAPC). Despite the difference in the number of inferred clusters, we observed the following distinctive patterns (Bedada *et al.*, 2014b). (1) Accessions from the drier, hot and rocky SFS of EC1 (EC1SFS) were uniquely clustered from the rest of the wild barley accessions across Israel. (2) Accessions from the humid NFS of EC1 (EC1NFS) were clustered with accessions from the northern part of Israel, which has similar environment. (3) Accessions from the northern and southern parts of Israel were unexpectedly coclustered even though the two regions have very contrasting environments. (4) At Evolution Canyon 1, accessions were clustered according to the features of the the Canyon. A recent transcriptome sequencing of one accession from each slope (Dai *et al.*, 2014) further confirmed our observation.

The observed population structure in wild barley and the strong transcriptome divergence between the two differentially adapted ecotypes indicate that both neutral and adaptive evolutionary forces are shaping the patterns of population differentiation across macro- and micro-environmental gradients in Israel. The coclustering of accessions from the northern region

with accessions from EC1NFS, the differential clustering of accessions from the two divergent slopes at EC1 and the high genetic divergence between the desert and Mediterranean ecotypes show the impacts of natural selection on the wild barley population clustering (Bedada *et al.*, 2014a; Bedada *et al.*, 2014b).

Neutral evolutionary forces such as geographical proximity (IBDL) and gene flow are also affecting the observed population structure. For instance, coclustering of accessions form the northern and southern parts of Israel show the presence of gene flow over long geographical distances probably through seed dispersal by animals and/or humans (Bedada *et al.*, 2014b). Similar coclustering was previously observed in different wild barley collections from the same regions (Hubner *et al.*, 2012). We also observed the presence of gene flow over short geographical distances among populations at EC1 despite strong genomic and environmental differentiation (Bedada *et al.*, 2014b). Such gene flow could be due to rare pollen dispersal and/or seed dispersal within and between slopes by different mechanisms. The similarities among accessions from geographically closer regions demonstrate the influence of IBDL.

The patterns of wild barley population structure we have observed over shorter and longer geographical scales and explained by both neutral and adaptive driving forces have also been documented in other studies (Russell *et al.*, 2014; Hubner *et al.*, 2013; Hubner *et al.*, 2012; Volis *et al.*, 2010; Hubner *et al.*, 2009; Morrell *et al.*, 2003). Hence, considering the strong adaptation potential to diverse and differentiating environments on one hand and the presence of gene flow and IBDL effects on the other hand, the selection-gene-flow-drift balance is likely shaping the dynamic of genomic divergence among wild barley populations (Volis *et al.*, 2010; Morrell *et al.*, 2003). Genome-wide analysis of large wild barley collections from broader geographical regions using high-throughput data is required to further dissect the patterns and genetic bases of adaptive divergence, and the effect of neutral evolutionary forces, such as gene flow.

## 3.2 Divergence among differentially adapted wild barley ecotypes (II)

### 3.2.1 Physiological divergence

To investigate the transcriptome divergence under drought stress between two differentially adapted wild barley ecotypes from the Negev desert B1K2 and the Mediterranean moist environment B1K30, we first analyzed and validated whether there is phenotypic divergence between the two ecotypes under drought stress. The phenotypic characterization was performed in Israel using two physiological traits, water use efficiency (WUE) and leaf relative water

content (RWC). WUE describes plant efficiency in biomass gain through photosynthesis (carbon assimilation) while minimizing water loss through transpiration. This is a commonly used parameter to evaluate plant adaptation potential to drought stress or water limited environments (Bramley *et al.*, 2013). RWC describes the pant water status and is associated with different leaf physiologies such as leaf turgor, stomatal conductance, transpiration, photosynthesis and growth.

Under both drought and well-irrigated conditions, the desert ecotype lost more water than the Mediterranean ecotype (Bedada *et al.*, 2014a). Nonetheless, the desert ecotype had a higher WUE and leaf RWC than the Mediterranean ecotype (Figure 10A and 10B). The results indicate that the desert ecotype B1K2 can efficiently assimilate more carbon into biomass (higher photosynthesis rate) per unit of lost water through transpiration than the Mediterranean B1K30 ecotype does. The change in the relative amount of water present on the plant tissue (RWC) under well-irrigated and drought stress was slight in the desert ecotype, but very high in the Mediterranean ecotype. Hence, shows the two ecotypes are phenotypically divergent and the desert ecotype has better adaptive response to drought stress.

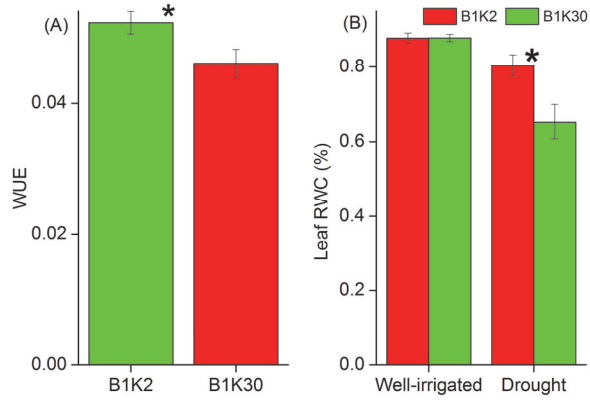


*Figure 10.* Physiological response of the desert B1K2 and the Mediterranean B1K30 ecotypes. (A) WUE of the desert and Mediterranean ecotype. (B) Leaf RWC of the desert and Mediterranean ecotypes under well-irrigated and drought conditions. *Shows significant differences between the two ecotypes.

### 3.2.2 Transcriptome divergence

To analyze the genomic divergence between the two phenotypically differentiating ecotypes, we performed transcriptome sequencing of normalized cDNA libraries from drought-stressed plants. Normalization of transcriptome libraries helps to remove and reduce the highly transcribed genes

and thereby get an even coverage to characterize as many transcripts as possible. Hence, it is the best approach to identify rare and novel transcripts and variants (Hirsch *et al.*, 2014; Ekblom *et al.*, 2012; Ekblom & Galindo, 2011; Good, 2011; Stapley *et al.*, 2010). When coupled with drought-stress treatment, it can help to uncover genes and genetic variation contributing to drought stress tolerance and adaptation.

We therefore used the 454 platform for transcriptome sequencing of normalized cDNA libraries from drought stressed desert and Mediterranean ecotypes. Over half-million processed reads from each ecotype were *de novo* assembled into 20,439 clustered putative unique transcripts (PUTs) for B1K2, 21,494 for B1K30 and 28,720 for joint assembly (denoted as B1K). To identify transcripts that are unique to each ecotype, we compared the PUTs and found that the majority of the total transcripts (71%) were not shared between ecotypes. Only 29% (9,546) of the total transcripts or 46% of B1K2 PUTs were shared between the ecotypes (Figure 11A). The transcriptome divergence between the two ecotypes could be due to one or more of the following reasons. (1) The non-shared transcripts may represent genes whose transcripts were lost during cDNA normalization or library preparation. (2) The divergence may represent presence/absence polymorphisms, meaning, ecotype-specific or non-shared transcripts that reflect transcriptome divergence due to differential loss or gain of transcripts. Such polymorphism has been documented in maize (Morgante *et al.*, 2007; Wang & Dooner, 2006; Morgante *et al.*, 2005). A recent transcriptome analysis of wild and domesticated barley by Dai *et al.* (2014) further supports our results in that they also found high transcriptome divergence in wild and domesticated barley and that wild barley had high transcript diversity. (3) The divergence among the ecotypes may be due to differential expression of genes in both accessions in response to the drought treatment.

To further identify how many of the transcripts are ecotype-specific novel transcripts and orthologous to barley genes, we further compared the PUTs against three cultivated barley sequence datasets ('high confidence' – HC genes, full-length cDNA – fl-cDNA and HarvEST) using a reciprocal BLAST hit (RBH) approach. We found that 16% (3,245) of B1K2 and 17% (3,674) of B1K30 transcripts were not orthologous to other wild barley ecotype and cultivated barley sequences (Figure 11B and 11C), and hence were considered as candidate ecotype-specific genes or novel transcripts (Bedada *et al.*, 2014a).

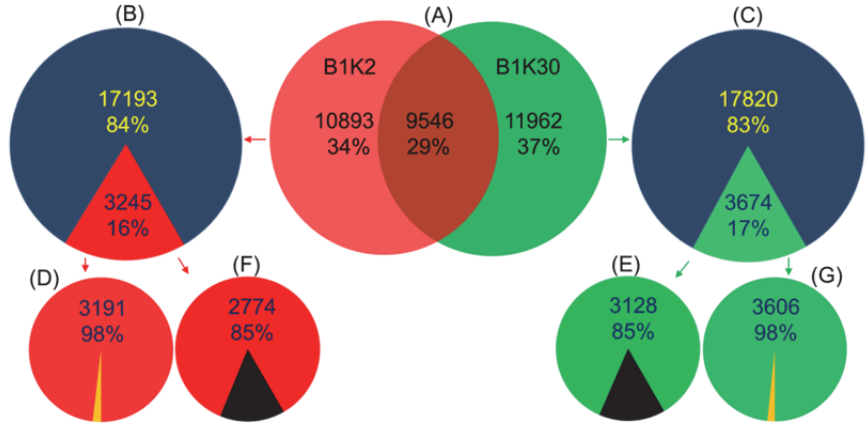


*Figure 11.* Homolog analysis for identification of novel transcripts. (A) A Venn diagram showing RBH analysis among differentially adapted B1K2 and B1K30 ecotypes. The two ecotypes shared 29% (9,546) of the total transcripts. RBH of (B) B1K2 and (C) B1K30 transcripts against barley sequence data from HC, fl-cDNA and HarvEST. 16% (3,245) of B1K2 and 17% (3,674) of B1K30 transcripts were without significant orthologous barley sequences. RBH of novel transcripts from (D) B1K2 and (E) B1K30 against five fully annotated plant genomes. 98% (3,191) of B1K2 and 98% (3,606) of B1K30 novel transcripts were without significant homologous hits in other plant genomes. (F) B1K2 and (G) B1K30 novel transcripts with predicted CDS ≥ 100 bp. 85% of both B1K2 and B1K30 novel transcripts have CDS ≥ 100 bp.

