

COMPARISON OF THE GENETIC DIVERSITY LEVELS OF WHEAT SPECIES
GROWN IN TURKEY & INVESTIGATION OF THE GENOMIC RESPONSES OF
LOCAL EINKORN WHEAT TO HEAT STRESS



by

Merve Sıvacı

B.S. in Molecular Biology and Genetics, Istanbul Technical University, 2011

M.S. in Molecular Biology and Genetics, Boğaziçi University, 2013

Submitted to the Institute of Environmental Sciences
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in
Environmental Sciences

Boğaziçi University

2023

COMPARISON OF THE GENETIC DIVERSITY LEVELS OF WHEAT SPECIES
GROWN IN TURKEY & INVESTIGATION OF THE GENOMIC RESPONSES OF
LOCAL EINKORN WHEAT TO HEAT STRESS

APPROVED BY:

Prof. Dr. Raşit Bilgin
Dissertation Advisor

Prof. Dr. Mahmut Çalışkan

Prof. Dr. Bahar İnce

Prof. Dr. Hakan Özkan

Assist. Prof. Dr. Berat Z. Haznedaroğlu

DATE OF APPROVAL: 24/07/2023

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my dissertation advisor, Prof. Dr. Raşit Bilgin, for his invaluable supervision, guidance, continuous support, and immense knowledge throughout this dissertation. His empathy and understanding have encouraged me in all the time of my academic research and daily life. None of this would have been possible without his help and support.

Besides, I would like to express my appreciation to my dissertation committee, Assist. Prof. Dr. Berat Z. Haznedaroğlu and Prof. Dr. Bahar İnce for their generous time, consideration, evaluation, and guidance during this study, and Prof. Dr. Mahmut Çalışkan, and Prof. Dr. Hakan Özkan for spending their valuable time in evaluating this dissertation.

I would also like to express my appreciation to Assist. Prof. Igor Kryvoruchko, and Assist. Prof. Steven Footitt for letting us use their growth chamber and generously providing their knowledge and expertise about plant growth and RNA isolation. I would also like to thank Dr. Alexey Yanchukov, Dr. Ortaç Çetintaş, and Halil Mert Solak for their support and help with ddRAD analyses, and Öncü Maracı and Cevahir Karaoğlu for their help in obtaining the samples used in the study.

I would deeply thank Dilek Şahin for her friendship and support, especially throughout all ddRAD experiments and data analyses, and Elif İçağasioğlu for her support during RNA-seq data analyses. I would also like to thank Maha Salameh, Onur Doğan, Anıl Örün, Aygün Karaçay, Nehir Sevim, Kaan Bayır, İdil Çağla Ertaş, and Elizabeth Somsen for their friendship, help, and support whenever I needed.

Last but not least, words cannot express my gratitude to my mother, my father, and my brother, for their endless love, patience, and support and for being there for me all the time. I feel really lucky to have them in my life.

Finally, I am grateful to YÖK for providing YÖK 100/2000 PhD scholarship during my dissertation. The data analyses were partially performed at TUBITAK ULAKBİM, High Performance and Grid Computing Center (TRUBA resources) that is also greatly acknowledged. This study was supported by a Boğaziçi University Research Fund grant (Project no: 15Y00R4) to RB.

ABSTRACT

COMPARISON OF THE GENETIC DIVERSITY LEVELS OF WHEAT SPECIES GROWN IN TURKEY & INVESTIGATION OF THE GENOMIC RESPONSES OF LOCAL EINKORN WHEAT TO HEAT STRESS

Global climate is changing rapidly, affecting nearly all organisms, including plants. Wheat is one of the major crops worldwide and, like many other species, is adversely affected by climate change. Therefore, it is important to determine how wheat species respond to stress conditions at the molecular level to understand their adaptive potential for their future breeding and conservation. In this study, through ddRAD sequencing, adaptive potentials of local einkorn, emmer, and bread wheat were compared in terms of their genetic diversity levels. Consequently, three species were clearly separated based on their diversity levels, and the highest genetic distance was found between old (1995) and new (2015) einkorn ($F_{ST} = 0.238923$), indicating ex-situ conservation is more advantageous for einkorn in a changing climate than in-situ conservation. However, no distinction was observed for old and new emmer populations, suggesting the protection of emmer from genetic bottlenecks caused by climate change, most probably because of its ploidy. Through RNA sequencing, old and new einkorn were also compared in terms of their response to heat stress. When the genes differentially expressed between old and new einkorn at stress temperature were investigated, it was found that genes and gene sets involved in energy and stress-responsive protein production were enriched as well as the genes related to the regulation of physiological processes, membrane stability, and photosynthesis, whereas at control temperature a more general response was observed. These findings suggest that transcriptomic responses of einkorn diverged in 20 years of climate change, providing evidence for rapid evolution.

ÖZET

TÜRKİYE’DE YETİŞEN BUĞDAY TÜRLERİNİN GENETİK ÇEŞİTLİLİK SEVİYELERİNİN KARŞILAŞTIRILMASI & YEREL SİYEZ BUĞDAYININ SICAKLIK STRESİNE OLAN GENOMİK CEVAPLARININ ARAŞTIRILMASI

Küresel iklim hızla değişmekte ve neredeyse bitkiler dahil tüm organizmaları etkilemektedir. Buğday, dünya çapında en temel gıda ürünlerinden biridir ve diğer türler gibi o da iklim değişiminden olumsuz bir şekilde etkilenmektedir. Bu yüzden, türlerin adaptasyon potansiyellerini anlamak için stres koşullarına moleküler düzeyde nasıl tepki verdiklerini belirlemek, ileride yapılacak olan yetiştirme ve koruma çalışmaları için oldukça önemlidir. Bu çalışmada, ddRAD dizileme aracılığıyla, yerel siyez, gernik ve ekmeklik buğday türlerinin adaptasyon potansiyelleri, genetik çeşitlilik seviyeleri aracılığıyla karşılaştırılmıştır. Sonuç olarak, üç tür, genetik çeşitlilik seviyeleri baz alınarak üç ayrı gruba net bir şekilde ayrılmıştır ve en fazla genetik uzaklık eski (1995) ve yeni (2015) arasında bulunmuştur ($F_{ST} = 0.238923$). Bu da siyez için, değişen iklim koşullarında, ex-situ korumanın in-situ korumadan daha avantajlı olduğunu göstermektedir. Diğer taraftan, eski ve yeni gernik popülasyonları için herhangi bir ayırım gözlemlenmemiştir ki bu da çok büyük ihtimalle gernik buğdayın ploidisi sayesinde, iklim değişiminin neden olduğu genetik darboğazlardan korunduğunu gösterir. RNA dizileme yoluyla, eski ve yeni siyez aynı zamanda sıcaklık stresine verdikleri tepkiler açısından da karşılaştırılmıştır. Eski ve yeni siyez arasında stres sıcaklığında farklı gen anlatımına sahip genler araştırıldığında, enerji ve stres kaynaklı protein üretimiyle alakalı yollara dahil olan gen ve gen setleriyle, fizyolojik süreçlerin düzenlenmesi, zar stabilitesi ve fotosentez ile ilgili genlerin fazla olduğu, ancak kontrol sıcaklığında daha genel bir cevap verildiği bulunmuştur. Bu bulgular, siyezin transkriptomik tepkilerinin 20 yıllık iklim değişimiyle farklılaştığına ve hızlı bir evrimsel sürece işaret etmektedir.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	iii
ABSTRACT.....	iv
ÖZET.....	v
TABLE OF CONTENTS.....	vi
LIST OF FIGURES.....	ix
LIST OF TABLES.....	xiv
LIST OF SYMBOLS/ABBREVIATIONS.....	xvi
1. INTRODUCTION.....	1
2. LITERATURE REVIEW.....	6
2.1. Global Warming.....	6
2.2. The Impact of Climate Change on Agriculture in Turkey.....	8
2.3. The Genus <i>Triticum</i>	10
2.4. Genetic Diversity of Wheat.....	12
2.5. Heat Stress Response of Plants.....	13
2.6. Genomic Response of Wheat to Heat Stress.....	15
2.7. Double Digest Restriction-Site Associated DNA Sequencing (ddRAD-seq).....	16
2.8. Next Generation RNA Sequencing.....	17
3. MATERIALS AND METHODS.....	18
3.1. Double Digest Restriction-Site Associated DNA Sequencing (ddRAD-seq).....	18
3.1.1. Plant Materials.....	18
3.1.2. DNA Extraction.....	20
3.1.3. Construction of ddRAD Libraries.....	20
3.1.4. Data Analyses.....	22
3.2. Heat Stress Experiment & RNA Sequencing.....	23
3.2.1. Plant Materials.....	23
3.2.2. Heat Stress Experiment.....	24
3.2.3. RNA Extraction.....	26
3.2.4. Library Preparation and Sequencing.....	26
3.2.5. Data Analyses.....	27

4. RESULTS.....	31
4.1. ddRAD-seq.....	31
4.1.1. Demultiplexing and Cleaning of Raw Reads.....	31
4.1.2. De novo Pipeline Parameter Optimization.....	32
4.1.3. De novo Pipeline Running of Einkorn Wheat.....	34
4.1.4. <i>Structure</i> Analysis of Einkorn Wheat.....	39
4.1.5. De novo Pipeline Running of Emmer Wheat.....	46
4.1.6. <i>Structure</i> Analysis of Emmer Wheat.....	49
4.1.7. De novo Pipeline Running of Bread Wheat.....	50
4.1.8. De novo Pipeline Running of All Populations.....	52
4.1.9. <i>Structure</i> Results of All Populations.....	56
4.2. RNA-Seq.....	56
4.2.1. RNA Quality.....	56
4.2.2. Raw Data Statistics.....	57
4.2.3. Transcriptome Assembly Statistics.....	58
4.2.4. Alignment Statistics.....	60
4.2.5. Pre-processing of Aligned Data.....	61
4.2.6. Differential Gene Expression Analysis.....	65
4.2.7. Functional Annotation.....	69
4.2.8. Functional Annotations of Old vs. New Einkorn.....	74
4.2.9. The Top 100 DE Contigs and Their Functional Annotations.....	85
4.2.10. Heat Stress Related Contigs.....	87
5. DISCUSSION.....	89
5.1. ddRAD-seq.....	89
5.2. RNA-seq.....	92
5.2.1. Functional Annotations of All Contigs.....	93
5.2.2. Functional Annotations of the DE Contigs Between Old and New Einkorn.....	95
5.2.3. Enriched GOs and Pathways in the GSEA of the DE Contigs Between Old and New Einkorn.....	97
5.2.4. Heat Stress Related Contigs.....	99
6. CONCLUSIONS AND RECOMMENDATIONS.....	101
REFERENCES.....	103
APPENDIX A: ddRAD-seq.....	134

APPENDIX B: RNA-seq.....147



LIST OF FIGURES

Figure 1.1. The monthly average temperature in Kastamonu for the periods of 1980-1999 and 2000-2015.....	4
Figure 2.1. Map of the Fertile Crescent (dashed red line)	8
Figure 2.2. Hybridization events of wheat species.....	11
Figure 3.1. Sampling locations of wheat populations from 2015 used for ddRAD-seq.....	19
Figure 3.2. The schematic representation of heat stress experiment. The numbers in green boxes indicate the RNA-seq libraries prepared and sequenced.....	27
Figure 3.3. The workflow of the RNA-seq data analyses.....	30
Figure 4.1. Retained reads obtained for each species after cleaning and demultiplexing.....	31
Figure 4.2. Plots of coverage (a) and assembled loci, polymorphic loci, and SNPs (b) varying with the increase of m	33
Figure 4.3. Plots of assembled loci, polymorphic loci and SNPs varying with the increase of $M=n$ (a) and the new polymorphic loci added with each increment of $M=n$ (b).....	34
Figure 4.4. Observed heterozygosity and nucleotide diversity values of einkorn populations calculated for variant positions.....	36
Figure 4.5. Matrices of pairwise F_{ST} values of einkorn populations calculated with default parameters with r80 (a), default parameters without r80 (b), optimized parameters with r80 (c), and optimized parameters without r80 (d).....	39

Figure 4.6. <i>Structure</i> results of the six datasets given in Table 4.7 for the determined K values.....	41
Figure 4.7. <i>Structure</i> results of the second, third, fourth, and fifth datasets (Table 4.4) for K=2.....	42
Figure 4.8. <i>Structure</i> results of the second (above) and sixth (below) datasets showing einkorn populations.....	43
Figure 4.9. Population and year-based <i>Structure</i> results with missing data encoded as -9 (a), and zero (b). The black dashed lines indicate the distinction between years.....	44
Figure 4.10. Batch-based <i>Structure</i> results with missing data encoded as -9 (a), and zero (b). The black dashed lines indicate the distinction between years.	46
Figure 4.11. Observed heterozygosity and nucleotide diversity values of emmer populations calculated for variant positions.....	47
Figure 4.12. Matrix of pairwise F_{ST} values of emmer populations.....	48
Figure 4.13. <i>Structure</i> result of emmer populations defined by the year of collection.....	49
Figure 4.14. <i>Structure</i> results of emmer populations defined by the location of the collection. Populations are both sorted by the year of collection (a) and the location of collection (b)	50
Figure 4.15. Observed heterozygosity and nucleotide diversity values of bread wheat populations calculated for variant positions.....	51
Figure 4.16. Matrix of pairwise F_{ST} values of bread wheat populations.....	51
Figure 4.17. Observed heterozygosity and nucleotide diversity values of all wheat populations calculated for variant positions.	52
Figure 4.18. Matrix of pairwise F_{ST} values of all wheat populations. In population names, EIN indicates einkorn, EMM indicates emmer and BRE indicates bread wheat.....	54

Figure 4.19. PCoA result of all wheat populations. The numbers in squares indicate the names of wheat populations without “WP”	55
Figure 4.20. <i>Structure</i> results of all populations defined by species.	56
Figure 4.21. RIN values recorded by SEM and MacroGen for 36 RNA samples.	57
Figure 4.22. rRNA ratios calculated by SEM and MacroGen for 36 RNA samples.	57
Figure 4.23. BUSCO analysis results.....	59
Figure 4.24. The contig length distribution of the de novo assembly.....	60
Figure 4.25. Sizes of 36 RNA-seq libraries before and after filtering.....	62
Figure 4.26. Expression distributions of the libraries before and after filtering and normalization.....	63
Figure 4.27. The density of log-CPM values of the libraries before and after filtering and normalization. The dotted line indicates the cutoff used for filtering.....	64
Figure 4.28. MDS plots of four einkorn groups for the first three dimensions.	65
Figure 4.29. The numbers of up-regulated and down-regulated contigs with an FDR cut-off of 0.05.....	66
Figure 4.30. MD plots of four comparisons.....	67
Figure 4.31. The numbers of up-regulated and down-regulated contigs with an FDR cut-off of 0.05 and logFC cut-off of ± 1	68
Figure 4.32. Volcano plots of four comparisons. The horizontal orange line indicates the $-\log_{10}(0.05)$, and the vertical orange lines indicate the log fold change $= \pm 1$	69
Figure 4.33. The number of hits obtained after each blast run.....	70

Figure 4.34. The distribution of contigs after Blast, InterProScan (IPS), and EggNOG analyses. The results of IPS and EggNOG were merged with Blast results, respectively.....	71
Figure 4.35. The direct GO counts of all contigs for the biological process (BP), molecular function (MF), and cellular component (CC) categories.....	72
Figure 4.36. The distribution of enzyme codes across all contigs.....	73
Figure 4.37. The EC subclasses within transferases and hydrolases across all contigs.....	73
Figure 4.38. The direct GO counts of biological process (BP) between the old and new einkorn at control (top panel) and stress (bottom panel) temperatures.....	75
Figure 4.39. The direct GO counts of molecular function (MF) between the old and new einkorn at control (top panel) and stress (bottom panel) temperatures.....	76
Figure 4.40. The direct GO counts of cellular component (CC) between the old and new einkorn at control (top panel) and stress (bottom panel) temperatures.....	77
Figure 4.41. The EC classes between the old and new einkorn at control (top panel) and stress (bottom panel) temperatures.....	78
Figure 4.42. The EC subclasses within hydrolases between the old and new einkorn at control (top panel) and stress (bottom panel) temperatures.....	79
Figure 4.43. The enriched GO categories between old and new einkorn after GSEA analyses for the control (top panel) and stress (bottom panel) temperatures.....	80
Figure 4.44. The photosynthesis – antenna proteins pathway enriched at the control temperature.....	83
Figure 4.45. The photosynthesis pathway that was enriched at stress temperature.....	84
Figure 4.46. The diterpenoid pathway that was enriched at stress temperature.....	85

Figure 4.47. The direct GO counts of the top 100 contigs for the biological process (BP), molecular function (MF), and cellular component (CC) categories.....	86
--	----



LIST OF TABLES

Table 3.1. Wheat populations used for ddRAD-seq.....	18
Table 3.2. Sequences of adapters and primers used for ddRAD library construction. xxxxxx represents the barcode, zzzzzz represents the index. Phosphorothiolate bonds are marked by *.....	21
Table 3.3. Sequences of barcodes and indexes embedded in adapter and primer sequences.....	21
Table 3.4. The number of ddRAD libraries sequenced for each population.....	22
Table 3.5. Einkorn populations used in heat stress experiment.....	24
Table 3.6. The numbers of seeds planted and germinated for the two einkorn populations.....	24
Table 3.7. The parameters set for wheat growth in the growth chamber.....	25
Table 4.1. The numbers of loci and SNPs obtained for each m	32
Table 4.2. Loci and SNP numbers, and the number of new polymorphic loci added for each iteration of $M(n)$	32
Table 4.3. Locus and SNP numbers obtained for einkorn with different de novo analyses.....	35
Table 4.4. The information about the datasets used in einkorn's <i>Structure</i> analyses.....	40
Table 4.5. Statistics of the de novo assembly generated by Trinity.....	59
Table 4.6. Statistics of the aligned reads.....	61
Table 4.7. The numbers of blasted, mapped, and annotated contigs.....	70

Table 4.8. The numbers of enriched pathways between old and new einkorn.....	80
Table 4.9. The top ten (top tag and bottom tag) enriched KEGG pathways for control temperature.....	81
Table 4.10. The top ten (top tag and bottom tag) enriched KEGG pathways for stress temperature.....	82
Table 4.11. Blast results of the top 100 DE contigs related to heat stress.....	87
Table 4.12. Blast results of the DE contigs between old and new einkorn at control temperature related to HSPs or HSFs.....	88
Table 5.1. Loci and SNP numbers obtained for de novo analyses.....	89

LIST OF SYMBOLS/ABBREVIATIONS

Symbol/Abbreviation	Explanation
μl	Microliter
μmol	Micromole
m^2	Square meter
mm	Millimeter
2-DE/MS	Two-dimensional gel electrophoresis/mass spectrometry
ABA	Absciscic acid
AFLP	Amplified fragment length polymorphism
APX	Ascorbate peroxidase
AT	Adenine-Thymine
ATP	Adenosine triphosphate
bp	Base pair
BP	Biological process
BR	Broad Range
BRE	Bread wheat
BUSCO	Benchmarking Universal Single-Copy Orthologs
Ca^{+2}	Calcium
CAT	Catalase
CC	Cellular component
cDNA	Complementary DNA
CDPK	Calcium-dependent protein kinase
Clp	Casein lytic proteinase
ClpB3	Casein lytic proteinase B3
CO_2	Carbon dioxide
CPM	Counts per million
CRISPR	Clustered regularly interspaced short palindromic repeats
DArT	Diversity arrays technology
ddRAD-seq	Double digested restriction-site associated DNA sequencing
DE	Differentially expressed
DNA	Deoxyribonucleic acid

dNTP	Deoxyribonucleotide triphosphate
dsDNA	Double stranded DNA
EC	Enzyme code
EIN	Einkorn
EMM	Emmer
ETAE	Ege Tarımsal Araştırma Enstitüsü
FDR	False discovery rate
F _{ST}	Fixation index
GA20ox	Gibberellin 20-oxidase
Gb	Gigabase
GBS	Genotyping by sequencing
GC	Guanine-Cytosine
GO	Gene ontology
GR	Glutathione reductase
GSEA	Gene set enrichment analyses
h	Hour
HS	High Sensitivity
H ₂ O ₂	Hydrogen peroxide
HSA32	Heat-stress-associated 32
HSE	Heat shock element
HSP	Heat shock protein
HSF	Heat stress transcription factor
kb	Kilobase
kDa	Kilodalton
KEGG	Kyoto Encyclopedia of Genes and Genomes
LB	Lithium borate
LHC	Light harvesting chlorophyll a/b binding protein
logFC	Log fold change
MAPK	Mitogen-activated protein kinase
MCMC	Markov Chain Monte Carlo
MD	Mean-difference
MDS	Multidimensional scaling
MF	Molecular function

mRNA	Messenger RNA
NADPH	Nicotinamide adenine dinucleotide phosphate
NES	Normalized enrichment score
NGS	Next generation sequencing
PCoA	Principal coordinate analysis
PCR	Polymerase chain reaction
pH	Potential of hydrogen
POX	Peroxidase
qPCR	Quantitative real-time PCR
RAD	Restriction-site associated DNA
RAD-seq	RAD sequencing
RAPD	Random amplification of polymorphic DNA
RFLP	Restriction fragment length polymorphism
RIN	RNA integrity number
RLK	Receptor-like protein kinase
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
ROS	Reactive oxygen species
rRNA	Ribosomal RNA
sec	Second
SNP	Single nucleotide polymorphism
SOD	Superoxide dismutase
SSR	Simple sequence repeat
T	Temperature
TAE	Tris-acetate-EDTA
TMM	Trimmed mean of M-values
V	Volt
WP	Wheat population

1. INTRODUCTION

Climate change is one of the most emphasized environmental challenges of today and the near future. Since the last century, global temperature has increased approximately by 1.09 °C, and adverse effects of this warming on natural and biological systems have already been recorded worldwide, indicating these systems' sensitivity to changing climate (IPCC, 2022). Considering the predictions of about 1.5-5 °C of temperature increase by the end of the 21st century, it is very likely to observe more adverse impacts in the future (IPCC, 2022).

With the continuous increase in the global population, there is a rising demand for agricultural products, especially for the major crops; wheat, rice, and maize (Perdomo, Conesa, Medrano, Ribas-Carbó, & Galmés, 2015). However, climate change affects the main croplands, negatively impacting agricultural production (Perdomo et al., 2015), culminating in a decrease at the rate of increase in crop yields (Dixon, Braun, Kosina, & Crouch, 2009). While the increase in yield was 2-3 % for the main crops between 1966-1994, it was recorded as 0.9 %, 1.1 %, and 1.8 % for rice, wheat, and maize, respectively, from 1995 through 2005 (Dixon et al., 2009). Likewise, crop yield decreases because of the shortening of the grain-filling period or the detriment of pollen formation with temperature increases (Lawlor & Mitchell, 2000). In a study assessing the impact of temperature increase on crop yields, it is calculated that with each 1 °C increase, yield decreases by 6 %, 3.2 %, and 7.4 % for wheat, rice, and maize, respectively (Zhao et al., 2017).

On the other hand, not all plant populations are affected by climate change in the same way, with the effect also being associated with genetic diversity. If the diversity of a population is high, it is easier for it to adapt to altered environmental conditions (Frankham et al., 2017). However, for populations with low genetic diversity and subsequent low adaptive capacity, the climatic change would have more adverse impacts and can even cause extinction (Walther et al., 2002). Therefore, detection of the levels and distribution of genetic diversity of crop populations and investigation of the ecotypes that can adapt to stress conditions are essential for the conservation of plant genetic resources, future breeding of crop plants, and to help ensure global food security.

Wheat, as one of the world's major crops, is essential for human nutrition and subsistence. However, the increasing world population coupled with climate change is a concern about its production. In order to meet the demand for wheat, annual wheat production has to increase by 0.9 % by 2050 (Dixon et al., 2009). On the other hand, as wheat production is negatively affected by environmental alterations, it is predicted that the yield and quality of wheat will decrease with changing climate in the future (Liu, Asseng, Muller, Ewert, & Elliott, 2016). Hence, the determination of the adaptive traits of wheat species that are resistant to changes in climate is crucial for decreasing the impacts of climate change on wheat production.

Turkey has a great variety of wheat genetic resources, with Karacadağ, Şanlıurfa, where wheat has been first domesticated, being in its territory being one of the main reasons (Heun et al., 1997; Lev-Yadun, Gopher, & Abbo, 2000; Karagöz, 2014). It is also rich in landraces, which are cultivated varieties found in specific regions and have evolved through natural or artificial selection (Karagöz, 2014). As they grow in specific areas, landraces have high adaptive capacities for local environmental conditions (Jaradat, 2013); and among crop species, landraces are of great importance.

Einkorn (*Triticum monococcum*) and emmer (*Triticum dicoccum*) are the two ancestral wheat species (Harlan & Zohary, 1966; Van Zeist & De Roller, 1991), which are still grown in Turkey as landraces. Both species are known to exhibit high levels of genetic diversity and adaptive capacity for various stress conditions, such as high/low temperatures, drought, and salinity (Cooper, 2015; Atak, 2017). Therefore, they are invaluable genetic resources for future breeding activities and developing species that can adapt to climate change. Conversely, bread wheat (*Triticum aestivum*), as one of the modern wheat species, exhibits lower levels of genetic diversity than the ancestral wheat species, and it is not well adapted to altered environmental conditions because of the bottlenecks during its domestication and intensive breeding (Roussel, Leisova, Exbrayat, Stehno, & Balfourier, 2005; Charmet, 2011; Shewry & Hey, 2015; Özberk et al., 2016).

In this perspective, the first section of the study aims to compare the genetic diversity levels of einkorn, emmer, and bread wheat species from Turkey using single nucleotide polymorphisms (SNPs) obtained with double digested restriction-site associated DNA sequencing (ddRAD-seq). Genetic diversity is very effective for adaptation to stress conditions. Ancient wheat species, *i.e.*, einkorn and emmer, are excellent targets to investigate this aspect, as they have been shown to exhibit high levels of genetic diversity compared to other wheat species using various markers (Vierling & Nguyen, 1992;

Mori, Liu, & Tsunewaki, 1995; Nagaoka & Ogihara, 1997; Medini, Hamza, Rebai, & Baum, 2005; Cenkci, Yildiz, Konuk, & Eren, 2008; Jing et al., 2009). However, the markers used in these studies do not cover the whole genome. Therefore, as it makes genome-wide analysis without the need for a reference genome possible, ddRAD-seq is a good approach to investigate the genetic diversity levels of wheat species.

Einkorn, emmer, and bread wheat populations collected from the cities of Kastamonu and Kars in 1995 and 2015 are the interests of this analysis. In this way, it is aimed to demonstrate the possible adaptation capacities of different wheat species from different regions of Turkey through their genetic diversity levels. Since, as mentioned above, ancient wheat species are known to have higher levels of genetic diversity and adaptive potential compared to other wheat species, einkorn and emmer populations are expected to have higher diversity levels than bread wheat.

Comparison of wheat populations from 1995 and 2015 also makes it possible to investigate if there is any difference between the genetic diversity levels and genetic make-up of seed bank seeds (1995) and in-situ conservation seeds (2015), indicating a rapid evolution within the last 20 years due to climate change. Even though the 1980s are accepted as the onset of climate change, the warmest years have been recorded mainly in the late 1990s and 2000s, starting with an extreme global temperature increase caused by El Niño in 1998 (NOAA, 2020; IPCC, 2022). In the case of Turkey, the pattern is quite similar, with higher annual temperature increases since the late 1990s and 2000s, compared to the earlier years (Demir, Kılıç, Coşkun, & Sümer, 2008). There are also some recordings of heat waves in 2000 and 2007 in Turkey (WMO, 2001; WMO, 2008). Specifically, in Kastamonu, where wheat seeds from 1995 were collected, the maximum temperature values recorded for the period of 2000-2015 were found to be higher than the period of 1980-1999 (Bolat, Kara, & Tok, 2017). The average annual temperature increased from 9.6 to 10.2 °C between these two periods with increments per month (Figure 1.1, Bolat et al., 2017). Likewise, Ilker & Terzi (2021) showed that there is an increasing trend for the temperature in Kastamonu for the period of 1980-2017, particularly in spring and autumn. Therefore, the determination of the genetic variation levels of wheat seeds from 1995 and 2015 may help to understand the possible adaptation responses of wheat populations to global warming.

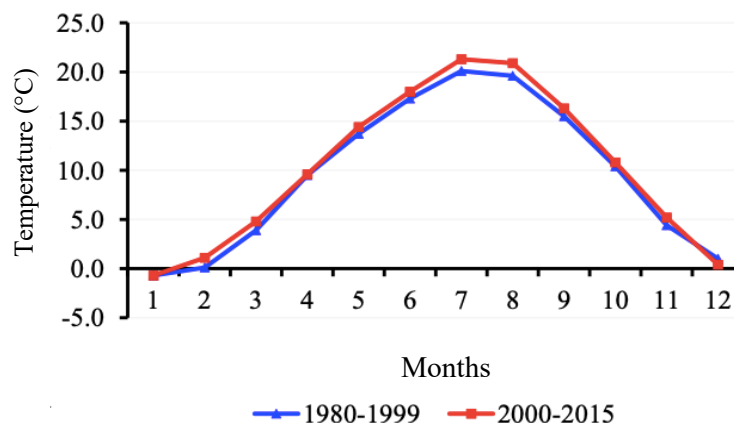


Figure 1.1. The monthly average temperature in Kastamonu for the periods of 1980-1999 and 2000-2015 (Bolat et al., 2017).

Indeed, plant populations' evolutionary adaptation against climate change is still under debate (Ravenscroft, Whitlock, & Fridley, 2015). Seed banks comprise the seeds of wild plant populations including ancestral crop species which are known to exhibit high genetic diversity levels (Jarvis, Lane, & Hijmans, 2008). Nevertheless, since they are conserved ex situ, they do not have the chance to improve their genetic variation against climate change through natural selection (Jarvis et al., 2008). On the contrary, natural populations may adapt to altering environmental conditions in situ by responding to the natural selection pressure of global warming (Hoffmann & Willi, 2008); and the adaptation capacities are mostly correlated with the genetic diversity levels that they possess (Jump, Marchant, & Peñuelas, 2009). However, high genetic variation does not always indicate that plant populations can adapt to changing climatic conditions (Jump & Penuelas, 2005). In one study, using SSRs to detect genetic variation, Nevo et al. (2012) found differences in allele frequencies of wild emmer (*T. dicoccoides*) populations from 1980 and 2008, showing that genetic divergence occurred over 28 years of climate change. Most of the populations from 2008 had fewer alleles than those of 1980, indicating a diversity loss through time, whereas three populations out of ten acquired new alleles possibly related to the adaptation to global warming (Nevo et al., 2012). In a similar vein, Fu et al. (2019) conducted the same analysis, but this time with exome capturing, and found similar results for *T. dicoccoides*: genetic bottleneck due to global warming, and the selection of new alleles with novel adaptive mutations to adapt to temperature increases. In this regard, for this study, for the selected wheat populations from 1995 and 2015, our expectations are as follows: i) seed bank seeds might have retained their high genetic diversity levels; whereas in-situ conservation seeds might have lower genetic diversity levels due to genetic bottlenecks caused by climate change or ii) in-situ conservation seeds might have acquired higher genetic diversity levels due to higher adaptive capacity associated with changing environmental conditions.

The second section of the study focuses on identifying differential genomic responses of local einkorn wheat populations to simulated climate change. Various biotic and abiotic stresses can affect wheat growth; however, heat stress, along with drought stress, is one of the main threats to wheat production (Zhang et al., 2017). Additionally, Turkey is one of the regions of the world expected to be the most severely affected by climate change in the future, with significant temperature increases (Demircan, Gürkan, Eskioğlu, Arabacı, & Coşkun, 2017). So, here the focus is investigating the genes showing differential expression under heat stress for the targeted wheat populations through RNA sequencing. The targeted populations are categorized as old (*i.e.*, collected in 1995) and new (*i.e.*, collected in 2015) einkorn populations from Kastamonu, Turkey. Using local populations from 1995 and 2015 allows the investigation of the possible effects of global warming on transcriptomic responses of the einkorn wheat in the last 20 years. Accordingly, since the new populations already experienced higher temperatures in the 2000s, they may have adapted to heat stress better than the old ones. On the other hand, if the genetic diversity levels of the new populations have decreased due to global warming, the old populations with potentially higher genetic diversity levels, see above, may respond to heat stress better than the new ones.

The reason for choosing the RNA-seq approach for differential gene expression analysis is that it is a very powerful technique for simultaneously investigating expression levels of thousands of genes (Mutz, Heilkenbrinker, Lönne, Walter, & Stahl, 2013). The genes of interest for this study are the ones showing the highest differentiation in expression when einkorn wheat populations are exposed to heat stress. In particular, the genes related to heat shock proteins (HSPs) and heat stress transcription factors (HSFs) play an essential role in plants' thermotolerance mechanisms (Bento, Pereira, Viegas, & Silva, 2017). So, they are particularly important in plants' heat stress studies, including wheat species. For instance, Rampino et al. (2009) showed that the expression of HSP genes is required for the reduction of negative impacts of heat stress in the durum wheat. In another study, using cDNA microarray for heat-sensitive and tolerant bread wheat, it was found that heat tolerance varies between different genotypes because of differentially expressed HSPs and HSFs (Qin et al., 2008). To investigate transcriptomic responses of bread wheat under normal and stress conditions (heat and drought), high throughput transcriptome sequencing was performed by Liu et al. (2015). Accordingly, it was shown that HSFs are involved in regulating the heat stress response of bread wheat (Liu et al., 2015). Therefore, even though all differentially expressed genes under elevated temperature are the focus of this analysis, the genes encoding HSPs and HSFs are of special interest as the potentially major genetic players of wheat's adaptation mechanism to rising temperatures.

2. LITERATURE REVIEW

2.1. Global Warming

The global average temperature has been rising since the late 19th century because of the increasing amounts of CO₂ and other greenhouse gases emitted into the atmosphere as a result of growing anthropogenic activities (Bolin, 1986). It has been reported that the total increase in global temperature was about 1.09 °C by 2020 relative to the preindustrial period (1850-1900) (IPCC, 2022). Moreover, if anthropogenic emissions continue to increase at the current rate, it is inevitable that the global temperature will be 1.5-5 °C higher than the current temperature at the end of the 21st century (IPCC, 2022). Therefore, it is very likely that global warming will continue to have negative impacts on natural and biological systems in the future, as it has until now. With an increase in temperature, precipitation dynamics change, sea levels rise because of increased melting of ice masses, warming and acidification of oceans take place, and many species, if they can, shift their geographic ranges or change their migration patterns, seasonal activities, or abundances (Trenberth, 2011; Mimura, 2013; Bijma, Portner, Yesson, & Rogers, 2013; Parmesan & Yohe, 2003). If the rate and magnitude of warming increase as predicted, the impacts will be even more severe, pervasive, and irreversible (IPCC, 2022). On the other hand, the risk of climate-related impacts depends on the exposure level of species and populations to changing climates and their ability to adapt to these alterations (Walther et al., 2002).

Multiple studies have investigated the effects of climate change on various species, from plants to animals (Walther et al., 2002; Root et al., 2003; Parmesan, 2006), and the common opinion is that global warming has already affected species significantly. For instance, the timing of spring activities changed for many species because of global warming, such as earlier breeding, earlier flowering, or earlier arrival of migrants (Beebee, 1995; Menzel & Estrella, 2001; Inouye, Barr, Armitage, & Inouye, 2000). Furthermore, adaptation to the altered environmental conditions through genetic evolution (Davis, Shaw, & Etterson, 2005) or geographic shifts were observed for many taxonomic groups in various locations, but the rates of alteration in distribution differed among and within species depending on their dispersal abilities and temperature and precipitation tolerances, with most of the shifts being towards higher latitudes and elevations (Walther et al., 2002). The abundance of many species, in addition to population dynamics, also changed through climate change (Parmesan, 2006). So, if such effects of climate change

have taken place even with an increase of only about 1.09 °C, the temperature rises predicted by the end of the 21st century will probably have even more severe impacts, including species-level-extinctions (IPCC, 2022). Indeed, the extinction of some species because of climate change has already been reported (Pounds, Fogden, & Campbell, 1999), and an evaluation of 131 published studies has suggested that about 8.5 % of species will go extinct with a 3 °C global increase in temperature (Urban, 2015). On the other hand, the probability of extinction depends on and varies with the levels of temperature increase, taxonomic groups, geographic regions, endemism, and the species' ability to disperse (Urban, 2015).

Although climate change has significant influences on many wild species (Root et al., 2003), agriculture is one of the most important fields to investigate its impacts in. Because of the rapid global population growth in a changing climate, it will be more difficult to access safe, sufficient, and nutritious food resources in the future (Wang, Vanga, Saxena, Orsat, & Raghavan, 2018). Therefore, investigation of the impacts of global climate change on plants is very important for the preparation of food security strategies in the near future (Moore, Baldos, Hertel, & Diaz, 2017).

Plants are affected by climate change mainly in terms of their development, growth, and yield (Tubiello, Soussana, & Howden, 2007). Generally, moderate temperature increases between 1-3 °C or increasing CO₂ emissions cause an increase in plant productivity by inducing photosynthesis (Cline, 2007; Jarvis et al., 2008). Consequently, the plant yield increases. However, major crops (wheat, maize, and rice) usually respond to global warming by yield reductions for all temperature ranges, and in almost all locations, with more damaging effects in regions that are already hot (Challinor et al., 2014; Moore et al., 2017). This is because when combined with other changes induced by climate change, such as high temperatures, elevated CO₂ does not have as a positive impact on yield as it would had it been the only factor that was increasing (Tubiello et al., 2007). For instance, high temperatures affect the yield negatively (Cline, 2007), causing a reduction in grain number, size, and quality by suppressing the positive effects of CO₂ (Tubiello et al., 2007). Similarly, high water demand generated by increased temperature and decreased precipitation reduces the positive effects of CO₂ (Xiao, Liu, Xu, Sun, & Wang, 2005). Furthermore, positive responses of plants to high CO₂ levels are diminished by low nutrient content in soils (Tubiello et al., 2007).

On the other hand, the salinity content of soils increases with the rising levels of oceans as a result of glacier melting, which adversely affects agricultural ecosystems (Newton, Johnson, Lyon, Hopkins, & Gregory, 2008). Additionally, the incidence and severity of extreme events such as droughts, floods,

or storms increase with the changes in climate, and they all have adverse impacts on yield (Ortiz, 2011). For example, in Europe, because of an extreme increase in temperature and a decrease in precipitation, a serious reduction in crop yield (the lowest yield with a 20 % decrease compared to 1960-1990) was recorded in 2003 (Ciais et al., 2005). Extreme climate events can also induce plant diseases and pest outbreaks (IPCC, 2022).

2.2. The Impacts of Climate Change on Agriculture in Turkey

Molecular evolutionary studies and the morphological characteristics of the ancestral plant species' remains found in the archaeological sites showed that agriculture has begun 10.000-12.000 years ago in the Fertile Crescent, which also encompasses southeast Turkey (Figure 2.1) (Smith, 1995; Zohary & Hopf, 2000; Moore, Hillman, & Legge, 2000; Gopher et al., 2002). Many plant species, such as cereals, used today as the major human and animal food resources, have emerged and have been domesticated in this region (Smith, 1995; Zohary & Hopf, 2000; Moore et al., 2000; Gopher et al., 2002).



Figure 2.1. Map of the Fertile Crescent (dashed red line) (Salamini, Özkan, Brandolini, Schäfer-Pregl, & Martin, 2002).

In addition to being a part of the Fertile Crescent, the Mediterranean and the Middle East gene centers are partly located and overlapped in Turkey (Vavilov, 1951), resulting in Turkey's plant genetic diversity being very high (Karagöz, 2003). Besides, high levels of endemism are observed in the country (Karagöz, 2003). Today, approximately 11.707 plant taxa are present in Turkey, and 3649 of these taxa are endemic (Güner, 2012; Özberk et al., 2016). Wild relatives of cereals, including wheat, still grow in the region, mostly in southeast Turkey, where their emergence and domestication have occurred (Atak,

2017). For instance, in a recent study conducted between 2009-2014, it was recorded that 95 varieties of three wheat species and six subspecies, including founder wheat species, einkorn and emmer, are grown in Turkey (FAO, 2015). Indeed, with its great variety of geomorphology, topography, and climate, it is not surprising for Turkey to be one of the regions where agriculture has arisen and to have such a high diversity of plant genetic resources (Özberk et al., 2016). However, many wild plant species in the region, including endemics, are under threat of growing population and increasing anthropogenic activities such as industrialization, urbanization, or agriculture, culminating in various environmental challenges (Karagöz, 2003).

Climatic changes such as rising temperatures causing an alteration in water availability or quality can also cause deficits in plant growth and development, as they adversely affect photosynthesis (Akter & Rafiqul Islam, 2017). Turkey, which is already a water-stressed country with recorded increases in its average temperatures since the 1960s (Dudu & Çakmak, 2018; Brohan, Kennedy, Harris, Tett, & Jones, 2006), is one of the global regions that will be most affected by climate change in the 21st century (Cline, 2007). It is predicted that the average temperature rise in Turkey might be 3-7 °C by 2070-2100, with higher increases in eastern regions (Öztürk, Türkeş, & Kurnaz, 2011). According to one projection, the temperature increment will be 3.5-4 °C in eastern Turkey by 2100, whereas central and southwestern regions will experience a temperature increase of 3-3.5 °C, and in the north, it will be 2.5-3 °C hotter than the current (MET, 2011). On the other hand, significant declines in precipitation are also predicted for the period of 2070-2100 (Öztürk et al., 2011), and the Mediterranean and the middle eastern regions will be the most affected areas over the country (MET, 2011).

There are several studies in which the effects of climate change on agricultural production in Turkey have been investigated (Yano, Aydin, & Haraguchi, 2007; Özdoğan, 2011; Giannakopoulos et al., 2009). The common finding is that climate change has already impacted agriculture in Turkey, and the effects are likely to increase over time (MET, 2011). In one study, temperature increases between 1.1-1.6 °C coupled with a precipitation decline of about 30 % were predicted for the period of 2070-2099, and it was suggested that these alterations would cause a decline of 11.8 % in average agricultural yield (Cline, 2007). In another study, the possible impacts of climate change on wheat yield in Çukurova basin were investigated using a regional climate model (Kapur, Topaloglu, Ozfidaner, & Koç, 2007). Accordingly, temperature increases up to 2.8 °C and precipitation declines of 35 % were estimated, although the possible changes in wheat yield due to these alterations were not reported quantitatively (Kapur et al., 2007). Furthermore, Özdoğan (2011) predicted that there would be a decrease (15-20 %) in wheat yield

in the Thrace region due to climate change in the 21st century. In a recent study investigating the yield change in Turkey for the period of 2010-2099, it was shown that the alterations in temperature and precipitation dynamics would be most effective in the second half of the 21st century (Dudu & Çakmak, 2018). In accordance with this finding, the impacts of climate change will become significant after 2035, and more severe temperature increases and precipitation decreases will be observed after 2060, causing yield change over time (Dudu & Çakmak, 2018). All in all, these studies showed that agriculture would be certainly expected to be impacted by climate change in Turkey during the rest of the century.

2.3. The Genus *Triticum*

One of the first plant species used by humans settled in the Fertile Crescent was wheat (Zohary & Hopf, 2000). The first wheat cultivation has occurred with the domestication of wild einkorn about 10.000 years ago in Karacadağ, Şanlıurfa, Turkey (Heun et al., 1997). In addition to einkorn, emmer has been cultivated and domesticated in the same region (Heun et al., 1997; Elias, Steiger, & Cantrell, 1996; Lev-Yadun et al., 2000). In time, with these domestication events, wheat has become one of the major food resources globally (Zohary, Hopf, & Weiss, 2012). Later, from the founder wheat species, modern wheat has evolved (Yaman, Ordu, Zencirci, & Kan, 2020), which is mostly preferred today over ancient wheat because of its high yield capacity and easier harvesting properties (Zohary & Hopf, 1993; Cooper, 2015).

Wheat belongs to the phylum Angiospermatophyta, class Monocotyledonopsida, order Poales, family Poaceae, tribe Triticeae, subtribe Triticinae, and genus *Triticum* (Bálint, Kovács, & Sutka, 2000). Like most cereals, it is polyploid, which takes its source from interspecific hybridization and whole genome duplication (Ramírez-González et al., 2018). A, B, D, and G genomes of diploid and tetraploid species of *Triticum* and *Aegilops* genera are responsible for the polyploidization of the species (Goncharov, 2011). The first hybridization event causing polyploidy in wheat is thought to be between *T. urartu* (AA) and *Ae. speltoides* (BB) (Dvorak & Akhunov, 2005; Bálint et al., 2000; Ogihara & Tsunewaki, 1988; Dvorak & Zhang, 1990). As a result of the hybridization between these diploid species, tetraploid emmer, which has both A and B genomes, has emerged (Figure 2.2) (Takumi, Nasuda, Liu, & Tsunewaki, 1993). Later, wild emmer (*T. dicoccoides* (AABB)) has been domesticated and formed *T. dicoccum* (Bálint et al., 2000). Hexaploid wheat, *i.e.*, *T. aestivum*, which contains A, B, and D genomes, has arisen because of the hybridization between tetraploid emmer and diploid *Ae. tauschii* (DD) (Figure 2.2) (Kihara, 1944; McFadden & Sears, 1946). On the other hand, einkorn (*T. monococcum*) was

domesticated from its wild progenitor, *T. boeoticum* (Harlan & Zohary, 1966; Van Zeist & De Roller, 1991).

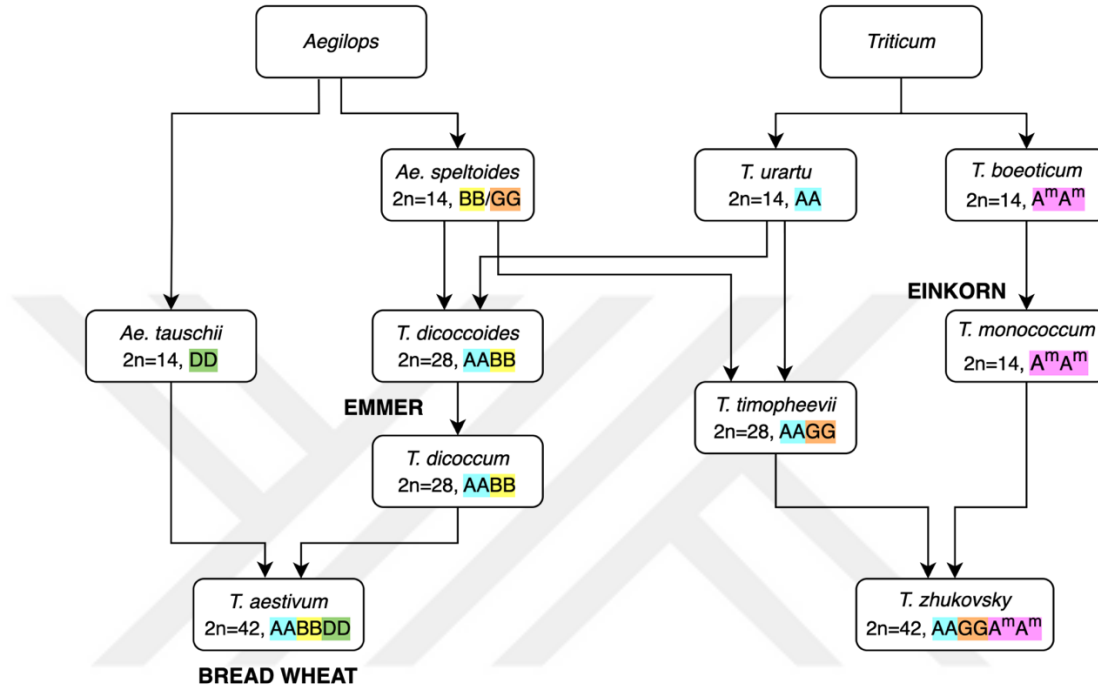


Figure 2.2. Hybridization events of wheat species.

In Turkey, the most cultivated crop species with the highest production amount is wheat (Atak, 2017). In addition to the domesticated forms, wild forms of the ancestral species and their close relatives still grow in the region, making Turkey a center of wheat genetic resources (Atak, 2017). A previous study based on the exploration and collection of wheat species growing in Turkey showed that the region is very rich in wheat species, so nearly all wheat varieties in the world can grow (Gökgöl, 1935). This is most likely because of the different climatic and environmental conditions of the region (Karagöz & Zencirci, 2005). Besides, Turkey has numerous wheat landraces (Karagöz & Zencirci, 2005). In general, landraces have been generated through the domestication of wild relatives (Zeven, 1998). They have been selected and grown by farmers, as well as naturally, through changing environmental conditions over the years (Zeven, 1998). It has been reported that landraces are even more diverse compared to their wild relatives, and hence their capacity for adaptation to biotic and abiotic stresses is higher than that of wild species (Eagles & Lothrop, 1994). In this regard, landraces are of great importance against climate change.

One of the wheat landraces grown in Turkey is einkorn (*T. monococcum*), which is also known as siyez among local people (Özberk et al., 2016). It is diploid and contains A^m genome with $2n=2x=14$ chromosomes (Figure 2.2) (Bálint et al., 2000). It has tough glumes that tightly enclose its grains (Cooper, 2015), making einkorn wheat more resistant to poor soils and adverse environmental conditions, including pests and diseases (Özberk et al., 2016). Hence, it is generally found in arid and semi-arid regions (Atak, 2017). Today, it is mostly planted in mountainous areas of Turkey as well as the Middle East, France, Spain, Germany, Switzerland, and Morocco (Atak, 2017; Brandolini, Volante, & Heun, 2016). In Turkey, it grows in the western Black Sea region, more specifically in Kastamonu, Bolu, Bilecik, and Sinop (Karagöz & Zencirci, 2005). However, Kastamonu is the place where it is harvested most abundantly (Atak, 2017).

The other wheat landrace grown in Turkey is emmer, gernik in Turkish (Özberk et al., 2016). Emmer (*T. dicoccum*) is an essential crop since all economically important domesticated wheat species, such as *T. aestivum*, originated from the wild form of this species (Figure 2.2) (Avni et al., 2017). It is tetraploid, consisting of A and B genomes with $2n=4x=28$ chromosomes (Figure 2.2) (Bálint et al., 2000). Like einkorn, because of its tough glumes, its tolerance for biotic and abiotic stresses is high (Cooper, 2015). It has been growing in the Middle East, Asia, and Europe (Nesbitt & Samuel, 1996). Today, it grows in Turkey, Caucasia, Ethiopia, Iran, Italy, Spain, and India (Zaharieva, Ayana, Al Hakimi, Misra, & Monneveux, 2010). In Turkey, it is mostly found in eastern regions, specifically in the provinces of Kars and Ardahan (Atak, 2017; Zaharieva et al., 2010).

Besides wheat landraces, breeding modern wheat (*T. aestivum*) is also common in Turkey (Atak, 2017). It grows throughout the country (Özberk et al., 2016). In addition to being common in Turkey, it is also the most cultivated wheat worldwide (Shewry, 2009). It is mostly used for making bread (Cooper, 2015). It is hexaploid with A, B, and D genomes and $2n=6x=42$ chromosomes (Figure 2.2) (Bálint et al., 2000). Unlike einkorn and emmer, it has fragile glumes making it more capable of germination (Cooper, 2015; Charmet, 2011). Even though this feature increases the yield of the species, the adaptation capacity of modern wheat to environmental alterations or diseases and pests is low (Atak, 2017).

2.4. Genetic Diversity of Wheat

As the climate changes so rapidly, crop improvement, including adaptation to biotic and abiotic stresses, becomes critical. In this respect, wheat species comprise useful organisms due to their high

levels of genetic variability (Dubcovsky & Dvorak, 2007). Especially progenitor species, such as einkorn and emmer, have advantageous features for breeding studies as they have high genetic diversity (Gürcan, Demirel, Tekin, Demirel, & Akar, 2017). Likewise, because these ancient wheat species grow in arid or semi-arid areas and, therefore, are well adapted to harsh environmental conditions with different stressors, they can be used for improvements in adaptation, plant productivity, and yield (Valkoun, 2001). Hence, the investigation of wheat genetic diversity in order to find valuable traits that may be used for crop improvement is very important (Seifolahpour, Bahraminejad, & Cheghamirza, 2017).

The genetic diversity of wheat species has been previously investigated using variable molecular markers, and it was shown that the genetic diversity levels of einkorn and emmer are greater than those of other wheat species. These studies employed various approaches, including restriction fragment length polymorphism (RFLP) (Mori et al., 1995; Nagaoka & Ogihara, 1997), amplified fragment length polymorphism (AFLP) (Nagaoka & Ogihara, 1997), random amplification of polymorphic DNA (RAPD) (Vierling & Nguyen, 1992; Nagaoka & Ogihara, 1997; Mantzavinou, Babal, & Kaltsikes, 2005; Aliyev, Abbasov, & Mammadov, 2007; Cenkci et al., 2008), simple sequence repeat (SSR) markers (Sourdille, Tavaud, Charvet, & Bernard, 2001; Medini et al., 2005), and diversity arrays technology (DArT) marker system (Jing et al., 2009).

2.5. Heat Stress Response of Plants

All organisms, including plants, register any temperature above their normal range as heat stress. Heat stress affects cellular homeostasis negatively and causes adverse impacts on growth and development (Kotak et al., 2007). In general, plants genetically respond to heat stress by producing heat shock proteins (HSPs) (Feder & Hoffman, 1999). When they are exposed to high temperatures, the expression of HSPs increases immediately, as a major part of their thermotolerance mechanisms, with the help of heat stress transcription factors (HSFs), which are the activators of HSPs (Bento et al., 2017; Banti, Mafessoni, Loreti, Alpi, & Perata, 2010). In the meantime, overall protein synthesis decreases, and hormone production is enhanced, and the levels of other protective molecules increase (Bento et al., 2017).

HSPs are molecular chaperones that are responsible for protein quality control to maintain cellular homeostasis under both normal and stress conditions (Kotak et al., 2007). They are classified into five families considering their molecular weights, which are HSP100, HSP90, HSP70, HSP60, and small

HSPs ranging from 15 to 30 kilodaltons (kDa) (Bento et al., 2017). Some of them are produced under stress conditions only, while others are synthesized in order to maintain proper protein folding under normal environmental conditions (Bita & Gerats, 2013). Because of their importance for survival under heat stress, they are highly conserved in plants, as well as in other organisms, from bacteria to animals (Vierling, 1991; Wang, Vinocur, Shoseyov, & Altman, 2004).

The HSP100 family contains proteins of 84-104 kDa, whose function is mostly to tolerate high temperatures in plants rather than to control plant growth (Grigorova, Vaseva, Demirevska, & Feller, 2011; Bita & Gerats, 2013). Under stress conditions, they prevent protein aggregation in collaboration with small HSPs and HSP70 (Gulli, Corradi, Rampino, Marmioli, & Perrotta, 2007). Therefore, heat stress studies usually focus on the HSP100 family as well as small HSPs and HSP70, since they are required for basal and acquired thermotolerance in plants (Kotak et al., 2007; Grigorova et al., 2011). As an example, the essential role of HSP101 in the thermotolerance of *Arabidopsis* plant was reported (Queitsch, Hong, Vierling, & Lindquist, 2000). Similarly, it was shown that the interaction between HSP101 and small HSP chaperon system in *Arabidopsis* causes dissolution of protein aggregates after heat stress (Lee et al., 2005). In another study, mitochondrial small HSPs were found to increase thermotolerance in tobacco when they were overexpressed (Sanmiya, Suzuki, Egawa, & Shono, 2004).

The expression of HSPs and other proteins related to heat stress is regulated by several HSFs (Kotak et al., 2007). Some members of the HSF family are responsible for the initial gene expression under heat stress, whereas others have a role in controlling the expression under prolonged heat stress and recovery (Lohmann, Eggers-Schumacher, Wunderlich, & Schoffl, 2004; Charng et al., 2007). In addition to HSPs and HSFs, multiple signaling pathways also contribute to the heat stress response of plants in order to help with their survival under high temperatures. For instance, heat stress causes oxidative stress resulting in the loss of membrane stability, and plants protect themselves from this damage through oxidative stress signaling (Kotak et al., 2007; Akter & Rafiqul Islam, 2017). When the temperature is high, the expression of heat stress responsive genes is induced by an H₂O₂ burst, and this burst is mediated by NADPH oxidase activity (Vacca et al., 2004; Volkov, Panchuk, Mullineaux, & Schöffl, 2006). As a result, the level of tolerance to heat stress increases (Larkindale & Huang, 2004). It is also known that mutations in NADPH oxidases decrease thermotolerance (Larkindale, Hall, Knight, & Vierling, 2005). Superoxide dismutase (SOD), ascorbate peroxidase (APX), catalase (CAT), glutathione reductase (GR), and peroxidase (POX) are some of the antioxidants known to be used by plants for coping with oxidative damage caused by heat stress (Suzuki, Miller, Morales, Shulaev, & Torres, 2011; Caverzan, Casassola,

& Brammer, 2016). Plant hormones, including abscisic acid, salicylic acid, and ethylene, are also related to heat stress signaling in various species as they help plants adapt to environmental changes (Kotak et al., 2007).

2.6. Genomic Response of Wheat to Heat Stress

Wheat is one of the most cultivated cereals in the world, and its production is affected negatively by abiotic stresses. So, it is very important to identify the genetic and molecular mechanisms underlying heat tolerance in wheat for future heat resistance breeding, especially in an environment where the temperature is continuously increasing (Zhang et al., 2017).

The heat stress response is not achieved by a single gene in cereals. There are different loci and genes associated with heat tolerance in various tissues, and they regulate the responses to heat stress at different stages of the life cycle (Maestri et al., 2002). However, since the HSPs and HSFs are the major components of plant thermotolerance (Bento et al., 2017), heat stress studies of wheat species generally focus on genes encoding HSPs and HSFs, as mentioned in the Introduction section (Rampino et al., 2009; Qin et al., 2008; Liu et al., 2015).