Similarly, 25% (7,102) of B1K transcripts from both ecotypes were without significant RBH in cultivated barley datasets, and hence are candidate wild barley-specific genes. Almost all (98%) novel transcripts were not similar to five fully sequenced and annotated plant genomes (Figure 11D and 11E). Further, 85% of the novel transcripts were *de novo* annotated with a CDS (coding sequencing) longer than 100 bp (Figure 11F and 11G). Our results are therefore indicating that 454 sequencing of normalized cDNA library is an efficient method to discover new genes. Other studies in the grass *Spartina* (Ferreira de Carvalho *et al.*, 2013), cultivated barley (Thiel *et al.*, 2012), zebra finch (Ekblom *et al.*, 2012) and wheat (Cantu *et al.*, 2011) have used a similar approach and identified novel transcripts or genes. The ecotype- and wild barley-specific novel transcripts without any orthologs in known barley sequences may be explained by one or more of the followings:

(1) The novel transcripts may represent genes that are found in wild but not in cultivated barley.
(2) The novel transcripts may represent unannotated barley genes – as only 86% (26,159) of the total 30,400 estimated barley genes were reported as HC genes (IBGS Consortium *et al.*, 2012). This is because 98% of the

reads could be mapped to the 'Morex' WGS, which is represent a draft genome assembly, using a local alignment method.

(3) The novel transcripts probably derived from genes affected by SVs and alternative splicing, which are prevalent in the barley genome (Munoz-Amatriain *et al.*, 2013; IBGS Consortium *et al.*, 2012). This is because the large proportion (98%) of mapped reads against WGS was achieved using a local alignment method, a method that trimmed the non-matching end of the reads for efficient mapping and thereby increased the proportion of mapped reads. Such trimmed reads are associated with SVs and alternative splicing.

(4) Some of the novel transcripts may represent untranslated region of the genome (originated from incompletely transcribed mRNA), non-coding RNAs or may be too short for significant RBH against known barley genes.

The transcripts generated from differentially adapted wild barley ecotypes can therefore contribute to further improvement of barely transcriptome and genome annotation. Furthermore, they are good genomic resources for the assembly and creation of a separate wild barley reference genome, which is an important and a required genomic data for several evolutionary and genomic studies. These are because some of our transcripts are longer than their orthologous barley genes and some are non-orthologous to all available barley sequences, but homologous to transcripts from other grasses and plant species. Functional and evolutionary conservation based analyses also indicated that the assembled transcripts were homologous to over 800 well-characterized stress-related genes and transcription factors. The generated transcripts are therefore a resource for further evolutionary and functional characterization of genes homologous to well-characterized and known stress-related genes and transcription factors. This is because the homologous transcripts may carry different and important variations, but it does not necessarily mean that they are involved in drought response.

## 3.3 SNVs identification and genomic distance analysis (II & III)

High-throughput NGS datasets generated from individual or pooled samples are source of high density and quality nucleotide variations (SNVs and InDels). Identification of high quality SNPs is, however, affected by several factors (Guo *et al.*, 2013; Kiani *et al.*, 2013; Lee *et al.*, 2012; Martin & Wang, 2011; Nielsen *et al.*, 2011). This is because high quality SNP discovery is multi-stage

processes involving several quality control measures. Several factors at one or more of the involved steps can therefore affect variant identification.

We have observed the impacts of different factors such as the algorithm implemented in different programs on SNP calling from transcriptome data of two differentially adapted wild barley ecotypes. To select the best high quality SNP detection method from our transcriptome data, we therefore selected three different tools (Bowtie-2, BWA-SW and GSMapper) and analyzed 454 reads from one ecotype (B1K2) by mapping against Hv fl-cDNA data. We observed 2.5-fold difference in the number of SNPs identified by global and local alignment methods, and 10-fold difference among SNP identified using GSMapper and Bowtie-2 mapping approaches. Only 47% (907) of the total high quality SNPs (1,937) identified by stringent filtering (i.e., SNP supported by ≥ 8x coverage of which a minimum of 4 reads each supporting reference and variant nucleotides) were detected by more than one of the three used mapping tools (Bedada *et al.*, 2014a). The rest 57% were unique to a single tool, and only 5.1% were detected by all three tools. The majority (84%) of the SNPs identified using Bowtie-2 mapping were, however, supported by at least one of the other two tools and hence Bowtie-2 was selected as a mapping tool.

By comparing the transcriptome of the desert B1K2 and Mediterranean B1K30 ecotypes, we identified 28,289 raw SNPs, of which 1,017 were high quality supported by ≥ 8x coverage (≥ 4 reads each supporting the two alleles) (Bedada *et al.*, 2014a). Similarly, by mapping B1K2, B1K30 and their merged data (B1K) against barley HC genes, we called 16,284, 14,509 and 24,446 raw SNPs from which we identified 1,184, 1,081 and 5,036 high quality SNPs (Table 2), respectively (Bedada *et al.*, 2014a). We applied a stringent filtering approach with the assumption that SNPs supported by high coverage are most likely true SNPs. Our filtering approach was, however, highly conservative and hence reduced the number of high quality filtered SNPs, which was 7% of the total called SNPs. The filtering had, however, relatively less effect on the combined data (B1K) in which 20% of the total raw SNPs were high quality for SNPs. This indicates the contribution of high coverage per nucleotide position and the high variability from combined dataset for identification of high quality SNPs. On the other hand, 25% (9,775) and 24% (8,682) of raw SNPs from B1K2 and B1K30 ecotypes, respectively overlapped with SNPs identified from the wild barley ecotype B1K4 (B1K-4-12) sequenced by the IBGS Consortium using different sequencing approaches (IBGS Consortium *et al.*, 2012) This indicates that a significant proportion of raw SNPs from transcriptome data was correctly inferred. The number of quality SNPs that can be identified from these highly divergent ecotypes could therefore be more than what we have filtered as high quality. Hence, as many as 4,220, 3,354 and

8,458 quality SNPs at depth of ≥ 10x and supported by reference and/or variant alleles and containing a fixed SNP (i.e., when the variant allele is a major allele) can be identified from B1K2, B1K30 and B1K, respectively (Table 2). This is over 3-fold higher than the high quality SNPs we have identified from each ecotype.

Table 2. *Summary of SNP from transcriptome sequencing data of wild barley ecotypes using different depth of coverage and filtering. The data is based on mapping against barley HC genes.*

| Data (raw SNPs) | Coverage at SNP position for | | | Filtered SNP | | Remark |
|---|---|---|---|---|---|---|
| | All (≥) | reference allele (≥) | variant allele (≥) | No. | (%) | |
| B1K2 (16,284) | 10x | | 2x | 4,220 | 25.9 | |
| | 15x | | 2x | 2,169 | 13.3 | |
| | 8x | 2x | 2x | 1,590 | 9.8 | |
| | 10x | 2x | 2x | 1,332 | 8.2 | |
| | 8x | 4x | 4x | 1,184 | 7.3 | applied |
| B1K30 (14,509) | 10x | | 2x | 3,354 | 23.1 | |
| | 15x | | 2x | 2,013 | 13.9 | |
| | 8x | 2x | 2x | 1,735 | 12.0 | |
| | 10x | 2x | 2x | 1,398 | 9.6 | |
| | 8x | 4x | 4x | 1,081 | 7.5 | applied |
| B1K (24,446) | 10x | | 2x | 8,458 | 34.6 | |
| | 15x | | 2x | 5,736 | 23.5 | |
| | 8x | 2x | 2x | 6,594 | 27.0 | |
| | 10x | 2x | 2x | 5,674 | 23.2 | |
| | 8x | 4x | 4x | 5,036 | 20.6 | applied |
| B1K2 (16,284) | 10x | 0x | 10x | 2,583 | 15.9 | potentially fixed SNPs |
| | 15x | 0x | 15x | 1,549 | 9.5 | |
| B1K30 (14,509) | 10x | 0x | 10x | 1,698 | 11.7 | potentially fixed SNPs |
| | 15x | 0x | 15x | 956 | 6.6 | |
| B1K (24,446) | 10x | 0x | 10x | 2,304 | 9.4 | potentially fixed SNPs |
| | 15x | 0x | 15x | 1,439 | 5.9 | |

By targeted capture Pool-seq of large number of wild (23) and Ethiopian barley (42) genotypes, we have identified 5,561 and 7,273 high quality SNPs and 654 and 739 InDels, respectively (Manuscript III). The SNVs identified from the two differentially adapted wild barley ecotypes and pooled sequenced wild and Ethiopian barley genotypes are therefore potential genomic resources.

Based on transcriptome sequencing, the SNP density within wild barley was 4.4 SNPs/kb (i.e., 1 SNP per 227 bp). The SNP density among wild and cultivated barley, however, varied between ecotypes. Hence, the density of the desert wild barley against cultivated barley was 1.9-fold higher than the density among the Mediterranean wild and cultivated barley. The targeted capture

analysis also reflects this pattern of variation in that the SNP density among wild and cultivated barley at novel transcripts from the desert ecotype was 1.6-fold higher than the density at novel transcripts from the Mediterranean ecotype. Likewise, targeted genes analysis further showed that the genomic variation among wild barley (from Mediterranean and Northern regions of Israel) and cultivated barley (4.7 SNPs/kb) was similar with the variation found within cultivated barley (4.4 SNPs/kb, Ethiopian barley against reference genome) (Manuscript III). The higher genomic variation in the desert barley and the more similarity among Mediterranean wild and cultivated barley indicate that (1) the barley domestication occurred in the northern part of Israel, (2) the accumulation of adaptive and linked neutral variation mostly through evolutionary adaptation to the desert environment causes higher divergence in the desert ectype, and/or (3) there is a gene flow between the Mediterranean wild barley and cultivated barley (Bedada *et al.*, 2014a). Like the larger genetic distance, higher phenotypic differentiation was observed between the desert and Mediterranean wild barley ecotypes, and the desert wild barley ecotype and cultivated barley than between the Mediterranean ecotype and cultivated barley for several quantitative traits (Hubner *et al.*, 2013). The distribution patterns of SNPs from transcriptome data showed the presence of high density at telomeric regions of the chromosomes, which is consistent with the patterns observed in the barley genome sequencing and mostly due to a higher gene density and/or increased recombination rate in the telomeric regions (Munoz-Amatriain *et al.*, 2013; IBGS Consortium *et al.*, 2012).

The nucleotide variations identified by transcriptome and targeted capture sequencing, particularly the variants from the stress-related and agronomically important genes, can therefore be used for different applications. These include high-throughput SNP-array for genomic analysis and identification of gene and genetic variation responsible for drought adaptation, genomic diversity analysis and characterization of large gene pools and detection of marker-trait association. The identified useful variations can be used for further introgression into barley breeding populations.

## 3.4 Genomic divergence in Ethiopian barley

The Ethiopian barley gene pool is unique with distinctive patterns of genomic diversity. It has been intensively used globally for several genetic and genomic studies such as mapping, identification and isolation of genes and genetic variations (Igartua *et al.*, 2013; Bjørnstad & Abay, 2010; Orabi *et al.*, 2007; and references therein; Pourkheirandish & Komatsuda, 2007; Piffanelli *et al.*, 2004; Bjornstad *et al.*, 1997). We analyzed the genomic divergence in 42

Ethiopian barley genotypes together with the wild barley accessions using customized targeted-enrichment Pool-seq. We found that the genomic variation in Ethiopian barley genotypes (4.41 SNPs/kb) was similar to the variation in wild barley (4.75 SNPs/kb), which is 93% of the variation found in wild barley (Manuscript III). According to window-based variation analysis, almost one-tenth (9 SNPs per 100 bp window) of the covered genomic regions in Ethiopian barley was variable. Further, the Ethiopian and wild barley genotypes shared large proportion of genomic similarity. About 58% (4,212) of SNPs identified from the Ethiopian barley genotypes were found in wild barley (Manuscript III). This indicates that 58% of the Ethiopian gene pool originates from wild barley and hence less than half of the gene pool was lost due to domestication. About 76% of the wild barley gene pool was found in Ethiopian barley. Moreover, the genomic differentiation within the Ethiopian barley pool (*Fst* = 0.047) and between the Ethiopian and wild barley gene pools (*Fst* = 0.046) was similar (Manuscript III).

The large overlap in the genomic background of wild and Ethiopian barley is in contrast to the recent publication by Dai *et al.* (2014), showing a significant loss of genetic diversity in cultivated barley through domestication and diversification events. The large proportion of shared variation among the wild and Ethiopian gene pools may indicate two things. First, high level of genomic divergence is most likely due to the adaptation to very diverse ecological habitats. This is because Ethiopia, particularly the areas where barley is cultivated and from where our genotypes were originally collected, is characterized by an extraordinary ecogeographical variation. Second, Ethiopian barley was probably domesticated directly from wild barley and the introgressed ancestral gene pool has been retained due to similar patterns of selection from the overlapping ecological habitats. Hence, the Ethiopian barley gene pool was probably less affected by domestication and diversification events. Our results therefore support the possibility that Ethiopia is one of the domestication and diversification centers, which was previously suggested based on information generated using different approaches (Igartua *et al.*, 2013; Orabi *et al.*, 2007; Molina-Cano *et al.*, 2005). Further studies based on whole genome sequence analysis of large Ethiopian and wild barley populations from different environments and geographical regions are highly required to further dissect the genomic composition of the Ethiopian barley and thereby perform in-depth analysis of potential signature of domestication and diversification events.

## 3.5 Adaptive selective sweeps in wild and domesticated barley (III)

To detect signature of adaptive selection in wild and Ethiopian barley, we used a pool-HMM method that uses allele frequency spectrum to identify the potential selective sweeps in Pool-seq datasets. The method estimates whether the patterns of allele frequency observed at each SNP is associated with one of three possible states: neutral, intermediate and selection. Based on stringent setting (-k 1E-7, defining SNP transition probability between the three states), we detected 1,202 selective sweeps in wild and 1,095 in Ethiopian barley in 40 genes (Manuscript III). Overall, 4.5% of the total identified SNPs from wild barley and 3.6% from Ethiopian barley showed signature of adaptive sweeps, whereas 26.8% and 16.8% were neutral and 68.7% and 79.6% were with signature of intermediate sweep for the respective species (Manuscript III).

The majority of the total selective sweeps were unique to wild or Ethiopian barley, while only 18% were shared among each other. One-third (32%) of adaptive selective sweeps in Ethiopian barley has originated from wild barley, while the majority (68%) was private selective sweeps. This indicates that the majority of the adaptive variation was lost due to and acquired after domestication and diversification events. As Ethiopian barley genotypes are collected from highly diverse ecogeographical environments, the observed large proportion of private selective sweeps most likely indicates the adaptive variation.

Large proportion of genes with signature of selection was private to wild (75%) and Ethiopian (63%) barley. Furthermore, the majority of selective sweeps, 62.8% in wild and 76.1% in Ethiopian barley (Manuscript III), were identified from genes that were differentially expressed among drought tolerant and sensitive wild barley ecotypes (Hubner S. *et al*., in preparation). The results indicate that adaptive genomic variation, rather than neutral variation due to random genetic drift, has most likely caused the observed differential gene expression among wild barley ecotypes under drought stress. Similar patterns have been observed in genes differentially expressed among wild and domesticated tomato (Koenig *et al.*, 2013). Detection of selective sweeps from genes that showed differential pattern of expression and have been previously characterized make the identified selective sweeps as potential candidates for adaptive selection to be further verified using other approaches such as high-throughput SNP-array system in different wild and domesticated barley collections. The results further show the presence of large proportion of adaptive genomic divergence in both wild and Ethiopian barley gene pools that can be used for introgression into breeding populations.

## 3.6 Targeted *BARE* capture reveal novel insertions (IV)

TEs are driving and shaping genome diversity and evolution (Bennetzen & Wang, 2014; Mirouze & Vitte, 2014; Vitte *et al.*, 2014). Large proportion of the barley genome is composed of TEs (IBGS Consortium *et al.*, 2012) in which the *BARE1* elements constitutes over 10% of the genome (Middleton *et al.*, 2012). We were interested to investigate the genome-wide patterns of *BARE* insertions in wild and domesticated barley populations from different environments. We therefore implemented a different TE-scanning method based on targeted-enrichment technique to detect genome-wide known and novel insertions from Pool-seq dataset.

Using the Pool-seq datasets from the wild and Ethiopian barley genotypes, we analyzed 6,789 and 33,666 known *BARE* CDS and LTRs insertions in the barley genome, respectively. We were able to detect 92% of known *BARE* CDS insertions in both wild and Ethiopian barley Pool-seqs (Manuscript IV). Similarly, 52% and 47% of the known *BARE* LTR insertion sites were detected in both pools, respectively. Over 97% of the longer (≥ 500 bp) CDS and LTR insertions were detected in both the Ethiopian and the wild barley pools. The difference in the proportion of detection among longer and shorter insertions indicates that (1) the targeted regions are most likely well represented and properly captured in the longer than in the shorter insertions, which probably contain non-targeted or only part of the targeted regions, and/or (2) the longer insertions are probably fixed or stable than undetected shorter insertions, which may represent unstable insertions that have been removed through purifying selection and hence absent in our samples. The proportion of detected *BARE* CDS insertions is 1.8-fold higher than the LTR insertions, which likely indicates that the CDS insertion sites are more stable than the dynamic LTR insertions.

To identify novel (non-reference) *BARE* insertions from the chromosomal genome, we used RetroSeq program, which relies on the discordantly mapped PE reads for the detection of novel insertions. Discordantly mapped reads further mapped against known *BARE* (*BARE1* and *BARE2*) sequences. We therefore detected 5,807 and 8,631 non-reference *BARE* LTR insertions in the wild and Ethiopian barley, respectively (Manuscript IV). After filtering out insertions that are closer to the known insertions sites, we identified 3,342 and 5,882 novel *BARE* LTR insertions in the wild and Ethiopian barley, respectively. We compared the novel insertions detected in the wild and the Ethiopian barley and found that only 3.8% (337) of the total 8,887 insertions were shared between them. That means that 6% of the novel *BARE* LTR insertions from Ethiopian barley were derived from wild barley and are hence mostly ancestral insertions. The small proportion of novel shared insertions

between wild and domesticated barley indicates that the majority of common ancestral insertions are in the reference genome. Over 90% (3,005 of wild and 5,545 Ethiopian) of the novel LTR insertions in both the wild and Ethiopian barley were unique, suggesting that they are either new insertions after the domestication and diversification events and/or undetected insertions in the reference genome.

Relatively more novel insertions were detected in Ethiopian than wild barley, where sample size normalized insertions of 184 per sample in Ethiopian and 145 insertions in wild barley were found. Large number of novel insertions in Ethiopian barley maybe indicate high genetic diversity since the Ethiopian barley genotypes were originally collected from diverse environments throughout the country, while the used wild barley accessions represent less differentiating Northern and Costal wild barley populations (Hubner *et al.*, 2013).

Our array-based targeted capturing approach is therefore an efficient method for genome-wide detection of both known and novel TE insertions from individual or pooled sample sequencing datasets. Hence, it can overcome the limitations associated with the two commonly practiced approaches for the analysis of known and novel insertions. Our approach can therefore be used for locus-specific (targeted) and genome-wide analysis of TE dynamics in individuals or large populations. Further, the approach can facilitate the genome-wide annotation and improvement of the barley reference genome.

# 4 Conclusions

Adaptive genomic divergence and high level of population structure exit in wild barley across environmental gradients in Israel. The genomic divergence is driven by both natural selection and neutral evolutionary forces.

The desert and Mediterranean wild barley ecotypes show strong physiological and genomic differentiation, and the Mediterranean ecotype is genetically closer to cultivated barley. The desert ecotype shows 2-fold higher genomic divergence and a larger proportion of deleterious mutations, indicating a differential adaptation to the stressful environment.

High genomic divergence is detected in novel transcripts identified from the desert ecotype and in genes differentially expressed in another drought-tolerant ecotype.

Potential candidate genes and genetic variations with signature of adaptive selection are identified in wild and Ethiopian barley.

High genomic divergence and a larger proportion of ancestral variation are detected in the Ethiopian barley gene pool. Further, low genomic differentiation is found between the Ethiopian barley and the Mediterranean wild barley gene pools.

In-solution targeted-enrichment method detected reference (known) and novel *BARE* insertions in Ethiopian and wild barley populations.

Large number of novel genes and nucleotide variations are identified from diverse wild and domesticated barely gene pools, which can be used as

genomic resources for improvement of barley genome annotations and different genomic applications.

# 5 Future perspectives

Large-scale genome-wide analysis of wild and domesticated barley populations adapted to diverse environments to address different research questions such as the genetic basis of adaptation and domestication and diversification processes are almost untouched areas of research in barley. Hence, further studies based on genome-wide analysis of systematically collected natural wild barley populations, landraces and improved barley cultivars are needed in the following areas to unlock the huge adaptive genomic potential in wild and domesticated barley adapted to different environments:

- Patterns of adaptive divergence and the genetic basis of adaptation in wild and cultivated barley populations.
- Patterns and impact of gene flow in different ecological environments, over shorter and longer geographical scales, and within and among wild barley populations as well as between wild and domesticated barley.
- Impact of natural and artificial selections on sequence and expressional divergences in wild and domesticated barley adapted to different environments.
- Transposable element dynamics and their adaptive role in natural populations and improved barley gene pool from diverse environments.
- Population epigenomics in barley – patterns and role of epigenetic variation in natural populations adapted to contrasting environments and the interaction among epigenetic and adaptive genomic variations.