In addition to the identification of the genes involved in plant thermotolerance, two-dimensional gel electrophoresis/mass spectrometry (2-DE/MS) is widely used to investigate proteins involved in heat stress response. A previous study showed that the amount of HSPs with molecular weights of 15-30 kDa increased in wheat cultivars after three days of heat treatment (Skylas et al., 2002). HSPs are also responsible for the formation of wheat gluten (Maestri et al., 2002). Wheat gluten, which is the main component of flour protein, is a storage protein, and it consists of two proteins: glutenin and gliadin (Dupont & Altenbach, 2003). The quality of wheat grain is determined by gluten, along with other proteins, including albumins and globulins, which are metabolic proteins (Yang et al., 2011), and heat stress affects the quality of wheat by changing the proportion of these proteins (Maestri et al., 2002). It was found that high temperature induces the development of wheat grain and increases the concentration of proteins associated with storage and defense, whereas it reduces the quantity of other proteins related to grain metabolism (Hurkman, Vensel, Tanaka, Whitehand, & Altenbach, 2009). Similarly, Yang et al. (2011) showed that high temperature caused an alteration in the fraction of several wheat gliadins and glutenins. In another study, an increase in high molecular weight wheat glutenin subunits was shown as a result of heat stress (Dupont et al., 2006). The differential expression of the proteins, which are related

to the stress response as well as glycolysis and carbohydrate metabolism, was also demonstrated by Laino et al. (2010).

2.7. Double Digest Restriction-Site Associated DNA Sequencing (ddRAD-seq)

Genotyping has been mostly performed using microarray-based methods. However, the suitability of these methods to only some targeted genomic regions and the requirement for a reference genome make them limited for population genotyping, especially for complex and non-model organisms (Peterson, Weber, Kay, Fisher, & Hoekstra, 2012). Restriction-site associated DNA (RAD) markers are short DNA fragments produced by restriction enzyme digestion, and they are used to screen a high number of targeted sites in the genome when they are used in conjunction with microarray methods (Miller, Dunham, Amores, Cresko, & Johnson, 2007). However, the microarray-based RAD method still has some limitations (Baird et al., 2008).

With the development of next generation sequencing technologies, genotyping by sequencing (GBS) has become a feasible method, especially for complex and diverse genomes (Elshire et al., 2011). So, in order to determine genotypes efficiently and accurately, sequencing of RAD libraries (RAD-seq) has recently begun to be used (Poland, Brown, Sorrells, & Jannink, 2012). In general, this method is based on producing RAD tags by restriction enzyme digestion, adapter ligation, random shearing, preparation, and multiplexed sequencing of RAD tag libraries (Baird et al., 2008). The reads obtained from sequencing are then aligned to a reference genome or clustered based on similarity *de novo* (Poland et al., 2012). The adapters contain amplification and sequencing sites as well as barcoded regions that allow sample tracking after sequencing (Baird et al., 2008). Additionally, rather than sequencing the whole genome, only the regions adjacent to restriction sites are sequenced, and this generates a reduced representation of the genome resulting in a decline in cost and time (Elshire et al., 2011).

However, sequencing DNA libraries containing only the regions adjacent to restriction sites still limits the number of targeted regions in the genome, as only one restriction enzyme digestion and a secondary random fragmentation are used to generate RAD tags, and the tags are size-selected using a broad range (Peterson et al., 2012). To overcome this problem, an expanded approach called double digest restriction-site associated DNA sequencing (ddRAD-seq) has been developed (Peterson et al., 2012). Accordingly, this method eliminates the random shearing of DNA and instead uses two restriction enzymes in order to cut the genome more precisely. It also includes a custom indexing method and

specific software tools for genotyping large numbers of individuals at low cost. In addition, the method does not require prior knowledge of the genome sequences. So, it is a promising approach to investigate genome-wide genetic differentiation of large and complex populations.

2.8. Next Generation RNA Sequencing

The classical sequencing method (Sanger sequencing) (Sanger, Nicklen, & Coulson, 1977), which is based on chain termination by enzymatic reaction, has been widely used for many years. However, the development of next generation sequencing (NGS) technologies has reduced the cost and time required for sequencing (Cullum, Alder, & Hoodless, 2011). The applicability of these technologies to many molecular analyses also makes them preferable (Metzker, 2010). The basic principle of NGS, in contrast to the Sanger method, is the detection of a signal following nucleotide incorporation (Mutz et al., 2013). Next, all detected signals are recorded in order as peaks, and the DNA sequences are obtained (Mutz et al., 2013). The approach allows parallel sequencing of millions of DNA fragments at the same time, and the read length of these fragments is generally between 30-400 bp (Rizzo & Buck, 2012). Although the range is higher for Sanger sequencing reads, NGS generates huge data sets, up to 50 Gb per run, because of its high throughput (Mutz et al., 2013).

NGS technology is also used to map and quantify transcriptomes (mRNAs, non-coding RNAs, and small RNAs), and this is called RNA sequencing (RNA-seq) (Metzker, 2010). In general, cDNA fragments containing adapters are synthesized from the total RNA molecules and fragmented into smaller pieces (Wang, Gerstein, & Snyder, 2009). Next, they are amplified and sequenced using the high throughput approach mentioned above (Wang et al., 2009). Because of the parallel sequencing runs, the transcriptional structure and expression levels of thousands of genes can be investigated at the same time (Mutz et al., 2013). In order to analyze the data obtained as a result of RNA-seq, first, the raw data is converted to short-read sequences (Mutz et al., 2013). Next, they are aligned to a reference genome or a transcriptome, and the numbers of mapped reads are quantified to calculate gene expression levels (Stark, Grzelak, & Hadfield, 2019). Because RNA-seq allows absolute quantification of transcripts, the changes at expression levels under different experimental conditions can also be detected (Zhou et al., 2010).

3. MATERIALS AND METHODS

3.1. Double Digest Restriction-Site Associated DNA Sequencing (ddRAD-seq)

3.1.1. Plant Materials

ddRAD sequencing was performed using einkorn (*T. monococcum*), emmer (*T. dicoccum*), and bread wheat (*T. aestivum*) populations collected from the cities of Kastamonu and Kars in 1995 and 2015 (Table 3.1). The seeds collected in 1995 were provided by Aegean Agricultural Research Institute (Ege Tarımsal Araştırma Enstitüsü – ETAE), whereas the seeds from 2015 were collected by one of our former lab members from different regions of Kastamonu and Kars (Figure 3.1). Since exact information about the collection sites of the seeds from 1995 is not known, only the location sites of the seeds from 2015 are shown on the map in Figure 3.1.

Table 3.1. Wheat populations used for ddRAD-seq.

	Population	Year of Collection	City	Village	Source
Einkorn (<i>T. monococcum</i>)	WP 22	1995	Kastamonu	NA	ETAE
	WP 81	2015	Kastamonu	Görpe	Collected
	WP 88	2015	Kastamonu	İnciğez	Collected
	WP 89	2015	Kastamonu	Enbiya	Collected
Emmer (<i>T. dicoccum</i>)	WP 15	1995	Kastamonu	NA	ETAE
	WP 20	1995	Kastamonu	NA	ETAE
	WP 90	2015	Kastamonu	Çatalyazı	Collected
	WP 92	2015	Kastamonu	Kapaklı	Collected
	WP 93	2015	Kastamonu	Yarışlar	Collected
	WP 95	2015	Kars	Merkez	Collected
	WP 96	2015	Kars	Çakmak	Collected
	WP 97	2015	Kars	Bayburt	Collected
Bread Wheat (<i>T. aestivum</i>)	WP 98	2015	Kastamonu	Enbiya	Collected
	WP 99	2015	Kastamonu	Görpe	Collected
	WP 103	2015	Kastamonu	İnciğez	Collected

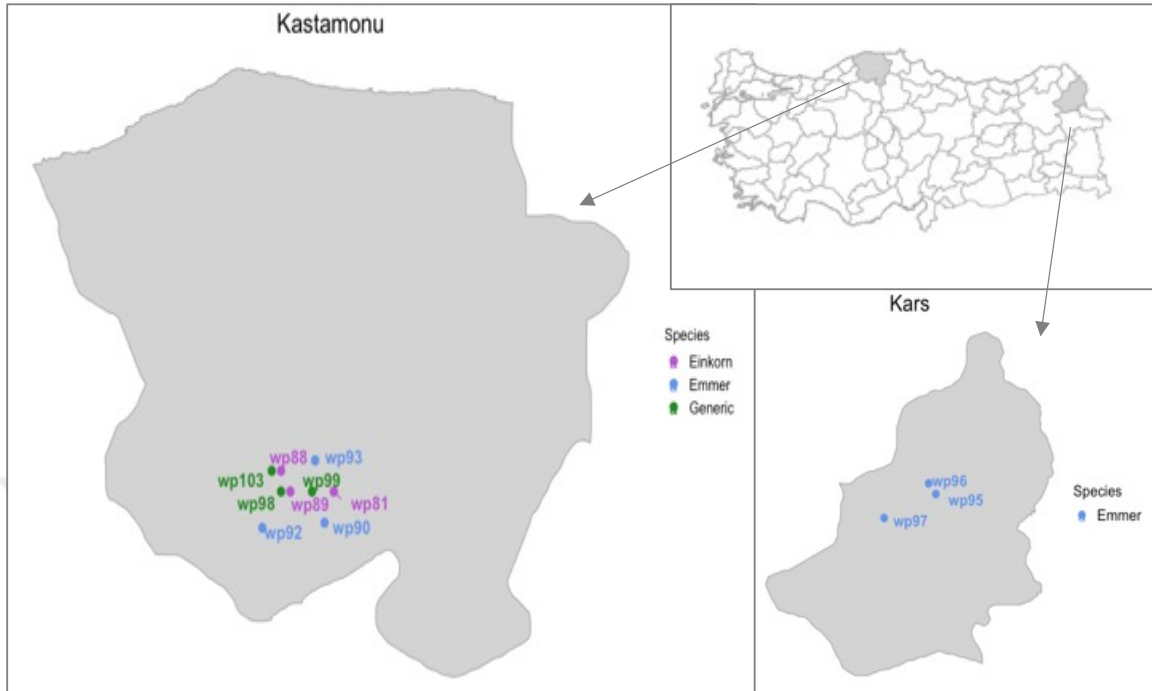


Figure 3.1. Sampling locations of wheat populations from 2015 used for ddRAD-seq.

For each population given in Table 3.1, at least 15 seeds were sown in pots with a capacity of 0.5 liters (93x93x104 mm) using peat soil produced in Turkey. Soils with high organic carbon content, 60-65 % moisture, and a pH between 6.5-7.5 provide the optimum conditions for wheat development (Mitрева, Krasteva, & Pankov, 2016). In this context, peat soil is a good candidate for growing wheat, as it is rich in organic matter (Kailla, 1956). In addition to the high organic carbon content, peat soils produced in Turkey have a pH between 5-7 and good water-holding capacity (50-70 % moisture content) (Sağlam, Bellitürk, & Kutlu, 2009), making them very suitable for wheat growth.

After sowing the seeds, the populations from 2015 were grown at our laboratory without any temperature, light, or humidity adjustments. On the other hand, to increase the chance of germination, the populations from 1995 were grown in the growth chamber located at the Department of Molecular Biology and Genetics of Boğaziçi University under optimum and controlled conditions for wheat growth (see section 3.2.2 for detailed explanation), since the number of seeds from 1995 was limited and one of those old-dated populations (WP22) was also in the scope of the second section of the study, see section 3.2.2.

For both groups (1995 and 2015), irrigation was done regularly (once every two or three days) to maintain soil humidity. Approximately after 14-20 days from planting, the first leaves grew adequately

(1-2 leaves stage) and were ready to harvest. Fresh leaves were collected into sterile microcentrifuge tubes and frozen immediately in liquid nitrogen. Since DNA isolation requires disrupted and homogenized tissue at the first step, leaves were ground using sterile pestles (Axygen, USA) attached to a hand drill while keeping the tubes submerged in liquid nitrogen. Homogenized tissues were stored at the -80 °C freezer (Thermo Scientific, USA) until DNA extraction.

3.1.2. DNA Extraction

For all populations except WP22, DNA extraction was performed with DNeasy Plant Mini Kit (Qiagen, Germany) following the manufacturer's instructions. In the case of WP22, since RNA was also required for the RNA-seq section of the study (see section 3.2.3), DNA and RNA were isolated simultaneously using Quick DNA/RNA Microprep Plus Kit (Zymo, USA). For the isolations, the instructions of the manufacturer were followed.

The concentrations of the isolated DNA samples were measured with a Qubit 3 Fluorometer (Invitrogen, USA) using dsDNA High Sensitivity (HS) Assay Kit (Invitrogen, USA). The purity ratios were calculated using a nanophotometer (Implen, USA). As integrity analysis, agarose gel electrophoresis was performed using 1 % agarose (PeqLab, Germany) in 1X TAE buffer. Four µl Redsafe (Intron Biotechnology, Korea) was used to stain the DNA. Five µl genomic DNA was mixed with one µl 6X agarose loading dye (AMBRD, Turkey). Six µl 1kb DNA ladder (Hibrogen, Turkey) was used as a size marker. The gels were run at 90 V for 30 minutes.

3.1.3. Construction of ddRAD Libraries

In general, ddRAD sequencing is performed using two restriction enzymes: one with a common recognition site that cuts DNA frequently, and one with a longer recognition site cutting DNA rarely (Poland et al., 2012; Peterson et al., 2012). So, before starting to prepare ddRAD libraries, these restriction enzymes were determined by in silico digestion of wheat reference genome using the software Fragmatic v3 (Chafin, Martin, Musmann, Douglas, & Douglas, 2018). The genome of *T. aestivum* (Triticum_4.0, GenBank: GCA_002220415.3) was used as a reference. To run Fragmatic, the chromosome sequences of *T. aestivum* were concatenated. Next, in silico digestion was performed using EcoRI – SbfI, SbfI – MspI, and PstI – MseI enzyme combinations. For each enzyme pair, the total number of fragments produced within the size range of 350-550 bp was recorded to find the enzyme combination

that gave the highest ddRAD loci for wheat. The size range was determined according to Brelsford, Dufresnes, & Perrin's protocol (2016).

After the selection of the enzyme combination, libraries were prepared following the instructions of the protocol of Brelsford et al. (2016), modified from the protocols of Parchman et al. (2012) and Peterson et al. (2012). Accordingly, the extracted genomic DNA was double digested with two restriction enzymes, the rare cutter PstI (NEB, UK) and the frequent cutter MseI (NEB, UK). Next, the adapters were ligated to the sticky ends of the double-digested DNA fragments with T4 DNA ligase (NEB, UK). After ligation of the adapters, short fragments were removed using Agencourt AMPure magnetic beads (Beckman Coulter, USA). The remaining fragments dissolved in the elution buffer of MinElute Gel Extraction Kit (Qiagen, Germany) were PCR amplified. Q5 Taq polymerase used for PCR was acquired from NEB (UK), and dNTPs were acquired from Sigma (USA) and AMBRD (Turkey). The sequences of adapters and primers used in library preparation are shown in Table 3.2, while the index/barcode codes are given in Table 3.3.

Table 3.2. Sequences of adapters and primers used for ddRAD library construction. xxxxxx represents the barcode, zzzzzz represents the index. Phosphorothiolate bonds are marked by *.

Name	Sequence (5' to 3')
MseI_ad1	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MseI_ad2	TAAGATCGGAAGAGCGAGAACAA
PstI_ad	ACACTCTTTCCCTACACGACGCTCTTCCGATCTxxxxxxTGCA
ILLPCR1	A*A*TGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG
ILLPCR2	C*A*AGCAGAAGACGGCATACGAGATzzzzzzGTGACTGGAGTTCAGACGTGTGC

Table 3.3. Sequences of barcodes and indexes embedded in adapter and primer sequences.

Barcode (xxxxxx)	Sequence (5' to 3')	Index (zzzzzz)	Sequence (5' to 3')
PstI_ad1	ACAGGT	ILLPCR2-01	ATCACG
PstI_ad2	ACGATG	ILLPCR2-02	CGATGT
PstI_ad3	AGAGGA	ILLPCR2-03	TTAGGC
PstI_ad4	AGTCCT	ILLPCR2-04	TGACCA
PstI_ad5	AGTCGA	ILLPCR2-05	ACAGTG
PstI_ad6	CACAGA	ILLPCR2-06	GCCAAT
PstI_ad7	CTCAGT	ILLPCR2-07	CAGATC
PstI_ad8	CTTCAG	ILLPCR2-08	ACTTGA
PstI_ad9	GACACA	ILLPCR2-09	GATCAG
PstI_ad10	GTGTCT	ILLPCR2-10	TAGCTT
PstI_ad11	TCTCGA	ILLPCR2-11	GGCTAC
PstI_ad12	TGTCCA	ILLPCR2-12	CTTGTA

Following PCR amplification, size selection was performed using 2 % ultrapure agarose (Invitrogen, USA) in 1X LB buffer. After running at 100 V for 165 minutes, the fragments in the 350-550 bp size range were cut out of the gel and purified using MinElute Gel Extraction Kit (Qiagen, Germany) following the instructions of the manufacturer. The quantities of the final products were measured by qPCR (Applied Biosystems, USA) using KAPA Library Quantification Kit (Roche Diagnostics, Swiss). For quantification, the instructions of the manufacturer were followed.

Based on their final concentrations, libraries were pooled in equimolar amounts. In total, 195 ddRAD libraries were prepared (Table 3.4) subdivided into four pools. Pools were sent to Macrogen (Seoul, Korea) and Gen-Era (Turkey) for sequencing in the Illumina platforms, HiSeq 2500 and NovaSeq, respectively.

Table 3.4. The number of ddRAD libraries sequenced for each population.

	Year	City	Population	Number of ddRAD Libraries
Einkorn (<i>T. monococcum</i>)	1995	Kastamonu	WP 22	31
	2015	Kastamonu	WP 81	11
			WP 88	11
			WP 89	11
Emmer (<i>T. dicoccum</i>)	1995	Kastamonu	WP 15	14
			WP 20	16
	2015	Kastamonu	WP 90	11
			WP 92	11
			WP 93	12
		Kars	WP 95	11
			WP 96	12
			WP 97	12
Bread Wheat (<i>T. aestivum</i>)	2015	Kastamonu	WP 98	9
			WP 99	13
			WP 103	10
				TOT: 195

3.1.4. Data Analyses

The sequencing data were processed remotely on TRUBA (TUBITAK ULAKBIM, High Performance and Grid Computing Center) using *Stacks* software v2.58 (Catchen, Amores, Hohenlohe, Cresko, & Postlethwait, 2011). First, the data were demultiplexed and cleaned with the *process_radtags.pl* program of *Stacks* based on the information about the data (paired-end), the enzymes used for creating ddRAD libraries (PstI – MseI), and barcode–index sequences (Table 3.3). Second, high-

quality reads (150 bp) with RAD tags and enzyme cut sites were processed with *denovo_map.pl* pipeline. The optimal parameters for wheat were determined before running the pipeline for all samples. The main parameters of *Stacks* used for de novo analysis are *m* (the minimum number of identical reads required to form a stack, which controls coverage), *M* (the maximum number of allowed mismatches between stacks to make a locus), and *n* (the number of allowed mismatches between stacks during merging loci to construct a ‘catalog’) (Catchen et al., 2013; Paris, Stevens, & Catchen, 2017). These parameters, which control locus formation and polymorphism, can be optimized for each organism before the analysis; since the species’ biology (e.g., ploidy level, polymorphism percentage) might affect stacks assembly (Paris et al., 2017). Therefore, to find the combination that gives the most accurate information for wheat, with a subset of samples, the de novo pipeline was run several times, varying only one parameter with each run, as described by Paris et al. (2017). Next, using the optimum parameters for wheat, *denovo_map.pl* pipeline was conducted. Accordingly, stacks from each population were created de novo by *ustacks*; to estimate possible loci, a catalog from these stacks was generated by *cstacks*; and populations were matched to this catalog to determine the haplotypes at each locus for each population by *sstacks*. After determining the haplotypes, the *populations* program of the software was run to calculate the heterozygosity/nucleotide diversity values for both variant positions and all positions (variant & fixed) using Hardy-Weinberg equilibrium, and fixation index (F_{ST}) values of wheat populations. Using these values, pairwise F_{ST} matrices were drawn with R v4.2.2 (R Core Team, 2022), and principal coordinate analysis (PCoA) was performed using adegenet package (Jombart & Ahmed, 2011) in R. The *populations* program of *Stacks* was also used to obtain the final data in *Structure* format. These *Structure* data were used to determine clusters of wheat populations by *Structure* software v2.3 (Pritchard, Stephens, & Donnelly, 2000). The analysis was conducted using $K=1-5$ with five replicates of each K value. For each run, 25,000 burnin period and 25,000 Markov Chain Monte Carlo (MCMC) repeats were performed using admixture and correlated allele frequencies models. The true value of clusters (K) was determined using the approach described by Evanno, Regnaut, & Goudet (2005) by Structure Harvester v0.7 (Earl & vonHoldt, 2012). The bar plots were drawn by Pophelper v1.0.10 (Francis, 2017).

3.2. Heat Stress Experiment & RNA Sequencing

3.2.1. Plant Materials

The heat stress experiment was carried out with two einkorn (*T. monococcum*) populations (WP22 and WP89, Table 3.5), which were also used in the first section of the study. The seeds of WP22, which

were collected in 1995, were provided by Aegean Agricultural Research Institute (Ege Tarımsal Araştırma Enstitüsü – ETAE). The exact location of the collection, including latitude and longitude, is not known for this population. On the other hand, the seeds of WP89 were collected from the village of Enbiya, Kastamonu (Figure 3.1) by one of our former lab members in 2015.

Table 3.5. Einkorn populations used in heat stress experiment.

	Population	Year of Collection	City	Village	Source
Einkorn (<i>T. monococcum</i>)	WP 22	1995	Kastamonu	NA	ETAE
	WP 89	2015	Kastamonu	Enbiya	Collected

3.2.2. Heat Stress Experiment

The heat stress experiment was conducted in the growth chamber at the Department of Molecular Biology and Genetics of Boğaziçi University. The seeds of two einkorn populations were planted into pots with a capacity of 0.2 liters (65x55x78 mm) using peat soil produced in Turkey. For each population, there were three pots placed randomly. The number of seeds planted in each pot is shown in Table 3.6.

Table 3.6. The numbers of seeds planted and germinated for the two einkorn populations.

Population	Number of seeds planted			Number of seeds germinated/collected			Number of samples used for RNA-seq		
	Pot I	Pot II	Pot III	Pot I	Pot II	Pot III	Pot I	Pot II	Pot III
WP 22	18	18	19	14	12	11	6	6	6
WP 89	15	15	15	15	15	15	6	6	6

After planting, the parameters of the growth chamber were set to the values given in Table 3.7. These values, except stress temperature, were determined considering the optimum requirements for wheat growth. The optimum temperature range for wheat growth is 12-22 °C (Farooq, Bramley, Palta, & Siddique, 2011), and leaf initiation and sowing to emergence generally occur at 22 °C (Porter & Gawith, 1999). Since the plants were planned to grow until the 1-2 leaves stage for the heat stress experiment, 22 °C was selected as the control temperature. On the other hand, the stress temperature was obtained by climatic modeling as the maximum temperature (36.5 °C) for the year 2075 in Kastamonu. In Kastamonu, einkorn wheat is generally sown in October or in April (Hendek Ertop & Atasoy, 2019) when the natural photoperiod is around 12 h day/12 h night (WeatherOnline, n.d.). Likewise, this photoperiod has been primarily used in studies to grow wheat species (Qin et al., 2008; Fox et al., 2014; Perdomo et al., 2015; Zhang et al., 2017). So, a light/dark period of 12h/12h was used during the

experiment. Since the optimum relative humidity for wheat is 60-70 % (Hopper et al., 1997), it was set to 60 % in the growth chamber. Wheat's optimum light intensity is 250-450 $\mu\text{mol}/\text{m}^2/\text{sec}$ (Hopper et al., 1997). On the other hand, the light intensity mainly used in studies conducting wheat growth ranges from 150-500 $\mu\text{mol}/\text{m}^2/\text{sec}$ (Fox et al., 2014; Perdomo et al., 2015; Kumar et al., 2015; Su et al., 2019). Considering these values, a moderate intensity of 300 $\mu\text{mol}/\text{m}^2/\text{sec}$ was selected for the experiment.

Table 3.7. The parameters set for wheat growth in the growth chamber.

Parameter	Value
Control temperature	22 °C/18 °C (day/night)
Stress temperature	36.5 °C
Photoperiod	12 h day/12 h night
Humidity	60 %
Light intensity	300 $\mu\text{mol}/\text{m}^2/\text{sec}$

The temperature in the growth chamber was acclimated to 22 °C (light)/18 °C (dark) for the first 14 days. After 14 days, when the seedlings were at the 1-2 leaves stage, it was increased directly to the stress temperature, 36.5 °C, while all other parameters were set accordingly and kept constant during the whole experiment. Such a rapid and high temperature increase has been used in studies to investigate the impacts of heat stress on wheat (Chauhan et al., 2011; Liu et al., 2015; Liu et al., 2018). Following the setup of the parameters, the seeds were irrigated and watered regularly (once in two or three days) until the experiment was completed. Additionally, the pots were relocated periodically to eliminate the possible impact of location inside the growth chamber.

For each population, the number of seeds germinated is given in Table 3.6. For a statistically robust RNA sequencing analysis, the minimum number of replicates required per condition is three (Schurch et al., 2016). Since there were two populations with three pots of each, and two temperature conditions (control and stress) in this study, at least six samples per pot per population (3 replicates/population/pot/temperature x 1 population x 1 pot x 2 temperatures = 6) had to be included to the analysis. So, on the 14th day of growing, leaves from half of the seedlings from each pot were collected into sterile microcentrifuge tubes, six of them as control samples, and the remaining as backups. The collected leaves were frozen immediately in liquid nitrogen. After snap freezing, leaves were disrupted using sterile pestles (Axygen, USA) attached to a hand drill while keeping the tubes submerged in liquid nitrogen. Next, DNA/RNA shield (Zymo, USA) was added to the disrupted tissues to protect RNA from degradation. The samples were kept in dry ice until they were stored at the -80 °C freezer (Thermo Scientific, USA).

Following the first harvest (control samples and their backups), the temperature was increased to the stress value (36.5 °C) at the beginning of the light cycle on the 15th day. After exposure to heat stress for 24 hours, the second harvest (heat stress samples and their backups) was conducted by collecting leaves from the other half of the seedlings as described above. The disrupted tissues were stored at the -80 °C freezer (Thermo Scientific, USA) until RNA extraction.

3.2.3. RNA Extraction

Even though leaves from all germinated seeds were collected (Table 3.6), RNA was isolated from four samples per pot and per temperature following the instructions of Quick DNA/RNA Microprep Plus Kit's (Zymo, USA) manufacturer. In total, 48 RNA samples (4 samples/population/pot/temperature x 2 populations x 3 pots x 2 temperatures) were obtained. The concentrations were measured with a Qubit 3 Fluorometer (Invitrogen, USA) using RNA Broad Range (BR) Assay Kit (Invitrogen, USA). Purity ratios were calculated by a nanophotometer (Implen, USA). Next, the samples were analyzed for their integrities with 'Plant RNA Assay' using Eukaryote Total RNA 6000 Nano Chip (Agilent Technologies, USA) on a Bioanalyzer 2100 System (Agilent Technologies, USA). This service was purchased from the company SEM (Turkey).

3.2.4. Library Preparation and Sequencing

Out of four, three samples per pot and per temperature with the highest RINs and rRNA ratios were selected for RNA sequencing (Figure 3.2). Thus, 36 RNA samples (3 replicates/population/pot/temperature x 2 populations x 3 pots x 2 temperatures) were sent to Macrogen (Seoul, Korea) for library preparation and sequencing. Before library preparation, the integrity check was performed again with 'Plant RNA Assay' using Eukaryote Total RNA 6000 Nano Chip (Agilent Technologies, USA) on a Bioanalyzer 2100 System (Agilent Technologies, USA) by Macrogen (Seoul, Korea). Next, RNA libraries were prepared using TruSeq Stranded mRNA Library Prep Kit (Illumina, USA) following the instructions of TruSeq Stranded mRNA Reference Guide # 1000000040498 v00 protocol. The quality and quantity control of the libraries was performed by 2200 TapeStation system (Agilent Technologies, USA) using D100 ScreenTape (Agilent Technologies, USA) and by qPCR following the instructions of Illumina qPCR Quantification Protocol Guide. The libraries were sequenced in an Illumina platform, NovaSeq, producing 150 bp paired-end reads.

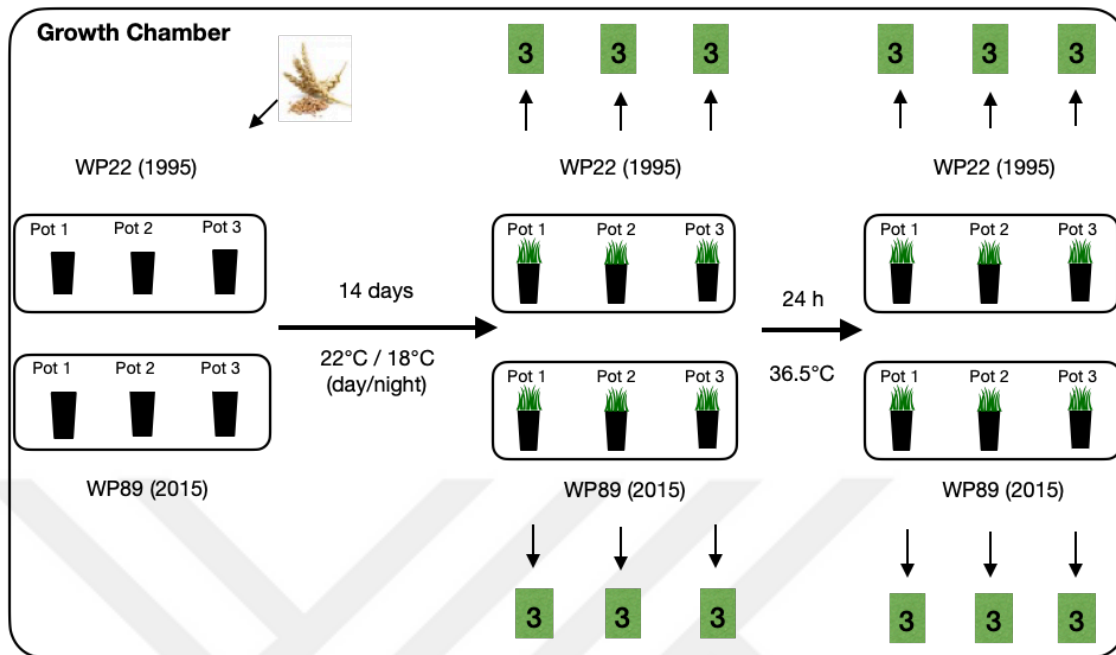


Figure 3.2. The schematic representation of heat stress experiment. The numbers in green boxes indicate the RNA-seq libraries prepared and sequenced.

3.2.5. Data Analyses

The workflow proceeded for the analyses of RNA-seq data is shown in Figure 3.3. Accordingly, raw reads were prepared for differential gene expression analysis by removing Illumina universal adapter sequences from the transcripts using Trimmomatic v0.36 (Bolger, Lohse, & Usadel, 2014). With Trimmomatic, transcripts below 25 bp in length were removed, as well. The quality control of the reads before and after trimming was performed with FastQC v0.11.9 (Andrews, 2010). After the quality check, all transcripts from 36 libraries were processed to construct a de novo assembly with Trinity software v2.14.0 (Grabherr et al., 2013). The statistics of the assembly were obtained with *TrinityStats.pl* script found in the Trinity software package. The completeness of the assembly was checked with BUSCO v3 (Seppey, Manni, & Zdobnov, 2019) through Omicsbox v3.0.30 (2019). To perform BUSCO analysis, the poales_odb10 reference database was used. Short contigs were filtered out from the assembly using seqkit command line toolkit v2.3.1 (Shen, Le, Li, & Hu, 2016). Redundancy removal was performed with CD-HIT-EST v4.8.1 (Li & Godzik, 2006).

After filtering out short contigs and redundancy removal, each transcriptomic datum pair (forward-reverse) for 36 libraries was aligned separately to the assembly. The alignment was performed using

BWA-MEM program (Li, 2013) of the RNA-seq module of Sequencer v5.4.1 (Gene Codes Corp., USA). Each SAM file generated by BWA-MEM was converted into a BAM file and sorted with Samtools v1.16.1 (Li et al., 2009). The statistics of these sorted BAM files were calculated with Samtools using *depth* and *flagstat* commands.

From the sorted alignment data, a count file encompassing quantification results of the fragments that align to each feature (WP22-controlT, WP22-stressT, WP89-controlT, WP89-stressT) was generated with featureCounts function of the Rsubread package (Liao, Smyth, & Shi, 2019) in R v4.2.2 (R Core Team, 2022). The data evaluation and pre-processing before differential expression analysis were conducted with edgeR (Robinson, McCarthy, & Smyth, 2010; McCarthy, Chen, & Smyth, 2012) and limma (Ritchie et al., 2015) packages in R. For this purpose, first, the library sizes along with the average and median values were calculated. Next, filtering was performed by removing unexpressed and lowly expressed contigs from the analysis. The lowly expressed contigs were determined as the contigs with read counts of less than ten in a minimum number of samples. For filtering, count data were transformed into counts per million (CPM), and FilterByExpr function of edgeR was used. After filtering, library sizes were calculated again. Following these calculations, normalization was carried out using the trimmed mean of M-values (TMM) method. Using filtered and normalized data, the multidimensional scaling (MDS) analysis was performed to explore the clustering patterns of the RNA-seq libraries.

After data pre-processing, differential gene expression analysis was performed by edgeR package in R. A multi-factor experimental design was generated, and quasi-likelihood F test was used to determine differentially expressed contigs for each comparison. Old vs. new einkorn at control and stress temperatures, and control vs. stress temperatures for old and new einkorn were the four comparisons used for the analysis. Up-regulated and down-regulated contigs were determined using a false discovery rate (FDR) cut-off value of 0.05. Following the differential gene expression analysis, the results were visualized with mean-difference (MD) plots. The differentially expressed contigs were also counted using both FDR and log fold change (logFC) cut-off values of 0.05 and ± 1 , respectively. Using these differentially expressed contigs, volcano plots were drawn and the top 100 significant genes (based on their FDR values) showing differentiation were identified.

Using the assembly obtained after filtering out short contigs and redundancy removal, functional analyses were also conducted simultaneously with differential gene expression analysis. Blastx (Blast+ v2.13.0) was used to determine the hits to *Triticum aestivum* and *Triticum urartu* reference genomes and

NCBI RefSeq Plant, Swiss-prot and NR databases with an e-value of 1×10^{-7} . After each blastx run, the contigs with no hits were blasted against the next database, while the hit results were appended to each other. Subsequently, gene ontology (GO) mapping of the retained blast hits and GO annotations were made using Blast2GO suite (Götz et al., 2008) through OmicsBox v3.0.30 (2019). For the blast hits with few GO annotations, blastx was run again with the same e-value cut-off, against NR database with a taxonomy filter using Diamond v2.1.6 (Buchfink, Reuter, & Drost, 2021). Taxonomy filtering was performed by selecting the genus *Triticum* and the species of *Triticum monococcum*, *Triticum urartu*, and *Triticum aestivum*. In addition to GO mapping and annotation with Blast2GO, InterProScan (Jones et al., 2014) and EggNOG (Huerta-Cepas et al., 2019) analyses were performed through Omicsbox. The GO counts in biological process, molecular function, and cellular component were determined to see the categories that most contigs were mapped to and annotated to. The enzyme classes with the most common GO terms within were also determined. Following annotation, combined pathway analysis which uses Reactome (Fabregat et al., 2018), Plant Reactome (Naithani et al., 2020), and KEGG (Kanehisa & Goto, 2000) databases was conducted through Omicsbox. Gene set enrichment analysis (GSEA) was made to determine enriched gene sets and pathways, using the same software.

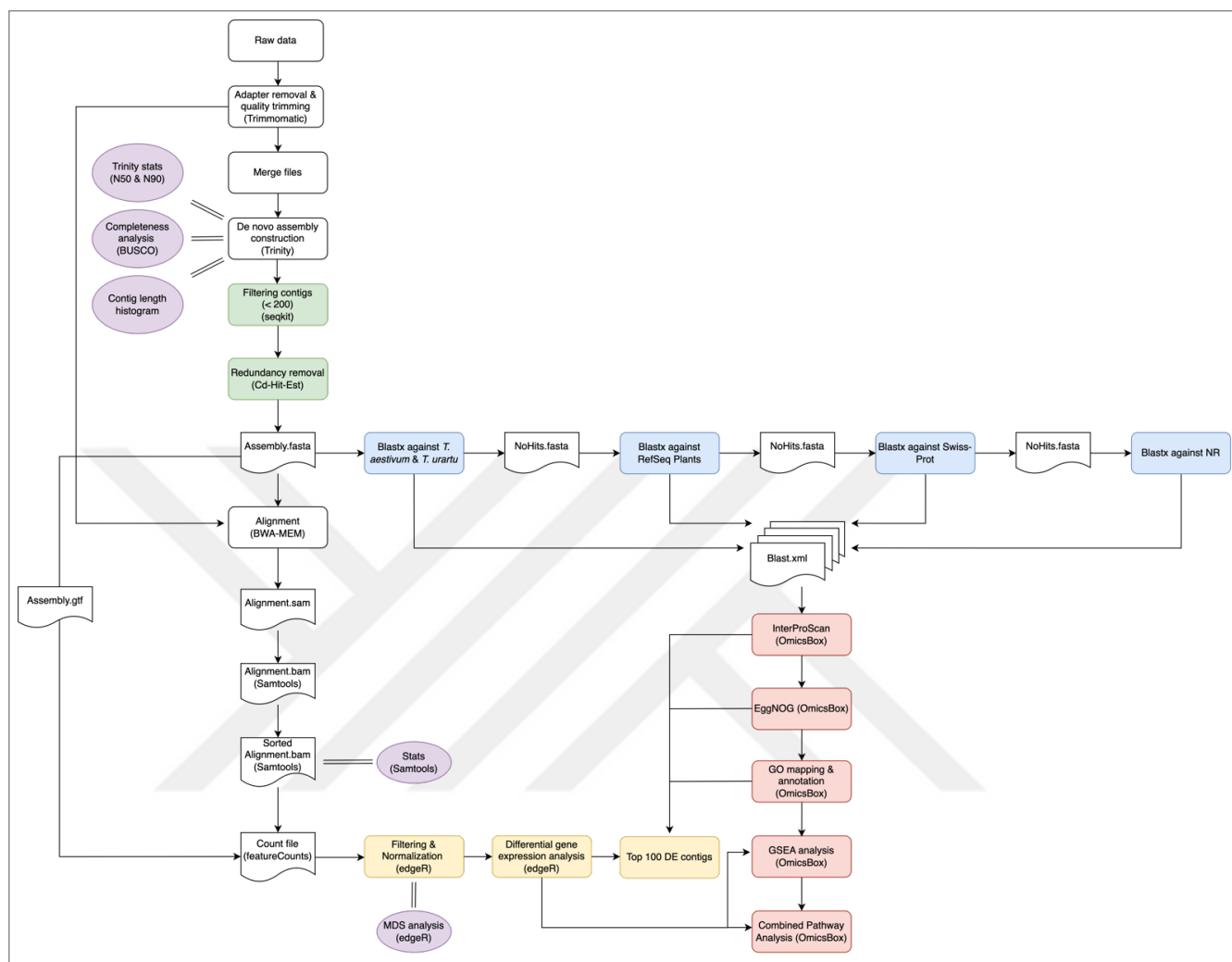


Figure 3.3. The workflow of the RNA-seq data analyses.

4. RESULTS

4.1. ddRAD-seq

4.1.1. Demultiplexing and Cleaning of Raw Reads

195 ddRAD libraries, multiplexed into four pools, were sequenced using a single lane for each pool. Pool1, Pool2, and Pool3, comprising 30, 56, and 75 libraries, respectively, were sequenced by Macrogen (Seoul, Korea) using the Illumina HiSeq 2500 platform. Pool4 containing 34 libraries, was sequenced by Gen-Era (Turkey) using NovaSeq platform. Raw reads, which were paired-end and 150 bp long, were cleaned and demultiplexed using the *process_radtags.pl* program of *Stacks* software. The number of retained reads obtained for each species is shown in Figure 4.1. The numbers of retained reads obtained for 33 libraries were lower than or around 5000, and the data of these libraries were excluded from the analysis. The total and retained reads obtained for the remaining 162 libraries are given in Table A.1 (Appendix A). For the total reads of these 162 libraries, mean value and standard deviation were calculated as 3940823.42 and 1677605.03, respectively (min=503426, max=10387490); whereas these values were 3525400.56 and 1546777.55 for retained reads, respectively (min=428234, max=10010588).

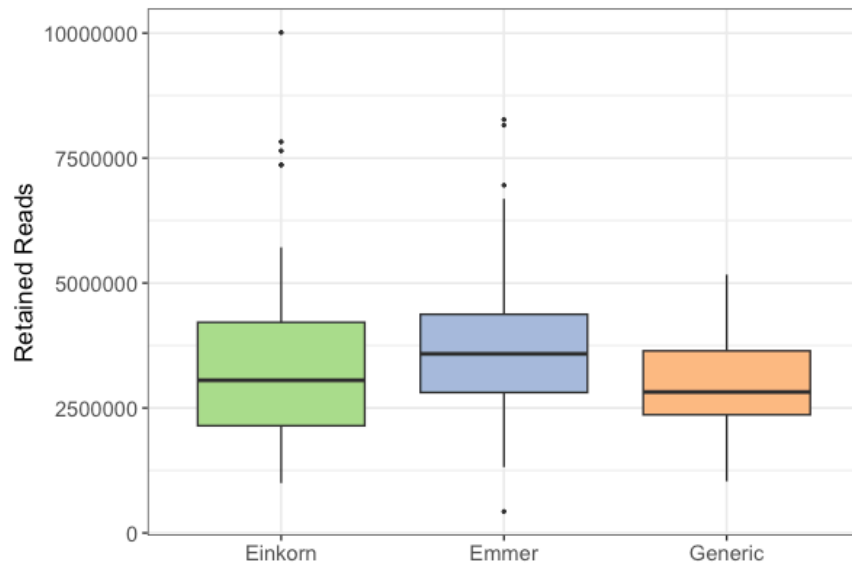


Figure 4.1. Retained reads obtained for each species after cleaning and demultiplexing.

4.1.2. De novo Pipeline Parameter Optimization

Before analyzing all ddRAD libraries, parameter optimization of *Stacks*' de novo pipeline for wheat was conducted with a subset of libraries following the instructions of Paris et al. (2017). As suggested in the article (Paris et al., 2017), the r80 method was used for the optimization in order to obtain loci shared by a minimum of 80 % of the samples in each population. m was tested from 2 to 7, and $M=n$ was tested from 0 to 9, according to the decision framework provided by Paris et al. (2017). The numbers of loci and SNPs found for each m and $M(n)$ value are given in Tables 4.1 and 4.2. The increment of polymorphic loci for each iteration of M is also given in Table 4.2.

Table 4.1. The numbers of loci and SNPs obtained for each m .

Parameter	Assembled_Loci	Polymorphic_Loci	SNPs
m2	10693	7586	18852
m3	6061	4448	12108
m4	3385	2512	7593
m5	1925	1459	4619
m6	1156	891	3076
m7	751	579	2183

Table 4.2. Loci and SNP numbers, and the number of new polymorphic loci added for each iteration of $M(n)$.

Parameter	Assembled_Loci	Polymorphic_Loci	SNPs		New Loci Added
M0	3093	1234	2649		
M1	6082	4268	10387	M0/M1	3034
M2	8125	6365	19669	M1/M2	2097
M3	9411	7759	29506	M2/M3	1394
M4	10280	8719	38257	M3/M4	960
M5	10912	9440	46744	M4/M5	721
M6	11346	9962	54384	M5/M6	522
M7	11680	10331	60865	M6/M7	369
M8	11930	10604	66273	M7/M8	273
M9	12077	10756	70315	M8/M9	152

As shown in Figure 4.2 (a), with increasing values of m , coverage increased as well. On the other hand, higher m values caused a decline in the numbers of assembled loci, polymorphic loci and SNPs (Figure 4.2 (b)). Even though m2 gave the highest numbers of loci and SNPs, the coverage for that value was the lowest. The coverage obtained with m4 was much higher than m2, but the numbers of loci and SNPs were very low. So, the value in between - m3, was determined as the optimal value to be used in the subsequent analyses.

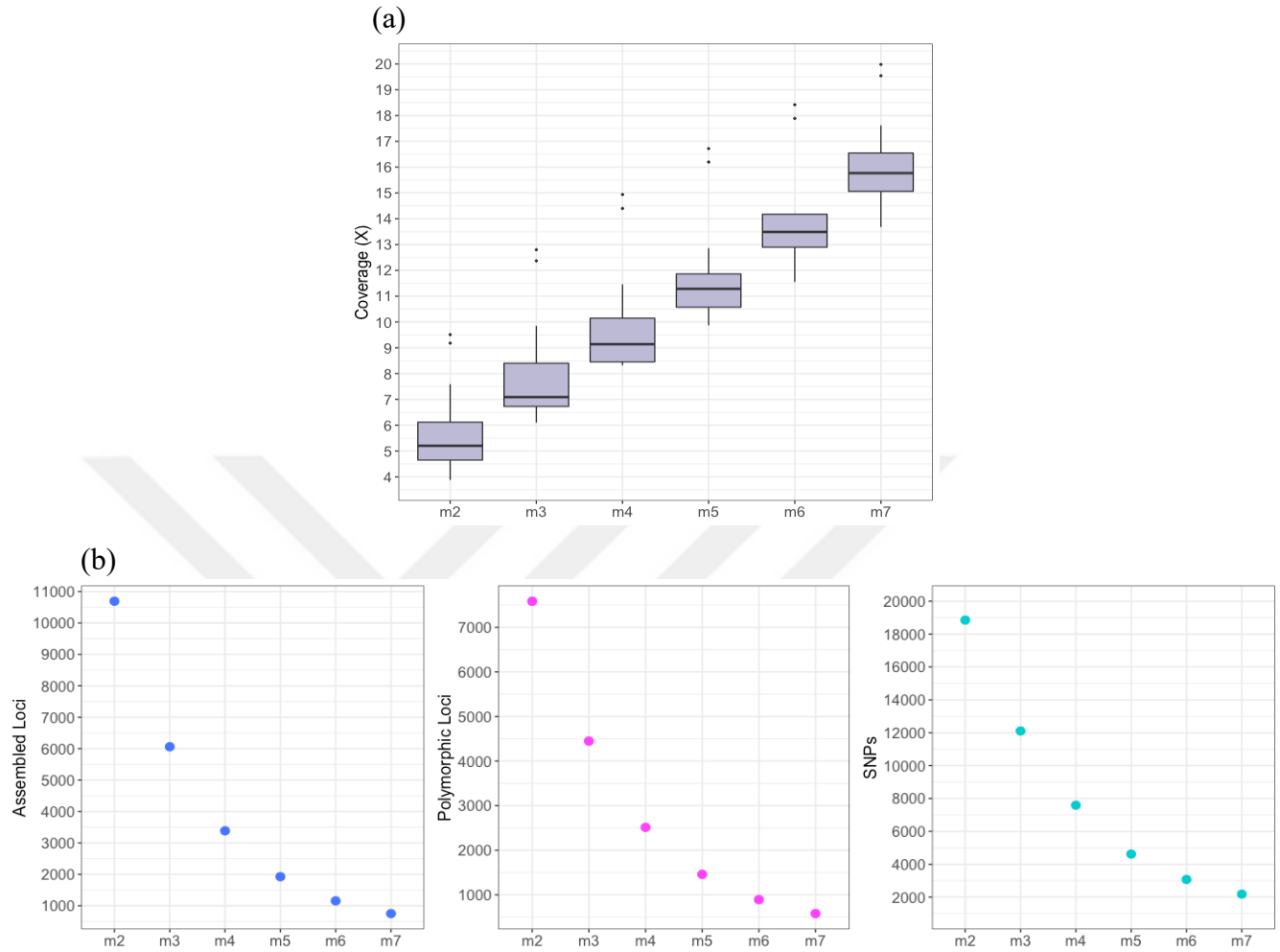


Figure 4.2. Plots of coverage (a) and assembled loci, polymorphic loci, and SNPs (b) varying with the increase of m .

In the case of M (which should be equal to n as suggested by Paris et al. (2017)), higher values of M increased the amounts of assembled loci, polymorphic loci, and SNPs (Figure 4.3 (a)). The number of loci (both assembled and polymorphic) began to plateau at approximately M5, whereas there was nearly a linear increase for the SNPs, which seemed to begin to flatten out after M9. On the other hand, new polymorphic loci added after approximately M5 were few (Figure 4.3 (b)). So, the optimal M and n values for wheat were determined as M5 and n5.

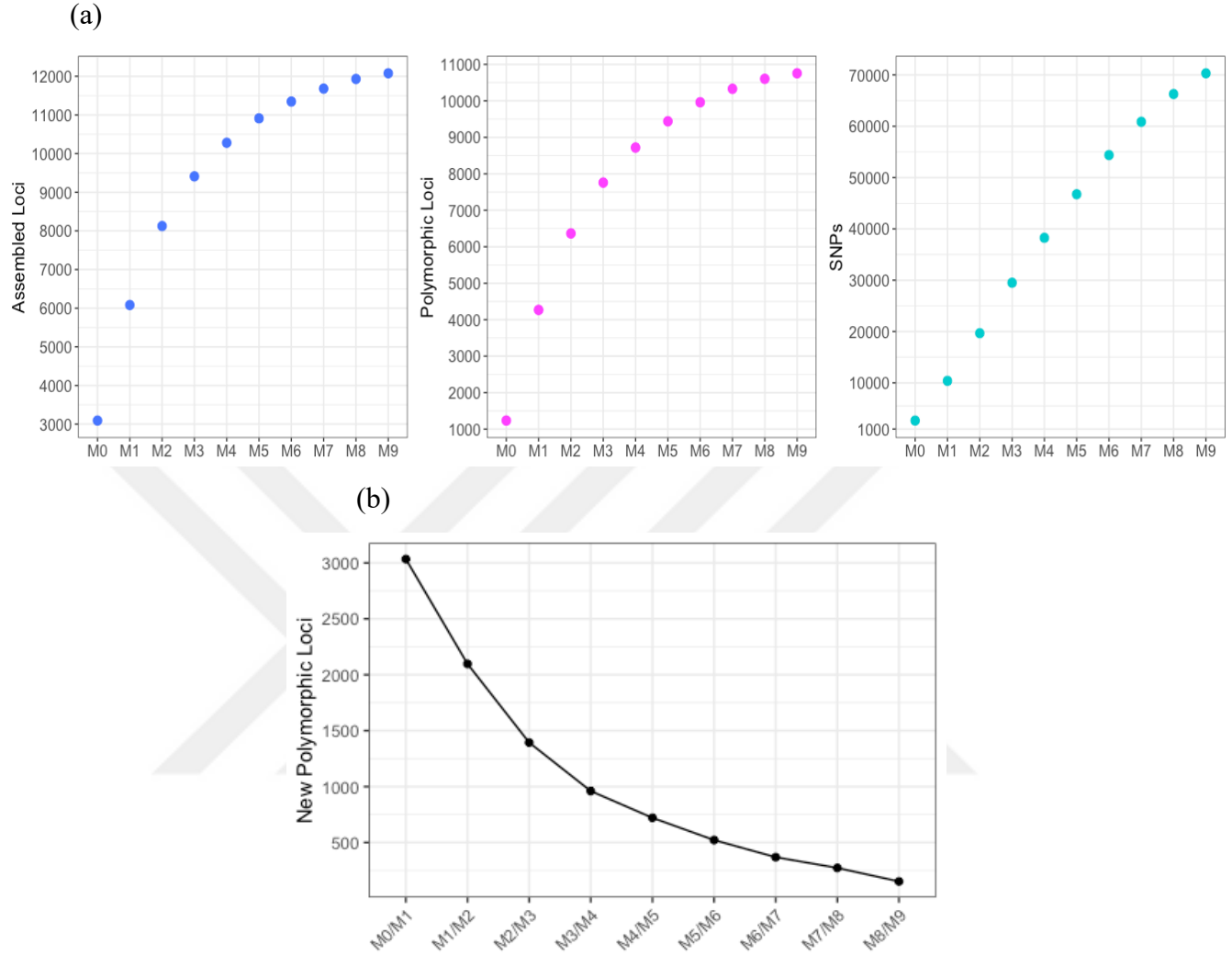


Figure 4.3. Plots of assembled loci, polymorphic loci and SNPs varying with the increase of $M=n$ (a) and the new polymorphic loci added with each increment of $M=n$ (b).

4.1.3. De novo Pipeline Running of Einkorn Wheat

The determined optimal parameters for wheat ($m=3$, $M=n=5$) were used for running the de novo pipeline of *Stacks* for each species separately. First, einkorn libraries were analyzed with and without the r80 method. In order to test the effect of using optimized parameters, de novo analysis of einkorn was also performed with default parameters ($m=3$, $M=2$, $n=1$) with and without the r80 method. The total numbers of loci and SNPs obtained with each de novo analysis for four einkorn populations (WP22 – Kastamonu, 1995; WP81, WP88, WP89 – Kastamonu, 2015) are given in Table 4.3. Accordingly, assembled loci numbers recorded for the analyses using both default and optimal parameters with the r80 method were similar, but there were more polymorphic loci and SNPs for the optimum parameters

indicating the success of parameter optimization. Without the r80 filtering, much more loci and SNPs were produced than the runs with the r80 method, as expected. Among the four analyses, the highest numbers of loci and SNPs were obtained using the optimum parameters without the r80 method, which is also suggested by Rivera-Colon & Catchen (2021) for the execution of the de novo pipeline with the whole dataset.

Table 4.3. Locus and SNP numbers obtained for einkorn with different de novo analyses.

De novo Analysis	Assembled Loci	Polymorphic Loci	SNPs
Default parameters with r80	38484	8480	15184
Default parameters without r80	491636	81094	194403
Optimized parameters with r80	37468	10076	25411
Optimized parameters without r80	345911	89389	309005

Figure 4.4 shows the observed heterozygosity and nucleotide diversity values recorded for variant sites (for more detail see Table A.2, Appendix A). For the first analysis using default parameters with the r80 method, the maximum observed heterozygosity was recorded for WP88 (0.19750), whereas WP22 had the minimum observed heterozygosity value (0.12492). The maximum expected heterozygosity was calculated for WP81 (0.23612), while the minimum value was recorded for WP22 (0.18518). Likewise, nucleotide diversity was found to be maximum for WP81 (0.25019) and minimum for WP22 (0.18963).

For the analysis using default parameters without the r80 method, the maximum observed heterozygosity was recorded as 0.28222 for WP22, and the minimum observed heterozygosity was recorded as 0.24794 for WP89. The maximum expected heterozygosity value was 0.26776 for WP22, while the minimum value was 0.19875 for WP89. Similarly, WP22 had the highest nucleotide diversity value (0.32569), and WP89 had the lowest (0.28047).

The third analysis, which was run with optimized parameters and the r80 method, resulted in a maximum observed heterozygosity for WP88 (0.22803) and a minimum observed heterozygosity for WP22 (0.15502). The maximum expected heterozygosity recorded for this analysis was for WP81 (0.26968), while WP22 had the minimum expected heterozygosity value (0.21362). Similarly, WP81 and

WP22 were found to have the highest (0.28576) and lowest (0.21877) nucleotide diversity values, respectively.

For the last analysis using the optimum parameters without the r80 method, the maximum observed heterozygosity was obtained for WP88 (0.20925), followed by WP22 (0.20587), whereas WP81 had the minimum observed heterozygosity (0.19740). The maximum and minimum expected heterozygosity values were recorded for WP22 (0.24649) and WP89 (0.18132), respectively. Likewise, WP22 had the highest nucleotide diversity value (0.29469), while WP88 had the lowest (0.24161).

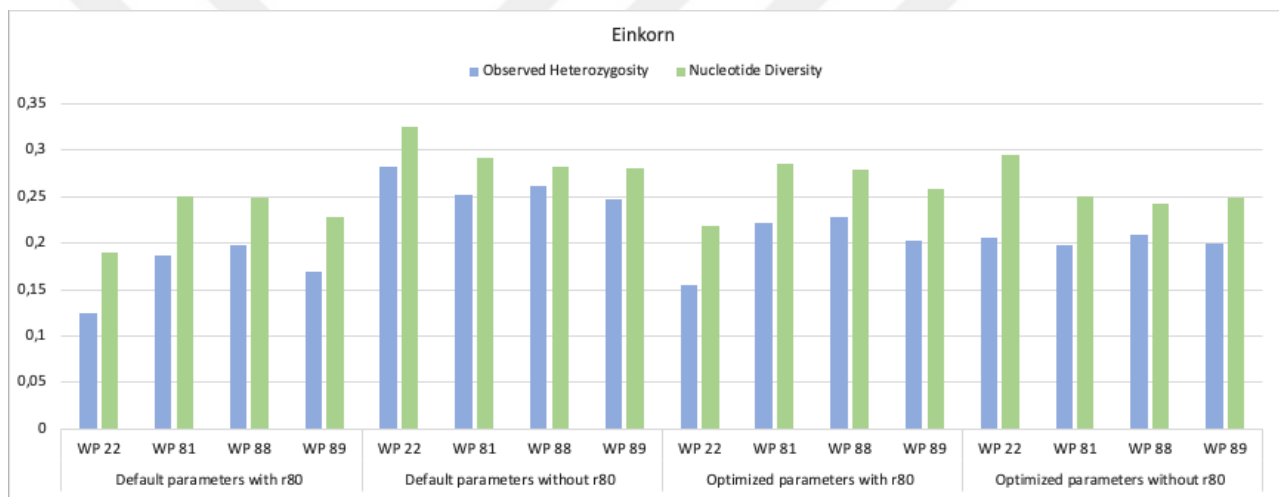


Figure 4.4. Observed heterozygosity and nucleotide diversity values of einkorn populations calculated for variant positions.

The same calculations were also done for all sites (variant & fixed) (Table A.3, Appendix A). For the analysis using default parameters with the r80 method, the maximum observed heterozygosity was calculated for WP88 (0.00031), and the minimum observed heterozygosity was calculated for WP22 (0.00026). The maximum expected heterozygosity was recorded for WP22 as 0.00038, while WP81 and WP88 had the lowest expected heterozygosity value (0.00036). The maximum nucleotide diversity was 0.00040 for WP89, whereas for WP22, WP81, and WP88, this value was 0.00039.