In addition to the above research areas, resource development is vital for different applications:

- Development of high-throughput SNP-array using SNVs with signature of adaptive selection and from differentially adapted wild barley ecotypes for further evaluation and verification of potential adaptive genes and genetic variants that can be used for screening of and introgression into breeding populations.

- Development of a high quality wild barley reference genome for efficient analyses and utilization of genomic resources in natural populations and domesticated barley.

# References


Abdel-Ghani, A.H., Parzies, H.K., Omary, A. & Geiger, H.H. (2004). Estimating the outcrossing rate of barley landraces and wild barley populations collected from ecologically different regions of Jordan. *Theor Appl Genet,* 109(3), pp. 588-95.

Aitken, S.N. & Whitlock, M.C. (2013). Assisted Gene Flow to Facilitate Local Adaptation to Climate Change. *Annual Review of Ecology, Evolution, and Systematics,* 44(1), pp. 367-388.

Alkan, C., Coe, B.P. & Eichler, E.E. (2011). Genome structural variation discovery and genotyping. *Nat Rev Genet,* 12(5), pp. 363-76.

Alonso-Blanco, C., Aarts, M.G., Bentsink, L., Keurentjes, J.J., Reymond, M., Vreugdenhil, D. & Koornneef, M. (2009). What has natural variation taught us about plant development, physiology, and adaptation? *Plant Cell,* 21(7), pp. 1877-96.

Anderson, C.D., Epperson, B.K., Fortin, M.J., Holderegger, R., James, P.M., Rosenberg, M.S., Scribner, K.T. & Spear, S. (2010). Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Mol Ecol,* 19(17), pp. 3565-75.

Anderson, J.T., Willis, J.H. & Mitchell-Olds, T. (2011). Evolutionary genetics of plant adaptation. *Trends Genet,* 27(7), pp. 258-66.

Andrews, K.R. & Luikart, G. (2014). Recent novel approaches for population genomics data analysis. *Mol Ecol,* 23(7), pp. 1661-7.

Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature,* 408(6814), pp. 796-815.

Baek, H.J., Beharav, A. & Nevo, E. (2003). Ecological-genomic diversity of microsatellites in wild barley, *Hordeum spontaneum*, populations in Jordan. *Theoretical and Applied Genetics,* 106(3), pp. 397-410.

Barnabas, B., Jager, K. & Feher, A. (2008). The effect of drought and heat stress on reproductive processes in cereals. *Plant Cell and Environment,* 31(1), pp. 11-38.

Barrett, R.D. & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends Ecol Evol,* 23(1), pp. 38-44.

Bedada, G., Westerbergh, A., Mueller, T., Galkin, E., Bdolach, E., Moshelion, M., Fridman, E. & Schmid, K.J. (2014a). Transcriptome sequencing of two wild

barley (*Hordeum spontaneum* L.) ecotypes differentially adapted to drought stress reveals ecotype-specific transcripts. *Bmc Genomics,* 15, p. 995.
Bedada, G., Westerbergh, A., Nevo, E., Korol, A. & Schmid, K.J. (2014b). DNA sequence variation of wild barley *Hordeum spontaneum* (L.) across environmental gradients in Israel. *Heredity (Edinb),* 112(6), pp. 646-55.
Bennetzen, J.L. & Wang, H. (2014). The Contributions of Transposable Elements to the Structure, Function, and Evolution of Plant Genomes. *Annu Rev Plant Biol*.
Bevan, M.W. & Uauy, C. (2013). Genomics reveals new landscapes for crop improvement. *Genome Biol,* 14(6), p. 206.
Bickhart, D.M. & Liu, G.E. (2014). The challenges and importance of structural variation detection in livestock. *Front Genet,* 5, p. 37.
Bjørnstad, A. & Abay, F. (2010). Multivariate patterns of diversity in Ethiopian barleys. *Crop Science,* 50(5), pp. 1579-1586.
Bjornstad, A., Demissie, A., Kilian, A. & Kleinhofs, A. (1997). The distinctness and diversity of Ethiopian barleys. *Theoretical and Applied Genetics,* 94(3-4), pp. 514-521.
Blattner, F., Pleines, T. & Jakob, S. (2010). Rapid Radiation in the Barley Genus *Hordeum* (Poaceae) During the Pleistocene in the Americas. In: Glaubrecht, M. (ed. *Evolution in Action* Springer Berlin Heidelberg, pp. 17-33.
Blattner, F.R. (2009). Progress in phylogenetic analysis and a new infrageneric classification of the barley genus *Hordeum* (Poaceae: Triticeae). *Breeding Science,* 59(5), pp. 471-480.
Blum, A. (2011). Drought resistance and its improvement. In: *Plant Breeding for Water-Limited Environments* Springer, pp. 53-152.
Blumenstiel, B., Cibulskis, K., Fisher, S., DeFelice, M., Barry, A., Fennell, T., Abreu, J., Minie, B., Costello, M., Young, G., Maquire, J., Kernytsky, A., Melnikov, A., Rogov, P., Gnirke, A. & Gabriel, S. (2010). Targeted exon sequencing by in-solution hybrid selection. *Curr Protoc Hum Genet,* Chapter 18, p. Unit 18 4.
Boitard, S., Kofler, R., Francoise, P., Robelin, D., Schlotterer, C. & Futschik, A. (2013). Pool-hmm: a Python program for estimating the allele frequency spectrum and detecting selective sweeps from next generation sequencing of pooled samples. *Molecular Ecology Resources,* 13(2), pp. 337-40.
Bolger, M.E., Weisshaar, B., Scholz, U., Stein, N., Usadel, B. & Mayer, K.F. (2014). Plant genome sequencing - applications for crop improvement. *Curr Opin Biotechnol,* 26, pp. 31-7.
Bonchev, G. & Parisod, C. (2013). Transposable elements and microevolutionary changes in natural populations. *Molecular Ecology Resources,* 13(5), pp. 765-75.
Bramley, H., Turner, N.C. & Siddique, K.H. (2013). Water Use Efficiency. In: *Genomics and Breeding for Climate-Resilient Crops* Springer, pp. 225-268.
Brown, A.H.D., Zohary, D. & Nevo, E. (1978). Outcrossing rates and heterozygosity in natural populations of *Hordeum spontaneum* Koch in Israel. *Heredity,* 41(1), pp. 49-62.

Buchmann, J.P., Keller, B. & Wicker, T. (2013). Transposons in Cereals: Shaping Genomes and Driving Their Evolution. In: *Cereal Genomics II* Springer, pp. 127-154.
Burke, M.K. (2012). How does adaptation sweep through the genome? Insights from long-term selection experiments. *Proc Biol Sci,* 279(1749), pp. 5029-38.
Cantu, D., Pearce, S.P., Distelfeld, A., Christiansen, M.W., Uauy, C., Akhunov, E., Fahima, T. & Dubcovsky, J. (2011). Effect of the down-regulation of the high Grain Protein Content (GPC) genes on the wheat transcriptome during monocarpic senescence. *Bmc Genomics,* 12, p. 492.
Carretero-Paulet, L. & Fares, M.A. (2012). Evolutionary dynamics and functional specialization of plant paralogs formed by whole and small-scale genome duplications. *Mol Biol Evol,* 29(11), pp. 3541-51.
Casacuberta, E. & Gonzalez, J. (2013). The impact of transposable elements in environmental adaptation. *Mol Ecol,* 22(6), pp. 1503-17.
Chang, W., Jaaskelainen, M., Li, S.P. & Schulman, A.H. (2013). BARE retrotransposons are translated and replicated via distinct RNA pools. *Plos One,* 8(8), p. e72270.
Chen, S.D., Krinsky, B.H. & Long, M.Y. (2013). New genes as drivers of phenotypic evolution. *Nature Reviews Genetics,* 14(10), pp. 745-745.
Chono, M., Honda, I., Zeniya, H., Yoneyama, K., Saisho, D., Takeda, K., Takatsuto, S., Hoshino, T. & Watanabe, Y. (2003). A semidwarf phenotype of barley *uzu* results from a nucleotide substitution in the gene encoding a putative brassinosteroid receptor. *Plant Physiol,* 133(3), pp. 1209-19.
Comadran, J., Kilian, B., Russell, J., Ramsay, L., Stein, N., Ganal, M., Shaw, P., Bayer, M., Thomas, W., Marshall, D., Hedley, P., Tondelli, A., Pecchioni, N., Francia, E., Korzun, V., Walther, A. & Waugh, R. (2012). Natural variation in a homolog of Antirrhinum CENTRORADIALIS contributed to spring growth habit and environmental adaptation in cultivated barley. *Nat Genet,* 44(12), pp. 1388-92.
Conant, G.C. & Wolfe, K.H. (2008). Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet,* 9(12), pp. 938-50.
Conrad, D.F. & Hurles, M.E. (2007). The population genetics of structural variation. *Nat Genet,* 39(7 Suppl), pp. S30-6.
Coop, G., Witonsky, D., Di Rienzo, A. & Pritchard, J.K. (2010). Using environmental correlations to identify loci underlying local adaptation. *Genetics,* 185(4), pp. 1411-23.
Cronin, J.K., Bundock, P.C., Henry, R.J. & Nevo, E. (2007). Adaptive climatic molecular evolution in wild barley at the *Isa* defense locus. *Proceedings of the National Academy of Sciences of the United States of America,* 104(8), pp. 2773-2778.
Dai, F., Chen, Z.H., Wang, X., Li, Z., Jin, G., Wu, D., Cai, S., Wang, N., Wu, F., Nevo, E. & Zhang, G. (2014). Transcriptome profiling reveals mosaic genomic origins of modern cultivated barley. *Proc Natl Acad Sci U S A,* 111(37), pp. 13403-8.
Doebley, J.F., Gaut, B.S. & Smith, B.D. (2006). The Molecular Genetics of Crop Domestication. *Cell,* 127(7), pp. 1309-1321.

Doolittle, W.F. & Sapienza, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. *Nature,* 284(5757), pp. 601-3.
Ekblom, R. & Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity,* 107(1), pp. 1-15.
Ekblom, R., Slate, J., Horsburgh, G.J., Birkhead, T. & Burke, T. (2012). Comparison between Normalised and Unnormalised 454-Sequencing Libraries for Small-Scale RNA-Seq Studies. *Comparative and Functional Genomics,* 2012, p. 8.
Ellstrand, N.C., Meirmans, P., Rong, J., Bartsch, D., Ghosh, A., de Jong, T.J., Haccou, P., Lu, B.-R., Snow, A.A. & Neal Stewart Jr, C. (2013). Introgression of crop alleles into wild or weedy populations. *Annual Review of Ecology, Evolution, and Systematics,* 44, pp. 325-345.
Eppel, A., Keren, N., Salomon, E., Volis, S. & Rachmilevitch, S. (2013). The response of *Hordeum spontaneum* desert ecotype to drought and excessive light intensity is characterized by induction of O2 dependent photochemical activity and anthocyanin accumulation. *Plant Science,* 201-202, pp. 74-80.
Excoffier, L., Hofer, T. & Foll, M. (2009). Detecting loci under selection in a hierarchically structured population. *Heredity (Edinb),* 103(4), pp. 285-98.
Fang, Z., Gonzales, A.M., Clegg, M.T., Smith, K.P., Muehlbauer, G.J., Steffenson, B.J. & Morrell, P.L. (2014). Two genomic regions contribute disproportionately to geographic differentiation in wild barley. *G3 (Bethesda),* 4(7), pp. 1193-203.
Ferreira de Carvalho, J., Poulain, J., Da Silva, C., Wincker, P., Michon-Coudouel, S., Dheilly, A., Naquin, D., Boutte, J., Salmon, A. & Ainouche, M. (2013). Transcriptome de novo assembly from next-generation sequencing and comparative analyses in the hexaploid salt marsh species *Spartina maritima* and *Spartina alterniflora* (Poaceae). *Heredity (Edinb),* 110(2), pp. 181-93.
Fitzgerald, T.L., Shapter, F.M., McDonald, S., Waters, D.L., Chivers, I.H., Drenth, A., Nevo, E. & Henry, R.J. (2011). Genome diversity in wild grasses under environmental stress. *Proc Natl Acad Sci U S A,* 108(52), pp. 21140-5.
Foll, M. & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics,* 180(2), pp. 977-93.
Fournier-Level, A., Korte, A., Cooper, M.D., Nordborg, M., Schmitt, J. & Wilczek, A.M. (2011). A map of local adaptation in *Arabidopsis thaliana*. *Science,* 334(6052), pp. 86-9.
Franks, S.J. & Hoffmann, A.A. (2012). Genetics of climate change adaptation. *Annu Rev Genet,* 46, pp. 185-208.
Fujii, M., Yokosho, K., Yamaji, N., Saisho, D., Yamane, M., Takahashi, H., Sato, K., Nakazono, M. & Ma, J.F. (2012). Acquisition of aluminium tolerance by modification of a single gene in barley. *Nat Commun,* 3, p. 713.
Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E.M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D.B., Lander, E.S. & Nusbaum, C. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol,* 27(2), pp. 182-9.

Godfray, H.C., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M. & Toulmin, C. (2010). Food security: the challenge of feeding 9 billion people. *Science,* 327(5967), pp. 812-8.
Good, J.M. (2011). Reduced representation methods for subgenomic enrichment and next-generation sequencing. *Methods Mol Biol,* 772, pp. 85-103.
Guo, Y., Ye, F., Sheng, Q., Clark, T. & Samuels, D.C. (2013). Three-stage quality control strategies for DNA re-sequencing data. *Brief Bioinform*.
Hancock, a.M., Brachi, B., Faure, N., Horton, M.W., Jarymowycz, L.B., Sperone, F.G., Toomajian, C., Roux, F. & Bergelson, J. (2011). Adaptation to Climate Across the *Arabidopsis thaliana* Genome. *Science,* 334, pp. 83-86.
He, Z., Zhai, W., Wen, H., Tang, T., Wang, Y., Lu, X., Greenberg, A.J., Hudson, R.R., Wu, C.I. & Shi, S. (2011). Two evolutionary histories in the genome of rice: the roles of domestication genes. *Plos Genetics,* 7(6), p. e1002100.
Hendry, A.P. (2013). Key questions in the genetics and genomics of eco-evolutionary dynamics. *Heredity (Edinb),* 111(6), pp. 456-66.
Henry, R.J. & Nevo, E. (2014). Exploring natural selection to guide breeding for agriculture. *Plant Biotechnol J,* 12(6), pp. 655-62.
Hirsch, C.D., Evans, J., Buell, C.R. & Hirsch, C.N. (2014). Reduced representation approaches to interrogate genome diversity in large repetitive plant genomes. *Brief Funct Genomics*.
Hodges, E., Xuan, Z., Balija, V., Kramer, M., Molla, M.N., Smith, S.W., Middle, C.M., Rodesch, M.J., Albert, T.J., Hannon, G.J. & McCombie, W.R. (2007). Genome-wide in situ exon capture for selective resequencing. *Nat Genet,* 39(12), pp. 1522-7.
Hofinger, B.J., Russell, J.R., Bass, C.G., Baldwin, T., dos Reis, M., Hedley, P.E., Li, Y., Macaulay, M., Waugh, R., Hammond-Kosack, K.E. & Kanyuka, K. (2011). An exceptionally high nucleotide and haplotype diversity and a signature of positive selection for the *eIF4E* resistance gene in barley are revealed by allele mining and phylogenetic analyses of natural populations. *Mol Ecol,* 20(17), pp. 3653-68.
Hohenlohe, P.A., Phillips, P.C. & Cresko, W.A. (2010). Using Population Genomics to Detect Selection in Natural Populations: Key Concepts and Methodological Considerations. *Int J Plant Sci,* 171(9), pp. 1059-1071.
Holsinger, K.E. & Weir, B.S. (2009). FUNDAMENTAL CONCEPTS IN GENETICS Genetics in geographically structured populations: defining, estimating and interpreting F-ST. *Nature Reviews Genetics,* 10(9), pp. 639-650.
Huang, X. & Han, B. (2014). Natural variations and genome-wide association studies in crop plants. *Annu Rev Plant Biol,* 65, pp. 531-51.
Hubner, S., Bdolach, E., Ein-Gedy, S., Schmid, K.J., Korol, A. & Fridman, E. (2013). Phenotypic landscapes: phenological patterns in wild and cultivated barley. *J Evol Biol,* 26(1), pp. 163-74.
Hubner, S., Gunther, T., Flavell, A., Fridman, E., Graner, A., Korol, A. & Schmid, K.J. (2012). Islands and streams: clusters and gene flow in wild barley populations from the Levant. *Mol Ecol,* 21(5), pp. 1115-29.
Hubner, S., Hoffken, M., Oren, E., Haseneyer, G., Stein, N., Graner, A., Schmid, K. & Fridman, E. (2009). Strong correlation of wild barley (*Hordeum*

*spontaneum*) population structure with temperature and precipitation variation. *Molecular Ecology,* 18(7), pp. 1523-1536.
Hufford, M.B., Lubinksy, P., Pyhajarvi, T., Devengenzo, M.T., Ellstrand, N.C. & Ross-Ibarra, J. (2013). The genomic signature of crop-wild introgression in maize. *Plos Genetics,* 9(5), p. e1003477.
IBGS Consortium, Mayer, K.F., Waugh, R., Brown, J.W., Schulman, A., Langridge, P., Platzer, M., Fincher, G.B., Muehlbauer, G.J., Sato, K., Close, T.J., Wise, R.P. & Stein, N. (2012). A physical, genetic and functional sequence assembly of the barley genome. *Nature,* 491(7426), pp. 711-6.
Igartua, E., Moralejo, M., Casas, A., Torres, L. & Molina-Cano, J.-L. (2013). Whole-genome analysis with SNPs from BOPA1 shows clearly defined groupings of Western Mediterranean, Ethiopian, and Fertile Crescent barleys. *Genetic Resources and Crop Evolution,* 60(1), pp. 251-264.
Innan, H. & Kondrashov, F. (2010). The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet,* 11(2), pp. 97-108.
Jaaskelainen, M., Chang, W., Moisy, C. & Schulman, A.H. (2013). Retrotransposon BARE displays strong tissue-specific differences in expression. *New Phytol,* 200(4), pp. 1000-8.
Jaaskelainen, M., Mykkanen, A.H., Arna, T., Vicient, C.M., Suoniemi, A., Kalendar, R., Savilahti, H. & Schulman, A.H. (1999). Retrotransposon BARE-1: expression of encoded proteins and formation of virus-like particles in barley cells. *Plant J,* 20(4), pp. 413-22.
Jia, Q., Zhang, J., Westcott, S., Zhang, X.Q., Bellgard, M., Lance, R. & Li, C. (2009). GA-20 oxidase as a candidate for the semidwarf gene sdw1/denso in barley. *Funct Integr Genomics,* 9(2), pp. 255-62.
Jilal, A., Grando, S., Henry, R.J., Lee, L.S., Rice, N., Hill, H., Baum, M. & Ceccarelli, S. (2008). Genetic diversity of ICARDA's worldwide barley landrace collection. *Genetic Resources and Crop Evolution,* 55(8), pp. 1221-1230.
Juenger, T.E. (2013). Natural variation and genetic constraints on drought tolerance. *Curr Opin Plant Biol,* 16(3), pp. 274-81.
Jump, A.S. & Peñuelas, J. (2005). Running to stand still: adaptation and the response of plants to rapid climate change. *Ecol Lett,* 8(9), pp. 1010-1020.
Kaessmann, H. (2010). Origins, evolution, and phenotypic impact of new genes. *Genome Res,* 20(10), pp. 1313-26.
Kaessmann, H., Vinckenbosch, N. & Long, M. (2009). RNA-based gene duplication: mechanistic and evolutionary insights. *Nat Rev Genet,* 10(1), pp. 19-31.
Kalendar, R., Tanskanen, J., Immonen, S., Nevo, E. & Schulman, A.H. (2000). Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. *Proceedings of the National Academy of Sciences of the United States of America,* 97(12), pp. 6603-6607.
Keane, T.M., Wong, K. & Adams, D.J. (2013). RetroSeq: transposable element discovery from next-generation sequencing data. *Bioinformatics,* 29(3), pp. 389-90.

Kejnovsky, E., Hawkins, J.S. & Feschotte, C. (2012). Plant transposable elements: biology and evolution. In: *Plant Genome Diversity Volume 1* Springer, pp. 17-34.

Kiani, S., Akhunova, A. & Akhunov, E. (2013). Application of next-generation sequencing technologies for genetic diversity analysis in cereals. In: *Cereal Genomics II* Springer, pp. 77-99.

Kliebenstein, D.J., West, M.A., van Leeuwen, H., Kim, K., Doerge, R.W., Michelmore, R.W. & St Clair, D.A. (2006). Genomic survey of gene expression diversity in *Arabidopsis thaliana*. *Genetics,* 172(2), pp. 1179-89.