The analysis using default parameters without the r80 method resulted in a maximum observed heterozygosity value of 0.00087 for WP88 and a minimum value of 0.00067 for WP22. The maximum expected heterozygosity was calculated for WP88 as 0.00068, whereas the minimum value was

calculated for WP89 (0.00062). The maximum and minimum nucleotide diversity values were recorded for WP88 (0.00093) and WP22 (0.00078), respectively.

The third de novo analysis, which used optimized parameters with the r80 method, generated maximum observed heterozygosity for WP81 (0.00058) and minimum observed heterozygosity for WP22 (0.00048). The maximum expected heterozygosity value obtained for WP81 was 0.00071, whereas WP22, WP88, and WP89 had a lower value (0.00066). For the nucleotide diversity records, the maximum value was for WP81 (0.00075), and the minimum value was for WP22 (0.00068).

The last analysis using the optimized parameters without the r80 method resulted in the highest observed heterozygosity value for WP88 (0.00120), while the minimum value was 0.00103 for WP22. The maximum and minimum expected heterozygosity values were obtained for WP22 (0.00124) and WP88 (0.00105). Likewise, WP22 had the maximum nucleotide diversity (0.00148) among four einkorn populations, and WP88 had the minimum (0.00139).

The fixation index (F_{ST}) is a measure that describes the genetic variation among populations (Holsinger & Weir, 2009). At the end of *Stacks*' de novo pipeline, F_{ST} values were calculated by *populations* program (Figure 4.5). The pairwise matrix generated by the analysis using default parameters with the r80 method is shown in Figure 4.5 (a). For this analysis, the highest F_{ST} was between WP81 and WP89 (0.0496883). The following values were 0.0492009 and 0.0459587 for WP88–WP89 and WP81–WP88, respectively. On the other hand, the lowest value was calculated between WP81 and WP22 (0.0212095), followed by WP89–WP22 (0.0220047) and WP88–WP22 (0.0248614). In concordance with these values, the pairs of the new einkorn populations (WP81, WP88, WP89, collected in 2015) resulted in dark blue, while the pairs of old (WP22, collected in 1995) and new populations produce a light purplish color in the matrix. According to the scale, as F_{ST} values decrease, the blue color becomes lighter, turning light purple or pink at the bottom of the scale. So, the color gradient observed for this analysis indicates a higher genetic distance within the new populations than the old and new ones. Since a genetic distinction between old and new populations was expected, the distinction within the new ones is inconsistent with the expectation.

The analysis performed with default parameters but not using the r80 method resulted in similar F_{ST} values for all population pairs, with an unclear color gradient (Figure 4.5 (b)). Nevertheless, the order of F_{ST} values still indicates a higher genetic distance between the new populations than the distance between

the old and new ones. The highest F_{ST} was 0.176716 between WP81–WP89, followed by 0.157958 for WP89–WP22, 0.152362 for WP81–WP88, and 0.151129 for WP88–WP89. The lowest F_{ST} was found for the pair of WP88–WP22, and WP81–WP22 (0.130530) followed that value.

The color scale observed for the analysis performed with optimal parameters and the r80 method (Figure 4.5 (c)) was similar to the color scale obtained for the analysis using default parameters and the r80 method (Figure 4.5 (a)). The highest F_{ST} was calculated for WP88–WP89 (0.0485338), and the values obtained for WP81–WP89 (0.0478997) and WP81–WP88 (0.046752) were close to that F_{ST} value. The lowest F_{ST} was obtained between WP81 and WP22 (0.0229849), followed by WP89–WP22 (0.0233563) and WP88–WP22 (0.0262587).

On the other hand, the pairwise F_{ST} matrix generated by the analysis using optimum parameters without the r80 method (Figure 4.5 (d)) was different from the previously mentioned matrices in population pairs of maximum and minimum values recorded. The maximum F_{ST} was calculated between WP89 and WP22 (0.238923), followed by WP81–WP22 (0.225718), and WP81–WP89 (0.193500). The value obtained for the pair of WP88–WP22 (0.192276) was also close to those values. The minimum value was 0.157580 between WP81 and WP88, followed by the value obtained for WP88–WP89 (0.163516). The F_{ST} results of this analysis were the fulfillment of a significant differentiation of populations, which requires F_{ST} values higher than 0.15 (Frankham, Ballou, & Briscoe, 2002). Additionally, although the F_{ST} value calculated for WP81–WP89 was closer to the values obtained for WP22 and the new populations, unlike the other three analyses, this *de novo* analysis' F_{ST} results indicate a higher genetic distance between WP22 (1995) and the populations from 2015 than the distance within the populations from 2015. This is in line with the expectation of populations' possible genetic distinction based on the years when they have been collected.

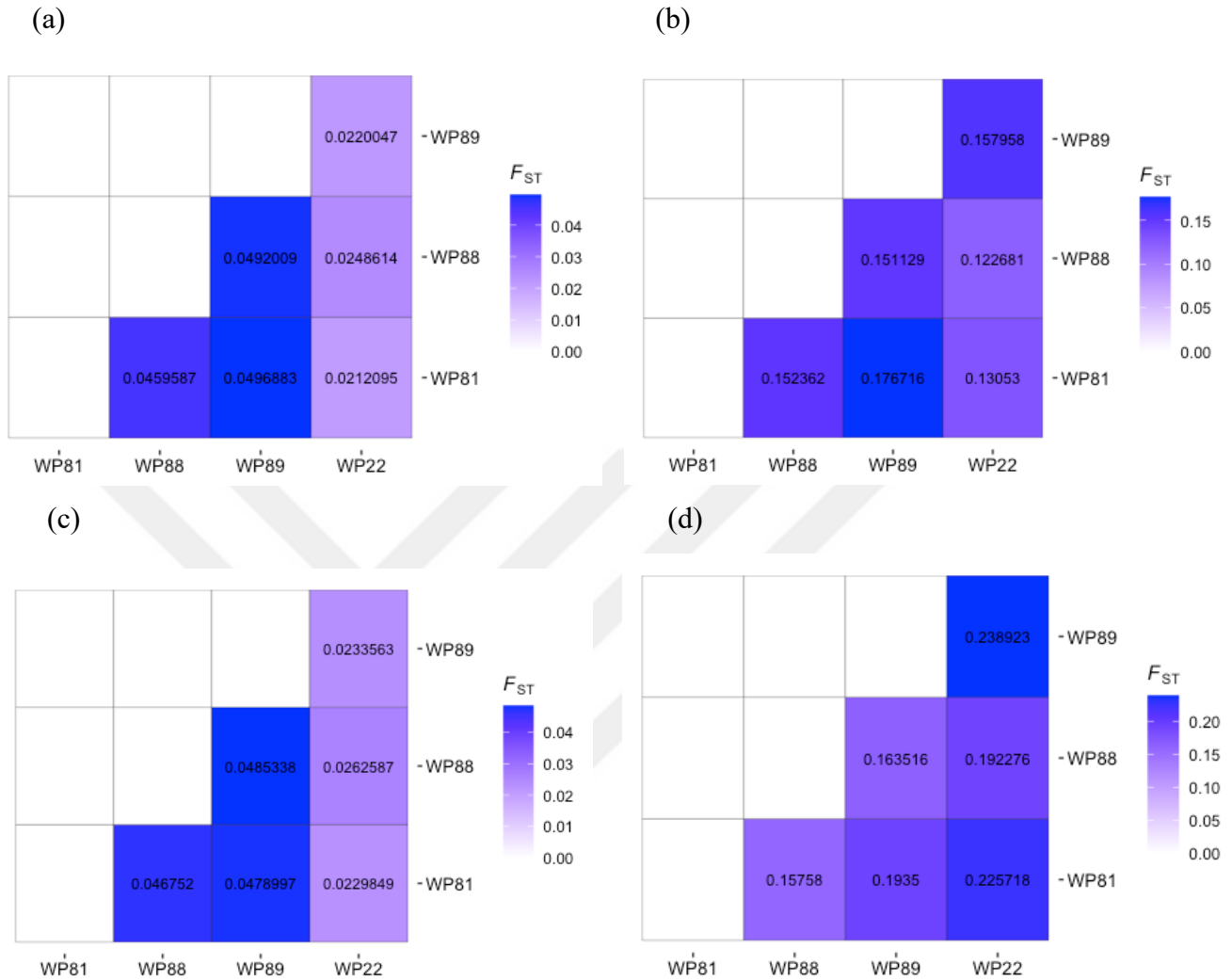


Figure 4.5. Matrices of pairwise F_{ST} values of einkorn populations calculated with default parameters with r80 (a), default parameters without r80 (b), optimized parameters with r80 (c), and optimized parameters without r80 (d).

4.1.4. Structure Analysis of Einkorn Wheat

The clustering of einkorn populations was also investigated with *Structure* software (Pritchard et al., 2000). The input data for the program were generated by *Stacks' populations* program using *write_single_snp* flag after de novo analyses. Since *Structure* can only deal with unlinked loci (Pritchard et al., 2000), *write_single_snp* flag was used for all datasets as it allows extracting only the first SNP of each RAD locus. The only exception was the dataset obtained with optimized parameters and the r80 method. The *populations* run with this dataset was also performed without *write_single_snp* flag. Additionally, for some datasets, after extracting the *Structure* input files, a filtering was applied again by

discarding loci having more than 20 % of missing data (in total, for all libraries). Thus, they contained only the loci present in more than or equal to 80 % of all ddRAD libraries. Since the r80 method used in the de novo pipeline is applied on a population basis, this filtering applied to the whole dataset provided an additional reduction of missing data. The datasets, which were used for *Structure* analysis, and the number of loci they contained are given in Table 4.4.

Table 4.4. The information about the datasets used in einkorn's *Structure* analyses.

	Dataset	Number of Loci	Determined K value
1	Default parameters, r80, <i>write_single_snp</i> , missing data filtering	1837	2
2	Default parameters, <i>write_single_snp</i>	81094	3
3	Optimized parameters, r80, <i>write_single_snp</i> , missing data filtering	2051	3
4	Optimized parameters, r80, missing data filtering	5016	4
5	Optimized parameters, r80, <i>write_single_snp</i>	10028	4
6	Optimized parameters, <i>write_single_snp</i>	89398	2

For each *Structure* analysis, the most likely number of clusters (K) was determined by Structure Harvester (Earl & vonHoldt, 2012) based on the method described by Evanno et al. (2005). The graphs showing the best K value for each structure run are shown in Figures A.1, A.2, A.3, and A.4 (Appendix A). As given in Table 4.4, for the first and sixth datasets, the value of K was calculated as two, while K=4 was the most likely cluster number for the fourth and fifth analyses. On the other hand, for the second and third datasets, the number of clusters for einkorn populations was found to be three.

Following the determination of the most likely K value for each dataset, the results were plotted with Pophelper (Francis, 2017). The bar plots are shown in Figure 4.6 for the determined K values. Since there were two groups within einkorn populations as old (seed bank seeds collected in 1995) and new (in-situ conservation seeds collected in 2015), the libraries were ordered in the plots based on this criterion to see if the clustering pattern is consistent with the expectation. For the datasets where K was calculated as another number rather than two, results were visualized for K2, as well (Figure 4.7).

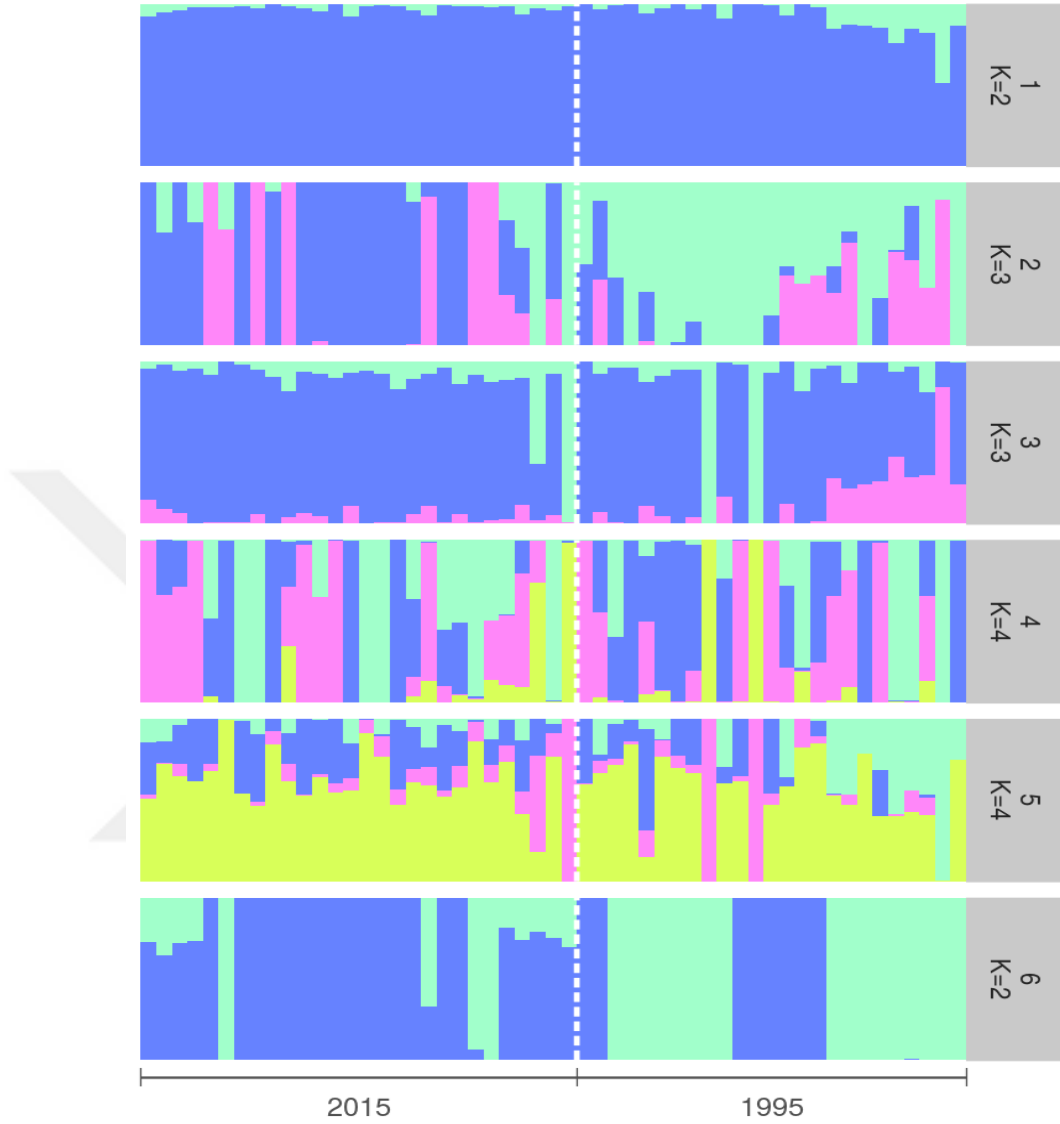


Figure 4.6. *Structure* results of the six datasets given in Table 4.7 for the determined K values.

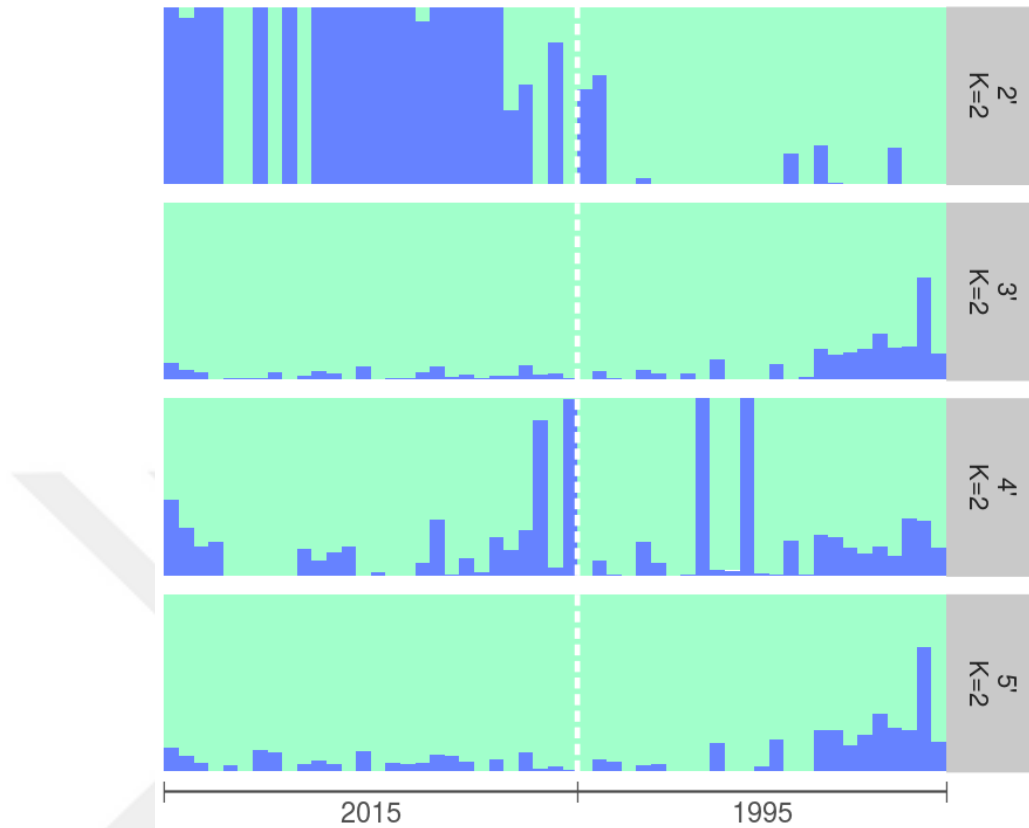


Figure 4.7. *Structure* results of the second, third, fourth, and fifth datasets (Table 4.4) for $K=2$.

Considering all the plots (Figures 4.6 and 4.7), the second and sixth datasets (default and optimized de novo parameters with *write_single_snp* flag) gave the most expected results, in which the expected distinction between the two einkorn groups (old vs. new) was seen. As seen in Figure 4.8, representing the same results of the two analyses on both population and year bases, the separation between the two groups was more complete for the second analysis visualized for $K=2$, than the one obtained for the sixth. However, for the second dataset, the most likely K value was determined as three (Figure 4.6). For the sixth dataset, when the libraries from 1995, which were clustered with the libraries from 2015, were evaluated, no plausible reason was detected for the differentiation. One possible reason for that pattern of clustering might be the higher number of loci obtained with the optimum parameters than defaults. Considering the results of the F_{ST} analyses mentioned above, the distinction between old and new einkorn populations was clearer for the dataset of optimized parameters than the one in which the default parameters were used (Figure 4.5). So, it was thought that the clusters observed for the sixth dataset are most probably the true grouping of einkorn. Even though further analyses were conducted with the two datasets, the results of the analyses using the sixth dataset are given below, while the results for the second dataset are shown in Appendix A (Figure A.5).

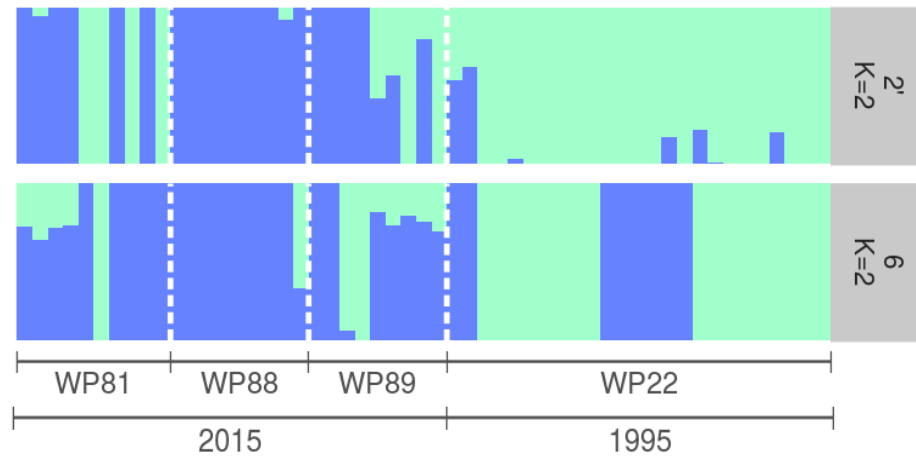


Figure 4.8. *Structure* results of the second (above) and sixth (below) datasets showing einkorn populations.

For the six *Structure* runs mentioned above (Table 4.4), each ddRAD library was defined according to which population it belongs to. In order to examine the possible batch effect caused by sequencing of ddRAD libraries in different pools and the effect of collection year specifically, additional *Structure* analyses were performed. The results given below are for the sixth dataset (optimized parameters with *write_single_snp*), which was modified by the libraries defined by the population, year of collection, and the sequencing batch, separately. Furthermore, the effect of missing data on the analysis was investigated using the same dataset. In general, *Stacks' populations* program generates *Structure* input files encoding missing data as zero. However, to run *Structure*, missing data is generally encoded as -9, which is a different number from the ones (1, 2, 3, 4) indicating the four bases (A, T, G, C). Hence, while running *Structure*, the missing data were defined as -9 as suggested, but they were encoded as zero and -9 in the input file to see if there is an effect of different codes of missing information on the cluster formation.

Figure 4.9 shows the results of *Structure* runs for the libraries defined by population. Figure 4.9 (a) was drawn for the missing data encoded as -9, while Figure 4.9 (b) was drawn for the missing data encoded as zero. For -9, the most likely K value was found to be two, as shown in the upper plot of Figure 4.9 (a). The clustering was relatively in line with the expectation of two einkorn groups (old vs. new), as mentioned above. Since there were four populations, the same result was also visualized for K=4 (the bottom plot), but no separation based on population was seen.

In Figure 4.9 (b), the results for missing data=0 are shown on the population bases. The upper plot was for the most likely K value (K=3), while the middle was for K=2, and the bottom was for K=4.

Accordingly, there was no clear separation between the populations for all three K values either by population or year. However, when compared with the bar plots in Figure 4.9 (a), it was seen that different results were obtained for the two values of missing data. When it was encoded as zero, an additional cluster was observed (Figure 4.9 (b), K=3, pink bars), which was almost evenly distributed among populations. These results suggest that there are no differences in terms of missing data for old and new populations, and most probably, any observed differences were due to the library preparation process but were not meaningful. On the other hand, for K=2, the distinction between old and new populations became less clear when the missing data were zero than when they were -9.

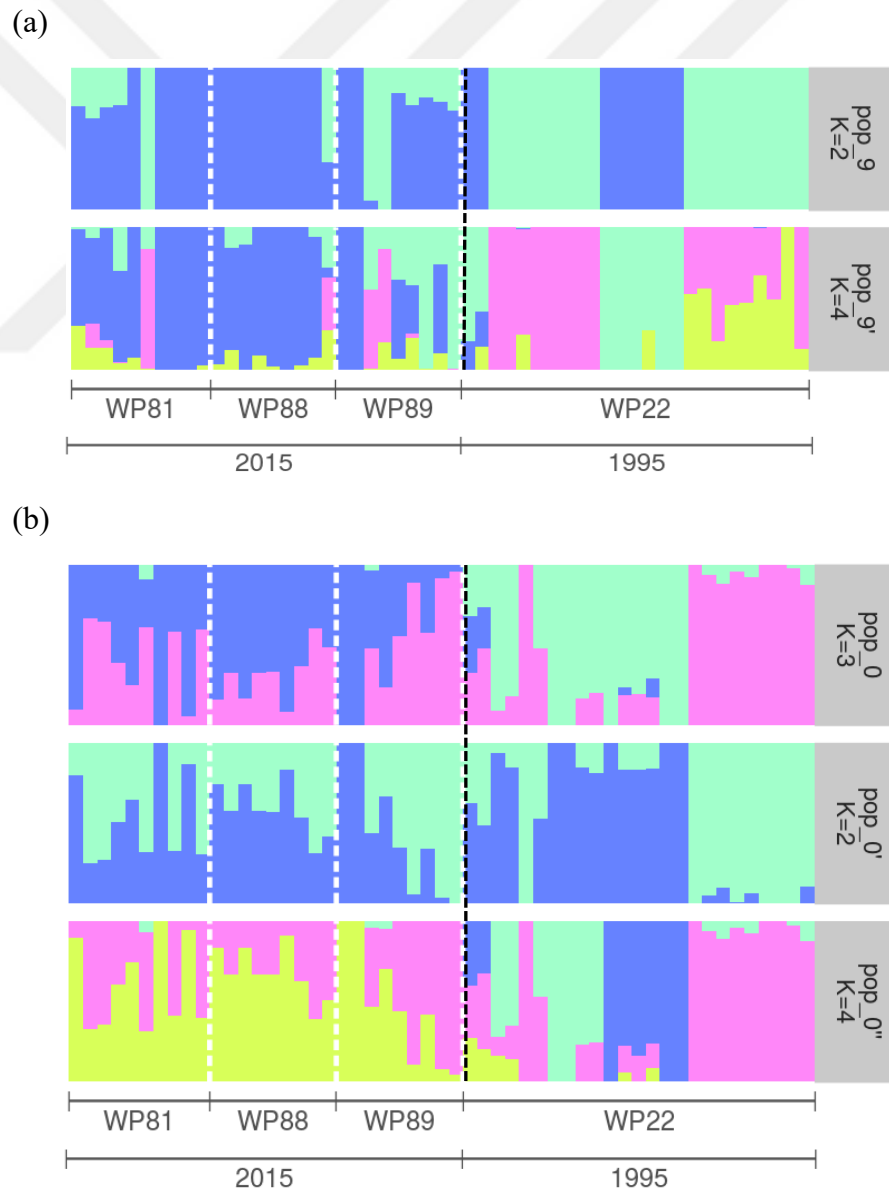


Figure 4.9. Population and year-based *Structure* results with missing data encoded as -9 (a), and zero (b). The black dashed lines indicate the distinction between years.

When the libraries were defined by the year of collection, the same results were obtained with the population-based analysis. Since the bar plots were exactly the same, instead of drawing them again, the separation by the year was shown in Figure 4.9 with extra black axis lines showing the distinction between the two years on the bar plots.

The *Structure* results of the libraries defined by the pools in which they were distributed for sequencing are shown in Figure 4.10. The results for the missing data encoded as -9 are shown in Figure 4.10 (a), while Figure 4.10 (b) shows the results for missing data=0. In Figure 4.10 (a), the upper plot was drawn for $K=2$, and the bottom plot was drawn for $K=4$. In the plot for $K=2$, a similar distribution was observed with the distribution of libraries obtained for population-based and year-based analyses. The fourth batch was grouped clearly, but for the other batches, the clustering was mixed. Likewise, for $K=4$, there was not a clear clustering among four batches. However, the first and second batches were relatively grouped together, whereas the grouping of the others was more mixed than the first two. This is not consistent, considering that the first three pools were sequenced by MacroGen, and the fourth was sequenced by Gen-Era.

In Figure 4.10 (b), the results of batch-based analysis with the missing data value of zero are shown. The upper plot was for the best K value determined for the dataset ($K=3$), the middle plot was for $K=2$, and the bottom one was for $K=4$. Accordingly, similar results were obtained with the population-based and year-based analyses for each K value. Even though the plot obtained for $K=2$ was consistent with the possible batch effect, it was thought that this was because of introducing the missing data to the program and did not affect the real clustering as seen in the upper plot in Figure 4.10 (a).

All in all, there were minor differences between the plots in Figures 4.9 and 4.10, but this was because of the sorting of the libraries by three different parameters (population, year, batch) while drawing the plots. In fact, the same results were obtained for each parameter regardless of the specification of the libraries. This indicates that the library specification during *Structure* analysis affected the clustering slightly. Using “zero” for the missing data instead of -9 also did not have a meaningful effect on the analysis, and there were no batch effects.

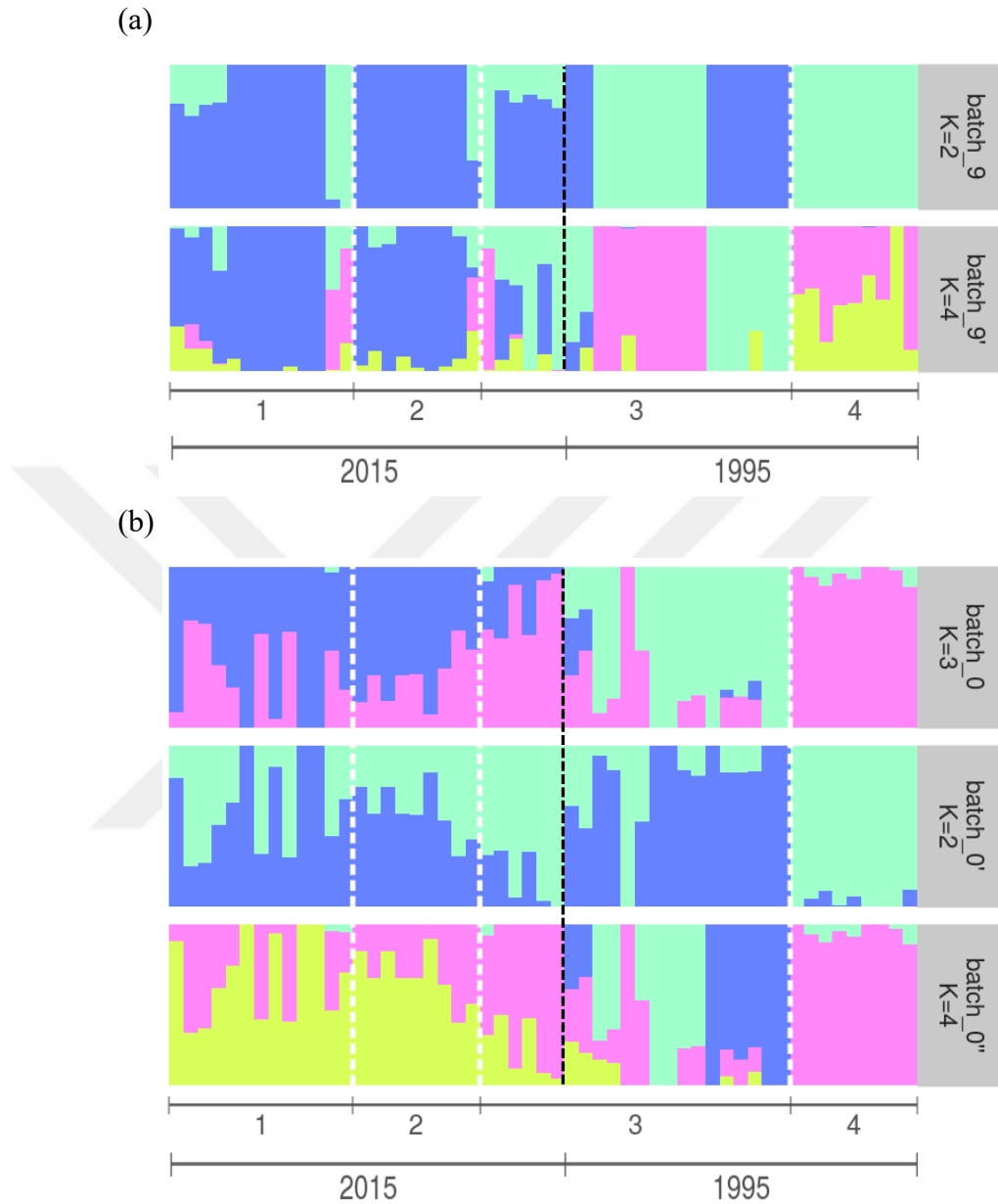


Figure 4.10. Batch-based *Structure* results with missing data encoded as -9 (a), and zero (b). The black dashed lines indicate the distinction between years.

4.1.5. De novo Pipeline Running of Emmer Wheat

Since the most meaningful results for einkorn were obtained with optimized parameters without the r80 method, the de novo pipeline was run under these conditions for emmer wheat, as well. As a result of de novo analysis of emmer populations, a total of 753,891 loci were assembled, of which 232,275 were polymorphic. The total number of SNPs obtained was 1,172,135.

Heterozygosity and nucleotide diversity values were calculated for the populations of WP90, WP92, WP93 (Kastamonu, 2015), WP95, WP96, WP97 (Kars, 2015), and WP15, WP20 (Kastamonu, 1995). According to the results of variant sites (Figure 4.11, also see Table A.4 in Appendix A), the maximum observed heterozygosity value was obtained for the population WP97 as 0.20071, while the minimum observed heterozygosity was found to be 0.07468 for WP20. The highest value of expected heterozygosity was recorded as 0.18141 for WP95, and the lowest value was recorded as 0.10359 for WP20. Among the nucleotide diversity values, the maximum diversity was obtained for WP95 (0.23475), whereas the minimum nucleotide diversity was 0.12864 for the population WP20.

For the calculations of all sites (Table A.5, Appendix A), the maximum observed heterozygosity value was 0.00262 for WP93, and the minimum observed heterozygosity value was 0.00105 for WP20. The expected heterozygosity was found to be maximum for WP95 as 0.00215, while the minimum value of that metric was 0.00145 for WP20. Similarly, the highest nucleotide diversity was calculated for the population WP95 as 0.00278, and the minimum nucleotide diversity was calculated as 0.00180 for WP20.

Considering the calculations mentioned above, the minimum values recorded for WP20 (from 1995) are in line with the expectation of two emmer groups (old vs. new). However, the calculations made for the other old population (WP15) produced closer values to the new populations rather than WP20. On the other hand, the values for the populations from Kastamonu and Kars were quite similar, indicating no genetic difference between these populations.

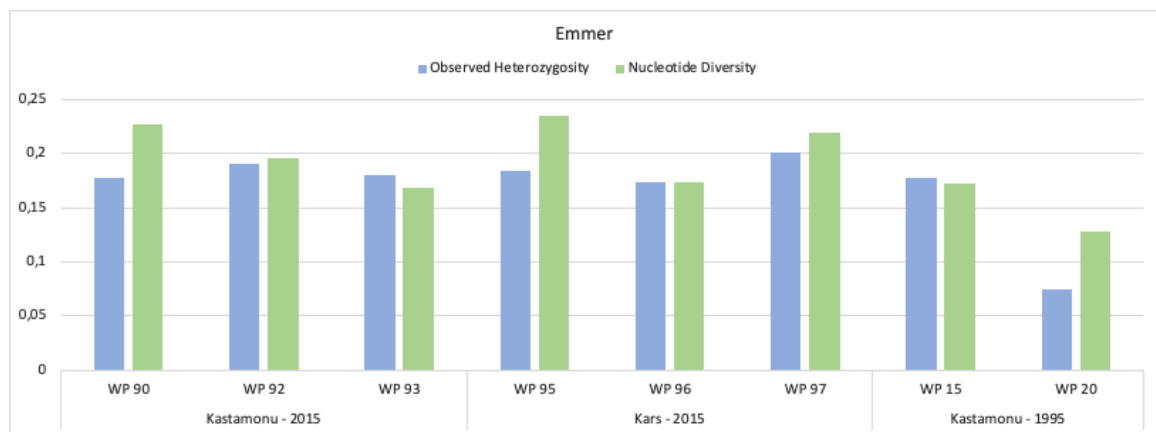


Figure 4.11. Observed heterozygosity and nucleotide diversity values of emmer populations calculated for variant positions.

As a result of de novo pipeline running, F_{ST} values giving insight of genetic distances between emmer populations were also calculated. As shown in Figure 4.12, the highest F_{ST} was obtained for the population pair of WP96–WP20 (0.592450). The F_{ST} values recorded for WP93–WP20 (0.583623) and WP92–WP20 (0.560807) followed that maximum value. The lowest F_{ST} was found to be between WP90 and WP15 (0.155917), followed by 0.164673 for WP90–WP93 and 0.166837 for WP95–WP96.

As can be seen from the gradient color scale in Figure 4.12, the highest genetic distance was found to be between WP20 (from 1995) and the other populations, including WP15 from 1995. Except for the high genetic distance with WP15, this is consistent with the expectation of two groups within emmer populations as old and new. Interestingly, the genetic distance between WP15 and the populations from 2015 collected in Kastamonu was one of the lowest, indicating the genetic similarity between these populations. The heterozygosity and nucleotide diversity values are also in concordance with this finding.

The genetic distance between emmer populations collected from different cities (Kastamonu and Kars) in the same year (2015) was higher than the distance within the populations from the same location. This indicates the effect of the location on genetic diversity, but it is lower than the divergence caused by the year. In the case of WP15, the F_{ST} values were closer to the populations from Kastamonu, supporting the information of its location of collection.

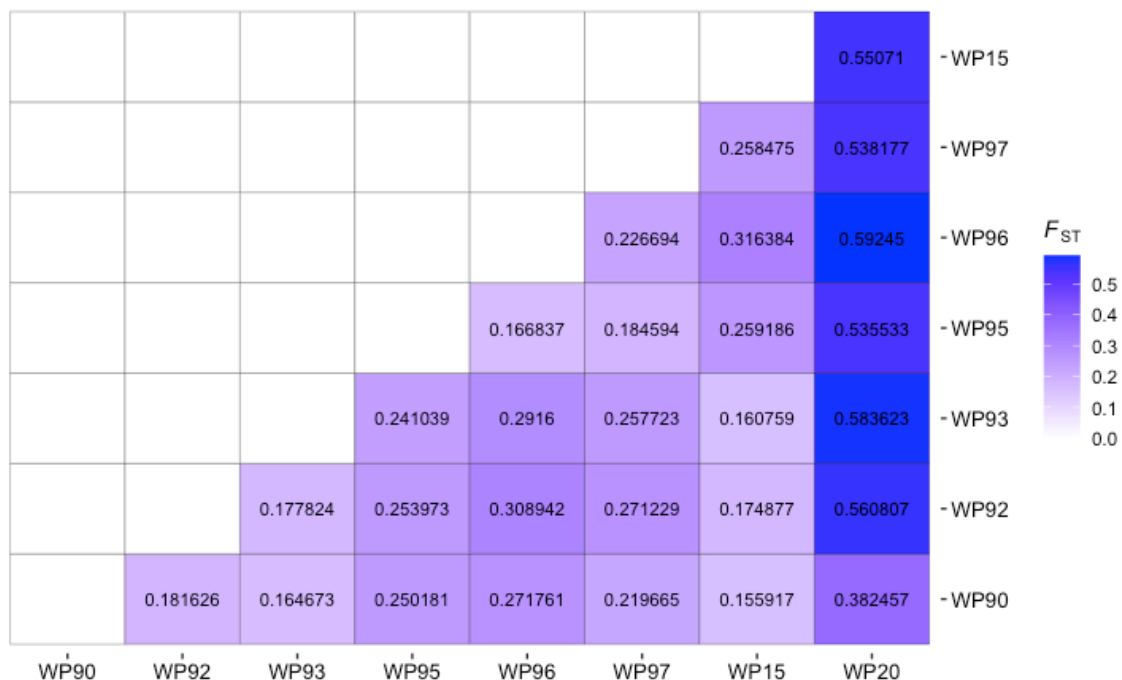


Figure 4.12. Matrix of pairwise F_{ST} values of emmer populations.

4.1.6. Structure Analysis of Emmer Wheat

After de novo pipeline running of emmer populations, clustering analysis was performed with *Structure* software (Pritchard et al., 2000) using the data obtained with *write_single_snp* flag of *populations* program. Since the most meaningful *Structure* results were obtained with missing data encoded and defined as -9, the analyses were conducted in this way. By defining populations based on the year and location of collection, two *Structure* runs were performed. In Figure 4.13, the result of the run with populations defined by the collection year is shown for the best K value determined (K=2). Accordingly, a relatively clear distinction between the first seven populations and WP20 was observed. As WP20 is one of the seed bank populations collected in 1995, this is in concordance with the expectation. However, the other population from 1995, WP15, was clustered with the in-situ conservation populations from 2015, which is consistent with the findings of F_{ST} analysis.

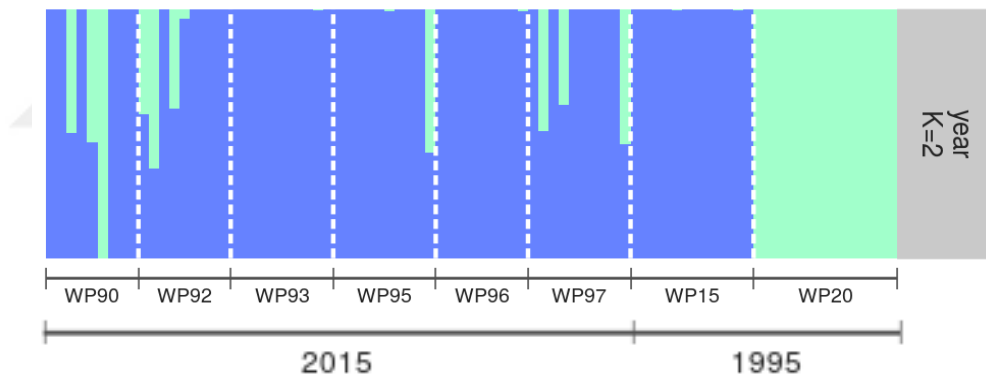


Figure 4.13. *Structure* result of emmer populations defined by the year of collection.

In Figure 4.14, the clustering of emmer populations defined by the location is shown for the best K value (K=2). Accordingly, there were two clusters: one with WP20 and one with the others. This is quite similar to the result obtained for the populations defined by the collection year (Figure 4.13), indicating that defining the libraries by the location where they have been collected did not have any effect on the analysis. The reason why location did not affect the clustering might be that it is ineffective in grouping the populations. However, as mentioned above (section 4.1.5), a moderate genetic distance between emmer populations from Kastamonu and Kars was observed. So, the absence of a grouping based on the location might be because of the relatively close F_{ST} values obtained for populations from different locations and the ones from the same location, and that *Structure* cannot detect it as a difference. Another possibility that prevents clustering by location is the insensitivity of the population tag parameter of *Structure* program. Since only one SNP per locus was used for *Structure* analyses, F_{ST} analyses were

also conducted with the *write_single_snp* flag, in order to see if it affected the results. The matrices are shown in Figures A.6, A.7, A.8, and A.9 (Appendix A). Accordingly, even though the F_{ST} values were slightly different, the same color scale was obtained, including emmer, showing that it could not be the reason for the unexpected clustering pattern obtained for emmer populations from Kastamonu and Kars.

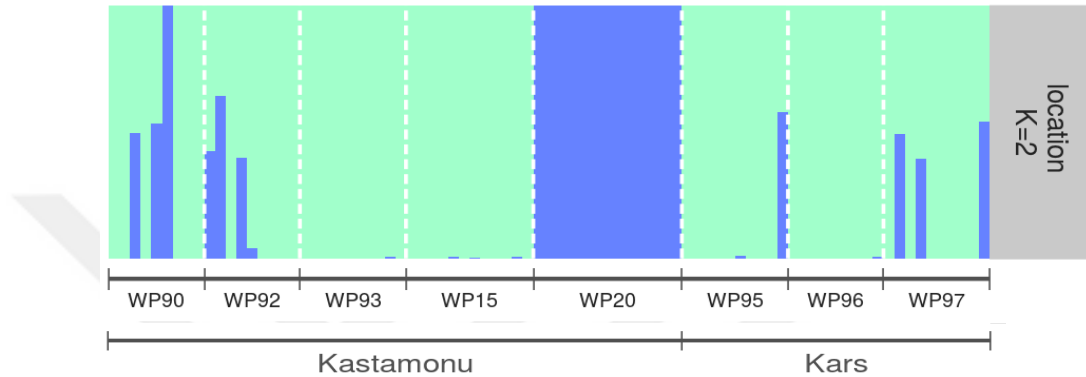


Figure 4.14. *Structure* results of emmer populations defined by the location of the collection. Populations are both sorted by the year of collection (a) and the location of collection (b).

4.1.7. De novo Pipeline Running of Bread Wheat

Like emmer, de novo pipeline was run for bread wheat populations using the optimized parameters and without the r80 method. Consequently, 533,475 loci were assembled in total. Out of 533,475 assembled loci, 116,347 were polymorphic, and 608,407 SNPs were obtained.

The values of observed heterozygosity and nucleotide heterozygosity obtained for variant sites of three bread wheat populations are shown in Figure 4.15 (see also Table A.6, Appendix A). Even though the values were close to each other, the maximum observed heterozygosity was 0.32658 recorded for WP99, and the minimum was 0.28682 for WP98. In the case of expected heterozygosity, the maximum value was obtained for WP99 (0.26664), and the minimum value was obtained for WP98 (0.20004). Similarly, nucleotide diversity was found to be the highest for WP99 (0.35762) and lowest for WP98 (0.28221).

As given in Table A.7 (Appendix A), the maximum observed heterozygosity value for variant and fixed positions was calculated as 0.00237 for WP103, and 0.00216 for WP98 as the minimum. The maximum expected heterozygosity was found to be 0.00188 for WP99, while the minimum value was

0.00150 for WP98. Likewise, for nucleotide diversity, the highest value was recorded for WP99 (0.002529), and the lowest value was recorded for WP98 (0.00212). However, like the values obtained for variant sites, the calculations made for three bread wheat populations for all sites were quite similar, which is normal considering the populations collected from the same city in the same year.

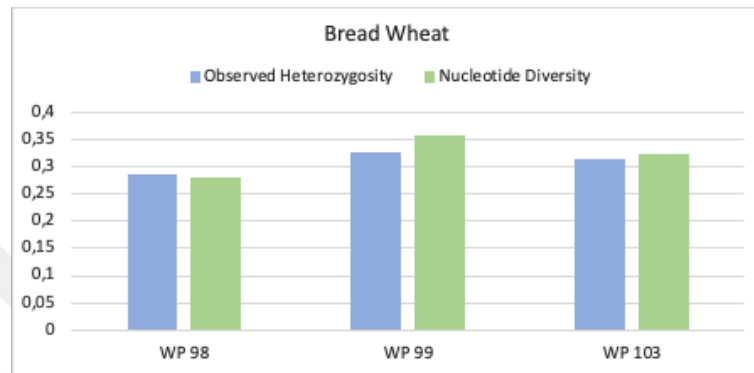


Figure 4.15. Observed heterozygosity and nucleotide diversity values of bread wheat populations calculated for variant positions.

The F_{ST} values were also calculated between the pairs of three bread wheat populations, as shown in Figure 4.16. The maximum F_{ST} was recorded as 0.320765 for the pair of WP98–WP103, and the minimum F_{ST} was 0.255655 recorded between WP99 and WP103. Like heterozygosity and nucleotide diversity values, the F_{ST} values of bread wheat populations were close to each other, indicating the absence of a prominent genetic distance between the populations, as expected.

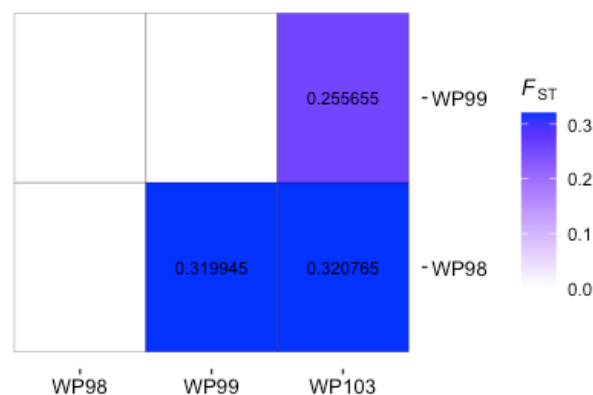


Figure 4.16. Matrix of pairwise F_{ST} values of bread wheat populations.

In opposition to einkorn and emmer populations, for bread wheat, *Structure* analysis was not performed after the de novo pipeline running, as there were only three populations collected in the same

year from the same city in Turkey. Rather than running *Structure* separately for these populations, they were included in the clustering analysis of all populations (see below, section 4.1.9).

4.1.8. De novo Pipeline Running of All Populations

After running the de novo pipeline for each species separately, another analysis was performed with all populations. Again, optimized parameters without the r80 method were used. For three species, 1,157,567 loci were assembled. Out of these, 361,082 were polymorphic. The number of SNPs was 1,891,799 in total.

For variant sites (Figure 4.17, Table A.8 (Appendix A)), the maximum observed heterozygosity was calculated for WP99 (0.19685), while the minimum observed heterozygosity was 0.06634, calculated for WP81. The highest expected heterozygosity value was also recorded for WP99 as 0.14725, and the lowest expected heterozygosity was 0.05956 for WP88. Likewise, the nucleotide diversity was found to be the highest for WP99 (0.20621) and lowest for WP88 (0.07867).

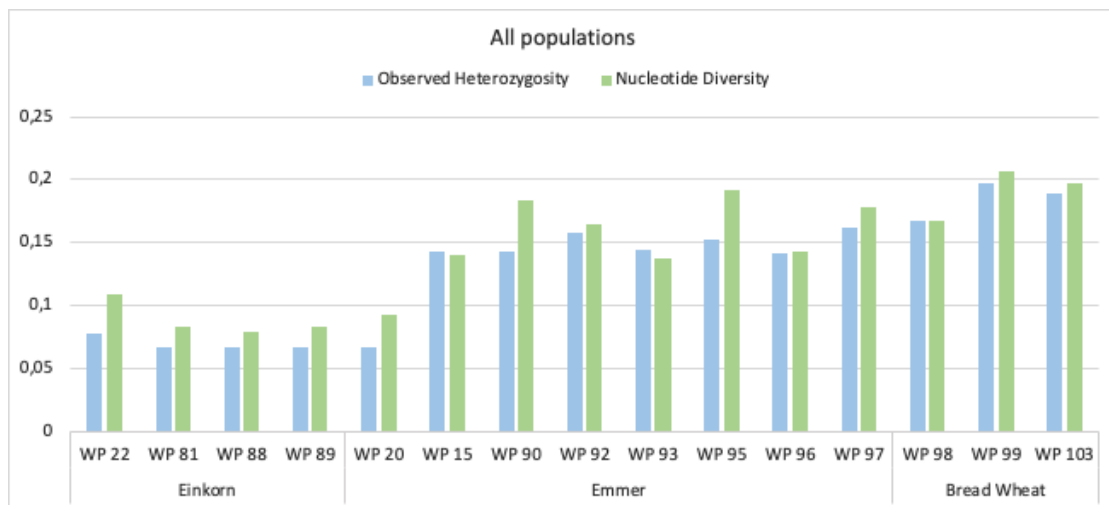


Figure 4.17. Observed heterozygosity and nucleotide diversity values of all wheat populations calculated for variant positions.

For both variant and fixed sites (Table A.9, Appendix A), 0.00269 was the maximum observed heterozygosity value calculated for WP99, and 0.00105 was the lowest (WP22). The expected heterozygosity was found to be the highest for WP95 (0.00215) and the lowest for WP89 (0.00104). For

nucleotide diversity, WP95 and WP99 were the populations for which the maximum value (0.00282) was recorded for, while WP81 and WP89 had the minimum value (0.00142).

For the pairs of einkorn, emmer, and bread wheat populations, F_{ST} values were calculated as shown in Figure 4.18. The maximum F_{ST} value was 0.611847 for the pair of WP96–WP89, followed by 0.609184 for WP93–WP89, and 0.607614 for WP96–WP81. The minimum F_{ST} value was recorded for the pair of WP88–WP81 (0.149724), and the values calculated for WP15–WP90 (0.159363) and WP89–WP88 (0.160335) followed that minimum value. As seen with the gradient color scale from high to low (from dark blue to light purple), the highest genetic distances were observed between einkorn and emmer populations, followed by einkorn and bread wheat. This indicates the high genetic divergence between einkorn and the other wheat species. In addition to einkorn, WP20, which is one of the old emmer populations, was also genetically far from emmer and bread wheat populations, as mentioned above in section 4.1.5. On the other hand, the genetic distance decreased for the populations belonging to the same species, as expected.

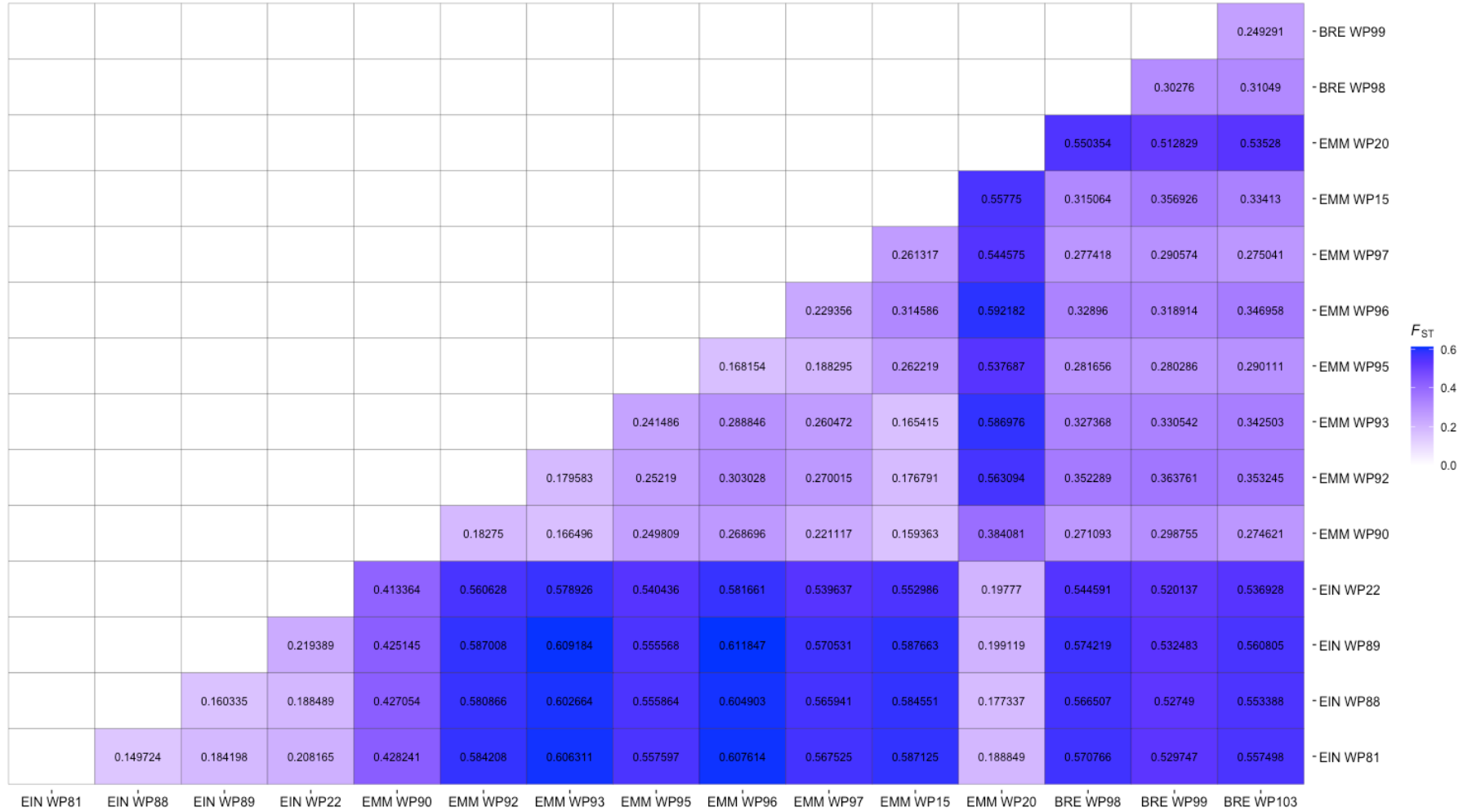


Figure 4.18. Matrix of pairwise F_{ST} values of all wheat populations. In population names, EIN indicates einkorn, EMM indicates emmer and BRE indicates bread wheat.

For all wheat populations, principal coordinate analysis (PCoA) was also performed using the SNP data. As seen in Figure 4.19, the clustering of the populations was mostly in line with the genetic distances calculated by F_{ST} analysis. Accordingly, there were three main groups encompassing einkorn, emmer, and bread wheat populations, respectively. The only exception was WP20, which is an emmer population but grouped with einkorn populations. Within the einkorn cluster, the separation between old (WP22) and new (WP81, WP88, WP89) populations was not so clear. WP22 seems to be located far from WP81 and WP88, but it was close to WP89. Likewise, emmer populations grouped together regardless of old or new, except WP20. WP15, which is an emmer population from 1995, Kastamonu, was closer to the emmer populations from 2015, Kastamonu (WP90, WP92, WP93). The emmer populations from Kars (WP95, WP96, WP97) were also clustered together and they were close to the emmer populations from Kastamonu. Lastly, bread wheat populations (WP98, WP99, WP103) were grouped together.

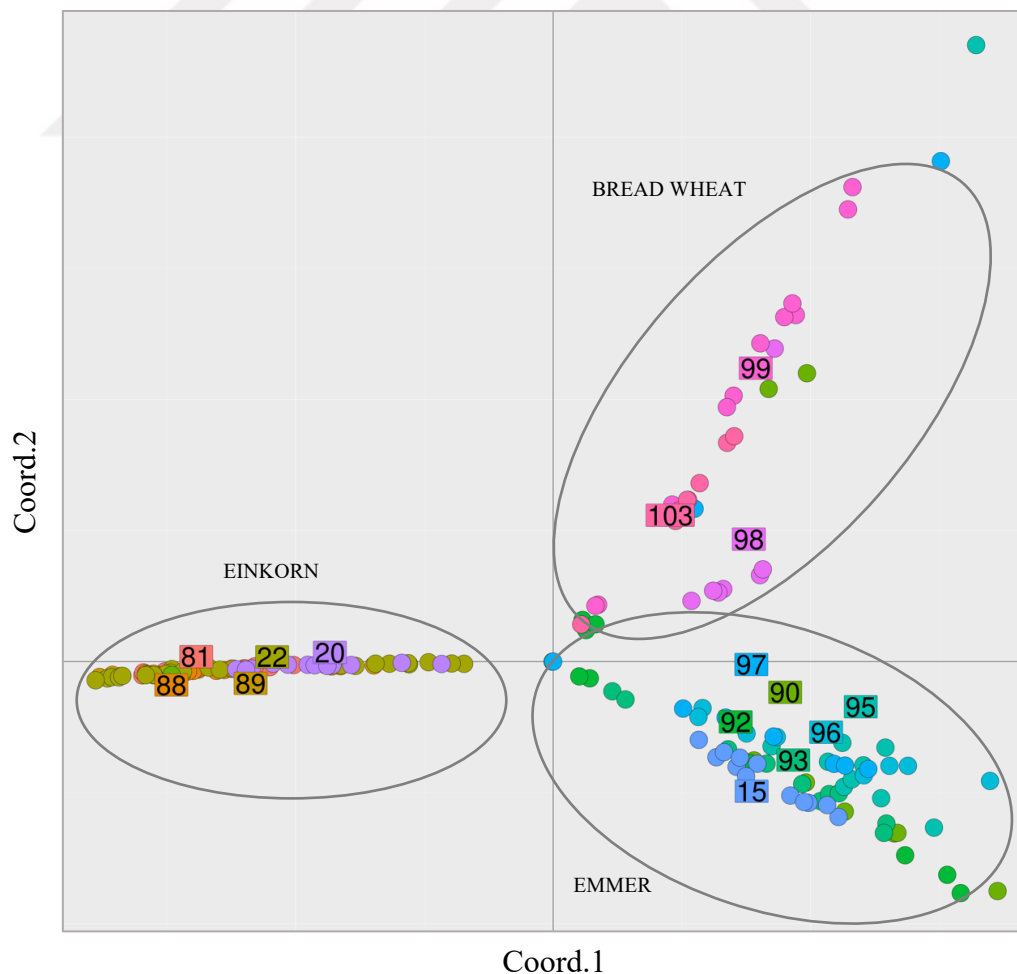


Figure 4.19. PCoA result of all wheat populations. The numbers in squares indicate the names of wheat populations without “WP”.