Koenig, D., Jimenez-Gomez, J.M., Kimura, S., Fulop, D., Chitwood, D.H., Headland, L.R., Kumar, R., Covington, M.F., Devisetty, U.K., Tat, A.V., Tohge, T., Bolger, A., Schneeberger, K., Ossowski, S., Lanz, C., Xiong, G., Taylor-Teeples, M., Brady, S.M., Pauly, M., Weigel, D., Usadel, B., Fernie, A.R., Peng, J., Sinha, N.R. & Maloof, J.N. (2013). Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proc Natl Acad Sci U S A,* 110(28), pp. E2655-62.

Kofler, R., Betancourt, A.J. & Schlotterer, C. (2012). Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in *Drosophila melanogaster*. *Plos Genetics,* 8(1), p. e1002487.

Komatsuda, T., Maxim, P., Senthil, N. & Mano, Y. (2004). High-density AFLP map of nonbrittle rachis 1 (btr1) and 2 (btr2) genes in barley (*Hordeum vulgare* L.). *Theor Appl Genet,* 109(5), pp. 986-95.

Komatsuda, T., Pourkheirandish, M., He, C., Azhaguvel, P., Kanamori, H., Perovic, D., Stein, N., Graner, A., Wicker, T., Tagiri, A., Lundqvist, U., Fujimura, T., Matsuoka, M., Matsumoto, T. & Yano, M. (2007). Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. *Proc Natl Acad Sci U S A,* 104(4), pp. 1424-9.

Kumar, S. & Blaxter, M.L. (2010). Comparing de novo assemblers for 454 transcriptome data. *Bmc Genomics,* 11, p. 571.

Lai, J., Li, R., Xu, X., Jin, W., Xu, M., Zhao, H., Xiang, Z., Song, W., Ying, K., Zhang, M., Jiao, Y., Ni, P., Zhang, J., Li, D., Guo, X., Ye, K., Jian, M., Wang, B., Zheng, H., Liang, H., Zhang, X., Wang, S., Chen, S., Li, J., Fu, Y., Springer, N.M., Yang, H., Wang, J., Dai, J., Schnable, P.S. & Wang, J. (2010). Genome-wide patterns of genetic variation among elite maize inbred lines. *Nat Genet,* 42(11), pp. 1027-30.

Langridge, P. & Fleury, D. (2011). Making the most of 'omics' for crop breeding. *Trends Biotechnol,* 29(1), pp. 33-40.

Lawniczak, M.K., Holloway, A.K., Begun, D.J. & Jones, C.D. (2008). Genomic analysis of the relationship between gene expression variation and DNA polymorphism in *Drosophila simulans*. *Genome Biol,* 9(8), p. R125.

Lee, H.C., Lai, K., Lorenc, M.T., Imelfort, M., Duran, C. & Edwards, D. (2012). Bioinformatics tools and databases for analysis of next-generation sequence data. *Brief Funct Genomics,* 11(1), pp. 12-24.

Linde-Laursen, I.B., Bothmer, R.V. & Jacobsen, N. (2008). Relationships in the genus *Hordeum*: Giemsa C-banded karyotypes. *Hereditas,* 116, pp. 111-116.

Lisch, D. (2013). How important are transposons for plant evolution? *Nat Rev Genet,* 14(1), pp. 49-61.

Long, M., VanKuren, N.W., Chen, S. & Vibranovski, M.D. (2013). New gene evolution: little did we know. *Annu Rev Genet,* 47, pp. 307-33.
Lundqvist, A. (1962). SELF-INCOMPATIBILITY IN DIPLOID HORDEUM BULBOSUM L. *Hereditas,* 48(1-2), pp. 138-152.
Manninen, I. & Schulman, A.H. (1993). BARE-1, a copia-like retroelement in barley (*Hordeum vulgare* L.). *Plant Mol Biol,* 22(5), pp. 829-46.
Marroni, F., Pinosio, S. & Morgante, M. (2014). Structural variation and genome complexity: is dispensable really dispensable? *Curr Opin Plant Biol,* 18C, pp. 31-36.
Martin, J.A. & Wang, Z. (2011). Next-generation transcriptome assembly. *Nat Rev Genet,* 12(10), pp. 671-82.
Mascher, M., Richmond, T.A., Gerhardt, D.J., Himmelbach, A., Clissold, L., Sampath, D., Ayling, S., Steuernagel, B., Pfeifer, M., D'Ascenzo, M., Akhunov, E.D., Hedley, P.E., Gonzales, A.M., Morrell, P.L., Kilian, B., Blattner, F.R., Scholz, U., Mayer, K.F., Flavell, A.J., Muehlbauer, G.J., Waugh, R., Jeddeloh, J.A. & Stein, N. (2013). Barley whole exome capture: a tool for genomic research in the genus *Hordeum* and beyond. *The Plant Journal,* 76(3), pp. 494-505.
Matsumoto, T., Tanaka, T., Sakai, H., Amano, N., Kanamori, H., Kurita, K., Kikuta, A., Kamiya, K., Yamamoto, M., Ikawa, H., Fujii, N., Hori, K., Itoh, T. & Sato, K. (2011). Comprehensive Sequence Analysis of 24,783 Barley Full-Length cDNAs Derived from 12 Clone Libraries. *Plant Physiology,* 156(1), pp. 20-8.
Mazaheri, M., Kianian, P., Mergoum, M., Valentini, G.L., Seetan, R., Pirseyedi, S.M., Kumar, A., Gu, Y.Q., Stein, N. & Kubaláková, M. (2014). Transposable element junctions in marker development and genomic characterization of barley. *The Plant Genome,* 7(1).
McClintock, B. (1956). Controlling elements and the gene. *Cold Spring Harb Symp Quant Biol,* 21, pp. 197-216.
Merila, J. & Hendry, A.P. (2014). Climate change, adaptation, and phenotypic plasticity: the problem and the evidence. *Evol Appl,* 7(1), pp. 1-14.
Messer, P.W. & Petrov, D.A. (2013). Population genomics of rapid adaptation by soft selective sweeps. *Trends Ecol Evol,* 28(11), pp. 659-69.
Meyer, R.S., DuVal, A.E. & Jensen, H.R. (2012). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytol,* 196(1), pp. 29-48.
Meyer, R.S. & Purugganan, M.D. (2013). Evolution of crop species: genetics of domestication and diversification. *Nat Rev Genet,* 14(12), pp. 840-52.
Middleton, C.P., Stein, N., Keller, B., Kilian, B. & Wicker, T. (2012). Comparative analysis of genome composition in Triticeae reveals strong variation in transposable element dynamics and nucleotide diversity. *Plant J*.
Mirouze, M. & Vitte, C. (2014). Transposable elements, a treasure trove to decipher epigenetic variation: insights from *Arabidopsis* and crop epigenomes. *J Exp Bot,* 65(10), pp. 2801-12.
Molina-Cano, J.L., Russell, J.R., Moralejo, M.A., Escacena, J.L., Arias, G. & Powell, W. (2005). Chloroplast DNA microsatellite analysis supports a polyphyletic origin for barley. *Theor Appl Genet,* 110(4), pp. 613-9.

Morey, M., Fernandez-Marmiesse, A., Castineiras, D., Fraga, J.M., Couce, M.L. & Cocho, J.A. (2013). A glimpse into past, present, and future DNA sequencing. *Molecular Genetics and Metabolism,* 110(1-2), pp. 3-24.

Morgante, M., Brunner, S., Pea, G., Fengler, K., Zuccolo, A. & Rafalski, A. (2005). Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. *Nature Genetics,* 37(9), pp. 997-1002.

Morgante, M., De Paoli, E. & Radovic, S. (2007). Transposable elements and the plant pan-genomes. *Curr Opin Plant Biol,* 10(2), pp. 149-55.

Morrell, P.L., Buckler, E.S. & Ross-Ibarra, J. (2011). Crop genomics: advances and applications. *Nat Rev Genet,* 13(2), pp. 85-96.

Morrell, P.L. & Clegg, M.T. (2007). Genetic evidence for a second domestication of barley (*Hordeum vulgare*) east of the Fertile Crescent. *Proceedings of the National Academy of Sciences of the United States of America,* 104(9), pp. 3289-3294.

Morrell, P.L., Lundy, K.E. & Clegg, M.T. (2003). Distinct geographic patterns of genetic diversity are maintained in wild barley (*Hordeum vulgare* ssp *spontaneum*) despite migration. *Proceedings of the National Academy of Sciences of the United States of America,* 100(19), pp. 10812-10817.

Morrell, P.L., Toleno, D.M., Lundy, K.E. & Clegg, M.T. (2005). Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite high rates of self-fertilization. *Proc Natl Acad Sci U S A,* 102(7), pp. 2442-7.

Munoz-Amatriain, M., Cuesta-Marcos, A., Hayes, P.M. & Muehlbauer, G.J. (2014). Barley genetic variation: implications for crop improvement. *Brief Funct Genomics*.

Munoz-Amatriain, M., Eichten, S.R., Wicker, T., Richmond, T.A., Mascher, M., Steuernagel, B., Scholz, U., Ariyadasa, R., Spannagl, M., Nussbaumer, T., Mayer, K.F., Taudien, S., Platzer, M., Jeddeloh, J.A., Springer, N.M., Muehlbauer, G.J. & Stein, N. (2013). Distribution, functional impact, and origin mechanisms of copy number variation in the barley genome. *Genome Biol,* 14(6), p. R58.

Nair, S.K., Wang, N., Turuspekov, Y., Pourkheirandish, M., Sinsuwongwat, S., Chen, G.X., Sameri, M., Tagiri, A., Honda, I., Watanabe, Y., Kanamori, H., Wicker, T., Stein, N., Nagamura, Y., Matsumoto, T. & Komatsuda, T. (2010). Cleistogamous flowering in barley arises from the suppression of microRNA-guided HvAP2 mRNA cleavage. *Proceedings of the National Academy of Sciences of the United States of America,* 107(1), pp. 490-495.

Narum, S.R. & Hess, J.E. (2011). Comparison of F(ST) outlier tests for SNP loci under selection. *Molecular Ecology Resources,* 11 Suppl 1, pp. 184-94.

Nevo, E. (2006). Genome evolution of wild cereal diversity and prospects for crop improvement. *Plant Genetic Resources Characterization and Utilization,* 4(1), pp. ISSN 1479-2621(print)|1479-263x(electronic).

Nielsen, R. (2005). Molecular signatures of natural selection. *Annu Rev Genet,* 39, pp. 197-218.

Nielsen, R., Paul, J.S., Albrechtsen, A. & Song, Y.S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nat Rev Genet,* 12(6), pp. 443-51.

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M.J., Clark, A.G. & Bustamante, C. (2005). Genomic scans for selective sweeps using SNP data. *Genome Research,* 15, pp. 1566-75.
Nosil, P., Funk, D.J. & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Mol Ecol,* 18(3), pp. 375-402.
Olsen, K.M. & Wendel, J.F. (2013). A bountiful harvest: genomic insights into crop domestication phenotypes. *Annu Rev Plant Biol,* 64, pp. 47-70.
Orabi, J., Backes, G., Wolday, A., Yahyaoui, A. & Jahoor, A. (2007). The Horn of Africa as a centre of barley diversification and a potential domestication site. *Theoretical and Applied Genetics,* 114(6), pp. 1117-1127.
Orgel, L.E., Crick, F. & Sapienza, C. (1980). Selfish DNA.
Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J. & De Meester, L. (2013). Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Mol Ecol,* 22(24), pp. 5983-99.
Piffanelli, P., Ramsay, L., Waugh, R., Benabdelmouna, A., D'Hont, A., Hollricher, K., Jorgensen, J.H., Schulze-Lefert, P. & Panstruga, R. (2004). A barley cultivation-associated polymorphism conveys resistance to powdery mildew. *Nature,* 430(7002), pp. 887-891.
Pourkheirandish, M. & Komatsuda, T. (2007). The importance of barley genetics and domestication in a global perspective. *Annals of Botany,* 100(5), pp. 999-1008.
Pritchard, J.K. & Di Rienzo, A. (2010). Adaptation - not by sweeps alone. *Nat Rev Genet,* 11(10), pp. 665-7.
Pybus, O., Shapiro, B., Lemey, P., Salemi, M. & Vandamme, A. (2009). Natural selection and adaptation of molecular sequences. *The phylogenetic handbook*, pp. 406-418.
Rebollo, R., Romanish, M.T. & Mager, D.L. (2012). Transposable elements: an abundant and natural source of regulatory sequences for host genes. *Annu Rev Genet,* 46, pp. 21-42.
Rensing, S.A. (2014). Gene duplication as a driver of plant morphogenetic evolution. *Current Opinion in Plant Biology,* 17(0), pp. 43-48.
Russell, J., Dawson, I.K., Flavell, A.J., Steffenson, B., Weltzien, E., Booth, A., Ceccarelli, S., Grando, S. & Waugh, R. (2011). Analysis of >1000 single nucleotide polymorphisms in geographically matched samples of landrace and wild barley indicates secondary contact and chromosome-level differences in diversity around domestication genes. *New Phytol,* 191(2), pp. 564-78.
Russell, J., van Zonneveld, M., Dawson, I.K., Booth, A., Waugh, R. & Steffenson, B. (2014). Genetic Diversity and Ecological Niche Modelling of Wild Barley: Refugia, Large-Scale Post-LGM Range Expansion and Limited Mid-Future Climate Threats? *Plos One,* 9(2), p. e86021.
Rutter, M.T., Cross, K.V. & Van Woert, P.A. (2012). Birth, death and subfunctionalization in the *Arabidopsis* genome. *Trends Plant Sci,* 17(4), pp. 204-12.
Saisho, D. & Purugganan, M.D. (2007). Molecular phylogeography of domesticated barley traces expansion of agriculture in the Old World. *Genetics,* 177(3), pp. 1765-76.

Sakuma, S., Pourkheirandish, M., Hensel, G., Kumlehn, J., Stein, N., Tagiri, A., Yamaji, N., Ma, J.F., Sassa, H., Koba, T. & Komatsuda, T. (2013). Divergence of expression pattern contributed to neofunctionalization of duplicated HD-Zip I transcription factor in barley. *New Phytol,* 197(3), pp. 939-48.

Sakuma, S., Salomon, B. & Komatsuda, T. (2011). The domestication syndrome genes responsible for the major changes in plant form in the Triticeae crops. *Plant Cell Physiol,* 52(5), pp. 738-49.

Salamini, F., Ozkan, H., Brandolini, A., Schafer-Pregl, R. & Martin, W. (2002). Genetics and geography of wild cereal domestication in the Near East. *Nature Reviews Genetics,* 3(6), pp. 429-441.

Savolainen, O., Lascoux, M. & Merila, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics,* 14(11), pp. 807-820.

Saxena, R.K., Edwards, D. & Varshney, R.K. (2014). Structural variations in plant genomes. *Brief Funct Genomics*.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F.S., Pasternak, S., Liang, C.Z., Zhang, J.W., Fulton, L., Graves, T.A., Minx, P., Reily, A.D., Courtney, L., Kruchowski, S.S., Tomlinson, C., Strong, C., Delehaunty, K., Fronick, C., Courtney, B., Rock, S.M., Belter, E., Du, F.Y., Kim, K., Abbott, R.M., Cotton, M., Levy, A., Marchetto, P., Ochoa, K., Jackson, S.M., Gillam, B., Chen, W.Z., Yan, L., Higginbotham, J., Cardenas, M., Waligorski, J., Applebaum, E., Phelps, L., Falcone, J., Kanchi, K., Thane, T., Scimone, A., Thane, N., Henke, J., Wang, T., Ruppert, J., Shah, N., Rotter, K., Hodges, J., Ingenthron, E., Cordes, M., Kohlberg, S., Sgro, J., Delgado, B., Mead, K., Chinwalla, A., Leonard, S., Crouse, K., Collura, K., Kudrna, D., Currie, J., He, R.F., Angelova, A., Rajasekar, S., Mueller, T., Lomeli, R., Scara, G., Ko, A., Delaney, K., Wissotski, M., Lopez, G., Campos, D., Braidotti, M., Ashley, E., Golser, W., Kim, H., Lee, S., Lin, J.K., Dujmic, Z., Kim, W., Talag, J., Zuccolo, A., Fan, C., Sebastian, A., Kramer, M., Spiegel, L., Nascimento, L., Zutavern, T., Miller, B., Ambroise, C., Muller, S., Spooner, W., Narechania, A., Ren, L.Y., Wei, S., Kumari, S., Faga, B., Levy, M.J., McMahan, L., Van Buren, P., Vaughn, M.W., Ying, K., Yeh, C.T., Emrich, S.J., Jia, Y., Kalyanaraman, A., Hsia, A.P., Barbazuk, W.B., Baucom, R.S., Brutnell, T.P., Carpita, N.C., Chaparro, C., Chia, J.M., Deragon, J.M., Estill, J.C., Fu, Y., Jeddeloh, J.A., Han, Y.J., Lee, H., Li, P.H., Lisch, D.R., Liu, S.Z., Liu, Z.J., Nagel, D.H., McCann, M.C., SanMiguel, P., Myers, A.M., Nettleton, D., Nguyen, J., Penning, B.W., Ponnala, L., Schneider, K.L., Schwartz, D.C., Sharma, A., Soderlund, C., Springer, N.M., Sun, Q., Wang, H., Waterman, M., Westerman, R., Wolfgruber, T.K., Yang, L.X., Yu, Y., Zhang, L.F., Zhou, S.G., Zhu, Q., Bennetzen, J.L., Dawe, R.K., Jiang, J.M., Jiang, N., Presting, G.G., Wessler, S.R., Aluru, S., Martienssen, R.A., Clifton, S.W., McCombie, W.R., Wing, R.A. & Wilson, R.K. (2009). The B73 Maize Genome: Complexity, Diversity, and Dynamics. *Science,* 326(5956), pp. 1112-1115.

Schoville, S.D., Bonin, A., François, O., Lobreaux, S., Melodelima, C. & Manel, S. (2012). Adaptive genetic variation on the landscape: methods and cases. *Annual Review of Ecology, Evolution, and Systematics,* 43, pp. 23-43.

Schulman, A.H. (2012). Hitching a ride: nonautonomous retrotransposons and parasitism as a lifestyle. In: *Plant Transposable Elements* Springer, pp. 71-88.

Schulman, A.H. (2013). Retrotransposon replication in plants. *Curr Opin Virol,* 3(6), pp. 604-14.
Slatkin, M. (1993). Isolation by Distance in Equilibrium and Non-Equilibrium Populations. *Evolution,* 47(1), pp. 264-279.
Slotkin, R.K., Nuthikattu, S. & Jiang, N. (2012). The impact of transposable elements on gene and genome evolution. In: *Plant Genome Diversity Volume 1* Springer, pp. 35-58.
Soleimani, V.D., Baum, B.R. & Johnson, D.A. (2006). Quantification of the retrotransposon BARE-1 reveals the dynamic nature of the barley genome. *Genome,* 49(4), pp. 389-396.
Stapley, J., Reger, J., Feulner, P.G., Smadja, C., Galindo, J., Ekblom, R., Bennison, C., Ball, A.D., Beckerman, A.P. & Slate, J. (2010). Adaptation genomics: the next generation. *Trends Ecol Evol,* 25(12), pp. 705-12.
Strasburg, J.L., Sherman, N.A., Wright, K.M., Moyle, L.C., Willis, J.H. & Rieseberg, L.H. (2012). What can patterns of differentiation across plant genomes tell us about adaptation and speciation? *Philosophical Transactions of the Royal Society B-Biological Sciences,* 367(1587), pp. 364-373.
Suoniemi, A., Narvanto, A. & Schulman, A.H. (1996). The BARE-1 retrotransposon is transcribed in barley from an LTR promoter active in transient assays. *Plant Mol Biol,* 31(2), pp. 295-306.
Suprunova, T., Krugman, T., Distelfeld, A., Fahima, T., Nevo, E. & Korol, A. (2007). Identification of a novel gene (Hsdr4) involved in water-stress tolerance in wild barley. *Plant Molecular Biology,* 64(1-2), pp. 17-34.
Sutton, T., Baumann, U., Hayes, J., Collins, N.C., Shi, B.J., Schnurbusch, T., Hay, A., Mayo, G., Pallotta, M., Tester, M. & Langridge, P. (2007). Boron-toxicity tolerance in barley arising from efflux transporter amplification. *Science,* 318(5855), pp. 1446-9.
Suzuki, Y. (2010). Statistical methods for detecting natural selection from genomic data. *Genes Genet Syst,* 85(6), pp. 359-76.
Tajima, F. (1989). Statistical-Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics,* 123(3), pp. 585-595.
Taketa, S., Amano, S., Tsujino, Y., Sato, T., Saisho, D., Kakeda, K., Nomura, M., Suzuki, T., Matsumoto, T., Sato, K., Kanamori, H., Kawasaki, S. & Takeda, K. (2008). Barley grain with adhering hulls is controlled by an ERF family transcription factor gene regulating a lipid biosynthesis pathway. *Proceedings of the National Academy of Sciences,* 105(10), pp. 4062-4067.
Tanskanen, J.A., Sabot, F., Vicient, C. & Schulman, A.H. (2007). Life without GAG: the BARE-2 retrotransposon as a parasite's parasite. *Gene,* 390(1-2), pp. 166-74.
Tester, M. & Langridge, P. (2010). Breeding technologies to increase crop production in a changing world. *Science,* 327(5967), pp. 818-22.
Thiel, J., Hollmann, J., Rutten, T., Weber, H., Scholz, U. & Weschke, W. (2012). 454 Transcriptome sequencing suggests a role for two-component signalling in cellularization and differentiation of barley endosperm transfer cells. *Plos One,* 7(7), p. e41867.