4.1.9. Structure Results of All Populations

The clustering analysis was performed for three wheat species (einkorn, emmer, bread wheat) using the *Structure* datum obtained with *write_single_snp* flag of *populations* program with missing data encoded as -9. The bar plot drawn for the *Structure* analysis result is shown in Figure 4.20. Accordingly, the separation between three species was clear, except for one emmer population, WP20. WP20 was clustered together with einkorn populations, which is in concordance with the PCoA and F_{ST} analysis results. On the contrary, there was no separation within einkorn and emmer populations indicating genetic differences caused by year or location.

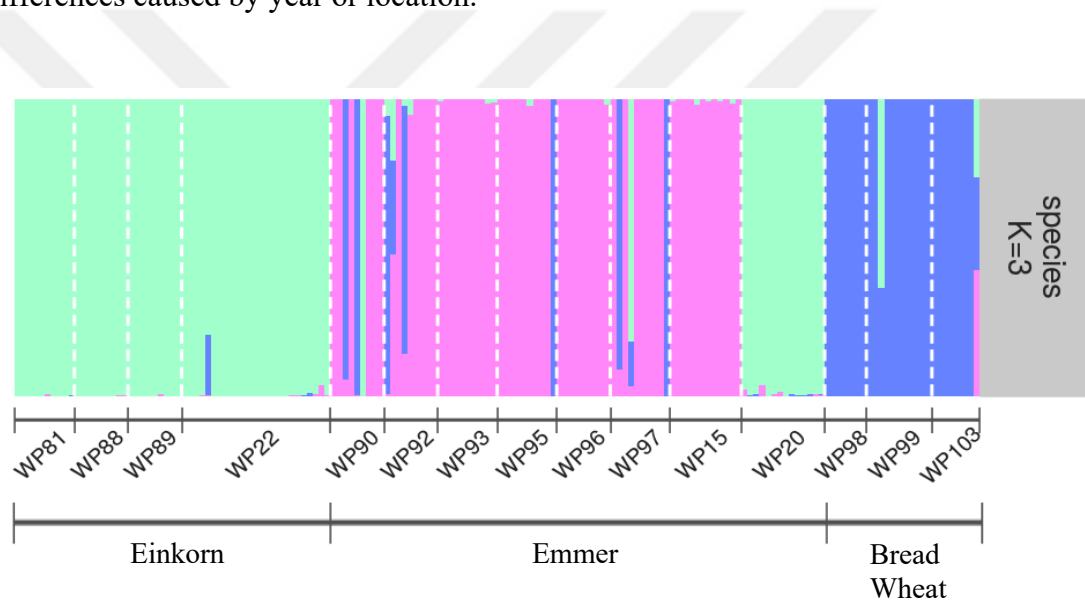


Figure 4.20. *Structure* results of all populations defined by species.

4.2. RNA-seq

4.2.1. RNA Quality

The RNA integrity numbers (RINs) and rRNA ratios measured by the companies SEM and MacroGen for 36 RNA samples are shown in Figure 4.21 and Figure 4.22 (see also Table B.1, Appendix B). The average RIN values were 7.65 and 8.01 for the Bioanalyzer analyses conducted by SEM and MacroGen, respectively. All RIN values were equal to or higher than 7.0 for both analyses. Likewise, all rRNA ratios were equal to or higher than one, and the average rRNA ratios were calculated as 1.41 and 1.61 for the analyses conducted by SEM and MacroGen, respectively.

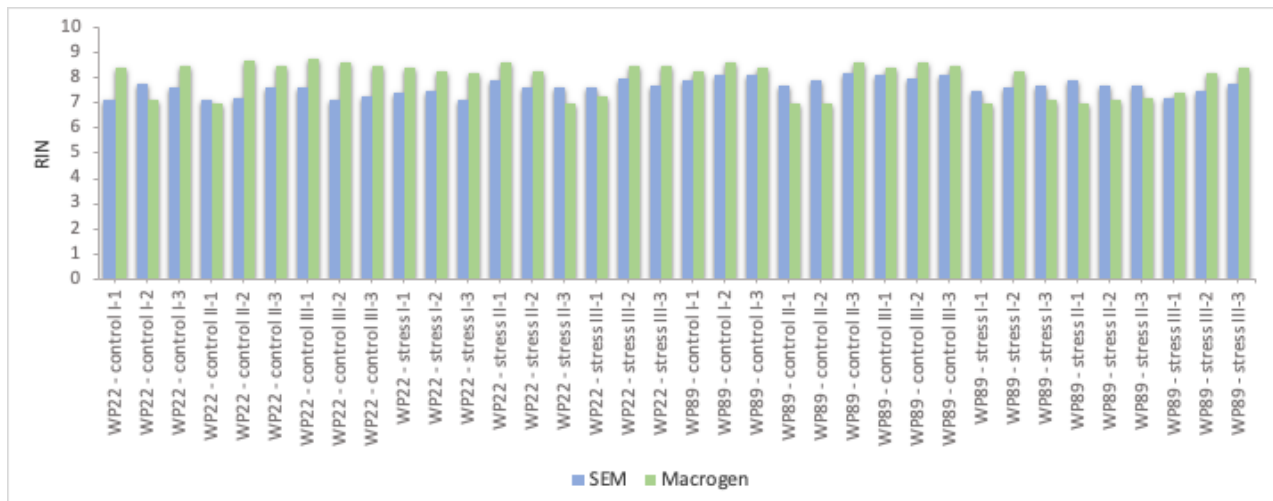


Figure 4.21. RIN values recorded by SEM and MacroGen for 36 RNA samples.

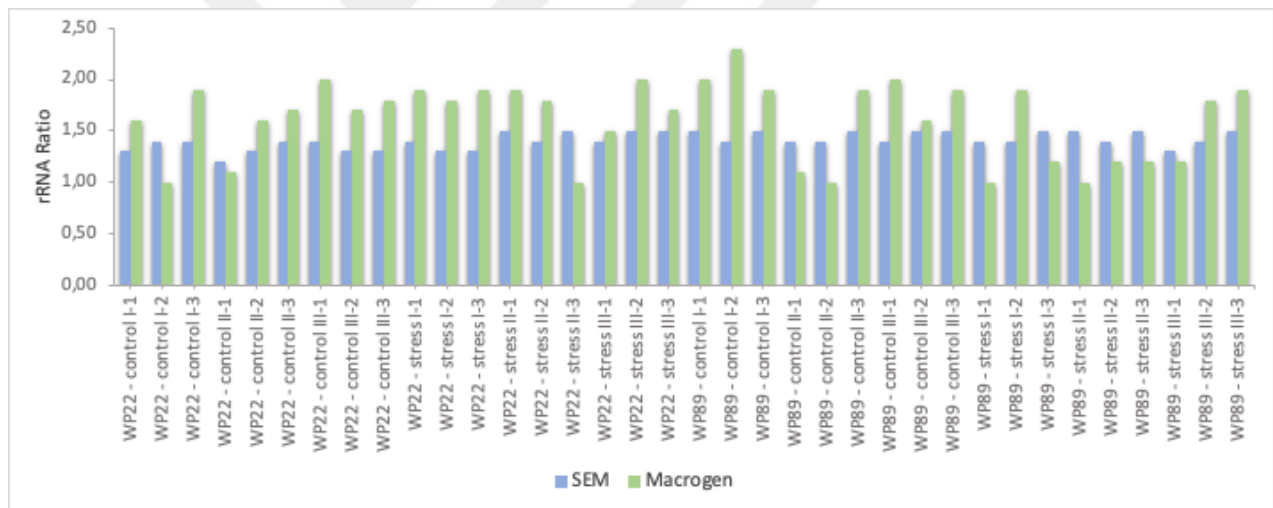


Figure 4.22. rRNA ratios calculated by SEM and MacroGen for 36 RNA samples.

4.2.2. Raw Data Statistics

The number of reads and bases obtained for each RNA-seq library sequenced in the Illumina HiSeqX platform by MacroGen, along with the GC and AT percentages, and the quality scores are given in Table B.2 (Appendix B). For 36 RNA-seq libraries, a total of ~2B 150 bp paired-end reads with ~274B bases were generated. Considering the number of reads produced for each library, an average of ~50M reads were obtained. The average GC content was 53.09 %. The average quality scores were 97.52 % and 94.05 % for Q20 and Q30, respectively. Each library had a Q20 score above 95 %, indicating high quality.

The quality control of the raw reads was also performed with FastQC v0.11.9 (Andrews, 2010). All but four sequence features passed the quality criteria. However, per base sequence content, per base GC content, sequence duplication levels, and adapter content failed, which are common for RNA-seq data. Even though there should be equal amounts of four bases in a non-biased library, RNA sequencing produces a bias in the first nine bases because of the sequencing primers (Külahoğlu & Bräutigam, 2014). So, it is normal to get an error about per base sequence content in RNA-seq studies. RNA sequencing may also produce shifts from the normal distribution of the GC content, which is considered as an error by FastQC even if the quality of the reads is high (Dudic, Banović Đeri, Pajić, & Pavlovic-Lazetic, 2021; Michigan State University, n.d.). As expression levels of sequences can change in RNA-seq studies, errors in sequence duplication levels are common in RNA-seq data, as well (Külahoğlu & Bräutigam, 2014). On the other hand, since there was a warning about the adapter content, Illumina universal adapter sequences were removed from each datum using Trimmomatic v0.36 (Bolger et al., 2014). After trimming, FastQC was run again in order to see if the adapter sequences were removed successfully.

4.2.3. Transcriptome Assembly Statistics

A de novo transcriptome assembly from the data of all 36 RNA-seq libraries was generated with Trinity software v2.14.0 (Grabherr et al., 2013). The statistics of this assembly are given in Table 4.5. Accordingly, the assembly contained 225712 genes and 430776 transcripts with an average contig length of 1020.71 bp. The N50 value is one of the measurements used for assembly evaluation (Li et al., 2014). It refers to the length of the longest contig, in which at least 50 % of the bases are assembled into contigs of this length (Miller, Koren, & Sutton, 2010). A high N50 value means an assembly with high quality (Li et al., 2014). However, for transcriptome assemblies, it is an indicator of the continuity of contigs rather than the accuracy of the assembly analysis since the assembler programs generating many isoforms maximize this value (Salzberg et al., 2012). As given in Table 4.5, for the transcriptome assembly constructed in this study, at least 50 % of all assembled bases were found in contigs, which had a length of 1802 bases. Similarly, at least 90 % of all assembled bases were found in contigs, which had a length of 396 bases. The N values for contigs were also computed using only the longest isoform for each gene to overcome the problem of maximized values produced by the assembler program. According to this calculation, the contig lengths decreased to 1194 and 293 bases for N50 and N90, respectively.

Table 4.5. Statistics of the de novo assembly generated by Trinity.

Trinity Statistics	
Total trinity 'genes': 225712 Total trinity transcripts: 430776 % GC: 48.32	
Stats based on ALL transcript contigs: Contig N10: 4473 Contig N20: 3410 Contig N30: 2746 Contig N40: 2229 Contig N50: 1802 Contig N60: 1397 Contig N70: 1014 Contig N80: 667 Contig N90: 396 Median contig length: 564 Average contig: 1020.71 Total assembled bases: 439696911	Stats based on ONLY LONGEST ISOFORM per 'GENE': Contig N10: 4227 Contig N20: 3102 Contig N30: 2342 Contig N40: 1719 Contig N50: 1194 Contig N60: 822 Contig N70: 571 Contig N80: 405 Contig N90: 293 Median contig length: 395 Average contig: 735.06 Total assembled bases: 165912979

The completeness of the assembly was checked with BUSCO analysis through OmicsBox v3.0.30 (2019). As shown in Figure 4.23, 93.18 % of BUSCO groups were found to have complete gene representation with 775 single-copy (15.83 %) and 3787 duplicated (77.35 %) orthologs. On the other hand, the numbers of fragmented and missing orthologs were 75 (1.53 %) and 259 (5.29 %), respectively.

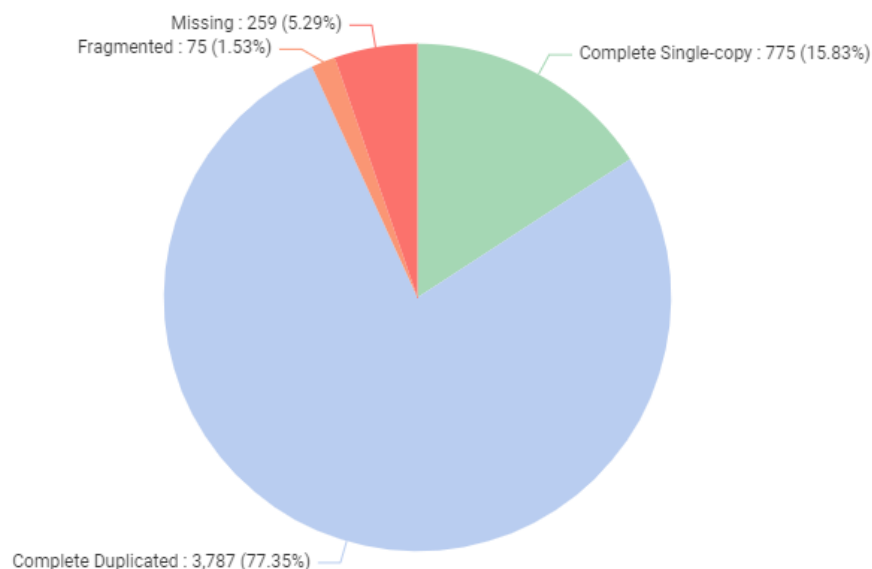


Figure 4.23. BUSCO analysis results.

The contig length distribution of the assembly was also investigated. The shortest contig had a length of 177 bp, while the longest contig had a length of 25835 bp. As shown in Figure 4.24, most of the contigs were in the range of 250-300 bp. There were 24 short contigs below 200 bp in length, which were removed from the assembly for the subsequent analyses.

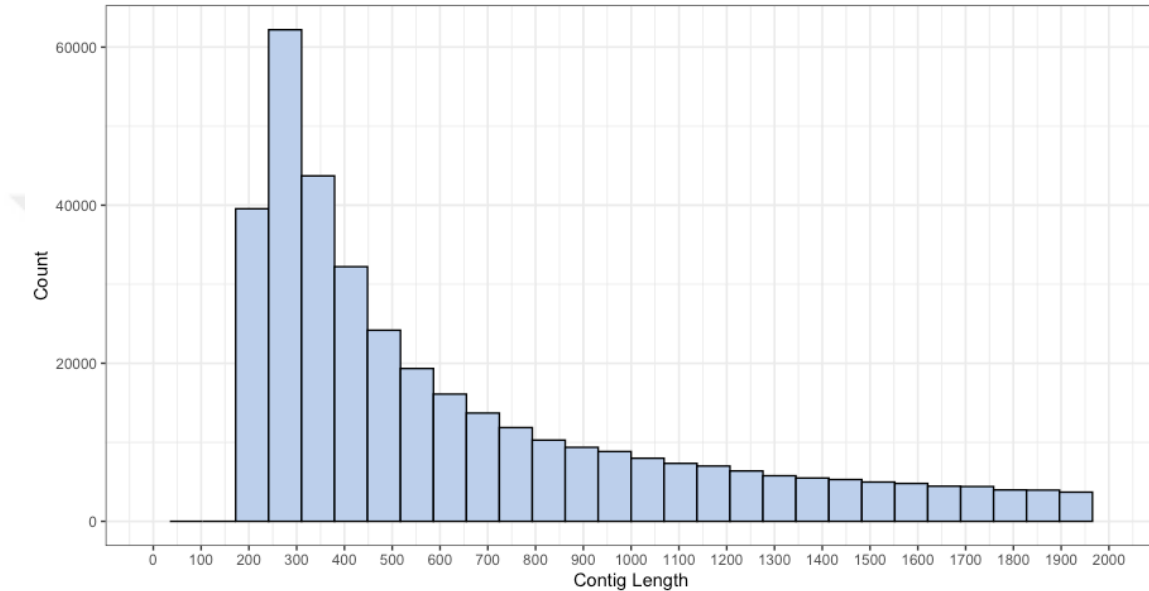


Figure 4.24. The contig length distribution of the de novo assembly.

After filtering short contigs, the remaining transcripts were clustered with CD-HIT-EST (Li & Godzik, 2006) to reduce redundancy. As a result, 281656 contigs were obtained out of 430752.

4.2.4. Alignment Statistics

After the alignment of each library to the de novo transcriptome assembly, the statistical analyses of the alignments were performed using Samtools v1.16.1 (Li et al., 2009). The mapping rate and the percentage of the reads properly paired for each library are given in Table 4.6. The average mapping rate was 99.31 %, and the average of the reads properly paired was 89.40 %. Table 4.6 also shows the coverage statistics. Accordingly, the mean read depth, which was calculated as the sum of the depths across all positions divided by the total number of positions, was found to be 31.74 % on average. The average breadth of coverage was 65.51 %, and it was calculated for each library as the total number of covered bases divided by the total number of positions. More detailed information about the statistical analyses can be found in Table B.3, in Appendix B.

Table 4.6. Statistics of the aligned reads.

Library	Mapping Rate (%)	Properly Paired (%)	Mean Read Depth	Breadth of Coverage
WP22 – Control I-1	99.48	87.24	33.42	65.24
WP22 – Control I-2	99.17	91.07	31.82	58.03
WP22 – Control I-3	99.30	86.44	29.36	63.28
WP22 – Control II-1	98.99	91.61	31.90	59.35
WP22 – Control II-2	99.55	92.65	33.19	67.06
WP22 – Control II-3	99.61	90.28	33.49	67.21
WP22 – Control III-1	99.22	87.14	30.19	66.21
WP22 – Control III-2	99.44	87.00	36.73	68.73
WP22 – Control III-3	99.37	86.72	31.75	66.49
WP22 – Stress I-1	99.54	90.06	32.52	69.90
WP22 – Stress I-2	99.27	88.01	27.58	70.36
WP22 – Stress I-3	99.06	84.76	27.85	63.82
WP22 – Stress II-1	99.56	87.61	30.98	66.71
WP22 – Stress II-2	99.44	90.43	34.36	69.86
WP22 – Stress II-3	99.06	91.21	29.51	62.52
WP22 – Stress III-1	99.31	87.33	32.86	67.38
WP22 – Stress III-2	99.42	86.12	32.06	68.29
WP22 – Stress III-3	99.56	89.58	30.35	67.91
WP89 – Control I-1	99.36	88.62	34.74	66.82
WP89 – Control I-2	99.53	88.85	34.94	67.11
WP89 – Control I-3	99.43	91.16	32.03	65.81
WP89 – Control II-1	99.15	93.14	27.98	60.59
WP89 – Control II-2	99.54	94.76	32.49	60.04
WP89 – Control II-3	99.57	92.02	34.14	69.00
WP89 – Control III-1	99.59	93.74	34.38	67.02
WP89 – Control III-2	99.33	84.79	32.45	63.15
WP89 – Control III-3	99.30	87.84	33.98	67.28
WP89 – Stress I-1	99.33	93.61	31.48	64.11
WP89 – Stress I-2	99.64	92.63	36.97	70.02
WP89 – Stress I-3	98.96	93.25	28.42	62.69
WP89 – Stress II-1	99.15	94.01	29.99	58.13
WP89 – Stress II-2	98.78	89.73	26.87	62.47
WP89 – Stress II-3	98.84	88.75	25.85	61.99
WP89 – Stress III-1	99.34	87.83	31.71	69.27
WP89 – Stress III-2	99.14	87.35	32.95	68.69
WP89 – Stress III-3	98.85	80.92	31.21	65.94

4.2.5. Pre-processing of Aligned Data

The data, which were aligned to the de novo assembly, were evaluated by edgeR package (Robinson et al., 2010; McCarthy et al., 2012) in R v4.2.2 (R Core Team, 2022). The sizes of 36 raw RNA-seq libraries were calculated as shown in Figure 4.25 (grey bars, see also Table B.4 in Appendix B). The average and median library sizes were 22.9 million and 23.2 million, respectively, with a maximum value of 27.5 million and a minimum value of 18.9 million.

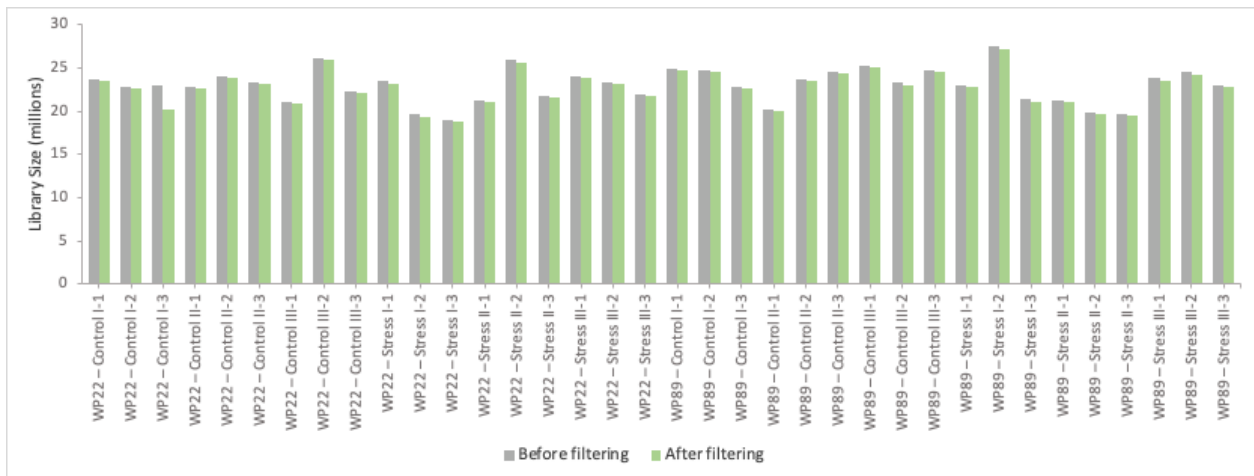


Figure 4.25. Sizes of 36 RNA-seq libraries before and after filtering.

In order to prepare aligned data for differential gene expression analysis, contigs unexpressed throughout all libraries, as well as the lowly expressed ones, were removed. The number of contigs with zero counts across 36 libraries was 182, constituting 0.06 % of all contigs (281656). The lowly expressed contigs were determined as the contigs with read counts of less than ten in a minimum number of samples. In this case, the minimum number of samples was nine, since each group had nine libraries. While filtering out lowly expressed contigs, count data transformed into CPM were used. In differential expression studies, this kind of data transformation is generally used, because raw read counts vary in the number of reads per sample or fragment length (Conesa et al., 2016). Therefore, contigs having a CPM value less than 0.43 ($10/23.2$, where 23.2 was the median library size in millions) were removed. As a result, 89578 contigs remained out of 281656.

The library sizes obtained after filtering is shown in Figure 4.25 (green bars, see also Table B.4 in Appendix B). The average and median library sizes decreased to 22.7 million and 22.9 million, respectively. The maximum library size was 27.2 million and the minimum library size was 18.8 million.

After filtering, normalization was performed to equalize the expression distributions of the libraries. The distributions before and after filtering and normalization are shown in Figure 4.26 and Figure 4.27. As shown in the graphs, log-CPM shifted towards higher values after filtering, and normalization caused a more similar pattern of expression distributions of the libraries. The only exception was the library of WP22 control I-3 (green peak in Figure 4.27), which still had a different distribution than the other libraries after normalization.

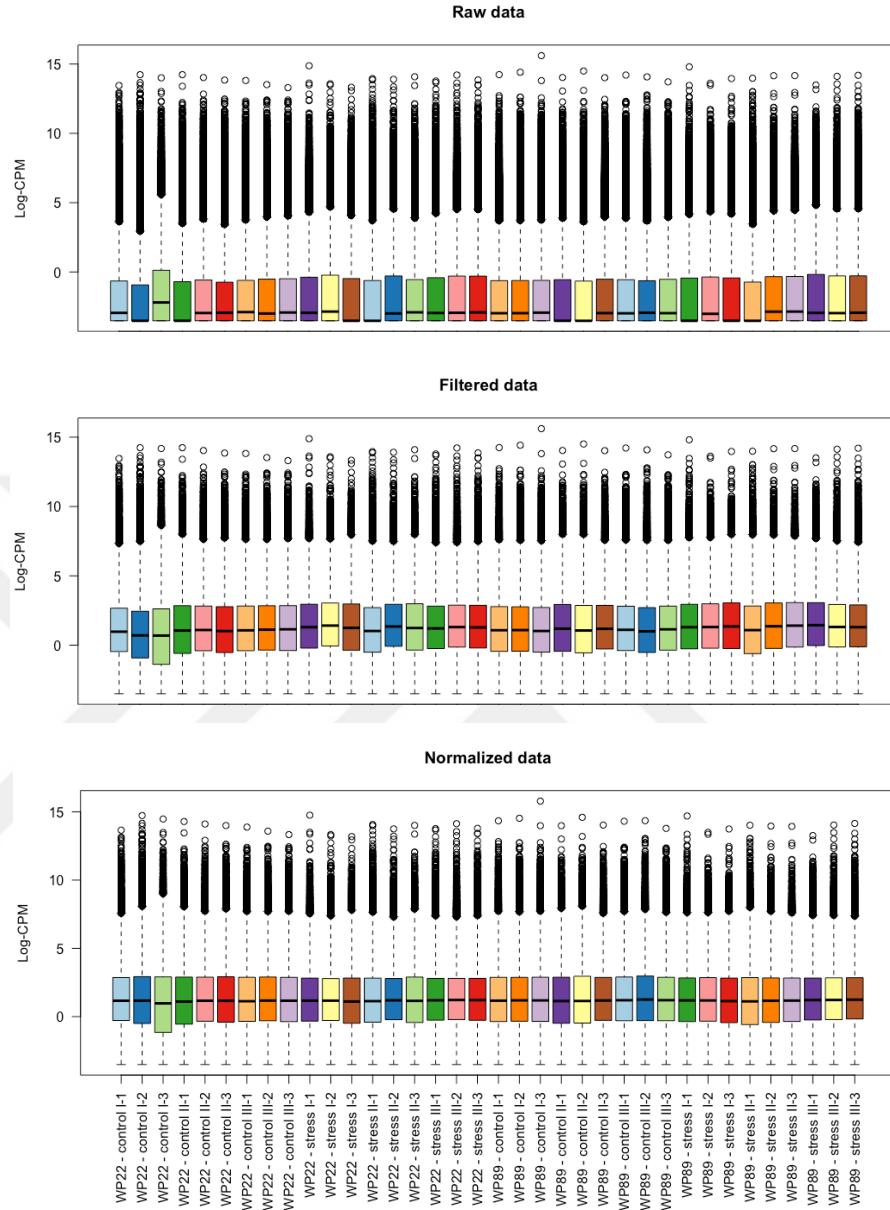


Figure 4.26. Expression distributions of the libraries before and after filtering and normalization.

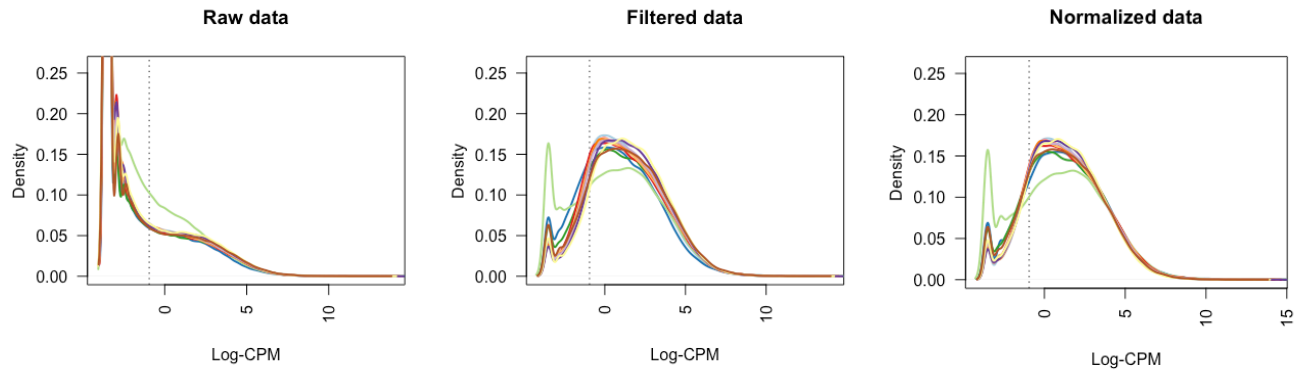


Figure 4.27. The density of log-CPM values of the libraries before and after filtering and normalization. The dotted line indicates the cutoff used for filtering.

Following filtering and normalization, the relations between four groups, old (WP22, 1995) and new (WP89, 2015) einkorn populations at control and stress temperatures, were investigated with MDS analysis. Since examining multiple dimensions rather than only the first two is recommended for studies with multiple factors (Law et al., 2016), the MDS plots for the first three dimensions were drawn (Figure 4.28). Accordingly, the control and stress libraries were clustered separately over the first dimension which indicates the success of the heat treatment. The grouping of old and new einkorn libraries was also partially separate, but the transcriptional difference between them was not as clear as between control and stress temperatures over all three dimensions. Additionally, within the old einkorn group, one library (WP22 control I-3) was located further from the other control libraries. The other two libraries from the same pot (I) were also clustered partially away from the control libraries for the second dimension.

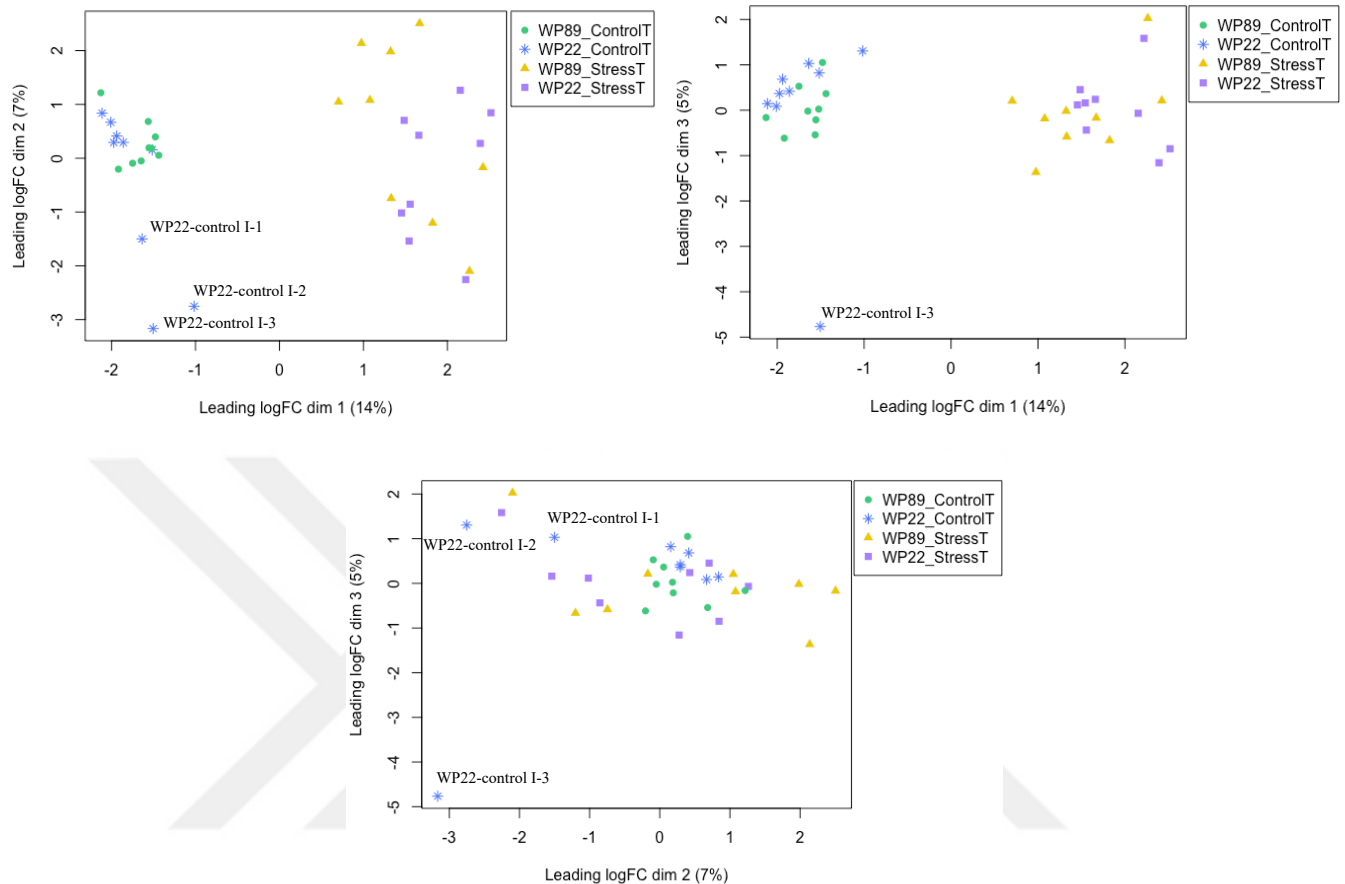


Figure 4.28. MDS plots of four einkorn groups for the first three dimensions.

MDS plots for the higher dimensions were also drawn to investigate if any of these dimensions caused a specific clustering of the libraries. As seen in the MDS plots in Figure B.1 (Appendix B), the libraries were not clustered differently, which makes the clustering with dimensions one and two the most plausible for four einkorn groups. Indeed, the variances explained by the first two dimensions (14 % and 7 % for dimensions one and two, respectively) were higher than the others, meaning that they explained the variation in the data best.

4.2.6. Differential Gene Expression Analysis

After pre-processing the data, differential gene expression analysis was performed using edgeR package (Robinson et al., 2010; McCarthy et al., 2012). As a result, 11025 contigs were found to be up-regulated and 11651 contigs were found to be down-regulated at an FDR cut-off of 0.05, making 22676 contigs differentially expressed (DE) in total. The numbers of significantly up-regulated and down-

regulated contigs for each comparison are shown in Figure 4.29. Out of 89578 contigs that remained after filtering, the total numbers of DE contigs between old (WP22) and new (WP89) einkorn were 1933 and 1148 at control and stress temperatures, respectively. On the other hand, greater numbers of contigs were found to be DE between control and stress temperatures (37499 for the old einkorn, 34983 for the new einkorn). These higher DE contig numbers counted for control vs. stress temperatures are also in line with the higher transcriptional differences obtained between control and stress temperatures with the MDS analysis (Figure 4.28).

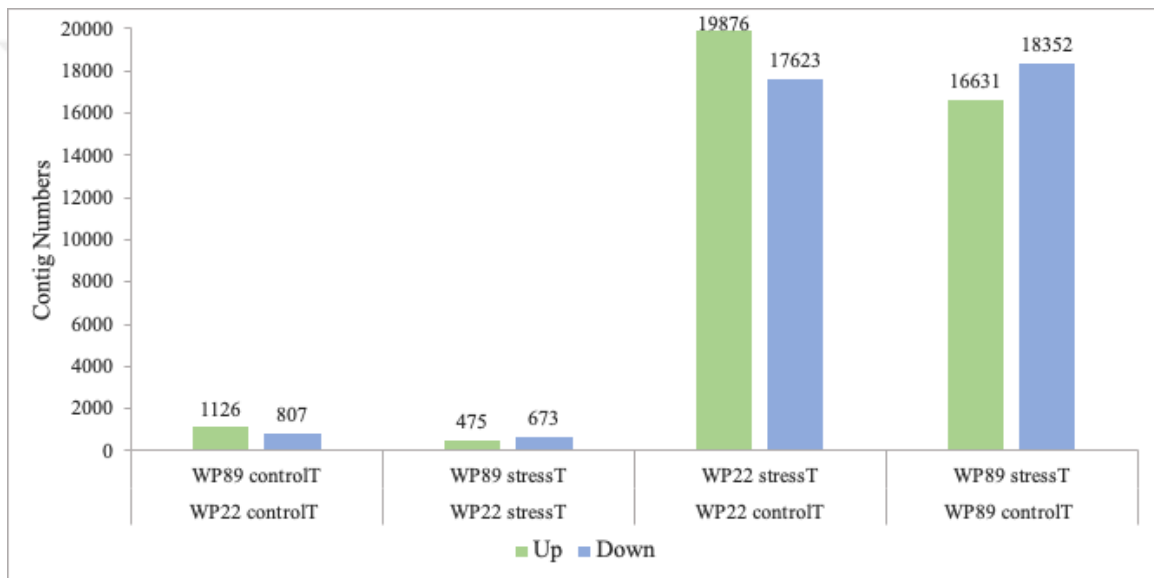


Figure 4.29. The numbers of up-regulated and down-regulated contigs with an FDR cut-off of 0.05.

The DE contigs determined for each comparison were visualized with an MD plot, as shown in Figure 4.30. The up-regulated and down-regulated contigs with an FDR cut-off of 0.05 are highlighted in the plots. Accordingly, the comparisons of old and new einkorn (upper plots) resulted in fewer DE contigs when compared to the comparisons of control and stress temperatures (bottom plots). In the case of old vs. new einkorn, there were more up-regulated contigs than down-regulated contigs at the control temperature, while at stress temperature, the number of down-regulated contigs was higher than that of up-regulated contigs. Similarly, between control and stress temperatures, old einkorn had more up-regulated contigs, while new einkorn had more down-regulated contigs.

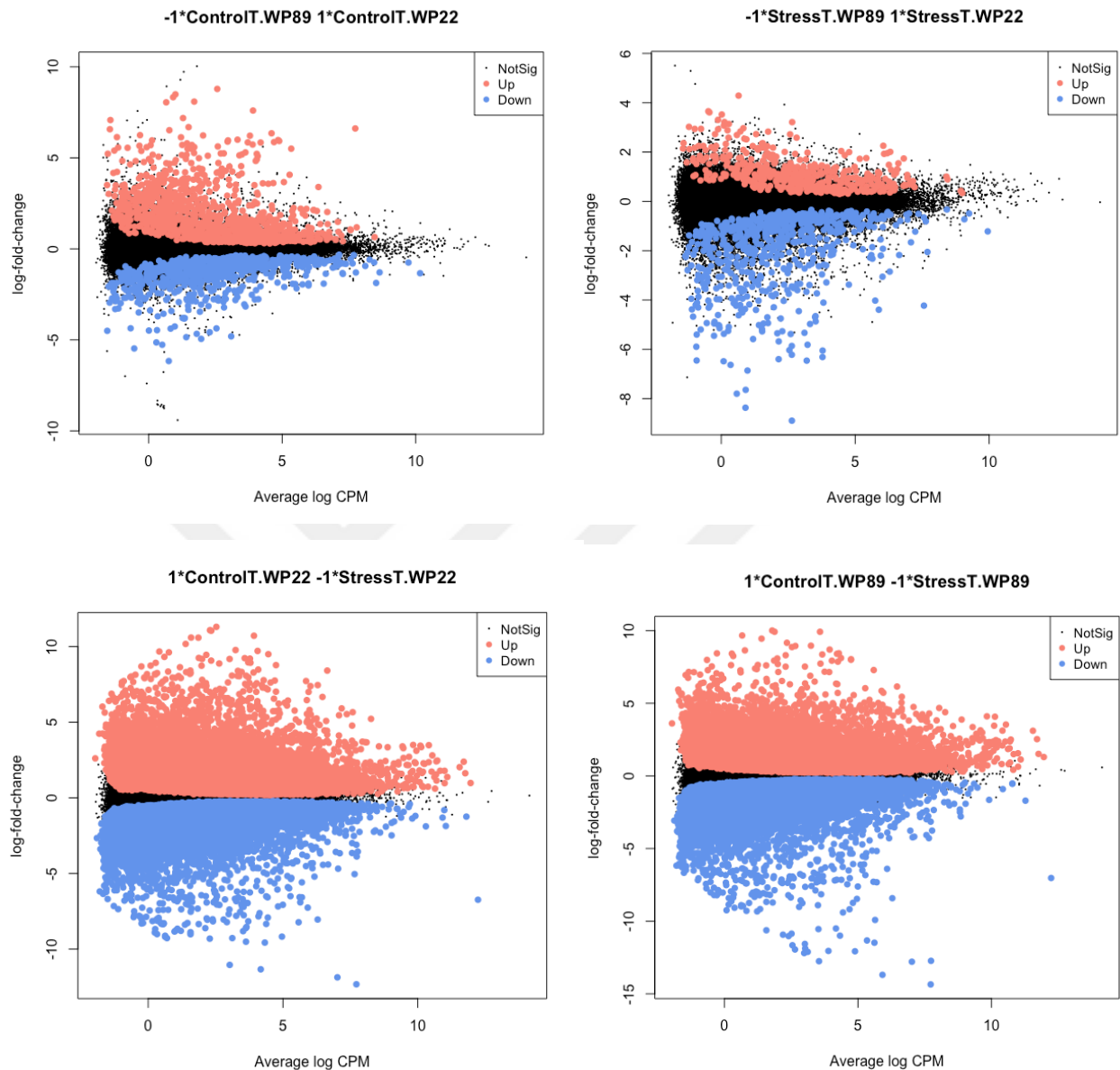


Figure 4.30. MD plots of four comparisons.

Next, the number of DE contigs was decreased using an additional metric, logFC, to get more biologically meaningful results. The numbers of DE contigs determined with logFC and FDR cut-off values of ± 1 and 0.05, respectively, are given in Figure 4.31. These DE contigs were used subsequently for the functional analyses of old vs. new einkorn at control and stress temperatures (see section 4.2.8).

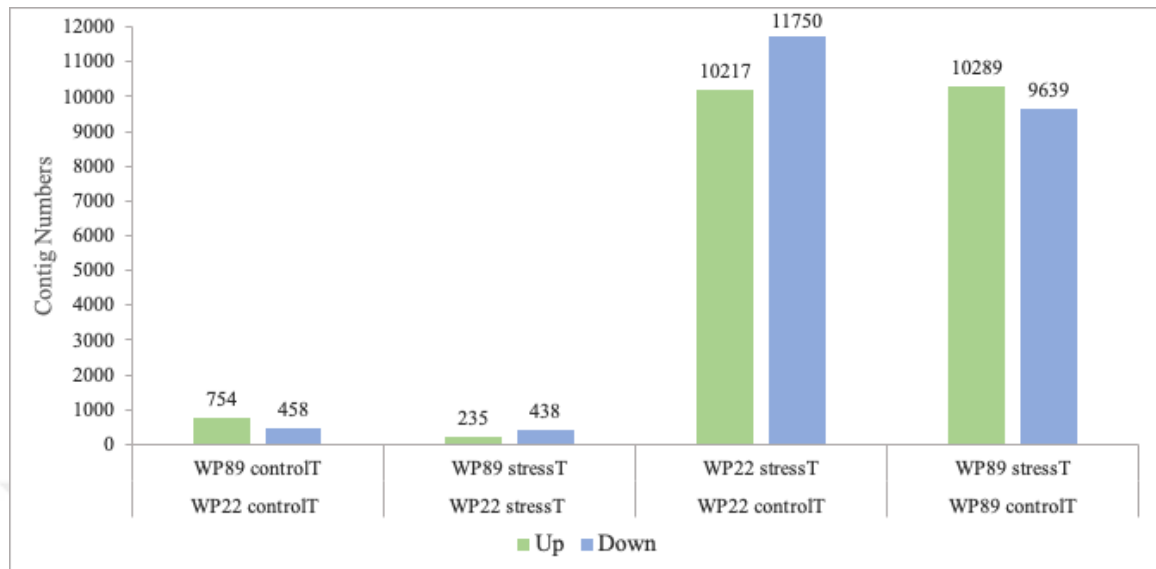


Figure 4.31. The numbers of up-regulated and down-regulated contigs with an FDR cut-off of 0.05 and logFC cut-off of ± 1 .

To demonstrate gene expression differences with logFC and FDR cut-off values, volcano plots were generated. Figure 4.32 shows the expression differences between old (WP22) and new (WP89) einkorn populations at the control or stress temperatures (upper plots), as well as the gene expression differences within the same populations (WP22 or WP89) but at different temperatures (control vs. stress) (bottom plots). The up-regulated (log fold change > 1) and down-regulated (log fold change < -1) contigs with an FDR cut-off of 0.05 are highlighted in the plots. Considering all plots, the comparisons of old vs. new einkorn gave lower numbers of up-regulated and down-regulated contigs than the control vs. stress comparisons.

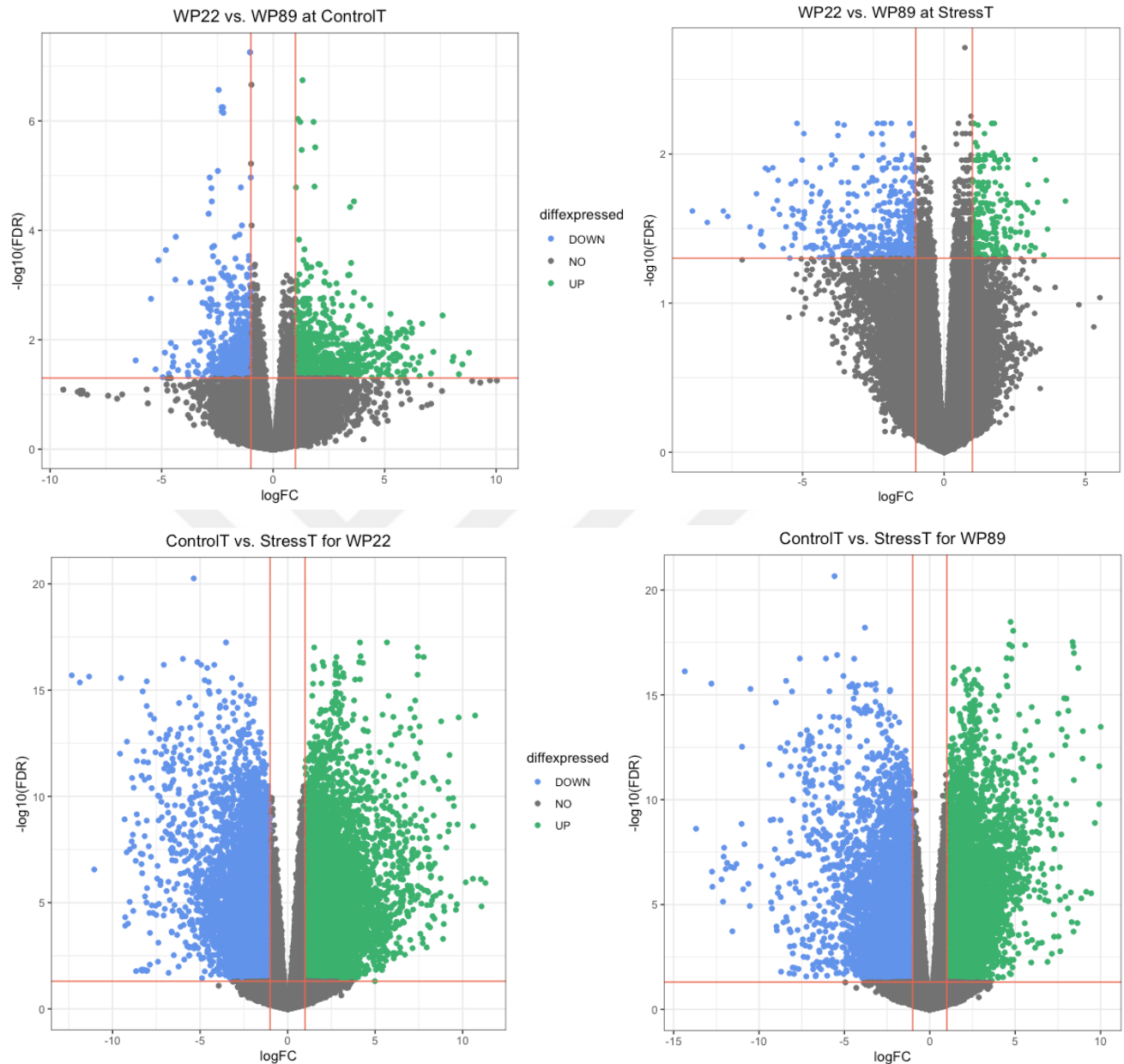


Figure 4.32. Volcano plots of four comparisons. The horizontal orange line indicates the $-\log_{10}(0.05)$, and the vertical orange lines indicate the log fold change $=\pm 1$.

4.2.7. Functional Annotation

For transcriptomic characterization, 281656 contigs of the de novo assembly remained after redundancy removal were blasted against *T. aestivum* and *T. urartu* reference genomes, and NCBI RefSeq Plant, Swiss-Prot, and NR databases. The numbers of blast hits obtained with each blastx run are shown in Figure 4.33. Accordingly, a total of 162313 contigs had blast hits after aligning against two

wheat genomes, and the blast runs against three NCBI databases resulted in a total of 17773 hits. In total, out of 281656 contigs, 180086 had blast hits and 101570 had no blast hits.

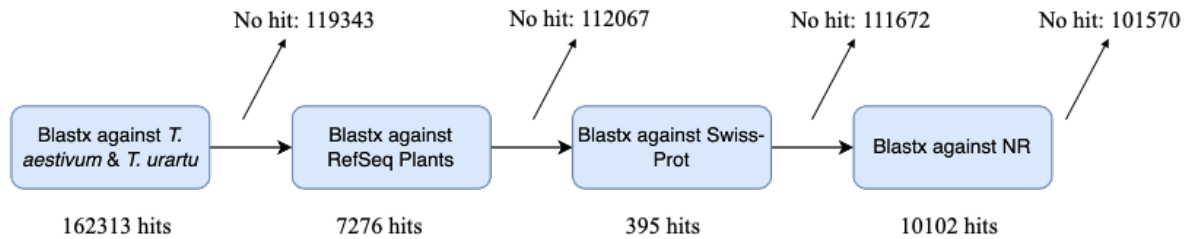


Figure 4.33. The number of hits obtained after each blast run.

After blast analyses, gene ontology (GO) mapping and annotation were performed with Blast2GO suite (Götz et al., 2008) through OmicsBox (2019). The numbers of mapped and annotated contigs as well as the numbers of contigs with blast hits that remained after importing to OmicsBox are given in Table 4.7. Since only 296 contigs were mapped out of all contigs with blast hits of two wheat reference genomes, 162313 sequences were blasted again, but this time against NCBI NR database with a taxonomy filter using Diamond Blast v2.1.6 (Buchfink et al., 2021). As a result, 138603 contigs had hits, and out of these, 89442 were mapped and 80756 were annotated. Since these numbers were much higher than the numbers recorded for *T. aestivum* and *T. urartu* reference genomes, Diamond blast results were used for subsequent steps. Uniquely mapped and annotated sequences, as well as blast results obtained with *T. aestivum* and *T. urartu* reference genomes, were included in the Diamond blast results. Thus, out of 162312 sequences with blast hits, 89510 were mapped and 80829 were annotated. Considering all blastx runs, out of 180085 contigs with blast hits, 94008 were mapped and 84645 were annotated.

Table 4.7. The numbers of blasted, mapped, and annotated contigs.

Databases used for blastx runs	Blast hits	Mapped	Annotated
<i>T. aestivum</i> & <i>T. urartu</i>	162312	296	285
NR with taxonomy filter	138603	89442	80756
TOT	162312	89510	80829
RefSeq Plant	7276	556	389
Swiss-Prot	395	388	364
NR	10102	3554	3063
TOT	180085	94008	84645

In addition to blast analyses and subsequent GO mapping and annotation, InterProScan and EggNOG analyses were also conducted using the whole dataset obtained after redundancy removal (281656 contigs). As a result, 49287 and 25708 contigs were annotated after InterProScan and EggNOG, respectively. As shown in Figure 4.34, after merging InterProScan results with the blast results, the number of annotated contigs increased from 84645 to 95107. Similarly, after the annotation with EggNOG, the number of annotated contigs increased from 95107 to 98043. Following EggNOG analysis, Enzyme Commission (EC) codes were also merged with the overall results. Accordingly, the number of ECs increased from 6792 to 73329.

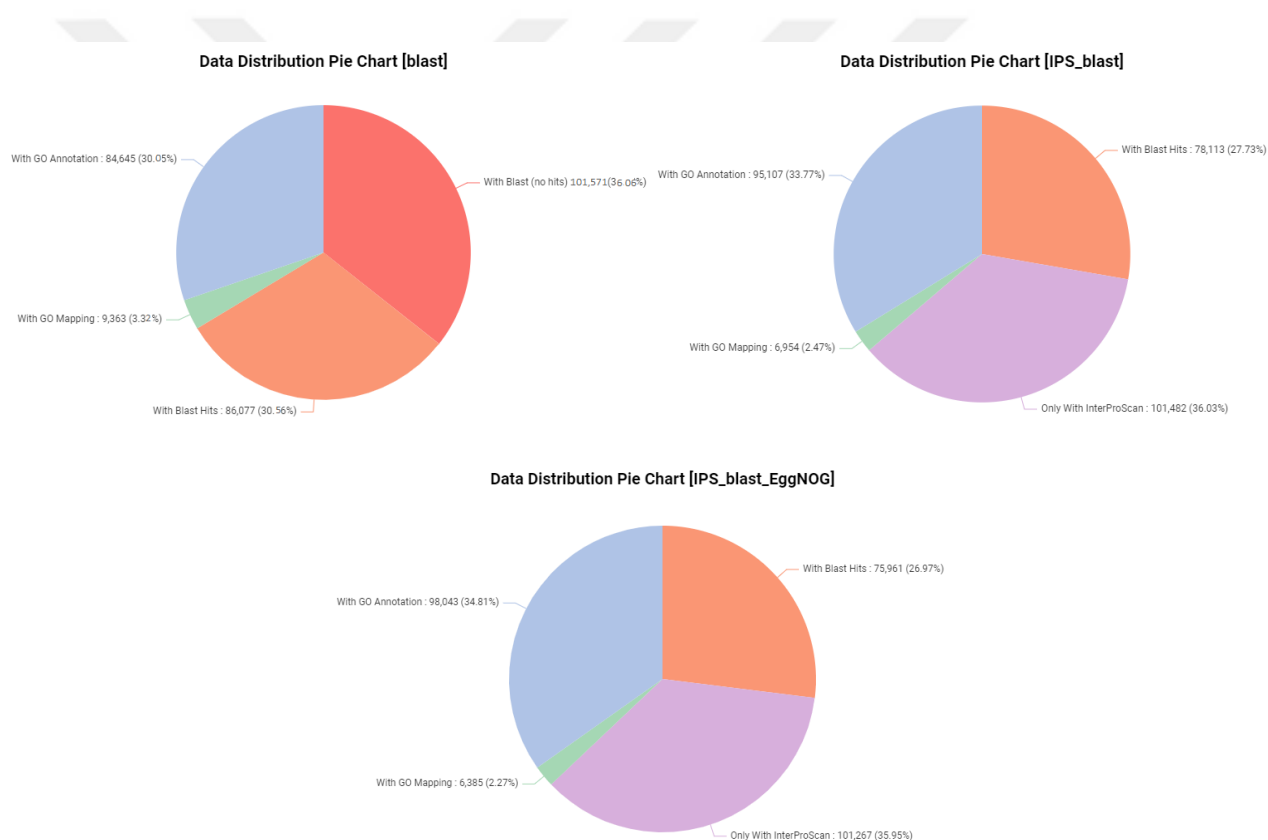


Figure 4.34. The distribution of contigs after Blast, InterProScan (IPS), and EggNOG analyses. The results of IPS and EggNOG were merged with Blast results, respectively.

Considering direct GO counts obtained for all annotated contigs, “protein phosphorylation” was the most common biological process (BP), followed by “regulation of DNA-templated transcription” (top panel in Figure 4.35). In the case of the molecular function (MF) category, “ATP binding” was the most frequent one, and “protein binding” was the second (middle panel in Figure 4.35). For the cellular component (CC) category, “membrane” was the most abundant (bottom panel in Figure 4.35).