Turner, A., Beales, J., Faure, S., Dunford, R.P. & Laurie, D.A. (2005). The pseudo-response regulator Ppd-H1 provides adaptation to photoperiod in barley. *Science,* 310(5750), pp. 1031-1034.
Verslues, P.E. & Juenger, T.E. (2011). Drought, metabolites, and *Arabidopsis* natural variation: a promising combination for understanding adaptation to water-limited environments. *Current Opinion in Plant Biology,* 14(3), pp. 240-245.
Via, S. (2012). Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philos Trans R Soc Lond B Biol Sci,* 367(1587), pp. 451-60.
Vicient, C.M., Kalendar, R. & Schulman, A.H. (2005). Variability, recombination, and mosaic evolution of the barley BARE-1 retrotransposon. *Journal of Molecular Evolution,* 61(3), pp. 275-291.
Vicient, C.M., Suoniemi, A., Anamthawat-Jonsson, K., Tanskanen, J., Beharav, A., Nevo, E. & Schulman, A.H. (1999). Retrotransposon BARE-1 and Its Role in Genome Evolution in the Genus *Hordeum*. *Plant Cell,* 11(9), pp. 1769-1784.
Vitte, C., Fustier, M.A., Alix, K. & Tenaillon, M.I. (2014). The bright side of transposons in crop evolution. *Brief Funct Genomics*.
Vitti, J.J., Grossman, S.R. & Sabeti, P.C. (2013). Detecting natural selection in genomic data. *Annu Rev Genet,* 47, pp. 97-120.
Volis, S., Yakubov, B., Shulgina, I., Ward, D., Zur, V. & Mendlinger, S. (2001). Tests for adaptive RAPD variation in population genetic structure of wild barley, *Hordeum spontaneum* Koch. *Biological Journal of the Linnean Society,* 74(3), pp. 289-303.
Volis, S., Zaretsky, M. & Shulgina, I. (2010). Fine-scale spatial genetic structure in a predominantly selfing plant: role of seed and pollen dispersal. *Heredity,* 105(4), pp. 384-393.
Wang, Q. & Dooner, H.K. (2006). Remarkable variation in maize genome structure inferred from haplotype diversity at the bz locus. *Proc Natl Acad Sci U S A,* 103(47), pp. 17644-9.
Watterson, G.A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology,* 7(2), pp. 256-76.
Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M. & Panaud, O. (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics,* 8(12), pp. 973-982.
Wicker, T., Taudien, S., Houben, A., Keller, B., Graner, A., Platzer, M. & Stein, N. (2009). A whole-genome snapshot of 454 sequences exposes the composition of the barley genome and provides evidence for parallel evolution of genome size in wheat and barley. *Plant Journal,* 59(5), pp. 712-722.
Wolf, J.B. (2013). Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. *Molecular Ecology Resources,* 13(4), pp. 559-72.
Yan, L., Fu, D., Li, C., Blechl, A., Tranquilli, G., Bonafede, M., Sanchez, A., Valarik, M., Yasuda, S. & Dubcovsky, J. (2006). The wheat and barley vernalization gene VRN3 is an orthologue of FT. *Proc Natl Acad Sci U S A,* 103(51), pp. 19581-6.

Yang, Z., Zhang, T., Bolshoy, A., Beharav, A. & Nevo, E. (2009). Adaptive microclimatic structural and expressional dehydrin 1 evolution in wild barley, *Hordeum spontaneum*, at 'Evolution Canyon', Mount Carmel, Israel. *Molecular Ecology,* 18(9), pp. 2063-2075.

Yang, Z.J., Zhang, T., Li, G.R. & Nevo, E. (2011). Adaptive microclimatic evolution of the dehydrin 6 gene in wild barley at "Evolution Canyon", Israel. *Genetica,* 139(11-12), pp. 1429-1438.

Zhang, W., Chen, J., Yang, Y., Tang, Y., Shang, J. & Shen, B. (2011). A practical comparison of de novo genome assembly software tools for next-generation sequencing technologies. *Plos One,* 6(3), p. e17915.

Zhuang, J., Wang, J., Theurkauf, W. & Weng, Z. (2014). TEMP: a computational method for analyzing transposable element polymorphism in populations. *Nucleic Acids Res,* 42(11), pp. 6826-38.

Zmienko, A., Samelak, A., Kozlowski, P. & Figlerowicz, M. (2014). Copy number polymorphism in plant genomes. *Theor Appl Genet,* 127(1), pp. 1-18.

Zohary, D., Hopf, M. & Weiss, E. (2012). *Domestication of plants in the Old World : the origin and spread of domesticated plants in south-west Asia, Europe, and the Mediterranean Basin*. Fourth edition. ed. Oxford: Oxford University Press.

# Acknowledgements

I would like to express my sincere gratitude to my advisor Docent Anna Westerbergh for her endless assistances and care and advice. Thank you Anna; for your encouragement, treatments and guidance throughout my PhD studies. I am very grateful for your critical readings and valuable comments and suggestions on my thesis, and for all the discussions we had.

I am very grateful to my supervisor Prof. Karl Schmid, for your guidance, supports and all your efforts. I am very thankful for having given me the opportunity to look and perform my works in the best possible directions I can. Thank you indeed for creating the opportunity to visit the natural habitat of wild barley in Israel, and for all your supports during our stay and for showing me that great environment.

My deepest gratitude to my mom, Alem: to whom I owe my life for her blessings, love, care and encouragements. Your life was for us and I would not have been to this level without your endless devotion! My utmost grateful to my lovely better-half Hiwot Amenu, I am blessed to have you! You are my boat throughout my academic and daily life. Your great and endless love, care and assistance since 1997 can speak loudly and fruitfully and brought me to this level. You dedicated your life for me and our life. You patently accepted my long hours in front of the computer and care for me and our daughter Soliana. Without you, all was unthinkable and will not! My little-angle Soliana - thank you for your patency and understanding.

I am very thankful to my brother Workneh who will soon finish his PhD, for all your supports and help. Our daughter are so lucky to have you, your presence covers my absence for her. I owe my deepest thanks to my father Bedada Chala. And Asku, Amenu, Anchuye, Dawi, Senaf, Demu, Shewarega, Qes Mekonnen, Tesfaye, Leta, Workwoha, Dr. Bezuayehu, Derje, and all my families, thank you for your encouragements throughout my study periods.

I would like to extend my sincere thanks to all people in my department for all the supports. I am very thankful to administration staff - Lotta, Mona, Birgitta,

Monica and Qing, and technical staff - Ingrid, Cecilia, Urban and Per, and department head – Prof. Eva.

My deepest thanks to Björn – for his genuine, timely and endless computer administration and supports. Thank you indeed Björn! I would like to express my sincere thanks to Niclas, Sara and Sun, for their critical evaluation, discussions and suggestions on my works.

I would extend my special thanks to my close colleagues in the department: Mohammad – for your kind, friendly and endless supports and help in my experiments and for your comments and suggestions on my thesis. Jim – thank you for your great friendship, endless supports and all the great discussions we had – be it science or others. We had great time together during teaching and while sharing office. Selcuk – we have started and ending our study sharing one office, thank you indeed for the nice times we had, all your help, the discussions and being the best office mate. Shah – thanks for your help, being a nice office colleague, and all the great discussions and times. Estelle and Ulrike – thank you for the nice time and your help; Philip – thanks man for the nice time, chats and discussions. Sultana, Malin, Pascal, Jia, Isabella and Anna – thank you for the good time we had during teaching courses.

I would like to thank our collaborator Prof. Abraham Korol, Sariel Hubner Eyal Fridman from Israel. I am thankful to Prof. Abraham Korol and Sariel – for all the supports in arranging my visit to Israel, and Sariel – for his kind guidance and supports in driving us throughout Israel with my supervisor to visit and familiarize myself with the extraordinary plant – wild barley and its natural habitat.

Lab members of Prof. Karl J. Schmid at University of Hohenheim, Stuttgart, Germany – thank you all for all the supports you gave me. It is unfortunate that I could not join you even for a few days! Secretary Bärbel Hessenauer – I would like to covey my special thanks to you; your support begins before I left my home-land and officially started my study. Fabian, Dieter and Thomas Brueckner – thanks a lot for your kind supports with the servers to smoothly perform all my analyses. Thomas Müller – thanks for help with the programs and for the nice discussions. Ivan – thanks a lot for your great help with our experiments. Elisabeth – my special thanks to you for your all efforts in sequencing my samples. Wosene and Trosten – thank you for all the supports.

I would like to convey my utmost and heartfelt thanks to Dr. Ararso Etana and his family – Ayelech Etefa and their great son Yohannes. My family and I can never forget the Fatherly and Motherly care, guidance, supports and affection you gave us through-out our stay in Sweden; you have been beyond our imagination! Having you around, my family and I feel at home in Uppsala. Your special cares are unforgettable in our life!

My deepest and heartfelt thanks to Dr. Teshoma, Argane and their great son Michael for the great time we had together in Uppsala and for being the best and intimate family-friend since then. Hope one day we will have a new time together.

My dear bro Muluneh Minta – my sincere gratefulness for all your encouragements and kind supports to us and my all families back-home on behalf of us. Thanks indeed for being there for us, we are extremely lucky to have you there! Keep going; you will soon defend your PhD. Great man Dr. Belay Moreda, thank you for everything - your great hospitality and supports will be unforgettable. I am very thankful to my best friends Abebe Chindi, Adugna Abera, Ibsa Fite Tesfaye Disasa, and all you families for your great friendship, hospitality and all helps and advices. All my friends and colleagues at EIAR Holetta Research Center thank you for all!

I would like to express my special grateful to Gezehagn Girma, his wife Tsega and their son Yonatan, for the great time we had together, and Taddesse Worku, his wife Belen and their son Emnet, for your advice, hospitality and friendship. My special thanks to Meaza Getachew and his family for his all encouragements and friendship.

*Help is not always direct …, rather in many different ways. A few seconds help matters a lot as it can save a day to week struggling in the lab and behind the computer screen. Beyond that, help to someone diffuse and reached to others – just like 'genetic-hitchhiking'. Thank you all!*

*Above all, Glory be to God for his daily guidance!*