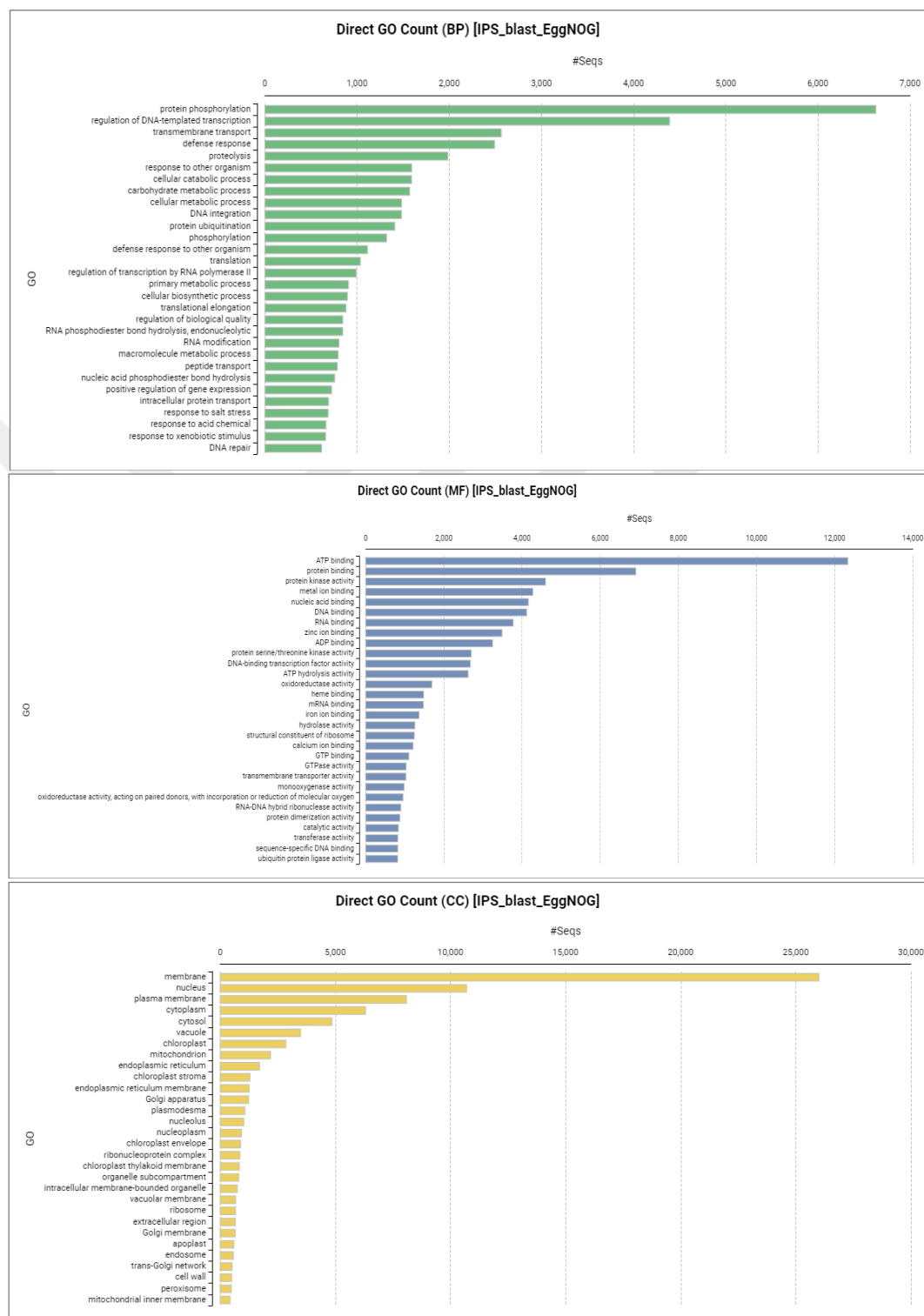


Figure 4.35. The direct GO counts of all contigs for the biological process (BP), molecular function (MF), and cellular component (CC) categories.

The enzyme code (EC) distribution of all contigs was also checked. Transferases were found to be the most frequent enzyme group, followed by hydrolases and oxidoreductases (Figure 4.36). Within

transferases, “transferring phosphorus-containing groups” was the most abundant, and within hydrolases, “acting on ester bonds” was the most common, as shown in Figure 4.37.

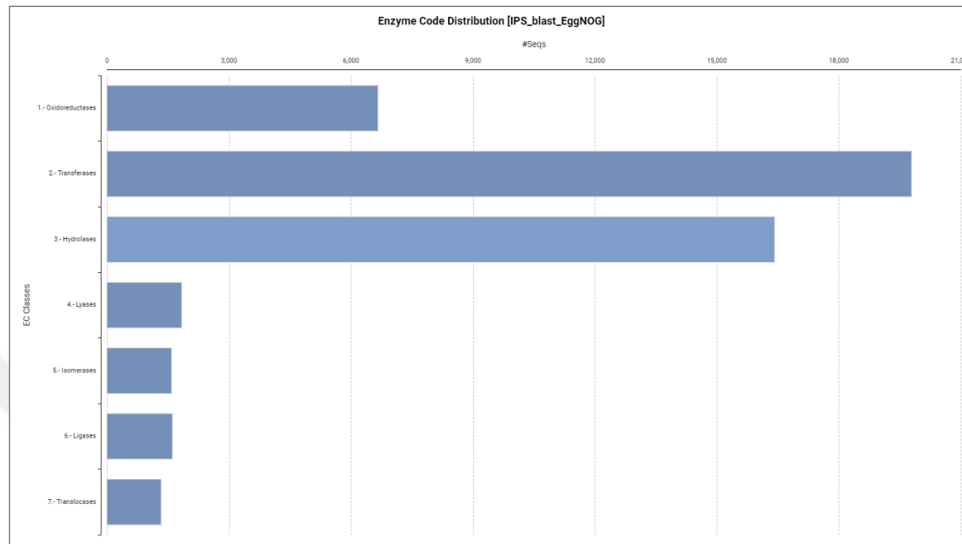


Figure 4.36. The distribution of enzyme codes (ECs) across all contigs.

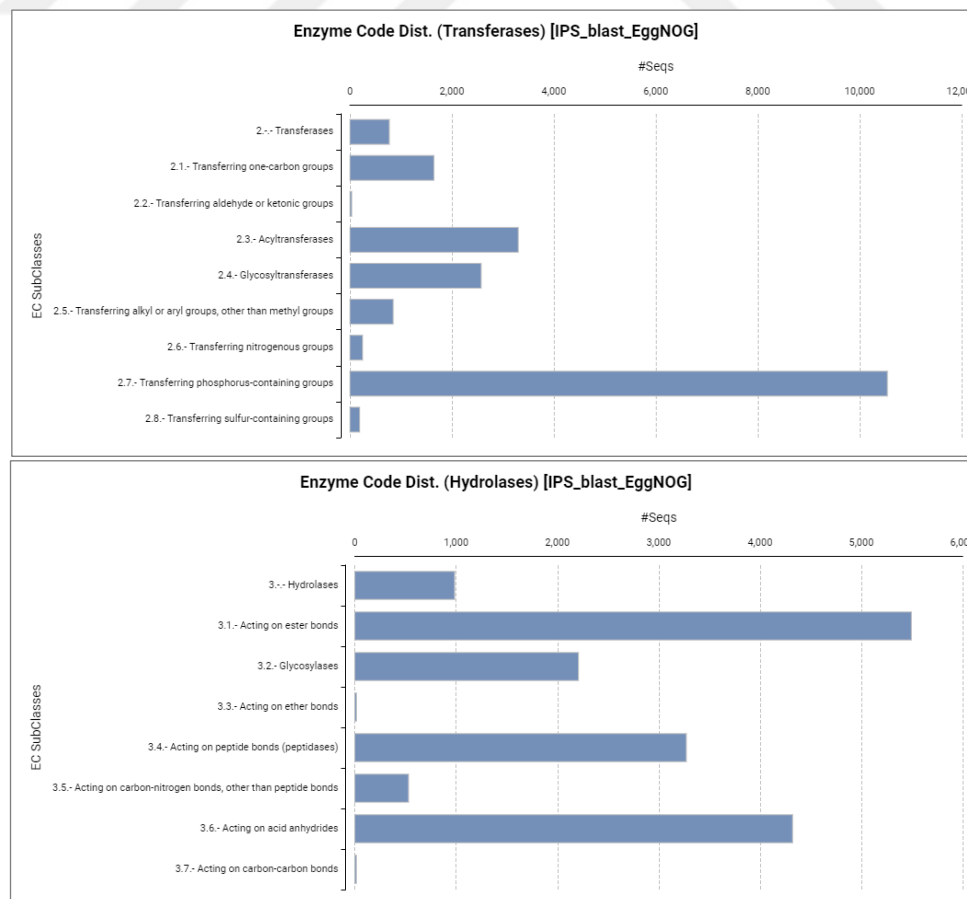


Figure 4.37. The EC subclasses within transferases and hydrolases across all contigs.

The most prominent orthologous groups found with EggNOG orthology assignment were root (17.85 %), Eukaryota (17.74 %), Viridiplantae (16.29 %), Streptophyta (16.21 %), Liliopsida (15.5 %), and Poales (14.86 %). Poales is the order to which wheat belongs (Bálint et al., 2000) and the others are all associated with wheat since they comprise eukaryotic green plants (Viridiplantae - Cavalier-Smith, 1981), land plants (Streptophyta – Lemieux, Otis, & Turmel, 2000), and monocotyledons (Liliopsida - Takhtajan, 2009).

Following GO annotation and orthology assignment, combined pathway analysis was conducted through OmicsBox (2019). As a result, 336, 2325, and 8558 pathways were found with KEGG, Reactome, and Plant Reactome databases, respectively.

4.2.8. Functional Annotations of Old vs. New Einkorn

In the previous section, the focus was the transcriptomic characterization of all contigs in the whole dataset. In this section, the results of the DE contigs (with cut-off values of FDR=0.05 and logFC=±1) between old and new einkorn are presented.

The top panel of Figure 4.38 shows the direct GO counts for the biological process category of DE contigs detected between the old (WP22) and new (WP89) einkorn at control temperature, whereas the same values are given in the bottom panel for DE contigs detected between the old and new einkorn at stress temperature. At control temperature, “protein phosphorylation” was the most common, along with “regulation of DNA-templated transcription” and “transmembrane transport” in the second and third spots, respectively. Similarly, at stress temperature, “protein phosphorylation” and “regulation of DNA-templated transcription” were in the first two spots, respectively. On the other hand, unlike control temperature, “carbohydrate metabolic process” was the third most common biological process, followed by “cellular catabolic process” and “transmembrane transport”.

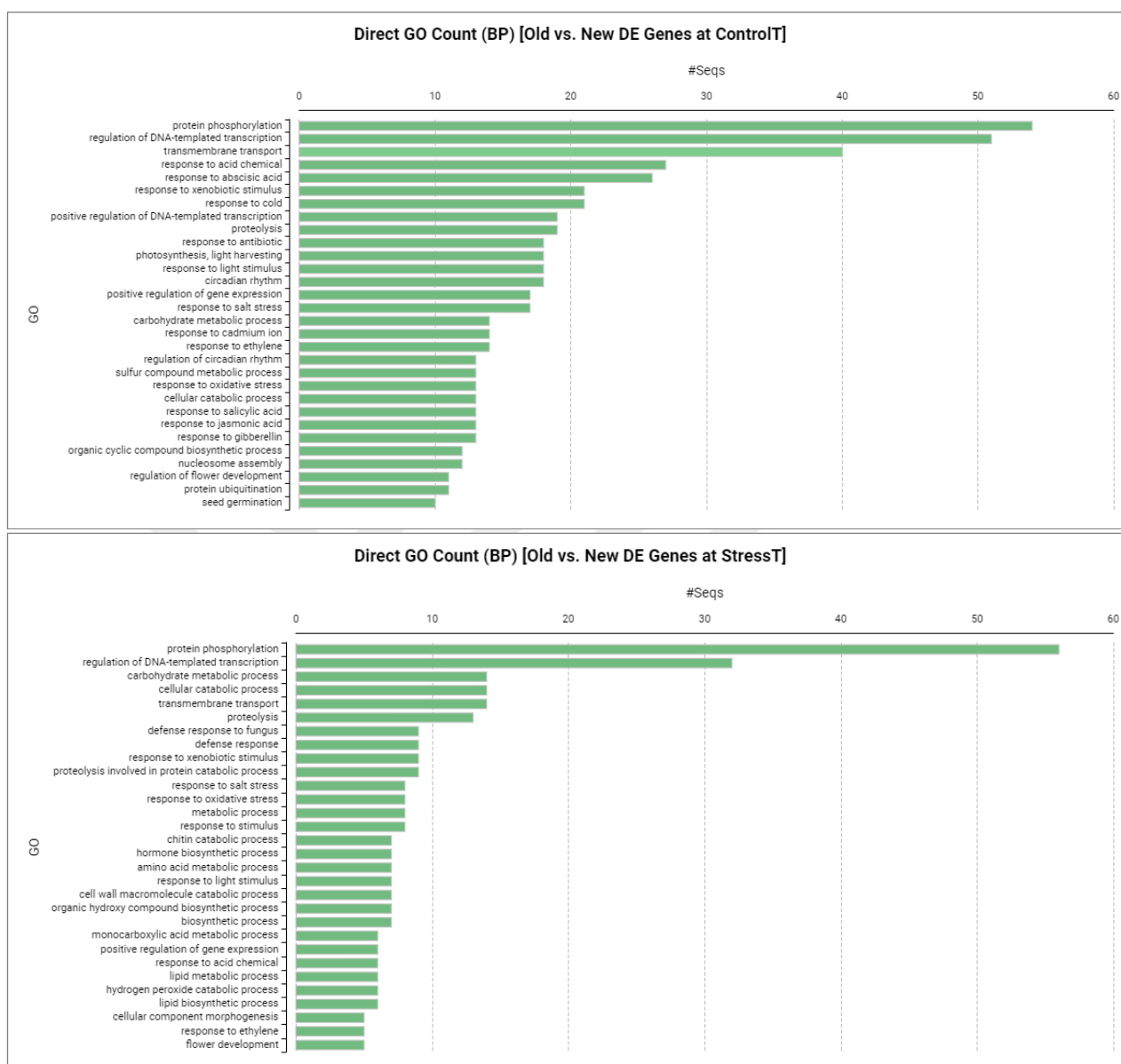


Figure 4.38. The direct GO counts of biological process (BP) between the old and new einkorn at control (top panel) and stress (bottom panel) temperatures.

Figure 4.39 (top panel) shows the direct GO counts for the molecular function category of DE contigs detected between the old and new einkorn at control temperature, whereas the same values are given in the bottom panel for the stress temperature. Accordingly, ATP binding related genes were the most common at both control and stress temperatures. On the other hand, “protein binding” and “DNA binding” followed “ATP binding” for control temperature, and “protein kinase activity” and “protein binding” were the second and third most common molecular functions at stress temperature.

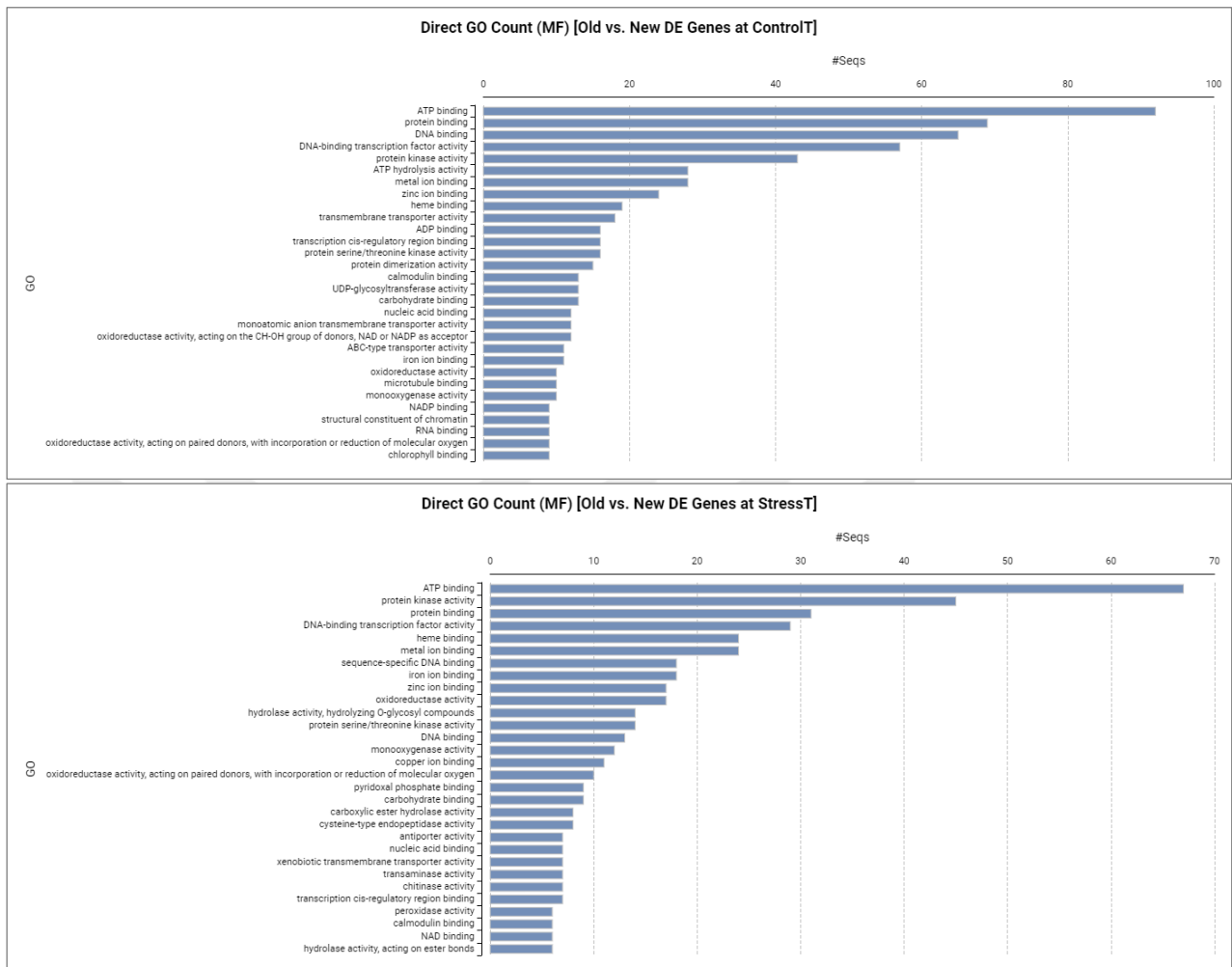


Figure 4.39. The direct GO counts of molecular function (MF) between the old and new einkorn at control (top panel) and stress (bottom panel) temperatures.

Again, in the top panel in Figure 4.40, the direct GO counts of contigs detected between the old and new einkorn at the control temperature are shown, and the same values are given in the bottom panel for the stress temperature, this time for the cellular component category. In both cases, “membrane” had the most contigs associated. “Nucleus” and “plasma membrane” were the second and third most common cellular components at control temperature, whereas these two components were in reverse order for stress temperature.

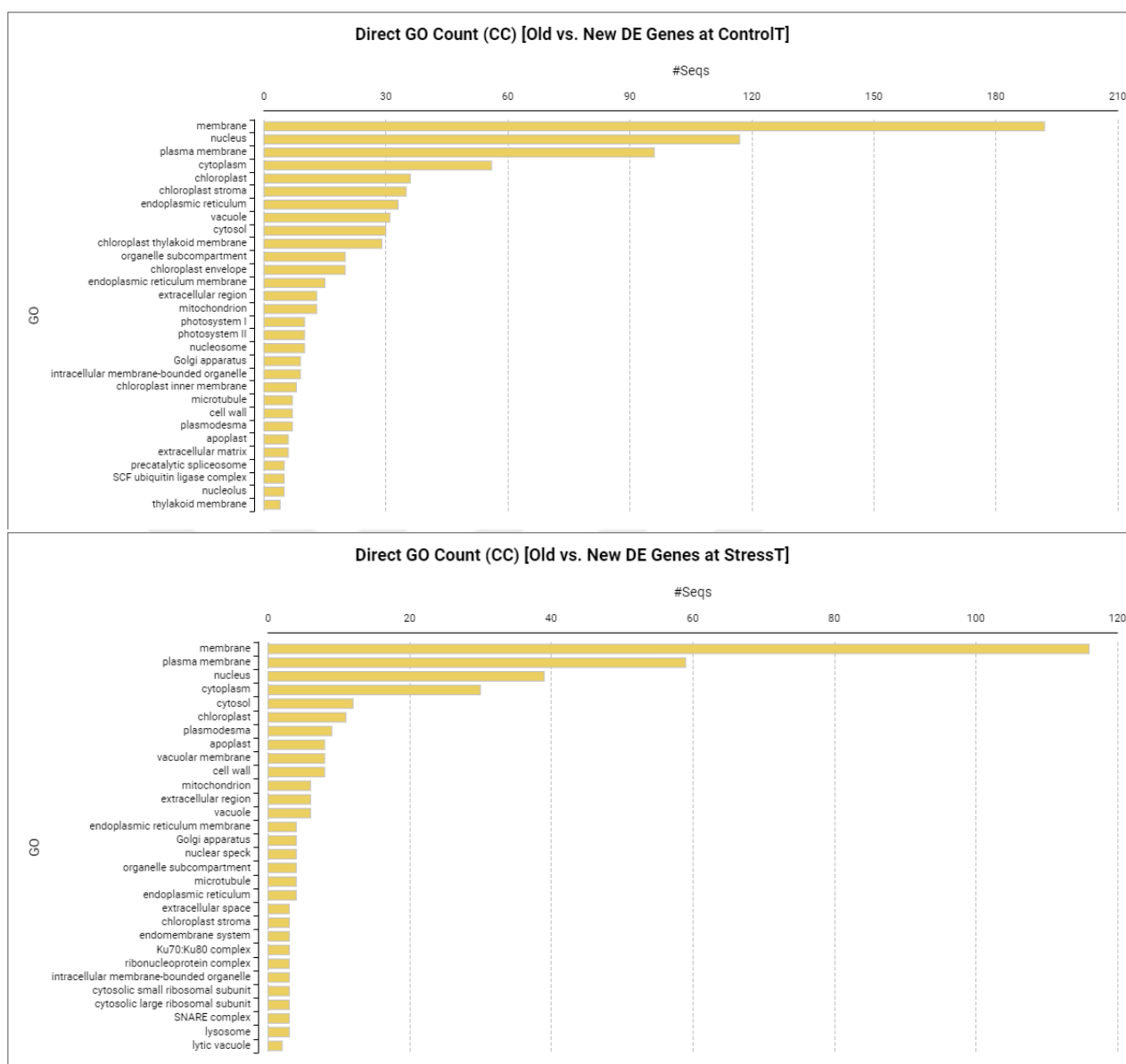


Figure 4.40. The direct GO counts of cellular component (CC) between the old and new einkorn at control (top panel) and stress (bottom panel) temperatures.

In terms of the enzyme code (EC) distribution of the contigs (Figure 4.41), transferases were found to be the most abundant at control temperature, followed by hydrolases (top panel), whereas a more equal distribution of transferases and hydrolases was found at stress temperature (bottom panel). In addition, within transferases, the enzyme code “transferring phosphorus-containing groups” had the most contigs associated with both control and stress temperatures. On the other hand, within hydrolases, “acting on acid anhydrides” was the most abundant at control temperature (top panel in Figure 4.42), and “acting on ester bonds” was the most common at stress temperature (bottom panel in Figure 4.42).

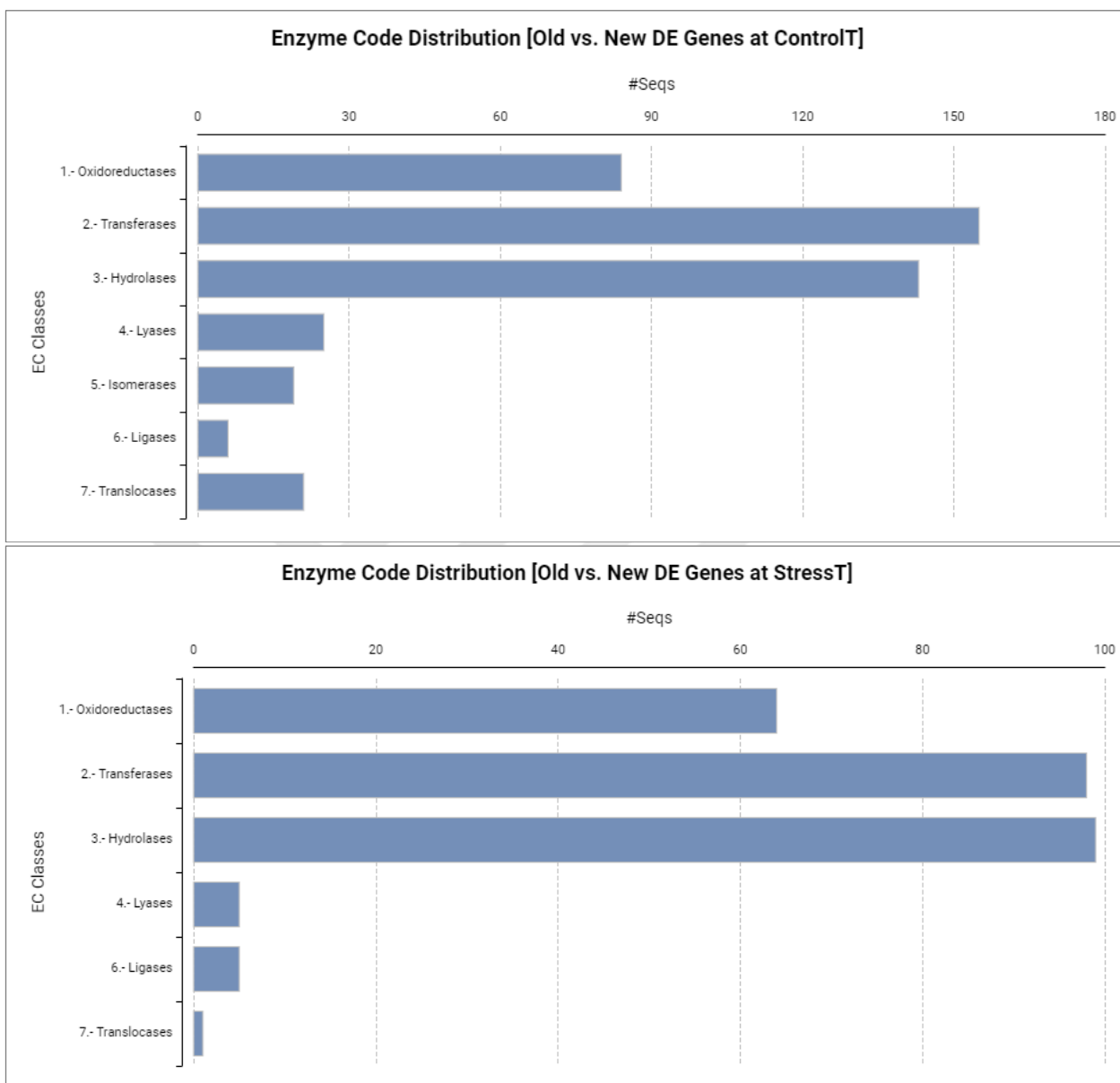


Figure 4.41. The EC classes between the old and new einkorn at control (top panel) and stress (bottom panel) temperatures.

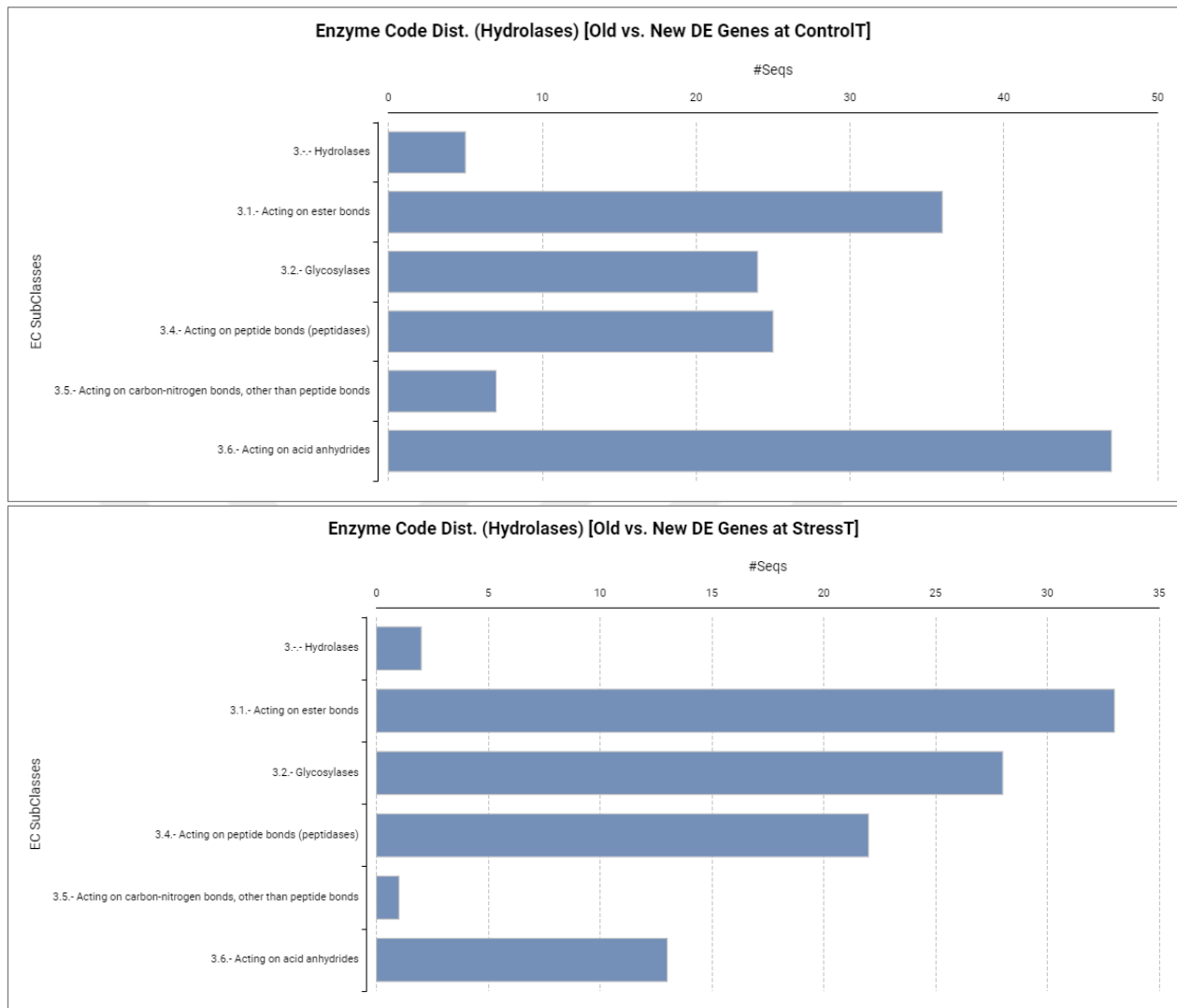


Figure 4.42. The EC subclasses within hydrolases between the old and new einkorn at control (top panel) and stress (bottom panel) temperatures.

In terms of the Gene Set Enrichment Analyses (GSEA), the top GOs enriched between the old and new einkorn at control and stress temperatures are shown in Figure 4.43. As shown in the top panel, “actin cytoskeleton organization”, “filament organization”, and “protein-DNA complex assembly” were the top three enriched GOs for control temperature. On the other hand, “photosystem”, “photosystem II”, and “preribosome” were the most enriched GOs at stress temperature (bottom panel).

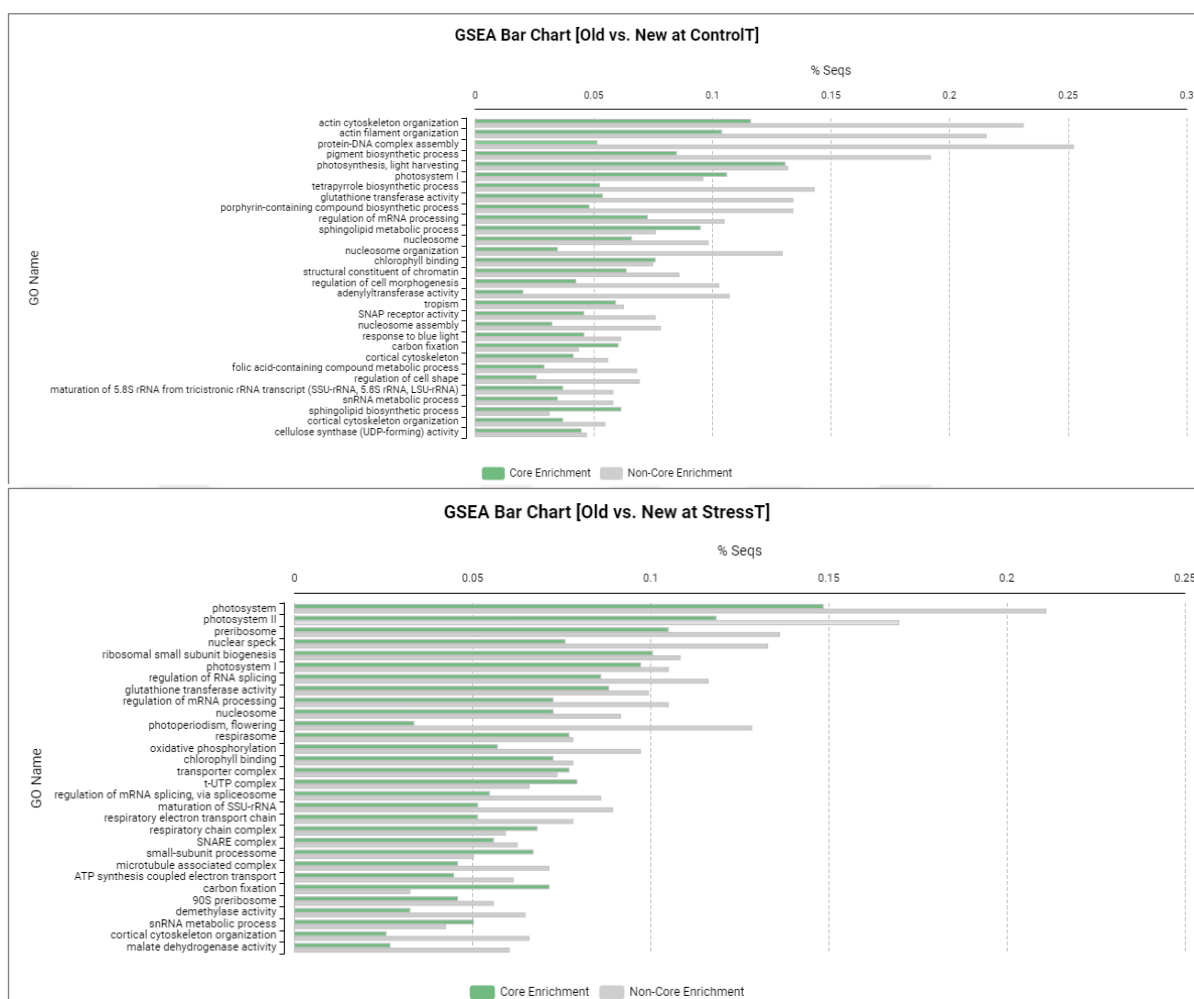


Figure 4.43. The enriched GO categories between old and new einkorn after GSEA analyses for the control (top panel) and stress (bottom panel) temperatures.

In terms of the pathways, out of 336 KEGG pathways, 128 and 99 pathways were enriched at control and stress temperatures, respectively (Table 4.8). Similarly, among 2325 Reactome pathways, 140 and 232 pathways were enriched at control and stress temperatures, respectively. On the other hand, as a result of Plant Reactome pathway analysis, no pathway was enriched.

Table 4.8. The numbers of enriched pathways between old and new einkorn.

	KEGG	Reactome	Plant Reactome
Found pathways	336	2325	8558
Enriched pathways at control temperature	128	140	0
Enriched pathways at stress temperature	99	232	0

Out of 128 KEGG pathways enriched at control temperature, the top ten pathways that were at the top and bottom of the ranked list (based on the normalized enrichment score (NES)) are given in Table 4.9. Accordingly, “photosynthesis – antenna proteins” was at the top of the list, whereas “salivary secretion” was at the bottom. Out of 140 Reactome pathways enriched at control temperature, the top ten pathways were all about “ABC-family proteins mediated transport” and “gluconeogenesis”. On the other hand, “pentose phosphate pathway” and “the proton buffering model” were recorded at the bottom of the list. The top 10 Reactome pathways enriched at control temperature are given in detail in Table B.5 in Appendix B.

Table 4.9. The top ten (top tag and bottom tag) enriched KEGG pathways for control temperature.

Pathway	NES
Photosynthesis - antenna proteins (ko00196)	2.1098883
Chemokine signaling pathway (ko04062)	2.0097375
B cell receptor signaling pathway (ko04662)	1.8692021
Circadian rhythm - plant (ko04712)	1.8536854
Fc epsilon RI signaling pathway (ko04664)	1.844425
Valine, leucine and isoleucine biosynthesis (ko00290)	1.8303883
Arginine biosynthesis (ko00220)	1.8254435
Signaling pathways regulating pluripotency of stem cells (ko04550)	1.8206339
Carbon fixation in photosynthetic organisms (ko00710)	1.8083938
Natural killer cell mediated cytotoxicity (ko04650)	1.8041666
Salivary secretion (ko04970)	-2.4648159
Toll and Imd signaling pathway (ko04624)	-2.1100423
Phototransduction (ko04744)	-2.0933492
Olfactory transduction (ko04740)	-2.0730717
NF-kappa B signaling pathway (ko04064)	-2.0313828
Inflammatory mediator regulation of TRP channels (ko04750)	-2.025334
Gastric acid secretion (ko04971)	-2.020499
Neurotrophin signaling pathway (ko04722)	-1.9044752
Drug metabolism - cytochrome P450 (ko00982)	-1.8944936
Renin secretion (ko04924)	-1.8893818

For the stress temperature, out of 99 KEGG pathways, the top ten pathways that were at the top and bottom of the ranked list are given in Table 4.10 with their NES values. “Diterpenoid biosynthesis” had the highest NES value, followed by “photosynthesis – antenna proteins”; and “styrene degradation” had the lowest NES value. Out of 232 Reactome pathways enriched at stress temperature, the top ten pathways were all about “fanconi anemia pathway”, and “passive transport by aquaporins”. On the other hand, at the bottom of the list, there was “ABC-family proteins mediated transport”. The top 10 Reactome pathways enriched at stress temperature are given in detail in Table B.6 in Appendix B.

Table 4.10. The top ten (top tag and bottom tag) enriched KEGG pathways for stress temperature.

Pathway	NES
Diterpenoid biosynthesis (ko00904)	2.5084507
Photosynthesis - antenna proteins (ko00196)	2.1099193
Insulin secretion (ko04911)	1.9936017
Biosynthesis of siderophore group nonribosomal peptides (ko01053)	1.8772817
ErbB signaling pathway (ko04012)	1.8438467
Protein digestion and absorption (ko04974)	1.8252279
Brassinosteroid biosynthesis (ko00905)	1.8031085
Cholinergic synapse (ko04725)	1.7404115
ECM-receptor interaction (ko04512)	1.7136899
RIG-I-like receptor signaling pathway (ko04622)	1.6688213
Styrene degradation (ko00643)	-2.084521
Synaptic vesicle cycle (ko04721)	-1.989848
Various types of N-glycan biosynthesis (ko00513)	-1.8810576
Glycosphingolipid biosynthesis - globo and isoglobo series (ko00603)	-1.8391048
Collecting duct acid secretion (ko04966)	-1.8133492
Biosynthesis of various secondary metabolites - part 3 (ko00997)	-1.80963
Arginine biosynthesis (ko00220)	-1.7140099
N-Glycan biosynthesis (ko00510)	-1.7124122
SNARE interactions in vesicular transport (ko04130)	-1.709887
Arachidonic acid metabolism (ko00590)	-1.7096734

Among pathways that were enriched at control and stress temperatures, “photosynthesis – antenna proteins” was the first and second top pathway, respectively. Within this pathway, contigs were found to

be DE between the old and new einkorn at control temperature, which took part in the synthesis of four light-harvesting chlorophyll proteins, namely LHCA2, LHCA3, LHCA4, and LHCB1 (Figure 4.44). On the other hand, none of the contigs were DE for the proteins involved in this pathway at the stress temperature.

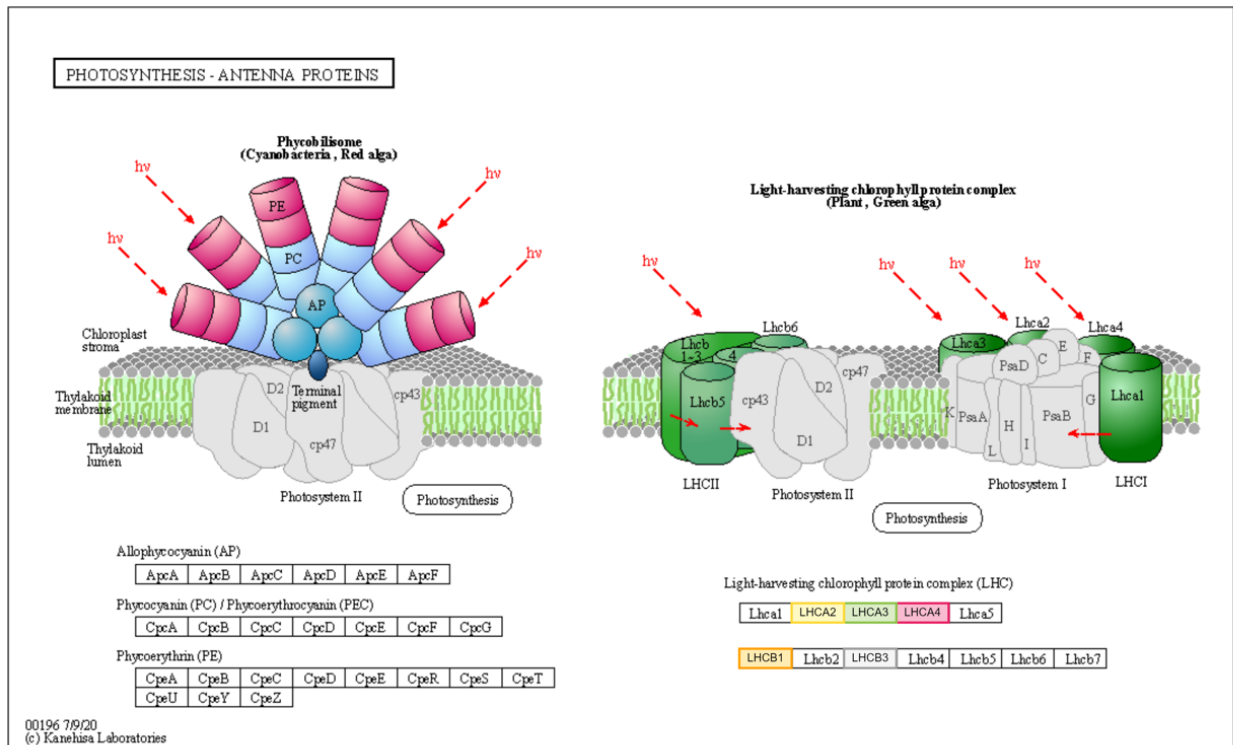


Figure 4.44. The photosynthesis – antenna proteins pathway enriched at the control temperature.

Even though it was not one of the top 10 enriched pathways, “photosynthesis” was also investigated. As shown in Figure 4.45, at stress temperature, contigs responsible for the synthesis of PetA (apocytochrome F), PetH (*i.e.*, EC 1.18.1.2, ferredoxin--NADP+reductase), and PsaC (photosystem I subunit VII) were found to be DE between old and new einkorn. On the other hand, for the control temperature, contigs related to photosystem proteins, PsaO and PsbW, were DE.

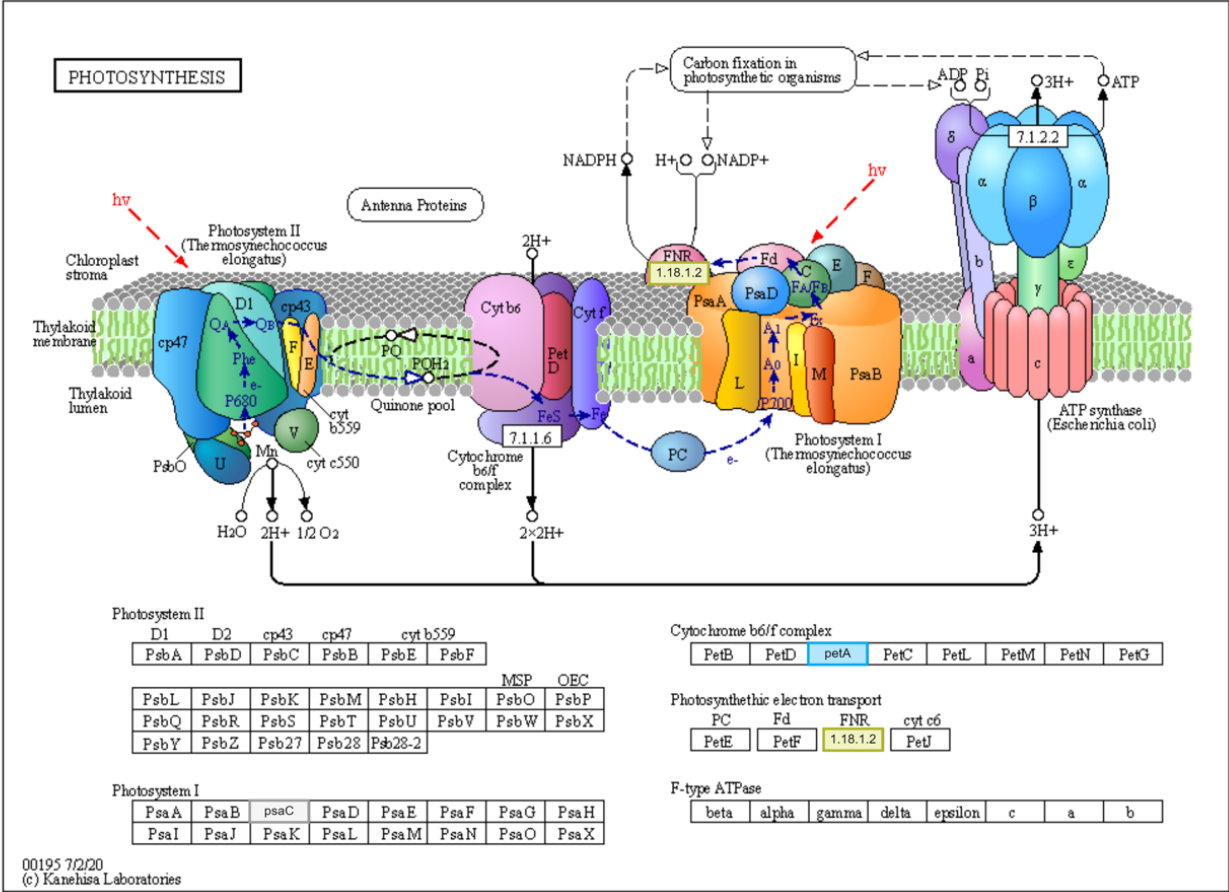


Figure 4.45. The photosynthesis pathway that was enriched at stress temperature.

The diterpenoid biosynthesis pathway was at the top of the enriched pathways at stress temperature. As shown in Figure 4.46, contigs associated with six GA20ox (Gibberellin 20-oxidase) proteins were found to be DE between old and new einkorn at stress temperature.

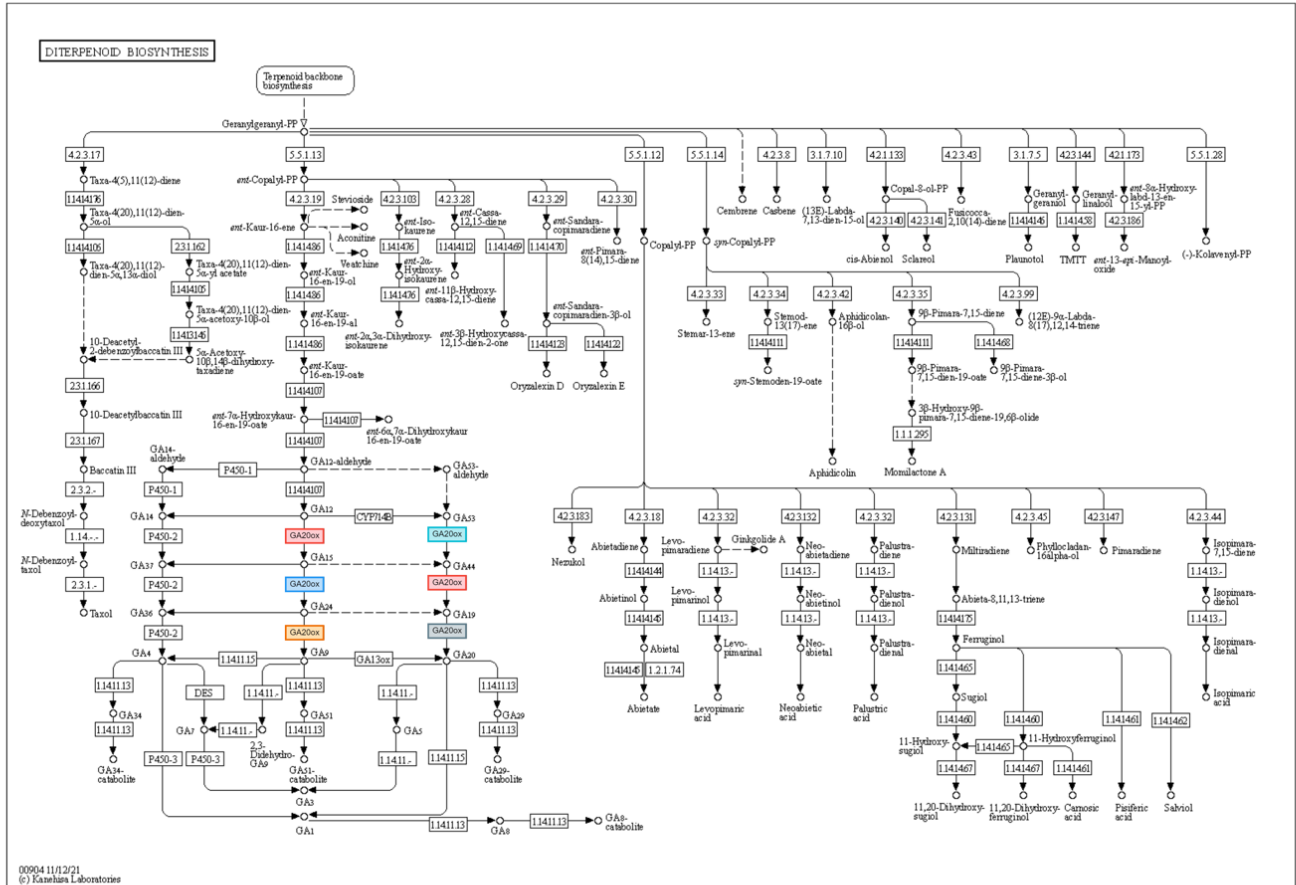


Figure 4.46. The diterpenoid pathway that was enriched at stress temperature.

4.2.9. The Top 100 DE Contigs and Their Functional Annotations

Out of all DE contigs sorted by their FDR values (adjusted p-value - from low to high), the top 100 contigs are given in Table B.7 (Appendix B). Accordingly, all 100 contigs were found for the comparison of control vs. stress temperatures. Out of 100, 30 were unique, meaning that they were observed in only old or new einkorn, and 35 were observed in both old and new einkorn. Among 30 unique contigs, 13 were found in the old einkorn, and 17 were found in the new einkorn. Considering 35 contigs that were observed twice, the FDR values of 23 contigs were higher in the old einkorn than in the new einkorn, whereas for 12 contigs, higher FDR values were recorded for the new einkorn than the old einkorn.

GO annotation results of the top 100 contigs are shown in Figure 4.47. For the biological process category (top panel), “regulation of DNA-templated transcription” was the most abundant process. Additionally, “response to heat stress” and “heat acclimation” were observed in the top 100 contigs. In terms of the molecular function category (middle panel), “ATP binding” was the most common function.

“Heat shock protein binding” was also observed in the top 100 contigs. For the cellular process category (bottom panel), “membrane” was the most frequent process.

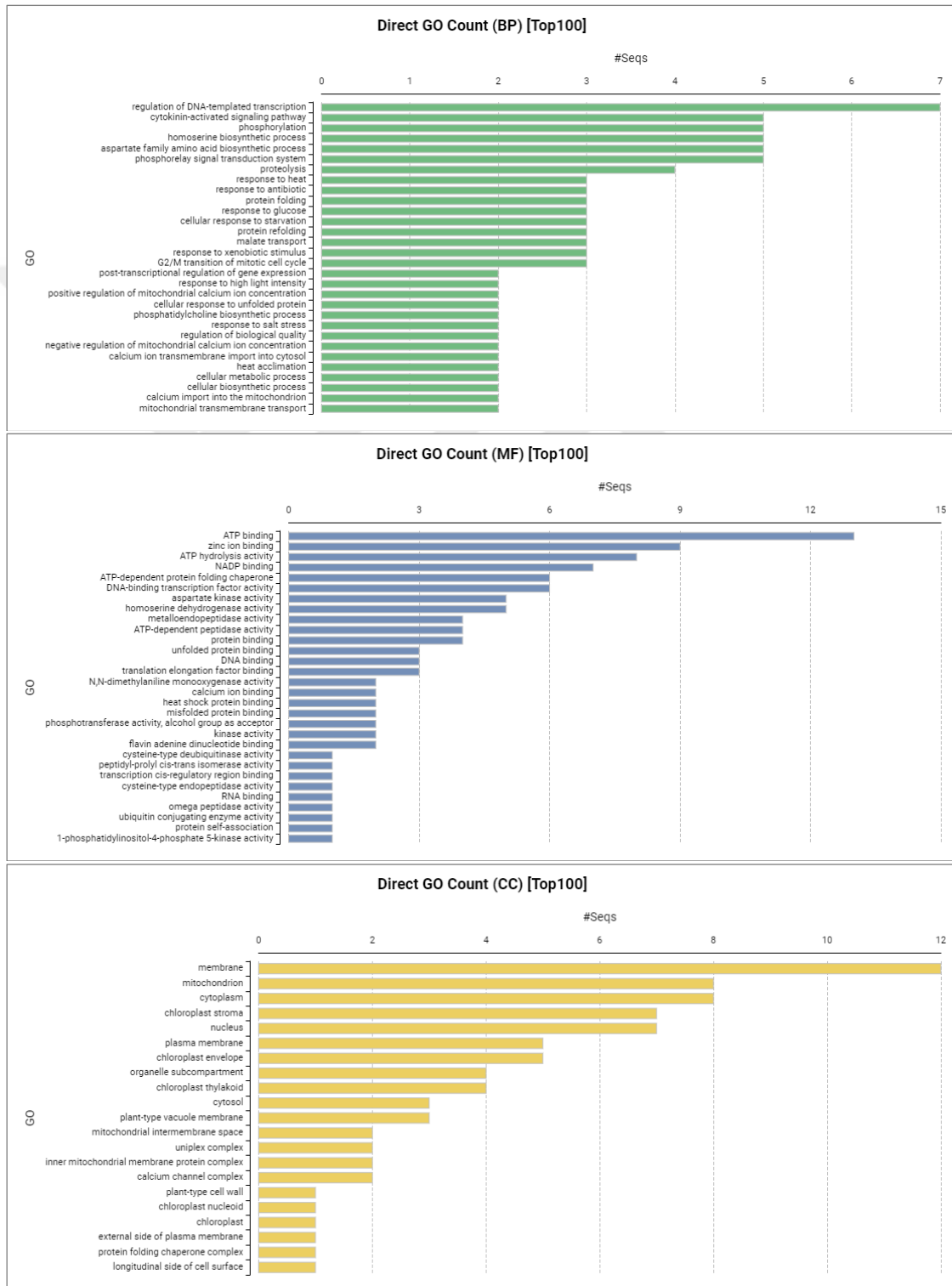


Figure 4.47. The direct GO counts of the top 100 contigs for the biological process (BP), molecular function (MF), and cellular component (CC) categories.

Considering all three categories, seven contigs were related to response to heat, heat acclimation, and/or heat shock protein binding according to the GO annotation results. The enzyme code distribution of the top 100 contigs was also checked. Like previous findings of the whole dataset and the comparison of old vs. new einkorn, hydrolases and transferases along with oxidoreductases were found to be the most common ones for the top 100 contigs.

4.2.10. Heat Stress Related Contigs

When the blast results were evaluated for the top 100 contigs mentioned in the previous section, eight contigs were found to be associated with heat shock proteins (HSPs) and heat stress. As given in Table 4.11, six of them were heat shock proteins with different molecular weights, and two of them were proteins associated with heat stress.

Table 4.11. Blast results of the top 100 DE contigs related to heat stress.

Contig	Accession Number	Gene Definition
TRINITY_DN1429_c3_g1_i1	XP_048572015	Heat shock 70 kDa protein 8
TRINITY_DN3465_c0_g2_i1	CAM96548	16.9 kDa heat shock protein
TRINITY_DN3465_c0_g2_i2	XP_044336482	16.9 kDa class I heat shock protein 2-like
TRINITY_DN548_c1_g1_i1	XP_04852387	Heat shock 70 kDa protein, mitochondrial-like
TRINITY_DN21206_c0_g1_i2	XP_037446456	Chaperone protein ClpB3, mitochondrial
TRINITY_DN2192_c0_g1_i2	XP_037465753	Protein heat-stress-associated 32-like
TRINITY_DN2690_c0_g1_i5	CAI96512	26.4 kDa heat shock protein
TRINITY_DN18960_c0_g3_i1	XP_037418584	Heat shock 70 kDa protein 4-like

In terms of HSPs and heat shock transcription factors (HSFs), DE contigs between old and new einkorn at control and stress temperatures were also examined. At control temperature, 10 contigs were found to be associated with HSPs or HSFs (Table 4.12), whereas at stress temperature none was related to these proteins. On the other hand, out of all DE contigs obtained with all four comparisons (old vs. new einkorn at control and stress temperatures, and control vs. stress temperatures for old and new einkorn), 200 contigs were found to be associated with heat, including HSPs and HSFs.

Table 4.12. Blast results of the DE contigs between old and new einkorn at control temperature related to HSPs or HSFs.

Contig	Accession Number	Gene Definition
TRINITY_DN7432_c1_g1_i2	XP_037480530	18.8 kDa class V heat shock protein-like
TRINITY_DN7432_c1_g2_i1	XP_037480530	18.8 kDa class V heat shock protein-like
TRINITY_DN12438_c0_g1_i4	XP_037461163	Heat stress transcription factor C-2b-like
TRINITY_DN28_c2_g2_i1	XP_044382889	Heat stress transcription factor A-2a-like
TRINITY_DN28_c2_g6_i1	XP_037468869	Heat stress transcription factor C-2b-like
TRINITY_DN1205_c4_g1_i1	XP_044375669	Heat shock cognate 70 kDa protein-like
TRINITY_DN15680_c1_g1_i1	XP_048531751	Heat stress transcription factor A-2a-like
TRINITY_DN15680_c1_g1_i2	XP_048531751	Heat stress transcription factor A-2a-like
TRINITY_DN15680_c1_g1_i3	XP_048531751	Heat stress transcription factor A-2a-like
TRINITY_DN15680_c1_g1_i4	XP_048531751	Heat stress transcription factor A-2a-like

5. DISCUSSION

5.1. ddRAD-seq

In this study, four de novo analyses were performed for three wheat species separately, and for all populations. The numbers of assembled loci, polymorphic loci, and SNPs obtained with each analysis are summarized in Table 5.1. Accordingly, among three species, the highest numbers of loci (assembled) were obtained for emmer (735,891), followed by bread wheat (533,475) and einkorn (345,911). Out of these assembled loci, the percentages of polymorphic loci from high to low were 31.6 %, 25.8 %, and 21.8 % for emmer, einkorn, and bread wheat, respectively. This suggests a high genetic diversity level for emmer, followed by einkorn and bread wheat. On the other hand, in the case of SNPs, the highest number was obtained for emmer (1,172,135), but in contrast to the polymorphic loci proportion, bread wheat had more SNPs (608,407) than einkorn (309,005), suggesting that the genetic diversity was the highest for emmer, followed by bread wheat and einkorn. In a way, this is consistent with the findings of many studies, showing that ancient wheat species have higher diversity levels than modern wheat (Roussel et al., 2005; Charmet, 2011; Shewry & Hey, 2015; Özberk et al., 2016). In particular, studies using SNPs for investigating the genetic diversity in wheat showed that the B genome has the highest proportion of polymorphism, followed by the A and D genomes (Ren et al., 2013; Kumar et al., 2020a; Yang et al., 2020; Tyrka et al., 2021). As einkorn has the A genome, while emmer has AB, and bread wheat has ABD, this supports the high diversity levels of einkorn and emmer found in this study, as well as the finding of emmer as being more diverse than einkorn. And the low number of SNPs in the einkorn genome can be explained by the findings of Deng et al. (2016), who showed that the A genome has the lowest number of SSRs, whereas there are more abundant SSRs in the B and D genomes ($B > D > A$). Similarly, according to the RFLP analysis conducted by Takumi et al. (1993), the number of polymorphisms was found to be higher in polyploid wheat than einkorn.

Table 5.1. Loci and SNP numbers obtained for de novo analyses.

De novo Analysis	Assembled Loci	Polymorphic Loci	Polymorphic Loci (%)	SNPs
Einkorn	345,911	89,389	25.8	309,005
Emmer	735,891	232,275	31.6	1,172,135
Bread wheat	533,475	116,347	21.8	608,407
All populations	1,157,567	361,082	31.2	1,891,799

When all populations were analyzed, the separation between the three species was clearly observed, showing the robustness of the analysis. The only exception was one emmer population, WP20, which was clustered with einkorn populations instead of emmer (discussed below). On the other hand, the nucleotide diversity calculations of all populations revealed lower diversity levels of einkorn than the other two species, and F_{ST} analysis showed the genetic distance between einkorn and the other species. Although all wheat species have the A genome, the A genome donor of the polyploid wheat is *T. urartu* (Dvorak & Akhunov, 2005; Bálint et al., 2000; Ogihara & Tsunewaki, 1988; Dvorak & Zhang, 1990). *T. urartu* and einkorn (*T. monococcum*) are both diploids and have the A genome, but they give rise to sterile hybrids (Johnson & Dhaliwal, 1976), meaning that they are different species. Besides, 15 % of the genome of einkorn does not match with the genome of *T. urartu* (Fox et al., 2014). So, the distance between einkorn and the other wheat species is most likely due to this difference in their A genomes. Also, as the seed exchange is less common for einkorn than the others, einkorn has mostly remained its genetic material as it was at the beginning of agriculture (Yaman et al., 2020). So, its genetic diversity might have become less than other wheat species over time, and this might have also caused a distance between einkorn and the other wheat species. Additionally, the cultivation and spreading rates of einkorn decreased during the Bronze Age because it was replaced by modern high-yielding cultivars (Zohary & Hopf, 1993; Yaman et al., 2020; FAO, 2015). Thus, the genetic distance between einkorn and the other wheat species increased over time (Jing et al., 2007), which may also explain the genetic distance between einkorn and polyploid wheat.

Considering the analyses of einkorn, a distinction by year was observed within the populations. *Structure* analysis generated a clustering pattern in which the old (WP22 - 1995) and new (WP81, WP88, WP89 - 2015) populations were separated by year. However, there were also similarities between old and new einkorn, making the separation relative. The genetic distance between old and new populations was observed by the F_{ST} analysis more precisely, with the highest distance for WP22 and WP89. In respect of nucleotide diversity values, the old einkorn population had a higher diversity level than the new ones. This means that ex-situ conserved seed bank populations (1995) retained their diversity levels, which made them more advantageous against altering climatic conditions than in-situ conservation populations (2015). The reason for having lower diversity levels for the new populations is most probably the genetic bottlenecks that they experienced caused by climate change within the last 20 years. A similar diversity loss in wheat populations through time (from 1980 to 2008) was also found by Nevo et al. (2012), and Fu et al. (2019).

In the case of emmer, one of the two old populations, WP20, was clustered in one group, and the new ones, as well as the other old population (WP15), were clustered in another group. The genetic distance between WP20 and the others was also high, indicating the divergence between the two groups (WP20 vs. the others). As mentioned above, in the clustering analyses of all wheat populations, WP20 was grouped with einkorn populations, suggesting that it is genetically closer to einkorn than emmer. So, this distinction between WP20 and the other emmer populations is about the resemblance of this old population to einkorn species rather than the distinction between old and new. This is most likely due to the misidentification of this population when collecting the seeds, as emmer and einkorn both have hulled seeds (Salamini et al., 2002), and they are physically alike. Or this might be due to introgression between einkorn and emmer populations, which made WP20 genetically close to einkorn.

In addition to the clear clustering pattern and the genetic distance between WP20 and the other emmer populations, the nucleotide diversity and heterozygosity values of WP20 also varied from the others. WP20 had lower genetic diversity levels than the new populations, as well as WP15 from 1995. This confirms the inference about genetic diversity levels made from the SNP numbers of each species (see above) because WP20 was found to be closer to the einkorn group than emmer.

When the results of emmer populations other than WP20 were evaluated, in contrast to einkorn, there was no evidence that climate change in the last 20 years caused a divergence between emmer populations, in opposition to our expectations. The nucleotide diversity, heterozygosity, and F_{ST} values were similar for WP15 and the new emmer populations, suggesting the absence of a rapid evolution from 1995 to 2015. Likewise, in a study investigating the genetic diversity levels of ancient and modern wheat, it was found that the level of diversity of some varieties was more or less the same in the period of 1930-2019 (Sthapit et al., 2020). One possible reason for that is the ploidy level of wheat. Polyploid plants have duplicated (homologous) genes, which allow them to respond to different environmental stresses with up or down-regulation of separate copies of homologous genes (Adams & Wendel, 2005; Liu et al., 2015). This partitioning expression pattern is very useful against biotic or abiotic stresses (Liu et al., 2015). In addition, although wheat genes are generally located in mobile sequences, and they are highly prone to replacements and mutations, it is known that as the ploidy level increases, the effect of null mutations in duplicate or triplicate homologous gene copies decreases (Dubcovsky & Dvorak, 2007). Because of these reasons, emmer, as a tetraploid, might have been protected from the genetic bottlenecks caused by climate change by its extra B genome.

On the other hand, even though the populations from Kars and Kastamonu had similar values of heterozygosity and diversity, the F_{ST} values showed that there was a certain distance between emmer populations from different locations. This is consistent with the findings of Gürcan et al. (2017), who showed a clear clustering of wheat populations based on the regions (including Kars and Kastamonu) which they have been collected from.

For bread wheat, due to the absence of populations from earlier years or from another location, only three new populations from Kastamonu were used in genetic diversity analyses. As they were from the same year and location, similar diversity and heterozygosity values, and genetic distances were recorded for all three populations, as expected. However, they were used for the comparisons with the other wheat species as described above.

5.2. RNA-seq

In the RNA-seq section of the study, transcriptome data obtained from 36 RNA-seq libraries of local einkorn populations from 1995 (WP22) and 2015 (WP89) were used to investigate their genomic responses under heat stress. When the de novo transcriptome assembly was examined with BUSCO analysis, a high percentage (93.18 %) of BUSCO groups was found to have complete gene representation indicating the success of the assembly. MDS analysis that was performed with the data obtained after redundancy removal, filtering, and normalization, resulted in two distinct clusters: one with the control libraries and the other with the stress libraries. There was a relatively weak separation between old and new einkorn, and the clear grouping based on the temperature indicates the success of the heat stress experiment and higher transcriptional differences of einkorn populations between control and stress temperatures than the difference between old (WP22, 1995) and new (WP89, 2015) einkorn.

As a result of differential gene expression analysis (with FDR and logFC cut-off values of 0.05 and ± 1), the highest proportions of DE contigs were obtained between control and stress temperatures within the old (24.4 %) and within the new (22.2 %) einkorn, respectively. On the other hand, for the comparisons of old and new einkorn at control temperature, and old and new einkorn at stress temperature, the proportions of DE contigs were 1.35 and 0.75 %, respectively. These are consistent with the MDS analysis results showing lower transcriptional differences between old and new einkorn than control and stress temperatures.

Even though the DE contigs recorded between old and new einkorn at both temperatures were relatively small in number, they still indicate that the genomic response of einkorn has changed in the past 20 years due to climate change. Since old and new einkorn populations were found to be distant in terms of their genetic diversity levels (see section 5.1), this also shows the relation between genetic diversity and adaptation capacity, as suggested by Jump et al. (2009). Easier adaptation to altered environmental conditions is assumed for populations with high genetic diversity levels, while populations with low genetic diversity are more prone to be adversely affected by climatic changes (Frankham et al., 2017; Walther et al., 2002). In the case of old and new einkorn investigated in this study, the genetic diversity level was found to be higher in the old einkorn than in the new einkorn. This might have resulted in old einkorn adapting to climate change, particularly to heat stress, through differential gene expression.

5.2.1. Functional Annotations of All Contigs

Transcriptomic characterization was performed with blast analysis and subsequent GO mapping and annotation, as well as InterProScan and EggNOG analyses. As a result, 98043 annotated contigs were obtained. Among those, direct GO counts were calculated for three categories: biological process, molecular function, and cellular component, as discussed below.

In terms of the most common biological process, the dominance of “protein phosphorylation” was observed. Protein phosphorylation is a process in which a phosphate group is added to a protein, altering its function. In plants, heat stress can cause changes in the phosphorylation state of specific proteins (Duncan & Hersey, 1989), leading to alterations in their activity. These include the activation of HSPs, which can help prevent damage to cellular components and maintain proper protein folding (Kalefetoğlu & Ekmekçi 2005). Protein phosphorylation can also help regulate ion channels, to maintain proper water balance, and prevent dehydration during heat stress (Aslam, Maqbool, & Cengiz, 2015; Zhang, Zhu, Gong, & Zhu, 2022). In addition to protein phosphorylation, “regulation of DNA-templated transcription” was also abundant in the dataset. In this regard, during heat stress, plants produce specific proteins such as HSPs to cope with the stress, and HSFs are the most important molecules that regulate the expression of the genes that produce those stress-responsive proteins (Ohama, Sato, Shinozaki, & Yamaguchi-Shinozaki, 2017; Guo et al., 2016).

For the molecular function category, “ATP binding” was the most frequent one. The ATP binding GO terms are associated with helping plants to generate and maintain high levels of ATP in their system. ATP is the primary source of energy for cells in plants and is used in a variety of cellular processes such as active transport and biosynthetic reactions, as well as dealing with heat stress, and proper ATP binding is necessary to maintain these functions during heat stress. ATP binding also helps to activate stress response pathways, including the activation of HSPs, which protect cellular components from damage, and these processes require ATP for their activation. In addition, ATP is also involved in regulating gene expression during heat stress by modulating the activity of transcription factors and other proteins involved in gene regulation (Song, Chen, Ci, Shao, & Zhang, 2014). The second most common function was “protein binding”. In general, molecular interactions such as binding to proteins are required for the regulation of physiological processes in plants and responding to abiotic stresses including heat stress is one of those processes (Struk et al., 2019). When plants are exposed to high temperatures, interactions between molecules such as proteins, nucleic acids, or lipids are made for the networks responsible for heat stress response (Struk et al., 2019). For instance, for the activation of HSPs, first, HSFs bind together, and then this HSF complex bind to heat shock elements (HSEs) (Zhao et al., 2020).

In terms of the cellular component, “membrane” category was the most common. Cellular components of membrane-related genes play a crucial role in helping cells respond to heat stress. Proteins that regulate the fluidity and stability of cell membranes, including HSPs, help protect cells from damage caused by changes in temperature. Additionally, membrane transporters also help regulate the influx and efflux of water into and out of cells during photosynthesis, allowing cells to maintain proper hydration levels and avoid thermal damage (Allakhverdiev et al., 2008; Saidi et al., 2010). These and other membrane-related genes are activated in response to heat stress and help cells survive and recover from heat-induced damage.

In addition to three categories, enzyme code distributions of all contigs were also investigated. In a similar vein to “ATP binding” having been detected as the most frequent molecular function, “transferring phosphorus-containing groups” was highly represented within the enzyme class transferases. Phosphorus is a vital component of ATP, the energy currency of cells. Heat stress is known to deplete energy reserves in various organisms (Klepsatel, Gálíková, Xu, & Kühnlein, 2016), including plants (Wahid, Gelani, Ashraf, & Foolad, 2017). Transferring phosphorus to cells under heat stress can help replenish their energy reserves, allowing them to maintain normal cellular functions and better cope with the stress, which can result in improved survival and recovery from heat stress. Additionally, “acting

on ester bonds” was the most abundant within hydrolases, which was the second common enzyme class. In plants, ester molecules are generally the source of signaling agents, plant hormones, and pathway intermediates, which can be released to the cells with their hydrolysis (Gershater & Edwards, 2007). For example, abscisic acid (ABA), which is a plant hormone, plays an important role in responding to environmental alterations (Lee et al., 2006), and it is normally found as an inactive glucose ester in plants (Hua, Sansenya, Saetang, Wakuta, & Ketudat Cairns, 2013). Under stress conditions, such as heat stress, the hydrolysis of these esters takes place, and the amount of ABA increases to deal with stress (Lee et al., 2006; Hua et al., 2013; Li et al., 2021).

Following the transcriptomic characterization of all contigs, the DE contigs between old and new einkorn were evaluated in terms of functional annotations. The results are discussed below.

5.2.2. Functional Annotations of the DE Contigs Between Old and New Einkorn

Similar to the results of all contigs, in terms of the comparisons of the DE contigs found in the old vs. new einkorn considering biological processes, contigs related to “protein phosphorylation” and “regulation of DNA-templated transcription” were the most common at both control and stress temperatures. As the contigs that contribute to these processes were found at both control and stress temperatures, this suggests that they may not be related to the heat stress response of einkorn. On the other hand, considering the other processes observed, “transmembrane transport” was the third most common process at control temperature, while there was a more equal distribution of “carbohydrate metabolic process”, “cellular catabolic process”, and “transmembrane transport” at stress temperature, suggesting that the combination of these three processes play an important role in the heat stress response of einkorn wheat.

As mentioned above, maintaining cell membrane stability when exposed to high temperatures is very important for plants to survive under heat stress. To maintain membrane stability, under heat stress, Ca^{2+} channels open so that Ca^{2+} can transport into the cell through transmembrane ion channels embedded in the plasma membrane (Saidi et al., 2010). These ions are involved in processes associated with the formation of HSPs, and the activation of plant hormones, such as protein kinase, which protect the plant's cellular structures and enzymes from heat damage (Al-Whaibi, 2011; Asthir, 2015; Gill et al., 2021; Mittler, Finka, & Goloubinoff, 2012), as expected. Indeed, the increase in calcium ions and subsequent increase in proteins induced by Ca^{2+} (e.g., HSPs) in wheat was shown by Liu et al. (2003). Additionally,

the membrane lipid composition changes in plants under heat stress with a decrease in the amount of unsaturated fatty acids to regulate membrane fluidity (Djanaguiraman, Boyle, Welti, Jagadish, & Prasad, 2018). Narayanan, Tamura, Roth, Prasad, & Welti (2016) also showed this to be the case in the wheat precursor, *Triticum aestivum*.

In a similar vein, “membrane” and “plasma membrane” cellular components were abundantly found at the stress temperature as associated with DE contigs between old and new einkorn, suggesting that membrane related genes may contribute to coping with heat stress in the old einkorn. On the other hand, “membrane” was dominant at the control temperature, too, whereas “plasma membrane” was at the third spot rather than being second, like at the stress temperature. So, genes involved in processes associated with membrane might be a common response of the old einkorn, but it may be more frequently related to coping with heat stress since membrane related components were in the first spots at stress temperature.

Regarding “cellular catabolic process”, proteins damaged by high temperatures and protein aggregates generated during heat stress are degraded by HSPs (Reis, Lima, & de Souza, 2012). Reactive oxygen species (ROS) generated excessively because of heat stress are also degraded since they are harmful to the cell (Hasanuzzaman et al., 2020). In addition, the regulation of carbohydrate metabolism is vital for plants to survive under stress conditions, since carbohydrates, which are the products of photosynthesis, are used for the main physiological processes in plants, such as energy metabolism, growth, and development (Herrmann, Schwartz, & Johnson, 2019). Sugars are also used as signaling molecules to activate genes that are related to stress response (Hasanuzzaman et al., 2013). So, when plants are exposed to heat stress, the amount of carbohydrates, including cell wall monosaccharides, increases (Kaplan & Guy, 2004). This supports the “carbohydrate metabolic process” being in the third spot of biological processes observed at the stress temperature.

In terms of molecular function, “ATP binding” was the most common GO at both control and stress temperatures, as found for all contigs. This suggests that energy production with ATP is a common process for einkorn populations. Focusing on the stress temperature, “protein kinase activity” and “protein binding” followed ATP binding. Protein binding was also at the second spot for the control temperature as well as for all contigs mentioned in the previous section. On the other hand, protein kinase activity was in the second spot at the stress temperature, while it was in the fifth at the control temperature, suggesting that it might have a role in the stress response of einkorn wheat. Like sugars,

protein kinases help activate stress-responsive genes in plants (Hasanuzzaman et al., 2013). They catalyze protein phosphorylation by transferring phosphorus from ATP to the proteins (Dissmeyer & Schnittger, 2011). Specifically, receptor-like protein kinases (RLKs), mitogen-activated protein kinases (MAPKs), and calcium-dependent protein kinases (CDPKs) are known to regulate gene expression when plants are exposed to heat stress (Huang, Zhou, Ding, & Zhu, 2022). Indeed, Kumar et al. (2021) showed an association between MAPK and HSPs, HSFs, and other heat stress related molecules and processes in a thermotolerant wheat variety under heat stress.

Similar to the “ATP binding” dominance observed at both temperatures, within the enzyme group transferases, “transferring phosphorus-containing groups” was the most abundant for both temperatures, supporting the general need for energy production in einkorn. On the other hand, at the stress temperature, there was nearly an equal distribution of transferases and hydrolases, and within hydrolases, “acting on ester bonds” was the most frequent one. This is in line with the results of all contigs mentioned above. Since the hydrolysis of ester bonds release signaling agents, plant hormones, and pathway intermediates, which are important for plants to cope with heat stress, it is meaningful to observe this category most at the stress temperature.

In summary, even though the most abundant biological processes, molecular functions, and cellular components recorded between old and new einkorn were mainly common for both control and stress temperatures, this might be due to the effectiveness of these GOs under both normal and stress conditions. When the frequent ones recorded for the stress temperature are considered, including transmembrane transport, ATP binding, protein kinase activity, plasma membrane, and transferring phosphorus-containing groups, this suggests a combined response of einkorn wheat to heat stress.

5.2.3. Enriched GOs and Pathways in the GSEA of the DE Contigs Between Old and New Einkorn

The gene set enrichment analysis aims to determine the genes in a set that contribute most to differences between two phenotypes and/or experimental states. The GSEA of the contigs differentially expressed between the old and new einkorn reflect gene sets that were found to be over-presented (top) or under-presented (bottom) at control or stress temperatures. The results show that the gene sets related to “photosystem”, “photosystem II”, and “preribosome” were at the top of the enriched GOs list at the stress temperature between old and new einkorn. When plants are exposed to heat stress, the first mechanism affected is photosynthesis (Nievola, Carvalho, Carvalho, & Rodrigues, 2017). With the

increase in temperature, the photosynthetic reactions decrease, mainly due to the photochemical stage of photosynthesis (Nievola et al., 2017). Photosystems, specifically thylakoids found in antenna complexes of the photosystems, are responsible for converting light energy to ATP and NADPH by electron transport (Taiz, Zeiger, Møller, & Murphy, 2015). Especially photosystem II is very sensitive to temperature increases, and if the photosynthesis reactions decrease, it is most probably because of the inhibited photosystem II by the damaged thylakoids under heat stress (Mathur et al., 2014). Ristic, Bukovnik, & Prasad (2007) showed that the chlorophyll content of *T. aestivum* decreases because of the thylakoid membrane damaged by high temperature. So, regulation of these systems during heat stress is crucial for plant growth and yield.

Similarly, “photosynthesis – antenna proteins” was the second common KEGG pathway observed at stress temperature, but there were no DE contigs involved in this pathway between old and new einkorn, suggesting that their genomic responses in terms of antenna proteins have not changed due to climate change between sample collection years, 1995 vs. 2015. However, when the photosynthesis pathway – the first pathway affected by heat stress (Nievola et al., 2017), was investigated for the stress temperature, contigs related to photosystem I, photosynthetic electron transport, and cytochrome b6/f complex, which are all about electron transport for energy production (Yamori, Shikanai, & Makino, 2015; Mathur et al., 2014), were found to be DE between old and new einkorn. This suggests that the heat stress response of einkorn might have changed over the years due to climate change in terms of photosynthesis. On the other hand, for the populations kept at the control temperature during the heat stress experiment, contigs related to some other photosystem proteins of the photosynthesis pathway, as well as multiple light-harvesting chlorophyll proteins of the photosynthesis - antenna proteins pathway, were DE between old and new einkorn, suggesting that the general mechanism of einkorn wheat populations for generating energy during photosynthesis might have also changed because of the effects of climate change over the years.

In addition to the photosystem, the contigs related to preribosome were also enriched at stress temperature. Since protein synthesis consumes a lot of energy, under stress conditions, ribosomes act rapidly to synthesize required proteins to respond to stresses (Wu et al., 2021). On the other hand, ribosomal RNA synthesis and biogenesis of ribosomes take place in the nucleolus (Oeffinger, Dlakic, & Tollervey, 2004). Under heat stress, the protein content and organization of the nucleolus change, and some of the proteins in the nucleolus, newly synthesized or already present in the nucleolus, respond to stress conditions such as high temperatures (Kalinina, Makarova, Makhotenko, Love, & Taliansky,

2018). In plants, it is reported that the ribosomal proteins are differentially expressed under environmental stresses, resulting in plant ribosome variation (Wu et al., 2021; Moin et al., 2016; Martinez-Seidel et al., 2021).

In contrast to the stress temperature, at the control temperature, “actin cytoskeleton organization”, “filament organization”, and “protein-DNA complex assembly” were enriched between old and new einkorn, which are all associated with the main physiological processes, as expected. Actins, which are found as filaments in cells under normal conditions, play an important role in major cellular processes in plants, including growth, development, and cellular transport (Szymanski & Staiger, 2018; Breuer et al., 2014). Similarly, protein-DNA complexes are responsible for many fundamental processes in plants, such as replication, regulation of transcription, and maintaining chromatin stability and modification (Hudson & Ortlund, 2014; Dickey, Altschuler, & Wuttke, 2013).

The pathway that was seen to be enriched mostly at the stress temperature was diterpenoid biosynthesis. In this pathway, a series of biochemical reactions convert geranylgeranyl diphosphate into diterpenoids (Zi, Mafu, & Peters, 2014). Diterpenoids have various functions in plants, including adaptation to biotic and abiotic stresses (Ding et al., 2019; Lu et al., 2018). Gibberellin, as a phytohormone mostly produced as a response to heat stress (Verma, Ravindran, & Kumar, 2016), is one of the most known and abundant diterpenoids (Zi et al., 2014). Between the old and new einkorn at stress temperature, this pathway involved DE contigs associated with the GA20ox proteins, which are responsible for the production of gibberellins, as expected.

5.2.4. Heat Stress Related Contigs

Heat stress related contigs found in this study (by examining the top 100 contigs and the contigs that were DE between old and new einkorn) were mostly associated with HSP70, or small HSPs, including HSP16.9, HSP18.8, HSP26.4, and Hsa32, or HSFs including HSFA2a and HSFC2b. In addition to these HSPs and HSFs, a contig related to chaperone protein ClpB3 was found to be associated with heat stress within the top 100 contigs.

The expression of HSP70 and small HSPs generally increases in response to heat stress in plants (Vierling, 1991; Kumar et al., 2020b). HSP70 is a chaperon, which regulates the proper folding of proteins and, if necessary, the degradation of misfolded ones in order to avoid protein aggregation under

stress conditions (Sheffield, Shore, & Randall, 1990; Mayer & Bukau, 2005). In respect of wheat, 114 HSP70 proteins were recorded for the hexaploid wheat, *T. aestivum*, by Kumar et al. (2020b), and these were mostly induced by abiotic stresses, including heat stress.

Small HSPs have a molecular weight of between 15 and 30 kilodaltons (Bento et al., 2017), sometimes being up to 42 kilodaltons (Waters, 2012). Like HSP70, they prevent misfolding of proteins and aggregate formation, and they do it as precursors under stress conditions (Hilton et al., 2013), as they do not depend on ATP to function and they can bind denatured proteins with high affinity (Nakamoto & Vigh, 2006; McHaourab, Godar, & Stewart, 2009). In *T. aestivum*, 169 small HSPs were reported by Kumar et al. (2020b) as being mostly induced at high temperatures. HSP16.9, which was recorded for einkorn wheat in this study, was one of the transcription factors observed in wheat (Waters, Lee, & Vierling, 1996).

Within HSFs, HSFA activates transcription (Scharf, Berberich, Ebersberger, & Nover, 2012). In particular, HSFA2 induces transcription of HSPs under heat stress in *Arabidopsis* (Li et al., 2017). In terms of wheat, Xue, Sadat, Drenth, & McIntyre (2014) showed the up-regulation of HSFA2 by high temperature along with HSFA6, as the main players of *T. aestivum* for the up-regulation of the genes associated with HSPs under heat stress. Similarly, HSFC2a and HSFC2b were up-regulated when wheat plants were exposed to heat stress (Xue et al., 2014).

Casein lytic proteinase (Clp) belongs to the HSP100 family, and it is crucial in terms of the thermotolerance mechanisms and heat stress response of plants, as it is very effective in degrading protein aggregates produced due to various environmental stresses (Mishra & Grover, 2014). In plants, ClpB can be found in mitochondria, chloroplast, or cytoplasm (Singh, Singh, Mittal, & Grover, 2010). In this study, the ClpB found in einkorn wheat was mitochondrial. Likewise, Lee et al. (2007) found that mitochondrial ClpB was induced by heat stress in *Arabidopsis*, and Singh et al. (2010) showed the high expression of organelle ClpB in rice at high temperatures. In wheat (*T. aestivum*), Muthusamy, Dalal, Chinnusamy, & Bansal (2016) also showed excessive expression of the cytosolic ClpB3 in response to heat stress.

6. CONCLUSIONS AND RECOMMENDATIONS

SNP data obtained from ddRAD sequencing analysis showed a clear separation between three local wheat species. Within the einkorn group, there was also a distinction between populations from 1995 and 2015. A higher level of diversity was observed for ex-situ conservation (1995) than in-situ (2015), and this is most likely because of the genetic bottlenecks that new einkorn experienced. On the contrary, for emmer, no distinction was found in terms of the collection year. This suggests that emmer retained its diversity level without impacting by the changes in climate in the last 20 years. It might be due to the extra B genome it contains, which also increased the diversity of emmer. All in all, these findings support the idea of wheat species have different genetic diversity levels, and their adaptation capacities might depend on these diversity levels, as well as the ploidy levels, that they possess.

Since the first two decades of the 21st century included some of the hottest years recorded in history, in this study, einkorn collected in 1995 and 2015 from Kastamonu were also compared in terms of their response to heat stress. The 20 years of temperature increase in Kastamonu and the subsequent selection seem to have modified the transcriptomics responses of einkorn to specialize in and enrich genes involved in various pathways in order to cope with heat stress. In this regard, genes and gene sets involved in increased energy production and the production of stress-responsive proteins including HSPs, along with the genes associated with the regulation of physiological processes, membrane stability, and photosynthesis were differentially expressed (DE) between old and new einkorn under heat stress. Likewise, at the control temperature, the genes associated with energy production, regulation of transcription, and transmembrane transport were again DE between old and new einkorn, suggesting that these have a role under both normal and stress conditions. Additionally, genes involved in the main physiological processes were DE between old and new einkorn under normal conditions. These suggest that the general response of einkorn to maintain cellular and physiological processes also changed within 20 years. This kind of drastic change in response suggests rapid transcriptomic evolution just within a span of 20 years, a blink in evolutionary timescales. In this regard, this study adds to the literature in terms of climate change being an instigator of rapid evolutionary change.

In the future, as a more comprehensive investigation, whole genome sequencing can be performed for einkorn and emmer, especially to more directly investigate potential insertion, deletion and copy

number changes of genes in a comparative manner (between old and new populations), and potentially gather more direct evidence on the hints of rapid evolution observed in this study. In addition, RNA-seq analyses of emmer can give further insights into its adaptation mechanisms under changing climate by demonstrating its transcriptomic response to heat stress. Investigating the response of old and new populations of einkorn and emmer to different abiotic stresses like drought and various combinations of stresses (e.g., heat and drought) may also help better understand if the rapid evolution we observed extends to different domains of wheat's stress response. Investigation of the methylation patterns of the compared populations may also provide a better understanding of the processes involved in stress response of einkorn and emmer from an epigenetics perspective. Using even older populations than the one we used here may also provide a longer-term comparison of the impacts of climate change on the genetic diversity levels of wheat. Finally, the broader genetic functionality observed in the old einkorn, and the different genomic responses of old and new einkorn may help with the development of wheat varieties that might be viable against climate change, using the CRISPR-cas method, and contribute to solving the food crisis awaiting humanity in the coming years of the 21st century.

REFERENCES

- Adams, K. L., & Wendel, J. F. (2005). Polyploidy and genome evolution in plants. *Current Opinion in Plant Biology*, 8(2), 135–41.
- Akter, N., & Rafiqul Islam, M. (2017). Heat stress effects and management in wheat. A review. *Agronomy for Sustainable Development*, 37(5).
- Aliyev, R. T., Abbasov, M. A., & Mammadov, A. C. (2007). Genetic identification of diploid and tetraploid wheat species with RAPD markers. *Turkish Journal of Biology*, 31(3), 173–180.
- Allakhverdiev, S. I., Kreslavski, V. D., Klimov, V. V., Los, D. A., Carpentier, R., & Mohanty, P. (2008). Heat stress: an overview of molecular responses in photosynthesis. *Photosynthesis Research*, 98(1-3), 541-550.
- Al-Whaibi, M. H. (2011). Plant heat-shock proteins: a mini review. *Journal of King Saud University-Science*, 23(2), 139-150.
- Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. Retrieved from <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Aslam, M., Maqbool, M. A., & Cengiz, R. (2015). *Drought stress in maize (Zea mays L.): Effects, resistance mechanisms, global achievements and biological strategies for improvement*.
- Asthir, B. (2015). Protective mechanisms of heat tolerance in crop plants. *Journal of Plant Interactions*, 10(1), 202-210.
- Atak, M. (2017). Buğday ve Türkiye Buğday Çeşitleri. *Journal of Agricultural Faculty of Mustafa Kemal University*, 22(2), 71-88.

Avni, R., Nave, M., Barad, O., Baruch, K., Twardziok, S.O., Gundlach, H., ... & Distelfeld, A. (2017). Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science*, 357(6346), 93-97.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., & Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*, 3(10), e3376.

Bálint, A. F., Kovács, G., & Sutka, J. (2000). Origin and taxonomy of wheat in the light of recent research. *Acta Agronomica Hungarica*, 48(3), 301-313.

Banti, V., Mafessoni, F., Loreti, E., Alpi, A., & Perata, P. (2010). The heat-inducible transcription factor HsfA2 enhances anoxia tolerance in *Arabidopsis*. *Plant Physiology*, 152(3), 1471–1483.

Beebee, T. J. C. (1995). Amphibian breeding and climate change. *Nature*, 219-220.

Bento, M., Pereira, S. G., Viegas, W., & Silva, M. (2017). Durum wheat diversity for heat stress tolerance during inflorescence emergence is correlated to TdHSP101C expression in early developmental stages. *PLoS One*, 12(12), e0190085.

Bijma, J., Portner, H. O., Yesson, C., Rogers, A. D. (2013). Climate change and the oceans - What does the future hold? *Marine Pollution Bulletin*, 74(2), 495–505.

Bitá, C. E., & Gerats, T. (2013). Plant tolerance to high temperature in a changing environment: Scientific fundamentals and production of heat stress-tolerant crops. *Frontiers in Plant Science*, 4(273), 1–18.

Bolat, I., Kara, O., & Tok, E. (2017). Kastamonu, Karabük ve Bolu’da 1980-1999 ile 2000-2015 yılları arasındaki sıcaklık ve yağışın değişimi. *Bartın Orman Fakültesi Dergisi*, 19(1), 276-289.

Bolger, A. M., Lohse, M., Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-2120.

Bolin, B. (1986). How much CO₂ will remain in the atmosphere? In: Bolin, B., Doos, B. R., Jager, J., Warrick, R. A. (Eds.), *The Greenhouse Effect, Climatic Change and Ecosystems*, 93-156, Scope 29. Chichester: Wiley 18.

Brandolini, A., Volante, A., & Heun, M. (2016). Geographic differentiation of domesticated einkorn wheat and possible Neolithic migration routes. *Heredity*, 117(3), 135-41.

Brelsford, A., Dufresnes, C., & Perrin, N. (2016). High-density sex-specific linkage maps of a European tree frog (*Hyla arborea*) identify the sex chromosome without information on offspring sex. *Heredity*, 116(2), 177–181.

Breuer, D., Ivakov, A., Sampathkumar, A., Hollandt, F., Persson, S., & Nikoloski, Z. (2014). Quantitative analyses of the plant cytoskeleton reveal underlying organizational principles. *Journal of The Royal Society Interface*, 11(97), 20140362.

Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B., & Jones, P. D. (2006). Uncertainty estimates in regional and global observed temperature changes: A new dataset from 1850. *Journal of Geophysical Research*, 111(D12).

Buchfink, B., Reuter, K., & Drost, H. G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods*, 18, 366–368.

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: Building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics*, 1(3), 171–182.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular ecology*, 22(11), 3124–3140.

Cavalier-Smith, T. (1981). Eukaryote kingdoms: Seven or nine? *Bio Systems*, 14(3–4), 461–81.

Caverzan, A., Casassola, A., & Brammer, S. A. (2016). Antioxidant responses of wheat plants under stress. *Genetics and Molecular Biology*, 39, 1–6.

- Cenkci, S., Yildiz, M., Konuk, M., & Eren, Y. (2008). RAPD analyses of some wild *Triticum* L. and *Aegilops* L. species and wheat cultivars in Turkey. *Acta Biologica Cracoviensia Series Botanica*, 5, 35–42.
- Chafin, T. K., Martin, B. T., Mussmann, S. M., Douglas, M. R., & Douglas, M. E. (2018). FRAGMATIC: In silico locus prediction and its utility in optimizing ddRADseq projects. *Conservation Genetics Resources*, 10(3), 325–328.
- Challinor, A. J., Watson, J., Lobell, D. B., Howden, S. M., Smith, D. R., & Chhetri, N. (2014). A meta-analysis of crop yield under climate change and adaptation. *Nature Climate Change*, 4, 287–291.
- Charmet, G. (2011), Wheat domestication: Lessons for the future, *Comptes Rendus - Biologies*, 334(3), 212–220.
- Charng, Y. Y., Liu, H. C., Liu, N. Y., Chi, W. T., Wang, C. N., Chang, S. H., & Wang, T. T. (2007). A heat-inducible transcription factor, HsfA2, is required for extension of acquired thermotolerance in *Arabidopsis*. *Plant Physiology*, 143(1), 251–262.
- Chauhan, H., Khurana, N., Tyagi, A. K., Khurana, J. P., & Khurana, P. (2011). Identification and characterization of high temperature stress responsive genes in bread wheat (*Triticum aestivum* L.) and their regulation at various stages of development. *Plant molecular biology*, 75, 35–51.
- Ciais, P., Reichstein, M., Viovy, N., Granier, A., Ogée, J., Allard, V., ... & Valentini, R. (2005). Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature*, 437(7058), 529–533.
- Cline, W. (2007). Global warming and agriculture: Impact estimates by country. Washington, DC: Peterson Institute for International Economics.
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Wojciech-Szczesniak, M., Gaffney, D. J., Elo, L. L., Zhang, X., & Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, 17, 13.

Cooper, R. (2015). Re-discovering ancient wheat varieties as functional foods. *Journal of Traditional and Complementary Medicine*, 5(3), 138-43.

Cullum, R., Alder, O., & Hoodless, P. A. (2011). The next generation: Using new sequencing technologies to analyze gene regulation. *Respirology*, 16, 210-222.

Davis, M. B., Shaw, R. G., & Etterson, J.R. (2005). Evolutionary responses to changing climate. *Ecology*, 86, 1704-1714.

Demir, I., Kılıç, G., Coşkun, M., & Sümer, U. M. (2008). Türkiye’de maksimum, minimum ve ortalama hava sıcaklıkları ile yağış dizilerinde gözlenen değişiklikler ve eğilimler. *TMMOB İklim Değişimi Sempozyumu, Bildiriler Kitabı*, 69-84. TMMOB Meteoroloji Mühendisleri Odası, 13-14 Mart 2008, Ankara.

Demircan, M., Gürkan, H., Eskioglu, O., Arabacı, H., & Coşkun, M. (2017). Climate change projections for Turkey: Three models and two scenarios. *Turkish Journal of Water Science & Management*, 1(1).

Deng, P., Wang, M., Feng, K., Cui, L., Tong, W., Song, W., & Nie, X. (2016). Genome-wide characterization of microsatellites in Triticeae species: Abundance, distribution and evolution. *Scientific Reports*, 6(1), 32224.

Dickey, T. H., Altschuler, S. E., & Wuttke, D. S. (2013). Single-stranded DNA-binding proteins: Multiple domains for multiple functions. *Structure*, 21, 1074-1084.

Ding, Y., Murphy, K. M., Poretsky, E., Mafu, S., Yang, B., Char, S. N., Christensen, S. A., Saldivar, E., Wu, M., Wang, Q., Ji, L., Schmitz, R. J., Kremling, K. A., Buckler, E. S., Shen, Z., Briggs, S. P., Bohlmann, J., Sher, A., Castro-Falcon, G., Hughes, C. C., ... & Schmelz, E. A. (2019). Multiple genes recruited from hormone pathways partition maize diterpenoid defences. *Nature plants*, 5(10), 1043–1056.

Dissmeyer, N., & Schnittger, A. (Eds.) (2011). *Plant kinases – Methods and protocols*. Springer Science and Humana Press.

Dixon, J., Braun, H. J., Kosina, P., & Crouch, J. (Eds.) (2009). *Wheat Facts and Futures 2009*. Mexico, D.F.: CIMMYT.

Djanaguiraman, M., Boyle, D. L., Welte, R., Jagadish, S. V. K., Prasad, P. V. V. (2018) Decreased photosynthetic rate under high temperature in wheat is due to lipid desaturation, oxidation, acylation, and damage of organelles. *BMC Plant Biology*, 18(55).

Dubcovsky, J., & Dvorak, J. (2007). Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science*, 316, 1862–1866.

Dudic, D., Banović Đeri, B., Pajić, V., & Pavlovic-Lazetic, G. (2021). Demystification of RNAseq quality control. *Journal of Information Technology and Applications*, 22.

Dudu, H., & Çakmak, E.H. (2018). Climate change and agriculture: An integrated approach to evaluate economy-wide effects for Turkey. *Climate and Development*, 10(3), 275-288.

Duncan, R. F., & Hershey, J. W. (1989). Protein synthesis and protein phosphorylation during heat stress, recovery, and adaptation. *The Journal of Cell Biology*, 109(4), 1467-1481.

Dupont, F. M., & Altenbach, S. B. (2003). Molecular and biochemical impacts of environmental factors on wheat grain development and protein synthesis. *Journal of Cereal Science*, 38, 133-146.

Dupont, F. M., Hurkman, W. J., Vensel, W. H., Tanaka, C. K., Kothari, K. M., Chung, O. K., & Altenbach, S. B. (2006). Protein accumulation and composition in wheat grains: Effects of mineral nutrients and high temperature. *European Journal of Agronomy*, 25, 96-107.

Dvorak, J., & Zhang, H. B. (1990). Variation in repeated nucleotide sequences sheds light on the phylogeny of the wheat B and G genomes. *Proceedings of the National Academy of Sciences*, 87(24), 9640-4.

Dvorak, J., & Akhunov, E. (2005). Tempos of gene locus deletions and duplications and their relationship to recombination rate during diploid and polyploid evolution in the *Aegilops-Triticum* alliance. *Genetics*, 171, 323–332.

Eagles, H. A., & Lothrop, J. E. (1994). Highland maize from central Mexico – its origin, characteristics, and use in breeding programs. *Crop Science*, 34, 11–19.

Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), 359–361.

Elias, E. M., Steiger, K. D., & Cantrell, R. G. (1996). Evaluation of lines derived from wild emmer chromosome substitutions. II. Agronomic traits. *Crop Science*, 36(2), 228–233.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, 6, 1–9.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, 14(8), 2611–2620.

Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., ... & D'Eustachio, P. (2018). The reactome pathway knowledgebase. *Nucleic acids research*, 46(D1), D649–D655.

FAO, (2015). Wheat landraces in farmers' fields in Turkey: National survey, collection, and conservation, 2009-2014, by Mustafa Kan, Murat Küçükçongar, Mesut Keser, Alexey Morgounov, Hafiz Muminjanov, Fatih Özdemir, Calvin Qualset. FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS, Ankara.

Farooq, M., Bramley, H., Palta, J.A., & Siddique, K.H. (2011). Heat stress in wheat during reproductive and grain-filling phases. *Critical Reviews in Plant Sciences*, 30, 491–507.

Feder, M.E., & Hoffman, G.E. (1999). Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. *Annual Review of Physiology*, 61, 243–282.

Fox, S.E., Geniza, M., Hanumappa, M., Naithani, S., Sullivan, C., Preece, J., ... & Jaiswal, P. (2014). De novo transcriptome assembly and analyses of gene expression during photomorphogenesis in diploid wheat *Triticum monococcum*. *PLoS One*, 9, 5.

Francis, R. M. (2017). Pophelper: An R package and web app to analyze and visualize population structure. *Molecular Ecology Resources*, 17, 27–32.

Frankham, R., Ballou, J. D., & Briscoe, D. A. (2002). Introduction to Conservation Genetics. Cambridge: Cambridge University Press.

Frankham, R., Ballou, J.D., Ralls, K., Eldridge, M.D.B., Dudash, M.R., Fenster, C.B., Lacy, R.C., & Sunnuck, P. (2017). Genetic Management of Fragmented Animal and Plant Populations. *Oxford University Press*, 65-86.

Fu, Y. B., Peterson, G. W., Horbach, C., Konkin, D. J., Beiles, A., & Nevo, E. (2019). Elevated mutation and selection in wild emmer wheat in response to 28 years of global warming. *Proceedings of the National Academy of Sciences of the United States of America*, 116(40), 20002–20008.

Gershater, M. C., & Edwards, R. (2007). Regulating biological activity in plants with carboxylesterases. *Plant Science*, 173(6), 579-588.

Giannakopoulos, C., Le Sager, P., Bindi, M., Moriondo, M., Kostopoulou, E., & Goodess, C. M. (2009). Climatic changes and associated impacts in the Mediterranean resulting from a 2 °C global warming. *Global and Planetary Change*, 68, 209-224.

Gill, R. A., Ahmar, S., Ali, B., Saleem, M. H., Khan, M. U., Zhou, W., & Liu, S. (2021). The role of membrane transporters in plant growth and development, and abiotic stress tolerance. *International Journal of Molecular Science*, 22(23), 12792.

Goncharov, N. P. (2011). Genus *Triticum* L. taxonomy: The present and the future. *Plant Systematics and Evolution*, 295, 1–11.

Gökgöl, M. (1935). Türkiye buğdayları - Tom: I. İstanbul: Tarım Bakanlığı İstanbul-Yeşilköy Tohum Yetiştirme Enstitüsü. *Devlet Press*, 7, 436.

Götz, S., Garcia-Gomez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talon, M., Dopazo, J., & Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic acids research*, 36(10), 3420-35.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... & Regev, A. (2013). Trinity: Reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology*, 29(7), 644-652.

Grigorova, B., Vaseva, I. I., Demirevska, K., & Feller, U. (2011). Expression of selected heat shock proteins after individually applied and combined drought and heat stress. *Acta Physiologiae Plantarum*, 33(5), 2041-2049.

Gulli, M., Corradi, M., Rampino, P., Marmioli, N., Perrotta, C. (2007). Four members of the HSP101 gene family are differently regulated in *Triticum durum* Desf. *FEBS Letters*, 581, 4841–4849.

Guo, M., Liu, J. H., Ma, X., Luo, D. X., Gong, Z. H., & Lu, M. H. (2016). The plant heat stress transcription factors (HSFs): Structure, regulation, and function in response to abiotic stresses. *Frontiers in Plant Science*, 7.

Güner, A. (2012). *Türkiye bitkileri listesi*. Nezahat Gökyiğit Botanik Bahçesi Yayınları, Flora Dizisi, ISBN: 978-605- 60425-7-7, İstanbul.

Gürçan, K., Demirel, F., Tekin, M., Demirel, S., & Akar, T. (2017). Molecular and agro-morphological characterization of ancient wheat landraces of Turkey. *BMC Plant Biology*, 17, 171.

Harlan, J. R., & Zohary, D. (1966). Distribution of wild wheats and barley. *Science*, 153, 1074–1080.

Hasanuzzaman, M., Nahar, K., Alam, M. M., Roychowdhury, R., & Fujita, M. (2013). Physiological, biochemical, and molecular mechanisms of heat stress tolerance in plants. *International journal of molecular sciences*, 14(5), 9643–9684.

Hasanuzzaman, M., Bhuyan, M. B., Zulfiqar, F., Raza, A., Mohsin, S. M., Mahmud, J. A., Fujita, M., Fotopoulos, V. (2020). Reactive oxygen species and antioxidant defense in plants under abiotic stress: Revisiting the crucial role of a universal defense regulator. *Antioxidants*, 9(8), 681.

Hendek Ertop, M., & Atasoy, R. (2019). Comparison of physicochemical attributes of einkorn wheat (*Triticum monococcum*) and durum wheat (*Triticum durum*) and evaluation of morphological properties using scanning electron microscopy and image analysis. *Journal of Agricultural Sciences*, 25(1), 93-99.

Herrmann, H. A., Schwartz, J. M., & Johnson, G. N. (2019) Metabolic acclimation - A key to enhancing photosynthesis in changing environments? *Journal of Experimental Botany*, 70(12), 3043– 3056.

Heun, M., Schafer-Pregl, R., Klawan, D., Castagna, R., Accerbi, M., Borghi B., & Salamini, F. (1997). Site of einkorn wheat domestication identified by DNA fingerprinting. *Science*, 278, 1321-1314.

Hilton, G. R., Lioe, H., Stengel, F., Baldwin, A. J., & Benesch, J. L. P. (2013). Small heat-shock proteins: Paramedics of the cell. *Topics in Current Chemistry*, 328, 69–98.

Hoffmann, A. A., & Willi, Y. (2008). Detecting genetic responses to environmental change. *Nature Reviews: Genetics*, 9, 421–432.

Holsinger, K., & Weir, B. (2009). Genetics in geographically structured populations: Defining, estimating and interpreting FST. *Nature Reviews Genetics*, 10, 639–650.

Hopper, D. A., Stutte, G. W., McCormack, A., Barta, D. J., Heins, R. D., Erwin, J. E., & Tibbitts, T. W. (1997). Crop growth requirements. *Plant Growth Chamber Handbook*. North Central Regional Research Publication, 340, 217-225.

Hua, Y., Sansenya, S., Saetang, C., Wakuta, S., & Ketudat Cairns, J. R. (2013). Enzymatic and structural characterization of hydrolysis of gibberellin A4 glucosyl ester by a rice β -d-glucosidase. *Archives of Biochemistry and Biophysics*, 537(1), 39-48.

Huang, L.-Z., Zhou, M., Ding, Y.-F., & Zhu, C. (2022). Gene networks involved in plant heat stress response and tolerance. *International Journal of Molecular Sciences*, 23(19), 11970.

Hudson, W.H., & Ortlund, E. A. (2014). The structure, function and evolution of proteins that bind DNA and RNA. *Nature Reviews Molecular Cell Biology*, 15, 749-760.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., von Mering, C., & Bork, P. (2019). eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic acids research*, 47(D1), D309–D314.

Hurkman, W. J., Vensel, W. H., Tanaka, C. K., Whitehand, L., & Altenbach, S. B. (2009). Effect of high temperature on albumin and globulin accumulation in the endosperm proteome of the developing wheat grain. *Journal of Cereal Science*, 49, 12–23.

Ilker, A., & Terzi, O. (2021). Sicaklik verilerinin trend analizi: Cankiri ve Kastamonu örneği. *Mühendislik Bilimleri ve Tasarım Dergisi*, 9 (4), 1339-1347.

Inouye, D.W., Barr, B., Armitage, K.B., & Inouye, B.D. (2000). Climate change is affecting altitudinal migrants and hibernating species. *PNAS*, 97(4), 1630-1633.

IPCC, 2022: Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Pörtner, H. O., Roberts, D. C., Tignor, M., Poloczanska, E. S., Mintenbeck, K., Alegría, A., Craig, M., Langsdorf, S., Löschke, S., Möller, V., Okem, A., Rama, B. (Eds.). Cambridge University Press, Cambridge, UK and New York, NY, USA, 3056 pp.

Jaradat, A. A. (2013). Wheat landraces: A mini review. *Emirates Journal of Food and Agriculture*, 25(1), 20-29.

Jarvis, A., Lane, A., & Hijmans, R.J. (2008). The effect of climate change on crop wild relatives. *Agriculture, Ecosystems and Environment*, 126(1–2), 13–23.

Jing, H. C., Kornukhin, D., Kanyuka, K., Orford, S., Zlatska, A., Mitrofanova, O.P., Koebner, R., & Hammond-Kosack, K. (2007). Identification of variation in adaptively important traits and genome-wide analysis of trait-marker associations in *Triticum monococcum*. *Journal of Experimental Botany*, 58(13), 3749-3764.

Jing, H. C., Bayon, C., Kanyuka, K., Berry, S., Wenzl, P., & Huttner, E. (2009). DArT markers: Diversity analyses, genomes comparison, mapping and integration with SSR markers in *Triticum monococcum*. *BMC Genomics*, 10, 458.

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27, 3070–3071.

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics (Oxford, England)*, 30(9), 1236–1240.

Johnson, B. L., & Dhaliwal, H. S. (1976). Reproductive isolation of *Triticum boeoticum* and *Triticum urartu* and the origin of the tetraploid wheats. *American Journal of Botany*, 63(8), 1088–1094.

Jump, A. S., & Peñuelas, J. (2005). Running to stand still: Adaptation and the response of plants to rapid climate change. *Ecology Letters*, 8, 1010-1020.

Jump, A.S., Marchant, R., & Peñuelas, J. (2009). Environmental change and the option value of genetic diversity. *Trends in Plant Science*, 14(1), 51–58.

Kailla, A. (1956). Phosphorus in virgin peat soils. *Agricultural and Food Science*, 28(1), 142-167.

Kalefetoğlu, T., & Ekmekci, Y. (2005). The effects of drought on plants and tolerance mechanisms. *Gazi University Journal of Science*, 18(4), 723-740.

Kalinina, N. O., Makarova, S., Makhotenko, A., Love, A. J., & Taliansky, M. (2018). The multiple functions of the nucleolus in plant development, disease and stress responses. *Frontiers in Plant Science*, 9.

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1), 27-30.

Kaplan, F., & Guy, C. L. (2004) Beta-amylase induction and the protective role of maltose during temperature shock. *Plant Physiology*, 135(3), 1674–1684.

Kapur, B., Topaloğlu, F., Ozfidaner, M., Koç, M. (2007). Çukurova bölgesinde küresel iklim değişikliği ve buğday verimliliği üzerine etkilerine genel bir yaklaşım. Küresel İklim Değişimi ve Çevresel Etkileri Sempozyumu, Konya, 18-20 October 2007.

Karagöz, A. (2003). Plant genetic resources conservation in Turkey. *Acta Horticulturae*, 598, 17-25.

Karagöz, A. (2014). Wheat landraces of Turkey. *Emirates Journal of Food and Agriculture*, 26(2), 149-156.

Karagöz, A., & Zencirci, N. (2005). Variation in wheat (*Triticum* spp.) landraces from different altitudes of three regions of Turkey. *Genetics Resources and Crop Evaluation*, 52, 6775- 785.

Kihara, H. (1944). Discovery of the DD analyzer, one of the ancestors of vulgare wheats. *Agriculture & Horticulture*, 19, 889–890.

Klepsatel, P., Gáliková, M., Xu, Y., & Kühnlein, R. P. (2016). Thermal stress depletes energy reserves in *Drosophila*. *Scientific reports*, 6(1), 33667.

Kotak, S., Larkindale, J., Lee, U., von Koskull-Döring, P., Vierling, E., & Scharf, K. D. (2007). Complexity of the heat stress response in plants. *Current Opinion in Plant Biology*, 10(3), 310-316.

Kumar, R. R., Goswami, S., Sharma, S. K., Kala, Y. K., Rai, G. K., Mishra, D. C., Grover, M., Singh, G., Pathak, H., Rai, A., Chinnusamy, V., & Rai, R. (2015). Harnessing next generation sequencing in climate change: RNA-Seq analysis of heat stress-responsive genes in wheat (*Triticum aestivum* L.). *OMICS A Journal of Integrative Biology*, 19(10), 632–647.

Kumar, D., Chhokar, V., Sheoran, S., Singh, R., Sharma, P., Jaiswal, S., Iquebal, M. A., Jaiswar, A., Jaisri, J., Angadi, U. B., Rai, A., Singh, G. P., Kumar, D., & Tiwari, R. (2020a). Characterization of genetic diversity and population structure in wheat using array based SNP markers. *Molecular biology reports*, 47(1), 293–306.

Kumar, A., Sharma, S., Chunduri, V., Kaur, A., Kaur, S., Malhotra, N., Kumar, A., Kapoor, P., Kumari, A., Kaur, J., Sonah, H., & Garg, M. (2020b). Genome-wide identification and characterization of heat shock protein family reveals role in development and stress conditions in *Triticum aestivum* L. *Scientific Reports*, 10(1), 7858.

Kumar, R. R., Dubey, K., Arora, K., Dalal, M., Rai, G. K., Mishra, D., Chaturvedi, K. K., Rai, A., Kumar, S. N., Singh, B., Chinnusamy, V., & Praveen, S. (2021). Characterizing the putative mitogen-activated protein kinase (MAPK) and their protective role in oxidative stress tolerance and carbon assimilation in wheat under terminal heat stress. *Biotechnology reports (Amsterdam, Netherlands)*, 29, e00597.

Külahoğlu, C., & Bräutigam, A. (2014). Quantitative transcriptome analysis using RNA-seq. In: Staiger, D. (Ed.) *Plant Circadian Networks*. Methods in Molecular Biology, 1158. Humana Press, New York, NY.

Laino, P., Shelton, D., Finnie, C., De Leonardis, A.M., Mastrangelo, A.M., Svensson, B., Lafiandra, D., & Masci, S. (2010). Comparative proteome analysis of metabolic proteins from seeds of durum wheat (cv. Svevo) subjected to heat stress. *Proteomics*, 10(12), 2359-68.

Larkindale, J., & Huang, B. (2004). Thermotolerance and antioxidant systems in *Agrostis stolonifera*: Involvement of salicylic acid, abscisic acid, calcium, hydrogen peroxide, and ethylene. *Journal of Plant Physiology*, 161, 405-413.

- Larkindale, J., Hall, J.D., Knight, M.R., & Vierling, E. (2005). Heat stress phenotypes of *Arabidopsis* mutants implicate multiple signaling pathways in the acquisition of thermotolerance. *Plant Physiology*, 138, 882-897.
- Law, C. W., Alhamdoosh, M., Su, S., Dong, X., Tian, L., Smyth, G. K., & Ritchie, M. E. (2016). RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Research*, 5.
- Lawlor, D. W., & Mitchell, R. A. (2000). Crop ecosystem responses to climatic change: Wheat. *Climate Change and Global Crop Productivity*, 57, 80.
- Lee, U., Wie, C., Escobar, M., Williams, B., Hong, S., & Vierling, E. (2005). Genetic analysis reveals domain interactions of *Arabidopsis* HSP100/ClpB and cooperation with the small heat shock protein chaperone system. *Plant Cell*, 17, 559–571.
- Lee, K. H., Piao, H. L., Kim, H. Y., Choi, S. M., Jiang, F., Hartung, W., Hwang, I., Kwak, J. M., Lee, I. J., & Hwang, I. (2006). Activation of glucosidase via stress-induced polymerization rapidly increases active pools of abscisic acid. *Cell*, 126(6), 1109-1120.
- Lee, U., Rioflorido, I., Hong, S. W., Larkindale, J., Waters, E. R., & Vierling, E. (2007). The *Arabidopsis* ClpB/Hsp100 family of proteins: chaperones for stress and chloroplast development. *The Plant Journal: for cell and molecular biology*, 49(1), 115–127.
- Lemieux, C., Otis, C., Turmel, M. (2000). Ancestral chloroplast genome in *Mesostigma viride* reveals an early branch of green plant evolution. *Nature*, 403(6770), 649–652.
- Lev-Yadun, A., Gopher, A., & Abbo, S. (2000). The cradle of agriculture. *Science*, 288, 1602-1603.
- Li, W., & Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22, 1658-1659.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25, 2078–9.

- Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv: Genomics.
- Li, B., Fillmore, N., Bai, Y., Collins, M., Thomson, J. A., Stewart, R., & Colin, N. (2014). Evaluation of de novo transcriptome assemblies from RNA-Seq data. *Genome Biology*, 15, 553.
- Li, X. D., Wang, X. L., Cai, Y. M., Wu, J. H., Mo, B. T., & Yu, E. R. (2017). *Arabidopsis* heat stress transcription factors A2 (HSFA2) and A3 (HSFA3) function in the same heat regulation pathway. *Acta Physiologiae Plantarum*, 39, 1-9.
- Li, N., Euring, D., Cha, J.Y., Lin, Z., Lu, M., Huang, L.-J., & Kim, W. Y. (2021). Plant hormone-mediated regulation of heat tolerance in response to global climate change. *Frontiers in Plant Science*, 11.
- Liao, Y., Smyth, G. K., & Shi, W. (2019). The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Research*, 47(8), e47.
- Liu, H. T., Li, B., Shang, Z. L., Li, X. Z., Mu, R. L., Sun, D. Y., & Zhou, R. G. (2003). Calmodulin is involved in heat shock signal transduction in wheat. *Plant physiology*, 132(3), 1186-1195.
- Liu, Z., Xin, M., Qin, J., Peng, H., Ni, Z., Yao, Y., & Sun, Q. (2015). Temporal transcriptome profiling reveals expression partitioning of homologous genes contributing to heat and drought acclimation in wheat (*Triticum aestivum* L.). *BMC Plant Biology*, 15, 152.
- Liu, B., Asseng, S., Muller, C., Ewert, F., & Elliott, J. (2016). Similar estimates of temperature impacts on global wheat yield by three independent methods. *Nature Climate Change*, 1130–6.
- Liu, Z., Qin, J., Tian, X., Xu, S., Wang, Y., Li, H., Wang, X., Peng, H., Yao, Y., Hu, Z., Ni, Z., Xin, M., & Sun, Q. (2018). Global profiling of alternative splicing landscape responsive to drought, heat and their combination in wheat (*Triticum aestivum* L.). *Plant Biotechnology Journal*, 16, 714-726.

- Lohmann, C., Eggers-Schumacher, G., Wunderlich, M., & Schoffl, F. (2004). Two different heat shock transcription factors regulate immediate early expression of stress genes in *Arabidopsis*. *Molecular Genetics Genomics*, 271, 11-21.
- Lu, X., Zhang, J., Brown, B., Li, R., Rodríguez-Romero, J., Berasategui, A., Liu, B., Xu, M., Luo, D., Pan, Z., Baerson, S. R., Gershenzon, J., Li, Z., Sesma, A., Yang, B., & Peters, R. J. (2018). Inferring Roles in Defense from Metabolic Allocation of Rice Diterpenoids. *The Plant cell*, 30(5), 1119–1131.
- Maestri, E., Klueva, N., Perrotta, C., Gulli, M., Nguyen, H. T., & Marmioli, N. (2002). Molecular genetics of heat tolerance and heat shock proteins in cereals. *Plant Molecular Biology*, 48(5-6), 667- 681.
- Mantzavinou, A., Babal, P. J., & Kaltsikes, P. J. (2005). Estimating genetic diversity in Greek durum wheat landraces with RAPD markers. *Australian Journal of Agricultural Research*, 56, 1355–1364.
- Martinez-Seidel, F., Suwanchaikasem, P., Nie, S., Leeming, M. G., Pereira Firmino, A. A., Williamson, N. A., ... & Boughton, B. A. (2021). Membrane-enriched proteomics link ribosome accumulation and proteome reprogramming with cold acclimation in barley root meristems. *Frontiers in Plant Science*, 12, 656683.
- Mathur, S., Agrawal, D., & Jajoo, A. (2014). Photosynthesis: Response to high temperature stress. *Journal of Photochemistry and Photobiology. B, Biology*, 137, 116–126.
- Mayer, M. P., & Bukau, B. (2005). Hsp70 chaperones: Cellular functions and molecular mechanism. *Cellular and Molecular Life Sciences*, 62, 670–684.
- McCarthy, D., J., Chen, Y., & Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, 40(10), 4288-4297.
- McFadden, E. S., & Sears, E. R. (1946). The origin of *Triticum spelta* and its free-threshing hexaploid relatives. *Journal of Heredity*, 37(81), 107.
- McHaourab, H. S., Godar, J. A., & Stewart, P. L. (2009). Structure and mechanism of protein stability sensors: Chaperone activity of small heat shock proteins. *Biochemistry*, 48, 3828–3837.

- Medini, M., Hamza, S., Rebai, A., & Baum, M. (2005). Analysis of genetic diversity in Tunisian durum wheat cultivars and related wild species by SSR and AFLP markers. *Genetic Resources and Crop Evolution*, 52, 21–31.
- Menzel, A., & Estrella, N. (2001). "Fingerprints" of climate change: Adapted behaviour and shifting species ranges. Walther, G.R., Burga, C.A., Edwards, P.J. (Eds.), Springer US, 123-137.
- MET, 2011. Climate: observations, projections and impacts (Turkey). Met Office, United Kingdom.
- Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nature Reviews*, 11, 31.
- Michigan State University, n. d. College of Natural Science Research Technology Support Facility. FastQC tutorial & FAQ. Retrieved from <https://rtsf.natsci.msu.edu/genomics/tech-notes/fastqc-tutorial-and-faq/> (accessed January 2023).
- Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., & Johnson, E. A., (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*, 17, 240-248.
- Miller, J. R., Koren, S., & Sutton, G. (2010). Assembly algorithms for next-generation sequencing data. *Genomics*, 95, 315–327.
- Mimura, N. (2013). Sea-level rise caused by climate change and its implications for society. *In Proceedings of the Japan Academy Series B: Physical and Biological Sciences*, 89(7), 281–301.
- Mishra, R. C., & Grover, A. (2016) ClpB/Hsp100 proteins and heat stress tolerance in plants. *Critical Reviews in Biotechnology*, 36(5), 862-874.
- Mitreva, Z., Krasteva, V., & Pankov, V. (2016). Method for evaluation of soil conditions suitable for growing emmer and einkorn wheat. *Bulgarian Journal of Soil Science*, 1, 2.
- Mittler, R., Finka, A., & Goloubinoff, P. (2012) How do plants feel the heat? *Trends in Biochemical Sciences*, 37(3), 118– 125.

- Moin, M., Bakshi, A., Saha, A., Udaya Kumar, M., Reddy, A. R., Rao, K. V., ... & Kirti, P. B. (2016). Activation tagging in indica rice identifies ribosomal proteins as potential targets for manipulation of water-use efficiency and abiotic stress tolerance in plants. *Plant, cell & environment*, 39(11), 2440-2459.
- Moore, A. M. T., Hillman, G. C., & Legge, A. J. (2000). Village on the Euphrates, from foraging to farming at Abu Hureyra. Oxford University Press, Oxford.
- Moore, F. C., Baldos, U., Hertel, T., & Diaz, D. (2017). New science of climate change impacts on agriculture implies higher social cost of carbon. *Nature Communications*, 8(1), 1607.
- Mori, N., Liu, Y. G., & Tsunewaki, K. (1995). Wheat phylogeny determined by RFLP analysis of nuclear DNA. 2. Wild tetraploid wheats. *Theoretical and Applied Genetics*, 90, 129–134.
- Muthusamy, S. K., Dalal, M., Chinnusamy, V., & Bansal, K. C. (2016). Differential regulation of genes coding for organelle and cytosolic ClpATPases under biotic and abiotic stresses in wheat. *Frontiers in Plant Science*, 7.
- Mutz, K. O., Heilkenbrinker, A., Lönne, M., Walter, J. G., & Stahl, F. (2013). Transcriptome analysis using next-generation sequencing. *Current Opinion in Biotechnology*, 24, 22-30.
- Nagaoka, T., & Ogihara, Y. (1997). Applicability of inter-simple sequence repeat polymorphisms in wheat for use as DNA markers in comparison to RFLP and RAPD markers. *Theoretical and Applied Genetics*, 94, 597–602.
- Naithani, S., Gupta, P., Preece, J., D'Eustachio, P., Elser, J. L., Garg, P., Dikeman, D. A., Kiff, J., Cook, J., Olson, A., Wei, S., Tello-Ruiz, M. K., Mundo, A. F., Munoz-Pomer, A., Mohammed, S., Cheng, T., Bolton, E., Papatheodorou, I., Stein, L., Ware, D., ... & Jaiswal, P. (2020). Plant Reactome: a knowledgebase and resource for comparative pathway analysis. *Nucleic acids research*, 48(D1), D1093–D1103.
- Nakamoto, H., & Vigh, L. (2007). The small heat shock proteins and their clients. *Cellular and Molecular Life Sciences*, 64, 294–306.

Narayanan, S., Tamura, P. J., Roth, M. R., Prasad, P. V. V., & Welti, R. (2016) Wheat leaf lipids during heat stress: I. High day and night temperatures result in major lipid alterations. *Plant Cell Environment*, 39, 787–803.

Nesbitt, M., & Samuel, D. (1996). From staple crop to extinction? The Archaeology and History of the Hulled Wheats. *Hulled Wheats*, 4, 41-100.

Newton, A. C., Johnson, S. N., Lyon, G. D., Hopkins, D. W., & Gregory, P. J. (2008). Impacts of climate change on arable crops – adaptation challenges. In Proceedings of the Crop Protection in Northern Britain Conference 2008. Dundee, UK, The Association for Crop Protection in Northern Britain.

Nevo, E., Fu, Y. B., Pavlicek, T., Khalifa, S., Tavasi, M., & Beiles, A. (2012). Evolution of wild cereals during 28 years of global warming in Israel. *Proceedings of the National Academy of Sciences of the United States of America*, 109(9), 3412–3415.

Nievola, C. C., Carvalho, C. P., Carvalho, V., & Rodrigues, E. (2017). Rapid responses of plants to temperature changes. *Temperature (Austin, TX)*, 4(4), 371-405.

NOAA National Centers for Environmental Information, State of the Climate: Monthly Global Climate Report for Annual 2020.

Oeffinger, M., Dlakic, M., & Tollervey, D. (2004). A pre-ribosome-associated HEAT-repeat protein is required for export of both ribosomal subunits. *Genes & Development*, 18(2), 196-209.

Ogihara, Y., & Tsunewaki, K. (1988). Diversity and evolution of chloroplast DNA in *Triticum* and *Aegilops* as revealed by restriction fragment analysis. *Theoretical & Applied Genetics*, 76, 321–332.

Ohama, N., Sato, H., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2017). Transcriptional regulatory network of plant heat stress response. *Trends in Plant Science*, 22(1), 53-65.

OmicsBox – Bioinformatics Made Easy, BioBam Bioinformatics, March 3, 2019, Retrieved from <https://www.biobam.com/omicsbox>

Ortiz, R. (2011). Agrobiodiversity management and climate change. In *Agrobiodiversity Management for Food Security: a Critical Review*, Lenné, J. M., Wood, D. (Eds.). CAB International, Wallingford, UK, 189–211.

Özberk, I., Atay, S., Altay, F., Cabi, E., Ozkan, H., & Atli, A. (2016). Türkiye'nin Buğday Atlası. Kalem, S., Dural, B., (Eds.), WWF Türkiye.

Özdoğan, M. (2011). Modeling the impacts of climate change on wheat yields in northwestern Turkey. *Agriculture, Ecosystems & Environment*, 141(1–2), 1–12.

Öztürk, T., Türkeş, M., & Kurnaz, M. L. (2011). Regcm4.3.5 iklim modeli benzetimleri kullanılarak Türkiye'nin gelecek hava sıcaklığı ve yağış klimatolojilerindeki değişikliklerin çözümlenmesi. *Ege Coğrafya Dergisi*, 20, 17-27.

Parchman, T., Gompert, Z., Mudge, J., Schilkey, F.D., Benkman, C.W., & Buerkle, C.A. (2012). Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology*, 21, 2991–3005.

Paris, J. R., Stevens, J. R., & Catchen, J. M. (2017). Lost in parameter space: A road map for Stacks. *Methods in Ecology and Evolution*, 8(10), 1360–1373.

Parmesan, C., & Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, 421, 37–42.

Parmesan, C. (2006). Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology, Evolution and Systematics*, 37, 637–669.

Perdomo, J. A., Conesa, M. À., Medrano, H., Ribas- Carbó, M., & Galmés, J. (2015). Effects of long-term individual and combined water and temperature stress on the growth of rice, wheat and maize: Relationship with morphological and physiological acclimation. *Physiologia Plantarum*, 155, 149- 165.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*, 7(5), e37135.

Poland, J. A., Brown, P. J., Sorrells, M. E., & Jannink, J. L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One*, 7, e32253.

Porter, J., & Gawith, M. (1999). Temperatures and the growth and development of wheat: A review. *European Journal of Agronomy*, 10, 23–36.

Pounds, J. A., Fogden, M. P. L., & Campbell, J. H. (1999). Biological response to climate change on a tropical mountain. *Nature*, 398.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics Society of America*, 155, 945-959.

Qin, D., Wu, H., Peng, H., Yao, Y., Ni, Z., Li, Z., Zhou, C., & Sun, Q. (2008). Heat stress-responsive transcriptome analysis in heat susceptible and tolerant wheat (*Triticum aestivum* L.) by using wheat genome array. *BMC Genomics*, 9, 432.

Queitsch, C., Hong, S., Vierling, E., & Lindquist, S. (2000). Heat shock protein 101 plays a crucial role in thermotolerance in *Arabidopsis*. *Plant Cell*, 12, 479–492.

R Core Team, 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ramírez-González, R. H., Borrill, P., Lang, D., Harrington, S. A., Brinton, J., Venturini, L., Davey, M., Jacobs, J., van Ex, F., Pasha, A., Khedikar, Y., Robinson, S. J., Cory, A. T., Florio, T., Concia, L., Juery, C., Schoonbeek, H., Steuernagel, B., Xiang, D., Ridout, C. J., ... & Uauy, C. (2018). The transcriptional landscape of polyploid wheat. *Science (New York, N.Y.)*, 361(6403), eaar6089.

- Rampino, P., Mita, G., Pataleo, S., De Pascali, M., Di Fonzo, N., & Perrotta, C. (2009). Acquisition of thermotolerance and HSP gene expression in durum wheat (*Triticum durum* Desf.) cultivars. *Environmental and Experimental Botany*, 66(2), 257-264.
- Ravenscroft, C. H., Whitlock, R., & Fridley, J. D. (2015). Rapid genetic divergence in response to 15 years of simulated climate change. *Global Change Biology*, 21(11), 4165–4176.
- Reis, S. P., Lima, A. M., & de Souza, C. R. B. (2012). Recent molecular advances on downstream plant responses to abiotic stress. *International Journal of Molecular Science*, 13, 8628–8647.
- Ren, J., Sun, D., Chen, L., You, F. M., Wang, J., Peng, Y., Nevo, E., Sun, D., Luo, M. C., & Peng, J. (2013). Genetic diversity revealed by single nucleotide polymorphism markers in a worldwide germplasm collection of durum wheat. *International Journal of Molecular Sciences*, 14(4), 7061–7088.
- Ristic, Z., Bukovnik, U., & Prasad, P. V. V. (2007). Correlation between heat stability of thylakoid membranes and loss of chlorophyll in winter wheat under heat stress. *Crop Science*, 47(5), 2067–2073.
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7), e47.
- Rivera-Colón, A. G., & Catchen, J. (2022). Population genomics analysis with RAD, reprised: Stacks 2. *Methods of Molecular Biology*, 2498, 99-149.
- Rizzo, J. M., & Buck, M. J. (2012). Key principles and clinical applications of "next-generation" DNA sequencing. *Cancer Prevention Research (Philadelphia, PA)*, 5(7), 887-900.
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139-140.
- Root, T. L., Price, J. T., Hall, K. R., Schneider, S. H., Rosenzweig, C., & Pounds, J. A. (2003). Fingerprints of global warming on wild animals and plants, *Nature*, 421, 57-60.

Roussel, V., Leisova, L., Exbrayat, F., Stehno, Z., & Balfourier, F. (2005). SSR allelic diversity changes in 480 European bread wheat varieties released from 1840 to 2000. *Theoretical and Applied Genetics*, 111(1), 162–170.

Sağlam, M. T., Bellitürk, K., & Kutlu, A. (2009). The comparison of the some physical and chemical properties of the peat samples sold in the market with the name of pot soil. *ADU Ziraat Fakultesi Dergisi*, 6(2), 25-31.

Saidi, Y., Peter, M., Finka, A., Cicekli, C., Vigh, L., & Goloubinoff, P. (2010) Membrane lipid composition affects plant heat sensing and modulates Ca^{2+} -dependent heat shock response. *Plant Signaling & Behavior*, 5(12), 1530-1533.

Salamini, F., Özkan, H., Brandolini, A., Schäfer-Pregl, R., & Martin, W. (2002). Genetics and geography of wild cereal domestication in the near east. *Nature reviews. Genetics*, 3(6), 429–441.

Salzberg, S. L., Phillippy, A. M., Zimin, A., Puiu, D., Magoc, T., Koren, S., Treangen, T. J., Schatz, M. C., Delcher, A. L., Roberts, M., Marçais, G., Pop, M., & Yorke, J. A. (2012). GAGE: A critical evaluation of genome assemblies and assembly algorithms. *Genome Research*, 22, 557–567.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12), 5463-5467.

Sanmiya, K., Suzuki, K., Egawa, Y., & Shono, M. (2004). Mitochondrial small heat-shock protein enhances thermotolerance in tobacco plants. *FEBS Letters*, 557(1-3), 265-268.

Scharf, K. D., Berberich, T., Ebersberger, I., & Nover, L. (2012). The plant heat stress transcription factor (Hsf) family: structure, function and evolution. *Biochimica et biophysica acta*, 1819(2), 104–119.

Schurch, N. J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., ... & Barton, G. J. (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA*, 22(6), 839-851.

Seifolahpour, B., Bahraminejad, S., & Cheghamirza, K. (2017). Genetic diversity of einkorn wheat (*Triticum boeoticum* Boiss.) accessions from the Central Zagros Mountains. *Zemdirbyste- Agriculture*, 104(1), 23-30.

Seppy, M., Manni, M., & Zdobnov, E. M. (2019). BUSCO: Assessing genome assembly and annotation completeness. *Methods in Molecular Biology*, 1962, 227-245.

Sequencher® version 5.4.1, DNA sequence analysis software, Gene Codes Corporation, Ann Arbor, MI USA, Retrieved from <http://www.genecodes.com>.

Sheffield, W. P., Shore, G. C., & Randall, S. K. (1990) Mitochondrial precursor protein. Effects of 70-kilodalton heat shock protein on polypeptide folding, aggregation, and import competence. *Journal of Biology & Chemistry*, 265, 11069–11076.

Shen, W., Le, S., Li, Y., & Hu, F. (2016). SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLOS One*, 11(10).

Shewry, P. R. (2009). Wheat. *Journal of Experimental Botany*, 60(6), 1537–1553.

Shewry, P. R., & Hey, S. (2015). Do “ancient” wheat species differ from modern bread wheat in their contents of bioactive components? *Journal of Cereal Science*, 65, 236-243.

Singh, A., Singh, U., Mittal, D., & Grover, A. (2010). Genome-wide analysis of rice ClpB/HSP100, ClpC and ClpD genes. *BMC genomics*, 11(1), 1-18.

Skylas, D. J., Cordwell, S. J., Hains, P. G., Larsen, M. R., Basseal, D. J., Walsh, B. J., ... & Wrigley, C. W. (2002). Heat shock of wheat during grain filling: proteins associated with heat-tolerance. *Journal of Cereal Science*, 35(2), 175-188.

Smith, B. D., 1995. The emergence of agriculture. Scientific American Library, New York.

Song, Y., Chen, Q., Ci, D., Shao, X., & Zhang, D. (2014). Effects of high temperature on photosynthesis and related gene expression in poplar. *BMC Plant Biology*, 14, 1-20.

Sourdille, P., Tavaud, M., Charmet, G., & Bernard, M. (2001). Transferability of wheat microsatellites to diploid Triticeae species carrying the A, B and D genomes. *Theoretical and Applied Genetics*, 103, 346–352.

Stark, R., Grzelak, M., & Hadfield, J. (2019). RNA sequencing: The teenage years. *Nature Review Genetics*, 20, 631–656.

Sthapit, S., Marlowe, K., Covarrubias, D., Ruff, T., Eagle, J., McGinty, E., Hooker, M., Duong, N., Skinner, D., & See, D. (2020). Genetic diversity in historical and modern wheat varieties of the US Pacific Northwest. *Crop Science*, 60.

Struk, S., Jacobs, A., Sanchez Martín-Fontecha, E., Gevaert, K., Cubas, P., & Goormachtig, S. (2019). Exploring the protein–protein interaction landscape in plants. *Plant Cell Environment*, 42, 387–409.

Su, P., Jiang, C., Qin, H., Hu, R., Feng, J., Chang, J., Yang, G., & He, G. (2019). Identification of potential genes responsible for thermotolerance in wheat under high temperature stress. *Genes*, 10(2).

Suzuki, N., Miller, G., Morales, J., Shulaev, V., & Torres, M. A. (2011). Respiratory burst oxidases: The engines of ROS signaling. *Current Opinion in Plant Biology*, 14, 691–699.

Szymanski, D., & Staiger, C. J. (2018). The actin cytoskeleton: Functional arrays for cytoplasmic organization and cell shape control. *Plant Physiology*, 176, 106–118.

Taiz, L., Zeiger, E., Møller, I. M., & Murphy, A. (2015). *Plant physiology and development* (No. Ed. 6). Sinauer Associates Incorporated.

Takhtajan, A. (Ed.) (2009), Class Liliopsida (Monocotyledons), In *Flowering Plants* (pp. 589–750). Dordrecht: Springer Netherlands.

Takumi, S., Nasuda, S., Liu, Y., & Tsunewaki, K. (1993). Wheat phytoeny determined by RFLP analysis of nuclear DNA. 1. Einkorn wheat. *The Japanese Journal of Genetics*, 68, 73-79.

Trenberth, K. E. (2011). Changes in precipitation with climate change. *Climate Research*, 47(1–2), 123–138.

Tubiello, F. N., Soussana, J. F., & Howden, S. M. (2007). Crop and pasture response to climate change. *Proceedings of the National Academy of Sciences of the United States of America*, 104(50), 19686–19690.

Tyrka, M., Mokrzycka, M., Bakera, B., Tyrka, D., Szeliga, M., Stojalowski, S., Matysik, P., Rokicki, M., Rakoczy-Trojanowska, M., & Krajewski, P. (2021). Evaluation of genetic structure in European wheat cultivars and advanced breeding lines using high-density genotyping-by-sequencing approach. *BMC Genomics*, 22(1), 81.

Urban, M. C. (2015). Accelerating extinction risk from climate change. *Science*, 348, 6234.

Vacca, R. A., de Pinto, M. C., Valenti, D., Passarella, S., Marra, E., & De Gara, L. (2004). Production of reactive oxygen species, alteration of cytosolic ascorbate peroxidase, and impairment of mitochondrial metabolism are early events in heat shock-induced programmed cell death in tobacco bright-yellow 2 cells. *Plant Physiology*, 134, 1100–1112.

Valkoun, J. J. (2001). Wheat pre-breeding using wild progenitors. *Euphytica*, 119, 17–23.

Van Zeist, W., & de Roller, G. J. (1991). The plant husbandry of aceramic Çayönü, SE Turkey. *Palaeohistoria*, 33/34, 65–96.

Vavilov, N. I. (1951). The origin, variation, immunity and breeding of cultivated plants. *The Chronica Botanica*, 13, 1–6.

Verma, V., Ravindran, P., & Kumar, P. P. (2016). Plant hormone-mediated regulation of stress responses. *BMC Plant Biology*, 16, 86.

Vierling, E. (1991). The roles of heat shock proteins in plants. *Annual Review of Plant Physiology and Plant Molecular Biology*, 42, 579–620.

- Vierling, R. A., & Nguyen, H. T. (1992). Use of RAPD markers to determine the genetic diversity of diploid wheat genotypes. *Theoretical and Applied Genetics*, 84, 835–838.
- Volkov, R. A., Panchuk, I. I., Mullineaux, P. M., & Schöffl, F. (2006). Heat stress-induced H₂O₂ is required for effective expression of heat shock genes in *Arabidopsis*. *Plant Molecular Biology*, 61, 733–746.
- Wahid, A., Gelani, S., Ashraf, M., & Foolad, M. R. (2007). Heat tolerance in plants: An overview. *Environmental and Experimental Botany*, 61(3), 199–223.
- Walther, G. R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T. J. C., Fromentin, J. M., Hoegh-Guldberg, O., & Bairlein, F. (2002). Ecological responses to recent climate change. *Nature*, 416, 389.
- Wang, W., Vinocur, B., Shoseyov, O., & Altman, A. (2004). Role of plant heat shock proteins and molecular chaperons in the abiotic stress response. *TRENDS in Plant Science*, 9(5), 1360–1385.
- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1), 57–63.
- Wang, J., Vanga, S. K., Saxena, R., Orsat, V., & Raghavan, V. (2018). Effect of climate change on the yield of cereal crops: A review. *Climate*, 6, 41.
- Waters, E. R., Lee, G. J., & Vierling, E. (1996). Evolution, structure and function of the small heat shock proteins in plants, *Journal of Experimental Botany*, 47, 325–338.
- Waters, E. R. (2013). The evolution, function, structure, and expression of the plant sHSPs. *Journal of experimental botany*, 64(2), 391–403.
- WeatherOnline, (n.d.). Retrieved from <https://www.havaturkiye.com/weather/maps/city> (Accessed December 2022).

WMO World Meteorological Organization, (2001). Statement on Status of the Global Climate in 2000, WMO-No. 920.

WMO World Meteorological Organization, (2008). Statement on Status of the Global Climate in 2007, WMO-No. 1031.

Wu, B., Qiao, J., Wang, X., Liu, M., Xu, S., & Sun, D. (2021). Factors affecting the rapid changes of protein under short-term heat stress. *BMC Genomics*, 22(1), 263.

Xiao, G., Liu, W., Xu, Q., Sun, Z., & Wang, J. (2005). Effect of temperature increase and elevated CO₂ concentration, with supplement irrigation, on the yield of rain-fed spring wheat in semi-arid region of China. *Agricultural Water Management*, 74, 243–255.

Xue, G. P., Sadat, S., Drenth, J., & McIntyre, C. L. (2014). The heat shock factor family from *Triticum aestivum* in response to heat and other major abiotic stresses and their role in regulation of heat shock protein genes. *Journal of Experimental Botany*, 65, 539–557.

Yaman, H. M., Ordu, B., Zencirci, N., & Kan, M. (2020). Coupling socioeconomic factors and cultural practices in production of einkorn and emmer wheat species in Turkey. *Environment, Development and Sustainability*, 22, 8079-8096.

Yamori, W., Shikanai, T., & Makino, A. (2015). Photosystem I cyclic electron flow via chloroplast NADH dehydrogenase-like complex performs a physiological role for photosynthesis at low light. *Scientific Reports*, 5, 13908.

Yang, F., Jørgensen, A. D., Li, H., Søndergaard, I., Finnie, C., Svensson, B., Jiang, D., Wollenweber, B., & Jacobsen, S. (2011). Implications of high-temperature events and water deficits on protein profiles in wheat (*Triticum aestivum* L. cv. Vinjett) grain. *Proteomics*, 11, 9.

Yang, X., Tan, B., Liu, H., Zhu, W., Xu, L., Wang, Y., ... & Kang, H. (2020). Genetic diversity and population structure of Asian and European common wheat accessions based on genotyping-by-sequencing. *Frontiers in Genetics*, 11, 580782.

Yano, T., Aydin, M., & Haraguchi, T. (2007). Impact of climate change on irrigation demand and crop growth in a Mediterranean environment of Turkey. *Sensors*, 7, 2297-2315.

Zaharieva, M., Ayana, N. G., Al Hakimi, A. S., Misra, C., & Monneveux, P. (2010). Cultivated emmer wheat (*Triticum dicoccon* Schrank), an old crop with promising future: A review. *Genetic Resources Crop Evaluation*, 57, 937–962.

Zeven, A. (1998). Landraces: A review of definitions and classifications. *Euphytica*, 104, 127–139.

Zhang, Y., Pan, J., Huang, X., Guo, D., Lou, H., Hou, Z., Su, M., Liang, R., Xie, C., You, M., & Li, B. (2017). Differential effects of a post-anthesis heat stress on wheat (*Triticum aestivum* L.) grain proteome determined by iTRAQ. *Scientific reports*, 7(1), 3468.

Zhang, H., Zhu, J., Gong, Z., & Zhu, J. K. (2022). Abiotic stress responses in plants. *Nature Reviews Genetics*, 23(2), 104-119.

Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., Huang, M., Yao, Y., Bassu, S., Ciais, P., Durand, J. L., Elliott, J., Ewert, F., Janssens, I. A., Li, T., Lin, E., Liu, Q., Martre, P., Müller, C., Peng, S., Peñuelas, J., Ruane, A. C., Wallach, D., Wang, T., Wu, D., Liu, Z., Zhu, Y., Zhu, Z., & Asseng, S. (2017). Temperature increase reduces global yields of major crops in four independent estimates, *PNAS*, 114(35), 9326-9331.

Zhao, P., Javed, S., Shi, X., Wu, B., Zhang, D., Xu, S., & Wang, X. (2020). Varying architecture of heat shock elements contributes to distinct magnitudes of target gene expression and diverged biological pathways in heat stress response of bread wheat. *Frontiers in Genetics*, 11.

Zhou, X., Ren, L., Meng, Q., Li, Y., Yu, Y., & Yu, J. (2010). The next-generation sequencing technology and application. *Protein Cell*, 1, 520-536.

Zi, J., Mafu, S., & Peters, R. J. (2014). To gibberellins and beyond! Surveying the evolution of (di)terpenoid metabolism. *Annual Review of Plant Biology*, 65, 259- 286.

Zohary, D., & Hopf, M. (1993). *Domestication of plants in the Old World: The origin and spread of cultivated plants in West Asia, Europe, and the Nile Valley*, 2nd edn. Oxford: Oxford University Press.

Zohary, D., & Hopf, M. (2000). *Domestication of plants in the Old World*. Oxford University Press, Oxford.

Zohary, D., Hopf, M., & Weiss, E. (2012). *Domestication of plants in the Old World: The origin and spread of domesticated plants in southwest Asia, Europe, and the Mediterranean basin* (4th ed.). Oxford University Press, Oxford, United Kingdom.



APPENDIX A: ddRAD-seq

Table A.1. Read numbers obtained for 162 libraries after cleaning and demultiplexing. Green indicates einkorn; blue indicates emmer, and orange indicates bread wheat populations.

Library	Total Reads	Retained Reads	Library	Total Reads	Retained Reads	Library	Total Reads	Retained Reads
WP81-1	4450670	4213371	WP90-3	5177182	4704814	WP15-1.1	5274444	4371794
WP81-2	2518920	2391428	WP90-4	5748160	5487260	WP15-1.3	4696324	4252394
WP81-3	2558148	2412530	WP90-5	5858874	4892371	WP15-1.4	5548910	4901464
WP81-4	3548048	3463731	WP90-6	3895846	3425250	WP15-1.5	3647322	3354712
WP81-5	3352994	3053326	WP90-7	3285120	3087889	WP15-2.1	4229736	3634241
WP81-10	2491448	2296814	WP90-8	2878974	2550269	WP15-2.2	3025492	2769632
WP81-11	8164340	7363247	WP90-9	3594318	3464606	WP15-2.3	3757042	3124238
WP81-12	3373106	3181755	WP92-1	2570098	2000062	WP15-3.1	4526784	4066814
WP81-13	6108990	5544805	WP92-3	4380348	3754688	WP15-3.2	6535610	5076388
WP81-14	3470500	3225432	WP92-4	1929810	1531057	WP15-3.3	4938444	4195268
WP88-1	3501852	3246130	WP92-5	1619108	1489133	WP15-3.4	5349916	4229610
WP88-2	2634554	2528578	WP92-6	3603470	2418781	WP15-3.5	4760990	4291123
WP88-3	3329776	3123138	WP92-8	5234480	4374643	WP20-1.1	4178964	3522694
WP88-4	2831024	2697191	WP92-10	6657626	5971623	WP20-1.3	5191058	4225567
WP88-6	3086574	2851973	WP92-12	9912802	8158628	WP20-1.4	4298410	3740520
WP88-7	3704962	3526021	WP92-14	7677402	6685144	WP20-1.5	4261232	3340576
WP88-10	3090838	2795591	WP93-1	3365256	3012008	WP20-2.1	2515878	2226784
WP88-11	2180326	2065113	WP93-1.1	4826606	4505643	WP20-2.2	3860050	3540671
WP88-12	3177734	2755345	WP93-2.2	3042020	2821833	WP20-2.3	4174410	3294535
WP89-5	5727490	4017781	WP93-2.3	3001188	2595058	WP20-2.4	4083948	3367444
WP89-7	7877516	7362658	WP93-3.2	2396228	2315404	WP20-2.7	3294366	2397971
WP89-8.1	8049264	7824436	WP93-3.3	1994530	1749556	WP20-2.8	3291114	2617172
WP89-8.3	1809874	1584579	WP93-4	3310102	3089365	WP20-3.1	3959162	3081099
WP89-9.1	6850106	5064653	WP93-5	2993666	2912799	WP20-3.3	2407242	1900256
WP89-9.3	3300934	2930324	WP93-6	3930262	3549989	WP20-3.4	1743476	1316278
WP89-12	1380918	1275415	WP93-7	3009634	2845479	WP20-3.5	1888890	1679304
WP89-13.1	4927088	4567376	WP95-1	6736620	6321034	WP98-1.3	2592770	2362317
WP89-13.3	1890342	1599676	WP95-2	3866522	3582850	WP98-1.4	3056188	2880465
WP22-1.2.1	2900458	2759387	WP95-3	4669266	4224236	WP98-1.5	3067536	2756554
WP22-1.3.1	2477762	2411731	WP95-4	4813962	4482230	WP98-2.1	2915408	2643261
WP22-1.5.1	2480738	1778743	WP95-5	5112706	4630490	WP98-2.4	3481016	3079319
WP22-1.7.1	2358320	1953563	WP95-6	5443046	4496016	WP98-2.5	3378438	3096840
WP22-2.1.1	5064472	4602200	WP95-7	4989490	4816158	WP98-3.2	3992426	3539101
WP22-2.3.1	4036738	3885420	WP95-9	7192226	6109691	WP99-2	3449960	3305576
WP22-2.4.1	1087728	997032	WP95-11	4192108	3994483	WP99-3	1955350	1889485
WP22-2.5.1	2661220	2086479	WP95-13	8559472	8270324	WP99-4	5107320	4953518
WP22-2.6.1	2511740	2017811	WP96-1.2	2215260	2079949	WP99-7	3822964	3608141
WP22-3.3.1	2529196	2441737	WP96-2	5976540	5702727	WP99-9	3905242	3761498
WP22-3.4.1	10387490	10010588	WP96-3	4377570	4264617	WP99-11	4813238	4594210
WP22-3.5.1	2309780	1741510	WP96-4	3283696	3006985	WP99-12	5339616	5168953
WP22-3.6.1	2565348	1982841	WP96-6	5103770	4792359	WP99-13	4787172	4235151
WP22-1.1.2	7890068	7644719	WP96-8	2980186	2795960	WP99-18	1225602	1031058
WP22-1.3.2	4811162	4220048	WP96-9	3939856	3691227	WP99-20	3986940	3652507
WP22-1.4.2	4313754	4058808	WP96-10	1747134	1672493	WP99-24	4937942	4399106
WP22-1.5.2	2612262	2144360	WP96-11	2858886	2616327	WP103-1.1	2675108	2518720
WP22-2.1.2	6163370	5713632	WP97-1	2870170	2618999	WP103-1.3	2013610	1696555
WP22-2.3.2	3926720	3717610	WP97-2	4464028	3776520	WP103-1.5	2653376	2364529
WP22-2.4.2	4267374	4000037	WP97-3	4263626	3761019	WP103-2.1	2830738	2464920
WP22-2.6.2	2883130	1982160	WP97-4	503426	428234	WP103-2.2	1730964	1634523
WP22-3.1.2	3453354	3299710	WP97-5	4552880	4120301	WP103-2.3	2851866	2525310
WP22-3.2.2	5059962	4777071	WP97-6	3966648	3321187	WP103-2.4	2618962	2317810
WP22-3.3.2	4725130	4519008	WP97-7	3481614	2956560	WP103-3.3	2450462	2165810
WP22-3.5.2	2025844	1775425	WP97-8	3838830	3615466			
WP90-1	4990672	4508608	WP97-10	3942910	3656915			
WP90-2	4119402	3804055	WP97-13	7496878	6955685			

Table A.2. Observed heterozygosity, expected heterozygosity, and nucleotide diversity values of einkorn populations calculated for variant positions.

de novo Analysis	Population	Observed Heterozygosity	Expected Heterozygosity	Nucleotide Diversity
Default parameters with r80	WP 22	0.12492	0.18518	0.18963
	WP 81	0.18640	0.23612	0.25019
	WP 88	0.19750	0.23411	0.24870
	WP 89	0.16953	0.21503	0.22863
Default parameters without r80	WP 22	0.28222	0.26776	0.32569
	WP 81	0.25154	0.21352	0.29140
	WP 88	0.26205	0.20539	0.28222
	WP 89	0.24794	0.19875	0.28047
Optimized parameters with r80	WP 22	0.15502	0.21362	0.21877
	WP 81	0.22172	0.26968	0.28576
	WP 88	0.22803	0.26222	0.27858
	WP 89	0.20290	0.24342	0.25885
Optimized parameters without r80	WP 22	0.20587	0.24649	0.29469
	WP 81	0.19740	0.18941	0.25004
	WP 88	0.20925	0.18266	0.24161
	WP 89	0.19910	0.18132	0.24908

Table A.3. Observed heterozygosity, expected heterozygosity, and nucleotide diversity values of einkorn populations calculated for all positions (variant & fixed).

de novo Analysis	Population	Observed Heterozygosity	Expected Heterozygosity	Nucleotide Diversity
Default parameters with r80	WP 22	0.00026	0.00038	0.00039
	WP 81	0.00029	0.00036	0.00039
	WP 88	0.00031	0.00036	0.00039
	WP 89	0.00029	0.00037	0.00040
Default parameters without r80	WP 22	0.00067	0.00064	0.00078
	WP 81	0.00078	0.00067	0.00091
	WP 88	0.00087	0.00068	0.00093
	WP 89	0.00077	0.00062	0.00087
Optimized parameters with r80	WP 22	0.00048	0.00066	0.00068
	WP 81	0.00058	0.00071	0.00075
	WP 88	0.00057	0.00066	0.00070
	WP 89	0.00055	0.00066	0.00070
Optimized parameters without r80	WP 22	0.00103	0.00124	0.00148
	WP 81	0.00114	0.00110	0.00145
	WP 88	0.00120	0.00105	0.00139
	WP 89	0.00116	0.00106	0.00145

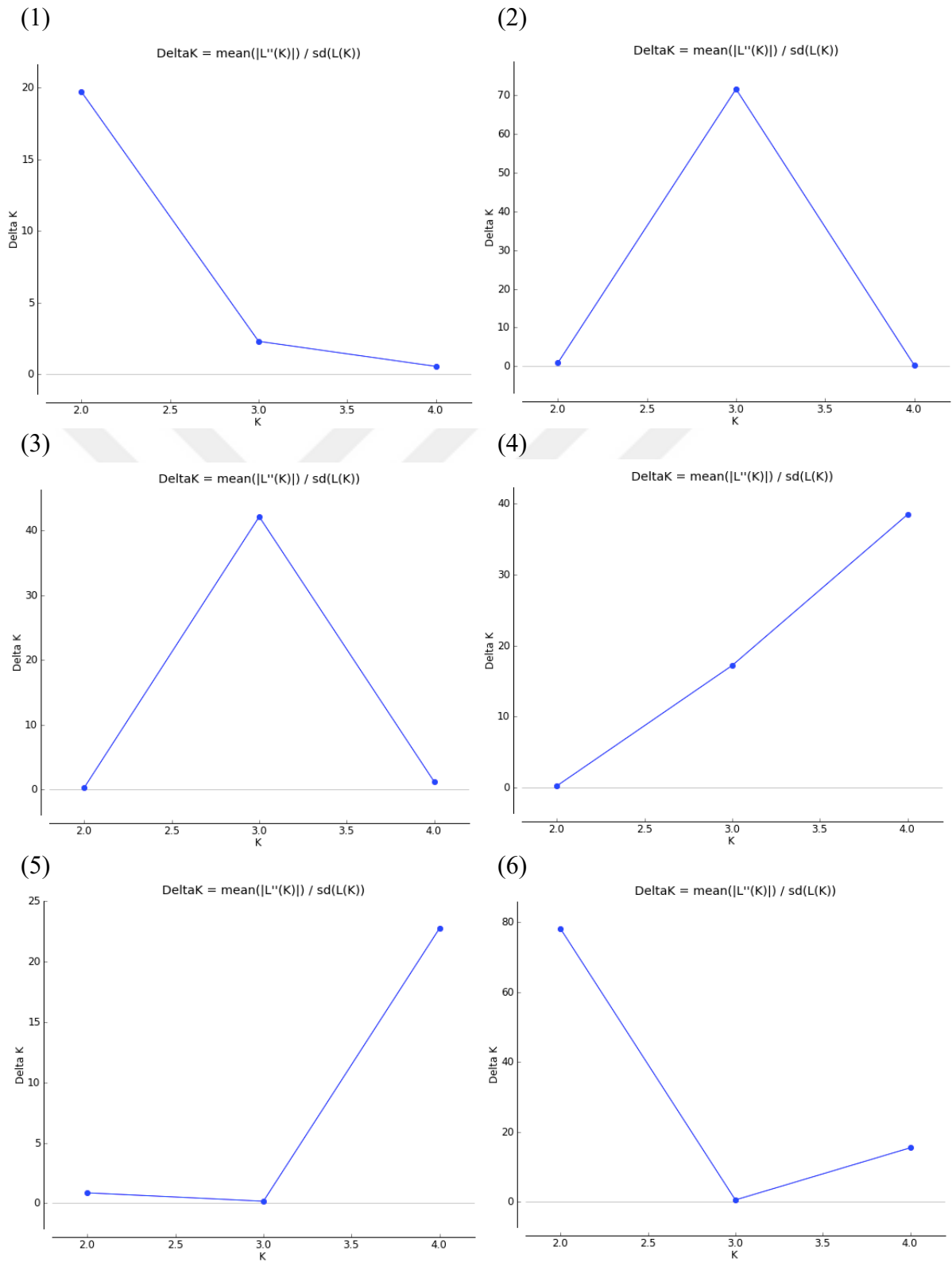


Figure A.1. Delta K calculated for the datasets in Table 4.4 used in einkorn's *Structure* analyses. The graphs were labeled as the datasets labeled in Table 4.4.

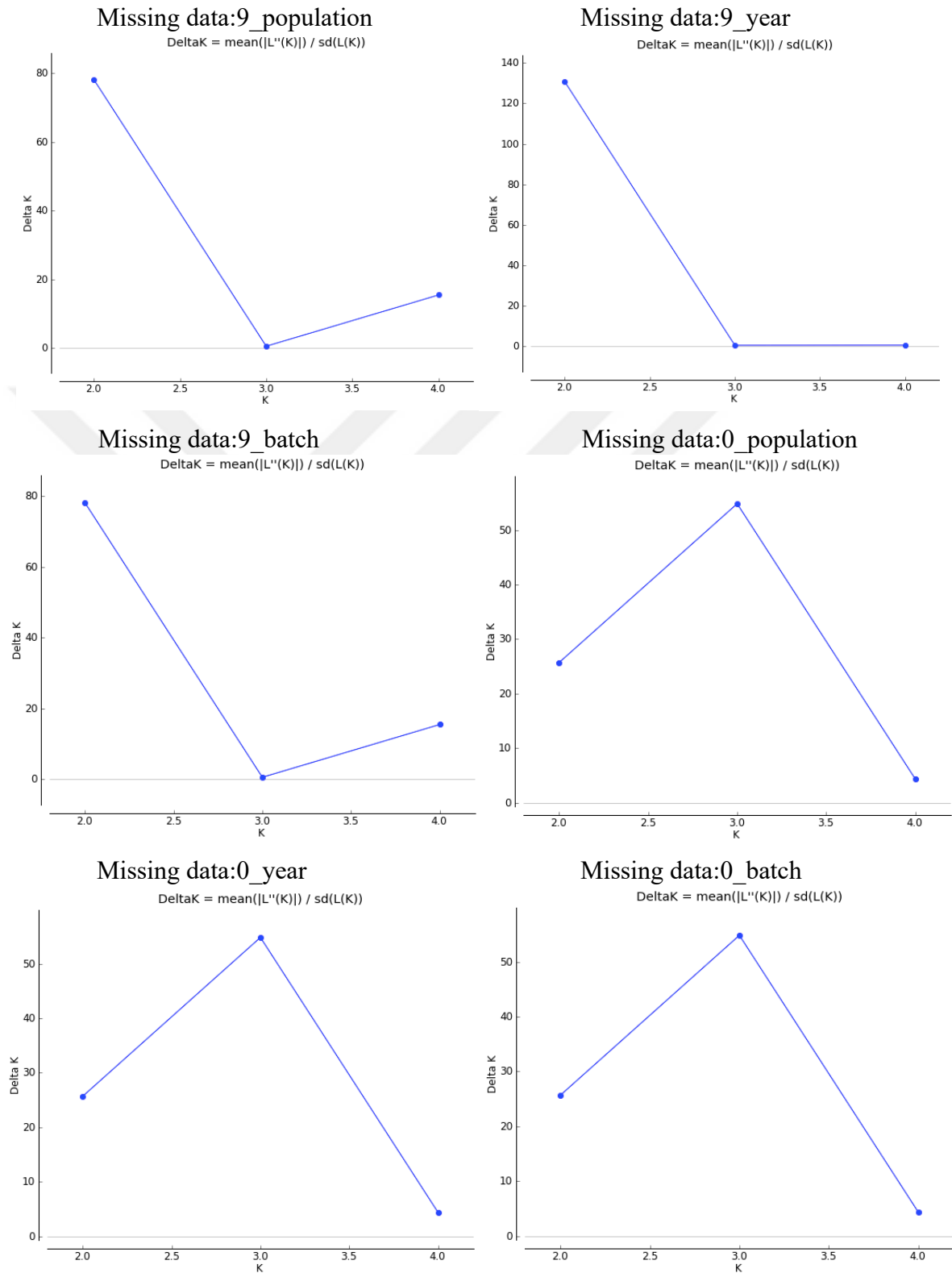


Figure A.2. Delta K calculated for einkorn's *Structure* analyses with optimized parameters and missing data encoded as -9 for population, year, and pool based runs.

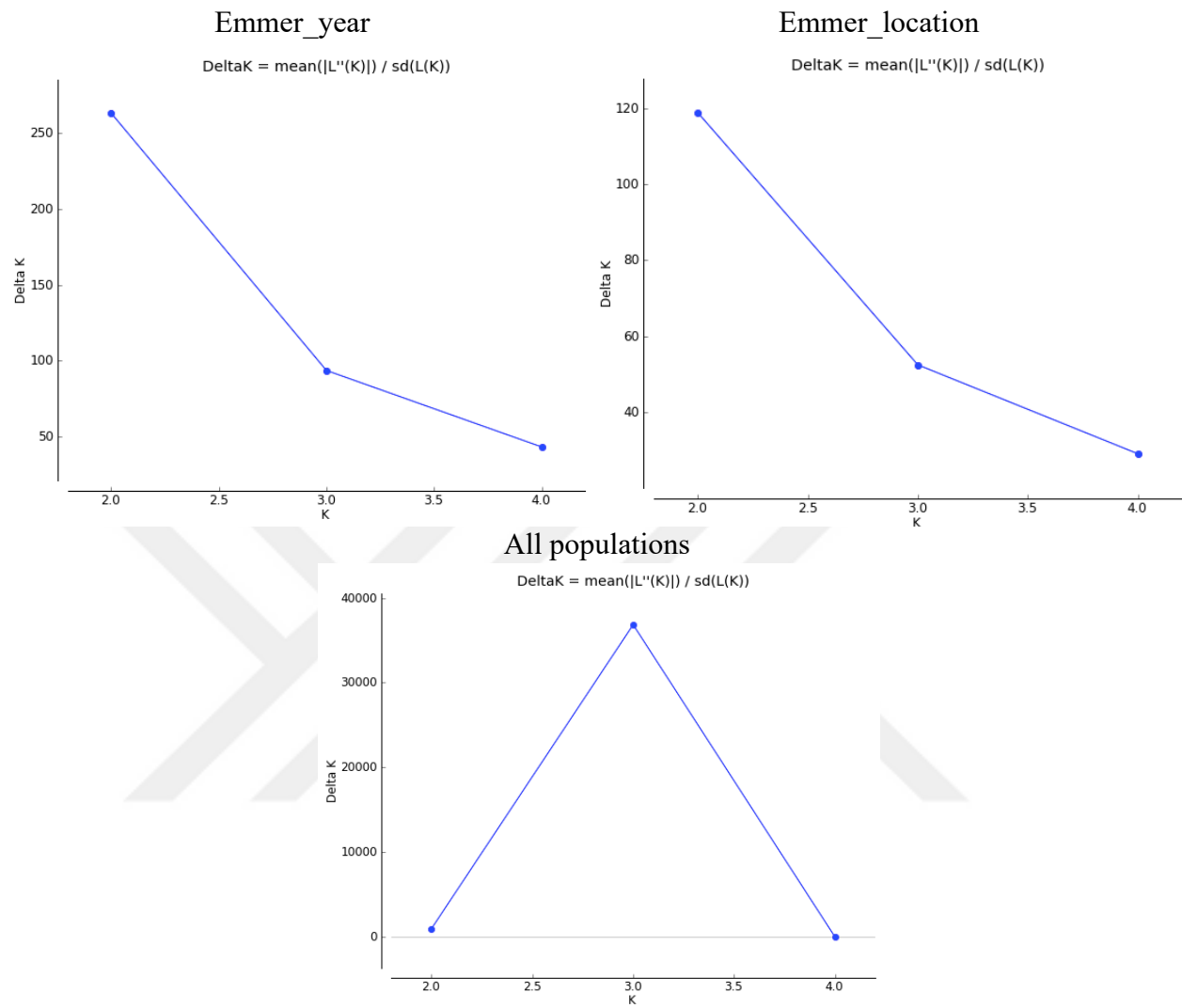


Figure A.3. Delta K calculated for *Structure* analyses of emmer and all populations.

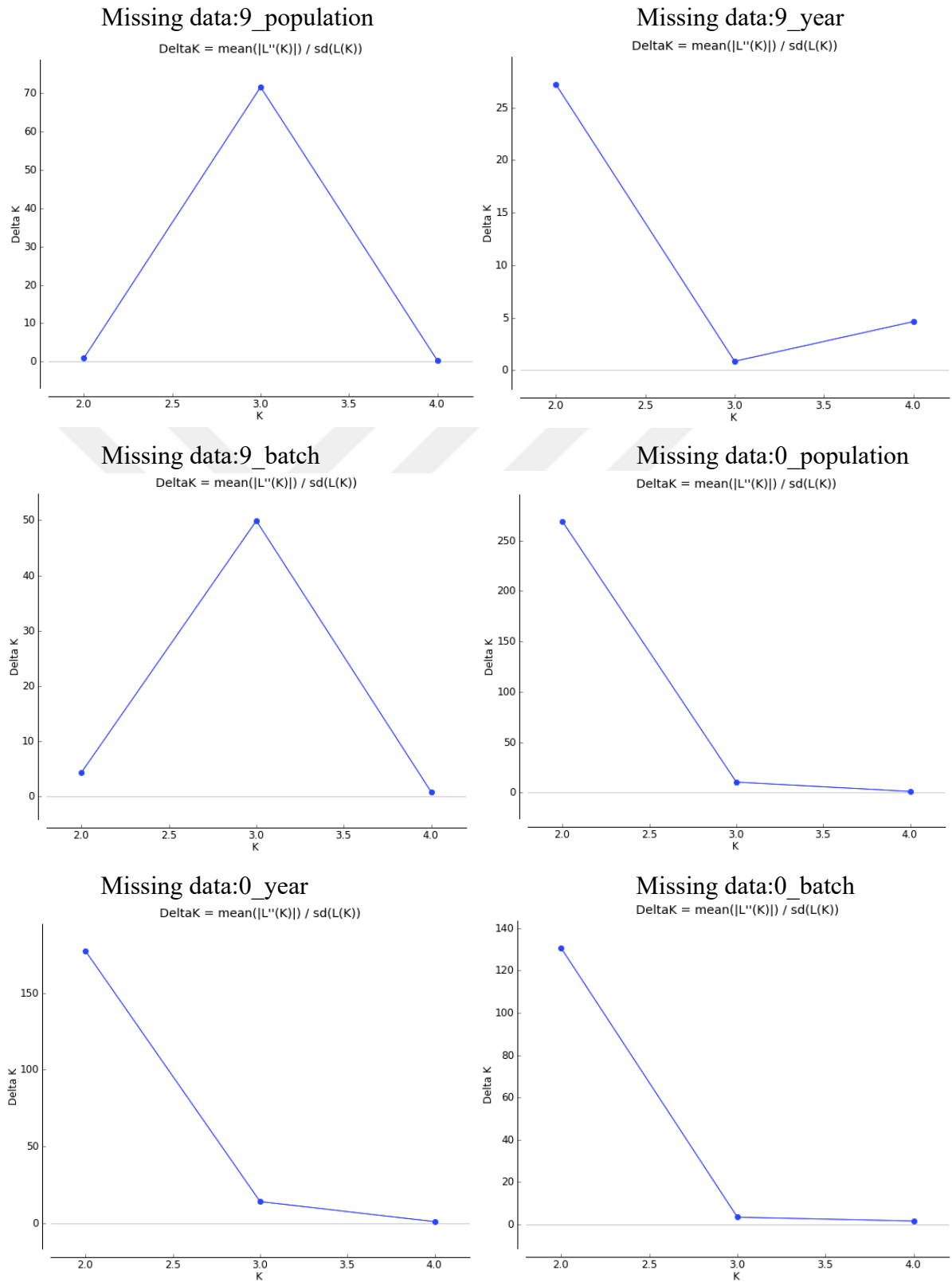


Figure A.4. Delta K calculated for *Structure* analyses of einkorn with default parameters and *write_single_snp* flag.

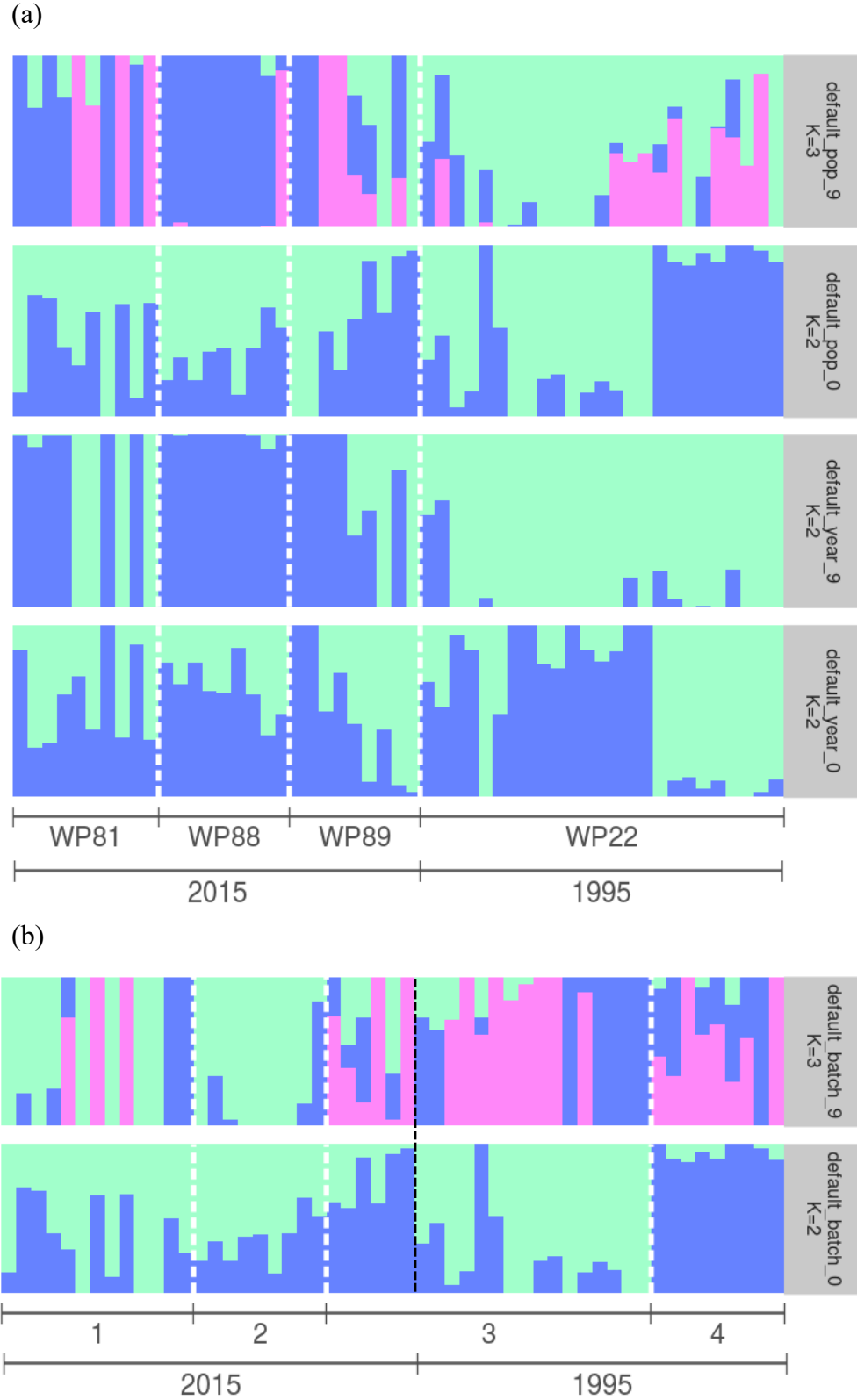


Figure A.5. *Structure* results of the runs with default parameters and missing data encoded as -9 and zero. The plots show the results of the libraries defined by population and year (a), and batch (b) for the most likely K value. The black dashed line indicates the distinction between years.

Table A.4. Observed heterozygosity, expected heterozygosity, and nucleotide diversity values of emmer populations calculated for variant positions.

	Population	Observed Heterozygosity	Expected Heterozygosity	Nucleotide Diversity
2015 Kastamonu	WP 90	0.17756	0.17371	0.22704
	WP 92	0.19050	0.13563	0.19621
	WP 93	0.18026	0.12399	0.16862
2015 Kars	WP 95	0.18396	0.18141	0.23475
	WP 96	0.17364	0.12969	0.17406
	WP 97	0.20071	0.15879	0.21867
1995 Kastamonu	WP 15	0.17738	0.13024	0.17187
	WP 20	0.07468	0.10359	0.12864

Table A.5. Observed heterozygosity, expected heterozygosity, and nucleotide diversity values of emmer populations calculated for variant & fixed positions.

	Population	Observed Heterozygosity	Expected Heterozygosity	Nucleotide Diversity
2015 Kastamonu	WP 90	0.00210	0.00205	0.00268
	WP 92	0.00207	0.00147	0.00213
	WP 93	0.00262	0.00180	0.00245
2015 Kars	WP 95	0.00218	0.00215	0.00278
	WP 96	0.00254	0.00189	0.00254
	WP 97	0.00246	0.00195	0.00268
1995 Kastamonu	WP 15	0.00261	0.00192	0.00253
	WP 20	0.00105	0.00145	0.00180

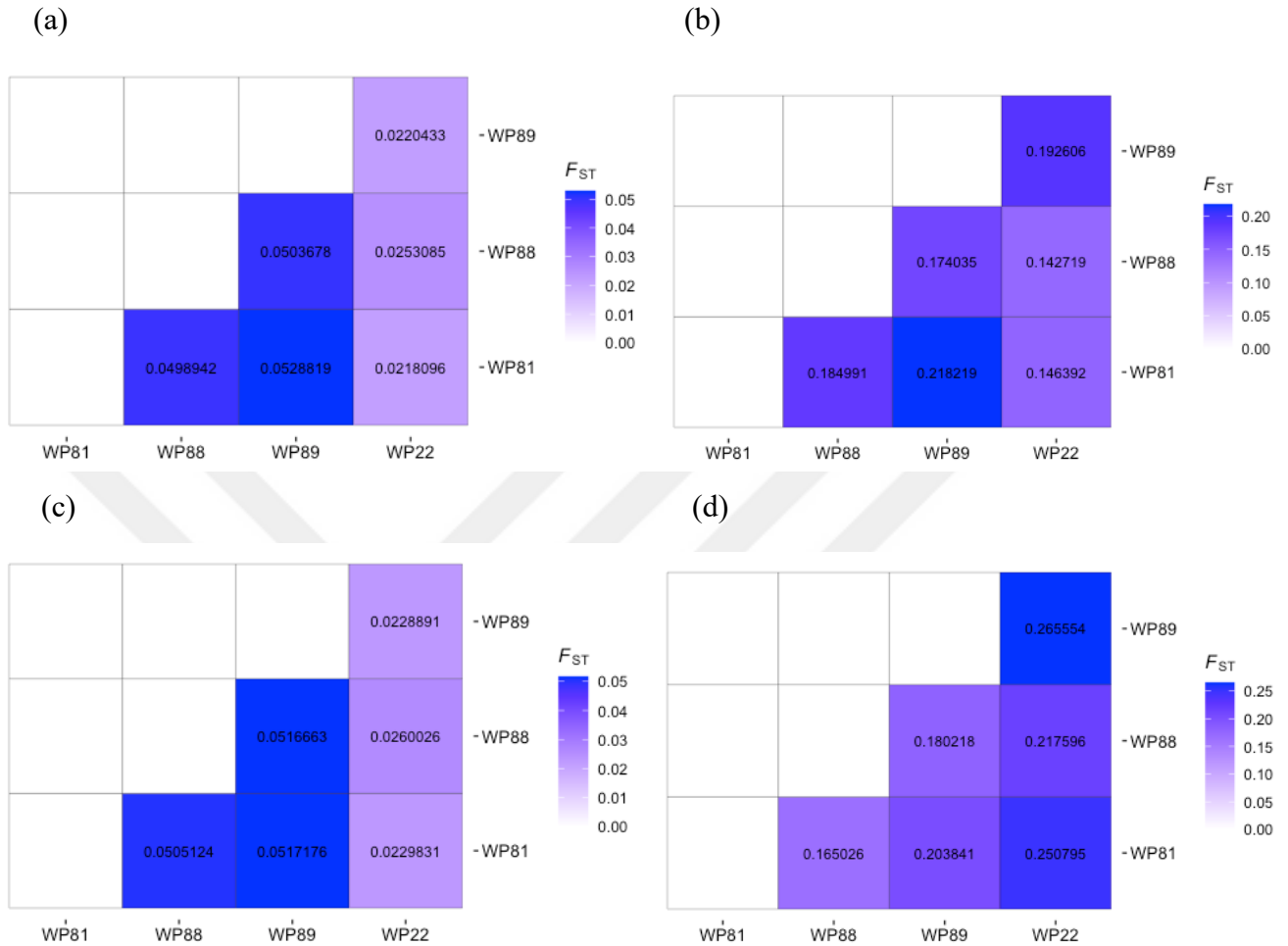


Figure A.6. Matrices of pairwise F_{ST} values of einkorn populations obtained with *write_single_snp* flag and default parameters with r80 (a), default parameters without r80 (b), optimized parameters with r80 (c), and optimized parameters without r80 (d).

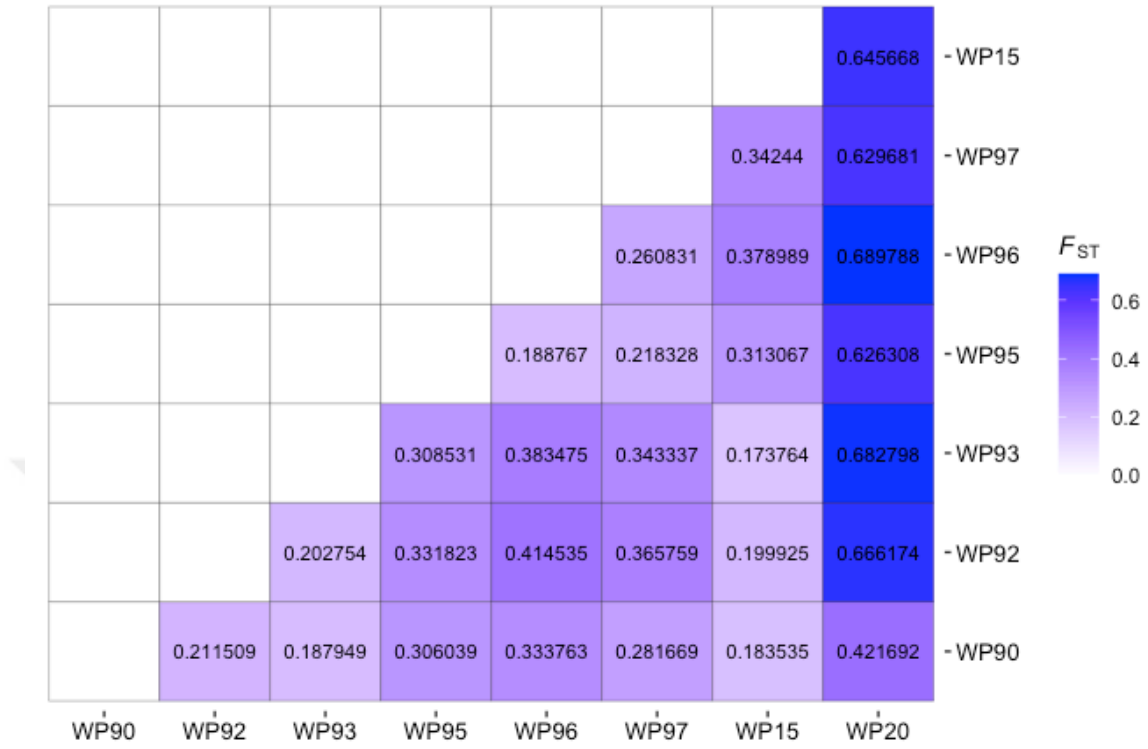


Figure A.7. Matrix of pairwise F_{ST} values of emmer populations obtained with *write_single_snp* flag.

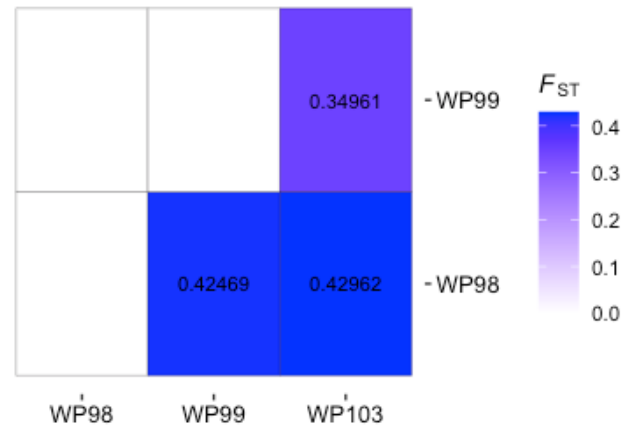


Figure A.8. Matrix of pairwise F_{ST} values of bread wheat populations obtained with *write_single_snp* flag.

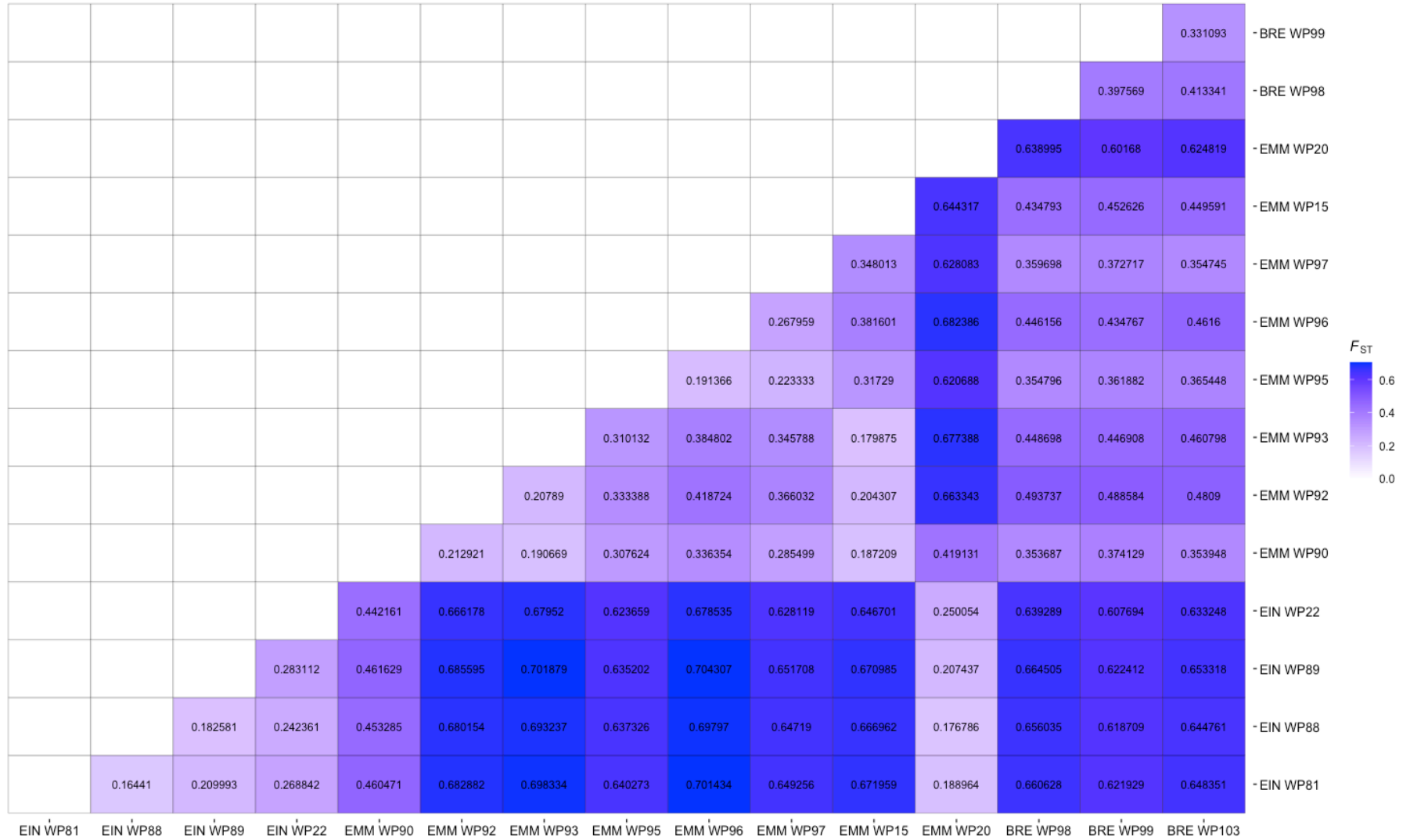


Figure A.9. Matrix of pairwise F_{ST} values of all wheat populations obtained with *write_single_snp* flag. In population names, EIN indicates einkorn, EMM indicates emmer and BRE indicates bread wheat.

Table A.6. Observed heterozygosity, expected heterozygosity, and nucleotide diversity values of bread wheat populations calculated for variant positions.

Population	Observed Heterozygosity	Expected Heterozygosity	Nucleotide Diversity
WP 98	0.28682	0.20004	0.28221
WP 99	0.32658	0.26664	0.35762
WP 103	0.31457	0.22041	0.32453

Table A.7. Observed heterozygosity, expected heterozygosity, and nucleotide diversity values of bread wheat populations calculated for variant & fixed positions.

Population	Observed Heterozygosity	Expected Heterozygosity	Nucleotide Diversity
WP 98	0.00216	0.00150	0.00212
WP 99	0.00230	0.00188	0.00252
WP 103	0.00237	0.00166	0.00244

Table A.8. Observed heterozygosity, expected heterozygosity, and nucleotide diversity values of all wheat populations calculated for variant positions.

Species		Population	Observed Heterozygosity	Expected Heterozygosity	Nucleotide Diversity
Einkorn	1995 Kastamonu	WP 22	0.07739	0.08825	0.10823
		WP 81	0.06634	0.06278	0.08337
	2015 Kastamonu	WP 88	0.06723	0.05956	0.07867
		WP 89	0.06672	0.06058	0.08325
Emmer	1995 Kastamonu	WP 15	0.14288	0.10599	0.14041
		WP 20	0.06645	0.06971	0.09201
	2015 Kastamonu	WP 90	0.14331	0.13972	0.18391
		WP 92	0.15824	0.11048	0.16494
		WP 93	0.14480	0.10122	0.13752
	2015 Kars	WP 95	0.15182	0.14641	0.19178
		WP 96	0.14164	0.10602	0.14306
		WP 97	0.16218	0.12859	0.17813
	Bread Wheat	2015 Kastamonu	WP 98	0.16758	0.11756
WP 99			0.19685	0.14725	0.20621
WP 103			0.18920	0.13222	0.19770

Table A.9. Observed heterozygosity, expected heterozygosity, and nucleotide diversity values of all wheat populations calculated for all positions.

Species		Population	Observed Heterozygosity	Expected Heterozygosity	Nucleotide Diversity
Einkorn	1995 Kastamonu	WP 22	0.00105	0.00120	0.00147
		WP 81	0.00113	0.00107	0.00142
	2015 Kastamonu	WP 88	0.00122	0.00108	0.00143
		WP 89	0.00114	0.00104	0.00142
Emmer	1995 Kastamonu	WP 15	0.00255	0.00189	0.00250
		WP 20	0.00115	0.00121	0.00160
	2015 Kastamonu	WP 90	0.00210	0.00205	0.00270
		WP 92	0.00233	0.00163	0.00243
		WP 93	0.00256	0.00179	0.00243
	2015 Kars	WP 95	0.00223	0.00215	0.00282
		WP 96	0.00252	0.00188	0.00254
		WP 97	0.00247	0.00196	0.00272
Bread Wheat	2015 Kastamonu	WP 98	0.00235	0.00165	0.00235
		WP 99	0.00269	0.00202	0.00282
		WP 103	0.00257	0.00180	0.00269

APPENDIX B: RNA-seq

Table B.1. RINs and rRNA ratios of 36 RNA samples.

Population	Treatment	Pot	Sample	RIN (SEM)	rRNA ratios (SEM)	RIN (Macrogen)	rRNA ratios (Macrogen)
WP 22	Control	I	1	7.1	1.3	8.4	1.6
			2	7.8	1.4	7.1	1.0
			3	7.6	1.4	8.5	1.9
		II	1	7.1	1.2	7.0	1.1
			2	7.2	1.3	8.7	1.6
			3	7.6	1.4	8.5	1.7
		III	1	7.6	1.4	8.8	2.0
			2	7.1	1.3	8.6	1.7
			3	7.3	1.3	8.5	1.8
	Stress	I	1	7.4	1.4	8.4	1.9
			2	7.5	1.3	8.3	1.8
			3	7.1	1.3	8.2	1.9
		II	1	7.9	1.5	8.6	1.9
			2	7.6	1.4	8.3	1.8
			3	7.6	1.5	7.0	1.0
		III	1	7.6	1.4	7.3	1.5
			2	8.0	1.5	8.5	2.0
			3	7.7	1.5	8.5	1.7
WP 89	Control	I	1	7.9	1.5	8.3	2.0
			2	8.1	1.4	8.6	2.3
			3	8.1	1.5	8.4	1.9
		II	1	7.7	1.4	7.0	1.1
			2	7.9	1.4	7.0	1.0
			3	8.2	1.5	8.6	1.9
		III	1	8.1	1.4	8.4	2.0
			2	8.0	1.5	8.6	1.6
			3	8.1	1.5	8.5	1.9
	Stress	I	1	7.5	1.4	7.0	1.0
			2	7.6	1.4	8.3	1.9
			3	7.7	1.5	7.1	1.2
		II	1	7.9	1.5	7.0	1.0
			2	7.7	1.4	7.1	1.2
			3	7.7	1.5	7.2	1.2
		III	1	7.2	1.3	7.4	1.2
			2	7.5	1.4	8.2	1.8
			3	7.8	1.5	8.4	1.9

Table B.2. Raw data statistics for 36 RNA-seq libraries sequenced by MacroGen.

Library	Total Reads	Total Bases (bp)	GC (%)	AT (%)	Q20 (%)	Q30 (%)
WP22 – Control I-1	50,339,390	7,601,247,890	53.3	46.7	97.8	94.4
WP22 – Control I-2	53,069,358	8,013,473,058	55.3	44.7	97.7	94.2
WP22 – Control I-3	50,235,078	7,585,496,778	51.7	48.3	97.4	94.0
WP22 – Control II-1	42,426,694	6,406,430,794	52.5	47.5	97.4	93.4
WP22 – Control II-2	51,537,048	7,782,094,248	55.0	45.0	97.6	94.1
WP22 – Control II-3	41,380,780	6,248,497,780	53.4	46.6	97.1	93.3
WP22 – Control III-1	50,554,924	7,633,793,524	50.2	49.8	97.3	93.9
WP22 – Control III-2	46,129,180	6,965,506,180	54.7	45.3	97.8	94.3
WP22 – Control III-3	52,504,296	7,928,148,696	54.3	45.7	97.8	94.2
WP22 – Stress I-1	50,456,956	7,619,000,356	54.8	45.2	97.8	94.2
WP22 – Stress I-2	55,513,550	8,382,546,050	53.5	46.5	97.6	93.9
WP22 – Stress I-3	48,497,282	7,323,089,582	50.0	50.0	97.2	93.5
WP22 – Stress II-1	45,679,214	6,897,561,314	53.9	46.1	97.3	93.4
WP22 – Stress II-2	52,991,570	8,001,727,070	54.3	45.7	97.5	93.9
WP22 – Stress II-3	51,564,674	7,786,265,774	54.2	45.8	97.6	94.1
WP22 – Stress III-1	56,812,824	8,578,736,424	54.5	45.5	97.7	94.2
WP22 – Stress III-2	48,165,532	7,272,995,332	54.0	46.0	97.8	94.3
WP22 – Stress III-3	48,797,950	7,368,490,450	54.1	45.9	97.6	94.1
WP89 – Control I-1	54,257,250	8,192,844,750	53.9	46.1	97.5	93.8
WP89 – Control I-2	50,546,664	7,632,546,264	50.5	49.5	97.8	94.4
WP89 – Control I-3	54,175,506	8,180,501,406	54.8	45.2	97.7	94.2
WP89 – Control II-1	59,765,144	9,024,536,744	53.2	46.8	97.8	94.1
WP89 – Control II-2	49,392,422	7,458,255,722	54.8	45.2	97.8	94.2
WP89 – Control II-3	47,401,898	7,157,686,598	50.4	49.6	97.1	93.3
WP89 – Control III-1	47,136,676	7,117,638,076	50.0	50.0	97.6	94.3
WP89 – Control III-2	44,757,840	6,758,433,840	50.0	50.0	97.5	94.1
WP89 – Control III-3	52,772,166	7,968,597,066	50.8	49.2	97.9	94.4
WP89 – Stress I-1	44,094,540	6,658,275,540	49.5	50.5	96.9	93.3
WP89 – Stress I-2	52,928,522	7,992,206,822	54.4	45.6	97.8	94.2
WP89 – Stress I-3	43,717,184	6,601,294,784	49.7	50.3	96.9	93.3
WP89 – Stress II-1	54,950,592	8,297,539,392	55.0	45.0	97.9	94.5
WP89 – Stress II-2	52,825,850	7,976,703,350	55.8	44.2	97.3	97.3
WP89 – Stress II-3	51,633,314	7,796,630,414	52.6	47.4	97.6	94.0
WP89 – Stress III-1	53,783,928	8,121,373,128	53.9	46.1	97.5	93.9
WP89 – Stress III-2	53,601,438	8,093,817,138	53.6	46.4	97.3	93.8
WP89 – Stress III-3	51,305,876	7,747,187,276	54.6	45.4	96.9	93.3
TOT	1,815,703,110	274,171,169,610				

Table B.3. The alignment statistics calculated by Samtools for each library.

WP22 – Control I-1	WP22 – Stress I-1
51550165 + 0 in total (QC-passed reads + QC-failed reads)	50219355 + 0 in total (QC-passed reads + QC-failed reads)
50224590 + 0 primary	49297828 + 0 primary
0 + 0 secondary	0 + 0 secondary
1325575 + 0 supplementary	921527 + 0 supplementary
0 + 0 duplicates	0 + 0 duplicates
0 + 0 primary duplicates	0 + 0 primary duplicates
51282466 + 0 mapped (99.48% : N/A)	49989583 + 0 mapped (99.54% : N/A)
49956891 + 0 primary mapped (99.47% : N/A)	49068056 + 0 primary mapped (99.53% : N/A)
50224590 + 0 paired in sequencing	49297828 + 0 paired in sequencing
25112295 + 0 read1	24648914 + 0 read1
25112295 + 0 read2	24648914 + 0 read2
43814304 + 0 properly paired (87.24% : N/A)	44399164 + 0 properly paired (90.06% : N/A)
49703682 + 0 with itself and mate mapped	48852688 + 0 with itself and mate mapped
253209 + 0 singletons (0.50% : N/A)	215368 + 0 singletons (0.44% : N/A)
5616000 + 0 with mate mapped to a different chr	4242624 + 0 with mate mapped to a different chr
2560538 + 0 with mate mapped to a different chr (mapQ>=5)	2106488 + 0 with mate mapped to a different chr (mapQ>=5)
WP22 – Control I-2	WP22 – Stress I-2
47856723 + 0 in total (QC-passed reads + QC-failed reads)	42630012 + 0 in total (QC-passed reads + QC-failed reads)
46728720 + 0 primary	41756638 + 0 primary
0 + 0 secondary	0 + 0 secondary
1128003 + 0 supplementary	873374 + 0 supplementary
0 + 0 duplicates	0 + 0 duplicates
0 + 0 primary duplicates	0 + 0 primary duplicates
47457543 + 0 mapped (99.17% : N/A)	42317776 + 0 mapped (99.27% : N/A)
46329540 + 0 primary mapped (99.15% : N/A)	41444402 + 0 primary mapped (99.25% : N/A)
46728720 + 0 paired in sequencing	41756638 + 0 paired in sequencing
23364360 + 0 read1	20878319 + 0 read1
23364360 + 0 read2	20878319 + 0 read2
42554028 + 0 properly paired (91.07% : N/A)	36749116 + 0 properly paired (88.01% : N/A)
45947810 + 0 with itself and mate mapped	41146882 + 0 with itself and mate mapped
381730 + 0 singletons (0.82% : N/A)	297520 + 0 singletons (0.71% : N/A)
3278682 + 0 with mate mapped to a different chr	4267102 + 0 with mate mapped to a different chr
1410213 + 0 with mate mapped to a different chr (mapQ>=5)	2082523 + 0 with mate mapped to a different chr (mapQ>=5)
WP22 – Control I-3	WP22 – Stress I-3
50275758 + 0 in total (QC-passed reads + QC-failed reads)	41765355 + 0 in total (QC-passed reads + QC-failed reads)
48913964 + 0 primary	40376270 + 0 primary
0 + 0 secondary	0 + 0 secondary
1361794 + 0 supplementary	1389085 + 0 supplementary
0 + 0 duplicates	0 + 0 duplicates
0 + 0 primary duplicates	0 + 0 primary duplicates
49924890 + 0 mapped (99.30% : N/A)	41372711 + 0 mapped (99.06% : N/A)
48563096 + 0 primary mapped (99.28% : N/A)	39983626 + 0 primary mapped (99.03% : N/A)
48913964 + 0 paired in sequencing	40376270 + 0 paired in sequencing
24456982 + 0 read1	20188135 + 0 read1
24456982 + 0 read2	20188135 + 0 read2
42279040 + 0 properly paired (86.44% : N/A)	34224358 + 0 properly paired (84.76% : N/A)
48268472 + 0 with itself and mate mapped	39601948 + 0 with itself and mate mapped
294624 + 0 singletons (0.60% : N/A)	381678 + 0 singletons (0.95% : N/A)
5704894 + 0 with mate mapped to a different chr	5201800 + 0 with mate mapped to a different chr
2964363 + 0 with mate mapped to a different chr (mapQ>=5)	2280629 + 0 with mate mapped to a different chr (mapQ>=5)

WP22 – Control II-1	WP22 – Stress II-1
47433223 + 0 in total (QC-passed reads + QC-failed reads) 46555230 + 0 primary 0 + 0 secondary 877993 + 0 supplementary 0 + 0 duplicates 0 + 0 primary duplicates 46954252 + 0 mapped (98.99% : N/A) 46076259 + 0 primary mapped (98.97% : N/A) 46555230 + 0 paired in sequencing 23277615 + 0 read1 23277615 + 0 read2 42650054 + 0 properly paired (91.61% : N/A) 45618498 + 0 with itself and mate mapped 457761 + 0 singletons (0.98% : N/A) 2767324 + 0 with mate mapped to a different chr 1072190 + 0 with mate mapped to a different chr (mapQ>=5)	46516918 + 0 in total (QC-passed reads + QC-failed reads) 45478176 + 0 primary 0 + 0 secondary 1038742 + 0 supplementary 0 + 0 duplicates 0 + 0 primary duplicates 46314199 + 0 mapped (99.56% : N/A) 45275457 + 0 primary mapped (99.55% : N/A) 45478176 + 0 paired in sequencing 22739088 + 0 read1 22739088 + 0 read2 39845102 + 0 properly paired (87.61% : N/A) 45084976 + 0 with itself and mate mapped 190481 + 0 singletons (0.42% : N/A) 5110120 + 0 with mate mapped to a different chr 2573093 + 0 with mate mapped to a different chr (mapQ>=5)
WP22 – Control II-2	WP22 – Stress II-2
50535536 + 0 in total (QC-passed reads + QC-failed reads) 49877962 + 0 primary 0 + 0 secondary 657574 + 0 supplementary 0 + 0 duplicates 0 + 0 primary duplicates 50307257 + 0 mapped (99.55% : N/A) 49649683 + 0 primary mapped (99.54% : N/A) 49877962 + 0 paired in sequencing 24938981 + 0 read1 24938981 + 0 read2 46213914 + 0 properly paired (92.65% : N/A) 49438908 + 0 with itself and mate mapped 210775 + 0 singletons (0.42% : N/A) 2959028 + 0 with mate mapped to a different chr 1467350 + 0 with mate mapped to a different chr (mapQ>=5)	54808146 + 0 in total (QC-passed reads + QC-failed reads) 53374052 + 0 primary 0 + 0 secondary 1434094 + 0 supplementary 0 + 0 duplicates 0 + 0 primary duplicates 54498594 + 0 mapped (99.44% : N/A) 53064500 + 0 primary mapped (99.42% : N/A) 53374052 + 0 paired in sequencing 26687026 + 0 read1 26687026 + 0 read2 48264386 + 0 properly paired (90.43% : N/A) 52778224 + 0 with itself and mate mapped 286276 + 0 singletons (0.54% : N/A) 4386370 + 0 with mate mapped to a different chr 2004329 + 0 with mate mapped to a different chr (mapQ>=5)
WP22 – Control II-3	WP22 – Stress II-3
50081329 + 0 in total (QC-passed reads + QC-failed reads) 49266352 + 0 primary 0 + 0 secondary 814977 + 0 supplementary 0 + 0 duplicates 0 + 0 primary duplicates 49884316 + 0 mapped (99.61% : N/A) 49069339 + 0 primary mapped (99.60% : N/A) 49266352 + 0 paired in sequencing 24633176 + 0 read1 24633176 + 0 read2 44476474 + 0 properly paired (90.28% : N/A) 48886076 + 0 with itself and mate mapped 183263 + 0 singletons (0.37% : N/A) 4196568 + 0 with mate mapped to a different chr 2146946 + 0 with mate mapped to a different chr (mapQ>=5)	45670906 + 0 in total (QC-passed reads + QC-failed reads) 44914212 + 0 primary 0 + 0 secondary 756694 + 0 supplementary 0 + 0 duplicates 0 + 0 primary duplicates 45239976 + 0 mapped (99.06% : N/A) 44483282 + 0 primary mapped (99.04% : N/A) 44914212 + 0 paired in sequencing 22457106 + 0 read1 22457106 + 0 read2 40964352 + 0 properly paired (91.21% : N/A) 44074044 + 0 with itself and mate mapped 409238 + 0 singletons (0.91% : N/A) 2874354 + 0 with mate mapped to a different chr 1138974 + 0 with mate mapped to a different chr (mapQ>=5)

WP22 – Control III-1	WP22 – Stress III-1
45745789 + 0 in total (QC-passed reads + QC-failed reads)	52157708 + 0 in total (QC-passed reads + QC-failed reads)
44571026 + 0 primary	50634638 + 0 primary
0 + 0 secondary	0 + 0 secondary
1174763 + 0 supplementary	1523070 + 0 supplementary
0 + 0 duplicates	0 + 0 duplicates
0 + 0 primary duplicates	0 + 0 primary duplicates
45386982 + 0 mapped (99.22% : N/A)	51796312 + 0 mapped (99.31% : N/A)
44212219 + 0 primary mapped (99.19% : N/A)	50273242 + 0 primary mapped (99.29% : N/A)
44571026 + 0 paired in sequencing	50634638 + 0 paired in sequencing
22285513 + 0 read1	25317319 + 0 read1
22285513 + 0 read2	25317319 + 0 read2
38840198 + 0 properly paired (87.14% : N/A)	44221620 + 0 properly paired (87.33% : N/A)
43867448 + 0 with itself and mate mapped	49930286 + 0 with itself and mate mapped
344771 + 0 singletons (0.77% : N/A)	342956 + 0 singletons (0.68% : N/A)
4882064 + 0 with mate mapped to a different chr	5518232 + 0 with mate mapped to a different chr
2315679 + 0 with mate mapped to a different chr (mapQ>=5)	2515927 + 0 with mate mapped to a different chr (mapQ>=5)
WP22 – Control III-2	WP22 – Stress III-2
57017206 + 0 in total (QC-passed reads + QC-failed reads)	51170858 + 0 in total (QC-passed reads + QC-failed reads)
55412368 + 0 primary	49583092 + 0 primary
0 + 0 secondary	0 + 0 secondary
1604838 + 0 supplementary	1587766 + 0 supplementary
0 + 0 duplicates	0 + 0 duplicates
0 + 0 primary duplicates	0 + 0 primary duplicates
56696536 + 0 mapped (99.44% : N/A)	50872562 + 0 mapped (99.42% : N/A)
55091698 + 0 primary mapped (99.42% : N/A)	49284796 + 0 primary mapped (99.40% : N/A)
55412368 + 0 paired in sequencing	49583092 + 0 paired in sequencing
27706184 + 0 read1	24791546 + 0 read1
27706184 + 0 read2	24791546 + 0 read2
48209532 + 0 properly paired (87.00% : N/A)	42702414 + 0 properly paired (86.12% : N/A)
54786426 + 0 with itself and mate mapped	49015974 + 0 with itself and mate mapped
305272 + 0 singletons (0.55% : N/A)	268822 + 0 singletons (0.54% : N/A)
6377888 + 0 with mate mapped to a different chr	6085610 + 0 with mate mapped to a different chr
3016198 + 0 with mate mapped to a different chr (mapQ>=5)	2752347 + 0 with mate mapped to a different chr (mapQ>=5)
WP22 – Control III-3	WP22 – Stress III-3
48747537 + 0 in total (QC-passed reads + QC-failed reads)	47217888 + 0 in total (QC-passed reads + QC-failed reads)
47340798 + 0 primary	46258164 + 0 primary
0 + 0 secondary	0 + 0 secondary
1406739 + 0 supplementary	959724 + 0 supplementary
0 + 0 duplicates	0 + 0 duplicates
0 + 0 primary duplicates	0 + 0 primary duplicates
48440429 + 0 mapped (99.37% : N/A)	47010610 + 0 mapped (99.56% : N/A)
47033690 + 0 primary mapped (99.35% : N/A)	46050886 + 0 primary mapped (99.55% : N/A)
47340798 + 0 paired in sequencing	46258164 + 0 paired in sequencing
23670399 + 0 read1	23129082 + 0 read1
23670399 + 0 read2	23129082 + 0 read2
41052134 + 0 properly paired (86.72% : N/A)	41438998 + 0 properly paired (89.58% : N/A)
46740576 + 0 with itself and mate mapped	45859136 + 0 with itself and mate mapped
293114 + 0 singletons (0.62% : N/A)	191750 + 0 singletons (0.41% : N/A)
5472202 + 0 with mate mapped to a different chr	4200352 + 0 with mate mapped to a different chr
2465721 + 0 with mate mapped to a different chr (mapQ>=5)	1965083 + 0 with mate mapped to a different chr (mapQ>=5)

WP89 – Control I-1	WP89 – Stress I-1
53536253 + 0 in total (QC-passed reads + QC-failed reads) 52154616 + 0 primary 0 + 0 secondary 1381637 + 0 supplementary 0 + 0 duplicates 0 + 0 primary duplicates 53191453 + 0 mapped (99.36% : N/A) 51809816 + 0 primary mapped (99.34% : N/A) 52154616 + 0 paired in sequencing 26077308 + 0 read1 26077308 + 0 read2 46221222 + 0 properly paired (88.62% : N/A) 51479860 + 0 with itself and mate mapped 329956 + 0 singletons (0.63% : N/A) 5054954 + 0 with mate mapped to a different chr 2335500 + 0 with mate mapped to a different chr (mapQ>=5)	47671431 + 0 in total (QC-passed reads + QC-failed reads) 47165280 + 0 primary 0 + 0 secondary 506151 + 0 supplementary 0 + 0 duplicates 0 + 0 primary duplicates 47352855 + 0 mapped (99.33% : N/A) 46846704 + 0 primary mapped (99.32% : N/A) 47165280 + 0 paired in sequencing 23582640 + 0 read1 23582640 + 0 read2 44149368 + 0 properly paired (93.61% : N/A) 46548742 + 0 with itself and mate mapped 297962 + 0 singletons (0.63% : N/A) 2119228 + 0 with mate mapped to a different chr 930059 + 0 with mate mapped to a different chr (mapQ>=5)
WP89 – Control I-2	WP89 – Stress I-2
53372639 + 0 in total (QC-passed reads + QC-failed reads) 52186998 + 0 primary 0 + 0 secondary 1185641 + 0 supplementary 0 + 0 duplicates 0 + 0 primary duplicates 53120111 + 0 mapped (99.53% : N/A) 51934470 + 0 primary mapped (99.52% : N/A) 52186998 + 0 paired in sequencing 26093499 + 0 read1 26093499 + 0 read2 46365920 + 0 properly paired (88.85% : N/A) 51697802 + 0 with itself and mate mapped 236668 + 0 singletons (0.45% : N/A) 5096222 + 0 with mate mapped to a different chr 2379381 + 0 with mate mapped to a different chr (mapQ>=5)	57689138 + 0 in total (QC-passed reads + QC-failed reads) 56898616 + 0 primary 0 + 0 secondary 790522 + 0 supplementary 0 + 0 duplicates 0 + 0 primary duplicates 57480595 + 0 mapped (99.64% : N/A) 56690073 + 0 primary mapped (99.63% : N/A) 56898616 + 0 paired in sequencing 28449308 + 0 read1 28449308 + 0 read2 52705932 + 0 properly paired (92.63% : N/A) 56503902 + 0 with itself and mate mapped 186171 + 0 singletons (0.33% : N/A) 3482984 + 0 with mate mapped to a different chr 1661535 + 0 with mate mapped to a different chr (mapQ>=5)
WP89 – Control I-3	WP89 – Stress I-3
48208655 + 0 in total (QC-passed reads + QC-failed reads) 47415492 + 0 primary 0 + 0 secondary 793163 + 0 supplementary 0 + 0 duplicates 0 + 0 primary duplicates 47933456 + 0 mapped (99.43% : N/A) 47140293 + 0 primary mapped (99.42% : N/A) 47415492 + 0 paired in sequencing 23707746 + 0 read1 23707746 + 0 read2 43222394 + 0 properly paired (91.16% : N/A) 46878286 + 0 with itself and mate mapped 262007 + 0 singletons (0.55% : N/A) 3445278 + 0 with mate mapped to a different chr 1689607 + 0 with mate mapped to a different chr (mapQ>=5)	43954761 + 0 in total (QC-passed reads + QC-failed reads) 43320140 + 0 primary 0 + 0 secondary 634621 + 0 supplementary 0 + 0 duplicates 0 + 0 primary duplicates 43497705 + 0 mapped (98.96% : N/A) 42863084 + 0 primary mapped (98.94% : N/A) 43320140 + 0 paired in sequencing 21660070 + 0 read1 21660070 + 0 read2 40394126 + 0 properly paired (93.25% : N/A) 42428980 + 0 with itself and mate mapped 434104 + 0 singletons (1.00% : N/A) 1912080 + 0 with mate mapped to a different chr 762690 + 0 with mate mapped to a different chr (mapQ>=5)

WP89 – Control II-1	WP89 – Stress II-1
41743273 + 0 in total (QC-passed reads + QC-failed reads) 41125096 + 0 primary 0 + 0 secondary 618177 + 0 supplementary 0 + 0 duplicates 0 + 0 primary duplicates 41388092 + 0 mapped (99.15% : N/A) 40769915 + 0 primary mapped (99.14% : N/A) 41125096 + 0 paired in sequencing 20562548 + 0 read1 20562548 + 0 read2 38302146 + 0 properly paired (93.14% : N/A) 40431560 + 0 with itself and mate mapped 338355 + 0 singletons (0.82% : N/A) 1952772 + 0 with mate mapped to a different chr 738992 + 0 with mate mapped to a different chr (mapQ>=5)	43753636 + 0 in total (QC-passed reads + QC-failed reads) 43224536 + 0 primary 0 + 0 secondary 529100 + 0 supplementary 0 + 0 duplicates 0 + 0 primary duplicates 43382160 + 0 mapped (99.15% : N/A) 42853060 + 0 primary mapped (99.14% : N/A) 43224536 + 0 paired in sequencing 21612268 + 0 read1 21612268 + 0 read2 40635790 + 0 properly paired (94.01% : N/A) 42504518 + 0 with itself and mate mapped 348542 + 0 singletons (0.81% : N/A) 1732974 + 0 with mate mapped to a different chr 675871 + 0 with mate mapped to a different chr (mapQ>=5)
WP89 – Control II-2	WP89 – Stress II-2
48764882 + 0 in total (QC-passed reads + QC-failed reads) 48032840 + 0 primary 0 + 0 secondary 732042 + 0 supplementary 0 + 0 duplicates 0 + 0 primary duplicates 48542074 + 0 mapped (99.54% : N/A) 47810032 + 0 primary mapped (99.54% : N/A) 48032840 + 0 paired in sequencing 24016420 + 0 read1 24016420 + 0 read2 45514622 + 0 properly paired (94.76% : N/A) 47608310 + 0 with itself and mate mapped 201722 + 0 singletons (0.42% : N/A) 1961858 + 0 with mate mapped to a different chr 789867 + 0 with mate mapped to a different chr (mapQ>=5)	41882087 + 0 in total (QC-passed reads + QC-failed reads) 40924040 + 0 primary 0 + 0 secondary 958047 + 0 supplementary 0 + 0 duplicates 0 + 0 primary duplicates 41372665 + 0 mapped (98.78% : N/A) 40414618 + 0 primary mapped (98.76% : N/A) 40924040 + 0 paired in sequencing 20462020 + 0 read1 20462020 + 0 read2 36721582 + 0 properly paired (89.73% : N/A) 39930014 + 0 with itself and mate mapped 484604 + 0 singletons (1.18% : N/A) 3048530 + 0 with mate mapped to a different chr 1165496 + 0 with mate mapped to a different chr (mapQ>=5)
WP89 – Control II-3	WP89 – Stress II-3
51858582 + 0 in total (QC-passed reads + QC-failed reads) 51145974 + 0 primary 0 + 0 secondary 712608 + 0 supplementary 0 + 0 duplicates 0 + 0 primary duplicates 51637035 + 0 mapped (99.57% : N/A) 50924427 + 0 primary mapped (99.57% : N/A) 51145974 + 0 paired in sequencing 25572987 + 0 read1 25572987 + 0 read2 47062160 + 0 properly paired (92.02% : N/A) 50718400 + 0 with itself and mate mapped 206027 + 0 singletons (0.40% : N/A) 3440630 + 0 with mate mapped to a different chr 1744604 + 0 with mate mapped to a different chr (mapQ>=5)	41516274 + 0 in total (QC-passed reads + QC-failed reads) 40222532 + 0 primary 0 + 0 secondary 1293742 + 0 supplementary 0 + 0 duplicates 0 + 0 primary duplicates 41035048 + 0 mapped (98.84% : N/A) 39741306 + 0 primary mapped (98.80% : N/A) 40222532 + 0 paired in sequencing 20111266 + 0 read1 20111266 + 0 read2 35696682 + 0 properly paired (88.75% : N/A) 39280452 + 0 with itself and mate mapped 460854 + 0 singletons (1.15% : N/A) 3474758 + 0 with mate mapped to a different chr 1318576 + 0 with mate mapped to a different chr (mapQ>=5)

WP89 – Control III-1	WP89 – Stress III-1
52577296 + 0 in total (QC-passed reads + QC-failed reads) 51924254 + 0 primary 0 + 0 secondary 653042 + 0 supplementary 0 + 0 duplicates 0 + 0 primary duplicates 52359269 + 0 mapped (99.59% : N/A) 51706227 + 0 primary mapped (99.58% : N/A) 51924254 + 0 paired in sequencing 25962127 + 0 read1 25962127 + 0 read2 48674936 + 0 properly paired (93.74% : N/A) 51505284 + 0 with itself and mate mapped 200943 + 0 singletons (0.39% : N/A) 2605528 + 0 with mate mapped to a different chr 1323560 + 0 with mate mapped to a different chr (mapQ>=5)	51059633 + 0 in total (QC-passed reads + QC-failed reads) 49353998 + 0 primary 0 + 0 secondary 1705635 + 0 supplementary 0 + 0 duplicates 0 + 0 primary duplicates 50720576 + 0 mapped (99.34% : N/A) 49014941 + 0 primary mapped (99.31% : N/A) 49353998 + 0 paired in sequencing 24676999 + 0 read1 24676999 + 0 read2 43346080 + 0 properly paired (87.83% : N/A) 48693658 + 0 with itself and mate mapped 321283 + 0 singletons (0.65% : N/A) 5164790 + 0 with mate mapped to a different chr 2235354 + 0 with mate mapped to a different chr (mapQ>=5)
WP89 – Control III-2	WP89 – Stress III-2
51334166 + 0 in total (QC-passed reads + QC-failed reads) 49499172 + 0 primary 0 + 0 secondary 1834994 + 0 supplementary 0 + 0 duplicates 0 + 0 primary duplicates 50988803 + 0 mapped (99.33% : N/A) 49153809 + 0 primary mapped (99.30% : N/A) 49499172 + 0 paired in sequencing 24749586 + 0 read1 24749586 + 0 read2 41971184 + 0 properly paired (84.79% : N/A) 48831604 + 0 with itself and mate mapped 322205 + 0 singletons (0.65% : N/A) 6586988 + 0 with mate mapped to a different chr 2886781 + 0 with mate mapped to a different chr (mapQ>=5)	52830768 + 0 in total (QC-passed reads + QC-failed reads) 51241858 + 0 primary 0 + 0 secondary 1588910 + 0 supplementary 0 + 0 duplicates 0 + 0 primary duplicates 52378457 + 0 mapped (99.14% : N/A) 50789547 + 0 primary mapped (99.12% : N/A) 51241858 + 0 paired in sequencing 25620929 + 0 read1 25620929 + 0 read2 44761906 + 0 properly paired (87.35% : N/A) 50356820 + 0 with itself and mate mapped 432727 + 0 singletons (0.84% : N/A) 5400278 + 0 with mate mapped to a different chr 2411743 + 0 with mate mapped to a different chr (mapQ>=5)
WP89 – Control III-3	WP89 – Stress III-3
53201555 + 0 in total (QC-passed reads + QC-failed reads) 51568080 + 0 primary 0 + 0 secondary 1633475 + 0 supplementary 0 + 0 duplicates 0 + 0 primary duplicates 52826504 + 0 mapped (99.30% : N/A) 51193029 + 0 primary mapped (99.27% : N/A) 51568080 + 0 paired in sequencing 25784040 + 0 read1 25784040 + 0 read2 45297586 + 0 properly paired (87.84% : N/A) 50833876 + 0 with itself and mate mapped 359153 + 0 singletons (0.70% : N/A) 5347030 + 0 with mate mapped to a different chr 2417651 + 0 with mate mapped to a different chr (mapQ>=5)	51436472 + 0 in total (QC-passed reads + QC-failed reads) 48849054 + 0 primary 0 + 0 secondary 2587418 + 0 supplementary 0 + 0 duplicates 0 + 0 primary duplicates 50844643 + 0 mapped (98.85% : N/A) 48257225 + 0 primary mapped (98.79% : N/A) 48849054 + 0 paired in sequencing 24424527 + 0 read1 24424527 + 0 read2 39528648 + 0 properly paired (80.92% : N/A) 47681640 + 0 with itself and mate mapped 575585 + 0 singletons (1.18% : N/A) 7908642 + 0 with mate mapped to a different chr 3383892 + 0 with mate mapped to a different chr (mapQ>=5)

Table B.4. The sizes of 36 RNA-seq libraries.

Library	Size	Size after filtering
WP22 – Control I-1	23623663	23435313
WP22 – Control I-2	22846052	22689806
WP22 – Control I-3	22939137	20271059
WP22 – Control II-1	22764395	22594634
WP22 – Control II-2	24109250	23896748
WP22 – Control II-3	23343474	23161933
WP22 – Control III-1	21013055	20855955
WP22 – Control III-2	26115383	25869279
WP22 – Control III-3	22335112	22116445
WP22 – Stress I-1	23442619	23218571
WP22 – Stress I-2	19611515	19397360
WP22 – Stress I-3	18972102	18794795
WP22 – Stress II-1	21217201	21060538
WP22 – Stress II-2	25917238	25649714
WP22 – Stress II-3	21768035	21550814
WP22 – Stress III-1	24073456	23825971
WP22 – Stress III-2	23322810	23072495
WP22 – Stress III-3	21981760	21751928
WP89 – Control I-1	24925085	24715020
WP89 – Control I-2	24723900	24525894
WP89 – Control I-3	22772388	22608441
WP89 – Control II-1	20198799	20047654
WP89 – Control II-2	23758809	23543201
WP89 – Control II-3	24558106	24334617
WP89 – Control III-1	25304433	25099095
WP89 – Control III-2	23280575	23060166
WP89 – Control III-3	24737504	24515910
WP89 – Stress I-1	23020679	22795501
WP89 – Stress I-2	27487950	27223453
WP89 – Stress I-3	21329357	21108127
WP89 – Stress II-1	21265397	21111614
WP89 – Stress II-2	19886000	19690001
WP89 – Stress II-3	19660583	19457999
WP89 – Stress III-1	23793507	23511883
WP89 – Stress III-2	24501605	24161196
WP89 – Stress III-3	23052041	22764636

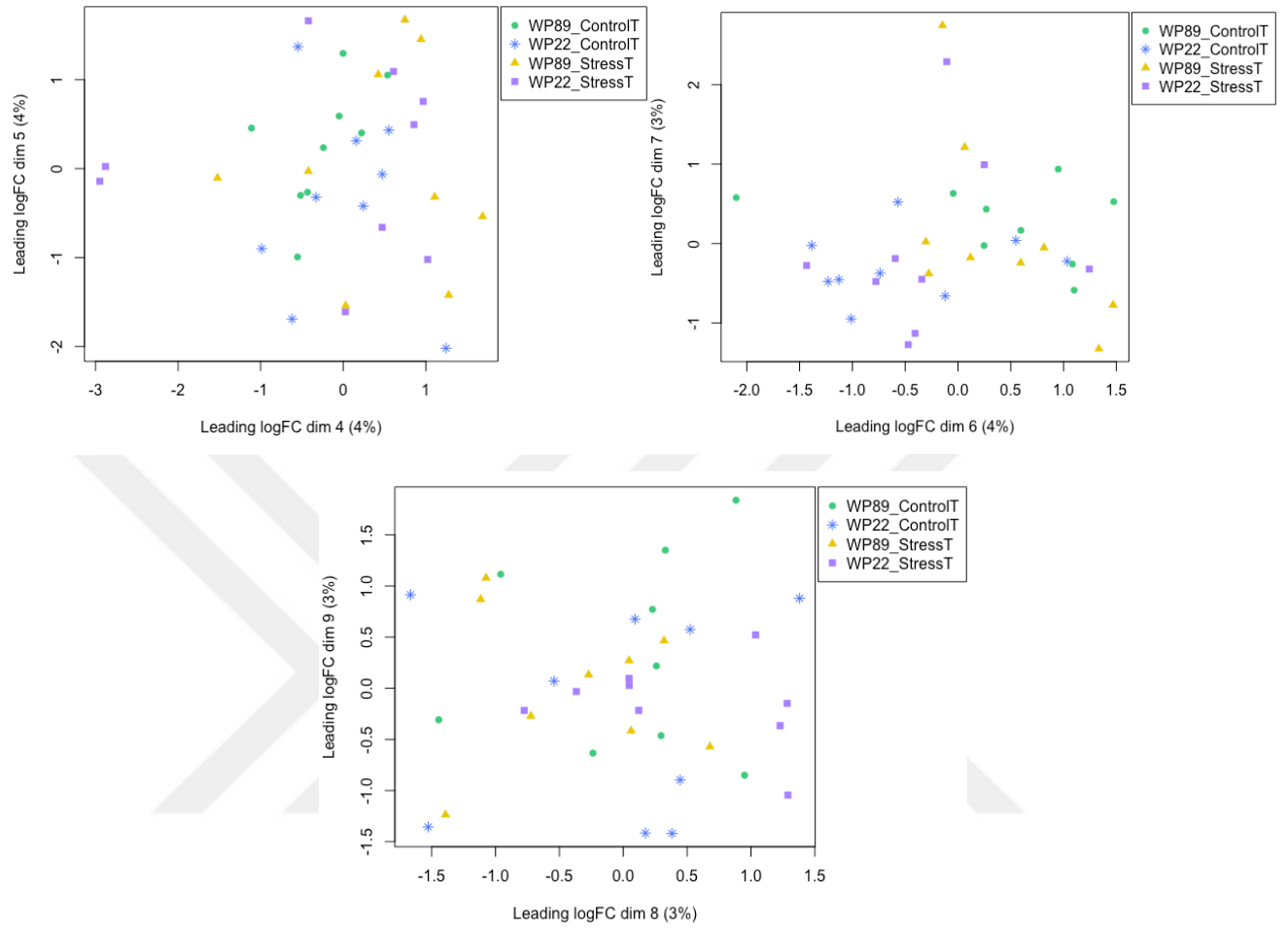


Figure B.1. MDS plots of four einkorn groups for higher dimensions.

Table B.5. The top tag and bottom tag enriched Reactome pathways for control temperature.

Pathway	NES
Gluconeogenesis (R-CFA-70263)	1.8811748
Gluconeogenesis (R-XTR-70263)	1.8348413
ABC-family proteins mediated transport (R-CEL-382556)	1.8290145
ABC-family proteins mediated transport (R-HSA-382556)	1.8192341
ABC-family proteins mediated transport (R-SCE-382556)	1.8169509
ABC-family proteins mediated transport (R-CFA-382556)	1.8168966
ABC-family proteins mediated transport (R-XTR-382556)	1.8131032
ABC-family proteins mediated transport (R-DME-382556)	1.812362
ABC-family proteins mediated transport (R-DRE-382556)	1.8106983
ABC-family proteins mediated transport (R-SSC-382556)	1.810178
Pentose phosphate pathway (R-XTR-71336)	-1.7622072
Pentose phosphate pathway (R-RNO-71336)	-1.748873
Pentose phosphate pathway (R-DME-71336)	-1.740556
Pentose phosphate pathway (R-SCE-71336)	-1.7396436
Pentose phosphate pathway (R-SPO-71336)	-1.7276226
Pentose phosphate pathway (R-DDI-71336)	-1.7198534
Pentose phosphate pathway (R-DRE-71336)	-1.709749
Pentose phosphate pathway (R-MMU-71336)	-1.7094065
The proton buffering model (R-XTR-167827)	-1.6552081
The proton buffering model (R-SSC-167827)	-1.6468319

Table B.6. The top tag and bottom tag enriched Reactome pathways for stress temperature.

Pathway	NES
Minus-strand DNA synthesis (R-HSA-164516)	2.163714
Fanconi Anemia Pathway (R-HSA-6783310)	1.962847
Fanconi Anemia Pathway (R-MMU-6783310)	1.9355344
Fanconi Anemia Pathway (R-DRE-6783310)	1.9310256
Fanconi Anemia Pathway (R-RNO-6783310)	1.9180901
Fanconi Anemia Pathway (R-CFA-6783310)	1.8611637
Passive transport by Aquaporins (R-DME-432047)	1.7391146
Passive transport by Aquaporins (R-GGA-432047)	1.7373463
Passive transport by Aquaporins (R-CFA-432047)	1.7358212
Passive transport by Aquaporins (R-DDI-432047)	1.7271397
ABC-family proteins mediated transport (R-XTR-382556)	-1.8214654
ABC-family proteins mediated transport (R-PFA-382556)	-1.8105714
Methionine salvage pathway (R-CEL-1237112)	-1.7925386
ABC-family proteins mediated transport (R-GGA-382556)	-1.7857283
Methionine salvage pathway (R-DDI-1237112)	-1.7796807
Mitochondrial fatty acid beta-oxidation of unsaturated fatty acids (R-MMU-77288)	-1.77685
Mitochondrial Fatty Acid Beta-Oxidation (R-DRE-77289)	-1.7760487
mitochondrial fatty acid beta-oxidation of unsaturated fatty acids (R-RNO-77288)	-1.7748717
Methionine salvage pathway (R-DRE-1237112)	-1.7728695
Methionine salvage pathway (R-SCE-1237112)	-1.7677262

Table B.7. The top 100 DE genes.

Contig	Contrast	logFC	P Value	FDR
TRINITY_DN548_c1_g1_i1	Control_Stress_New	-5.58304	2.42E-26	2.17E-21
TRINITY_DN548_c1_g1_i1	Control_Stress_Old	-5.36218	6.16E-26	5.52E-21
TRINITY_DN8908_c0_g1_i15	Control_Stress_New	4.729033	7.37E-24	3.30E-19
TRINITY_DN15822_c0_g1_i4	Control_Stress_New	-3.79848	2.08E-23	6.22E-19
TRINITY_DN8908_c0_g1_i16	Control_Stress_New	4.885646	3.91E-23	8.75E-19
TRINITY_DN951_c0_g1_i63	Control_Stress_New	8.361314	1.69E-22	3.04E-18
TRINITY_DN8908_c0_g1_i14	Control_Stress_New	4.652798	2.65E-22	3.96E-18
TRINITY_DN9888_c0_g1_i2	Control_Stress_New	5.584215	3.29E-22	4.21E-18
TRINITY_DN8908_c0_g1_i35	Control_Stress_New	4.815801	4.87E-22	4.85E-18
TRINITY_DN951_c0_g1_i65	Control_Stress_New	8.404321	4.88E-22	4.85E-18
TRINITY_DN15822_c0_g1_i4	Control_Stress_Old	-3.51918	1.75E-22	5.69E-18
TRINITY_DN9888_c0_g1_i2	Control_Stress_Old	5.677632	1.95E-22	5.69E-18
TRINITY_DN8908_c0_g1_i15	Control_Stress_Old	4.135666	2.54E-22	5.69E-18
TRINITY_DN3971_c0_g1_i3	Control_Stress_Old	1.520229	6.27E-22	9.80E-18
TRINITY_DN951_c0_g1_i63	Control_Stress_Old	7.426156	6.56E-22	9.80E-18
TRINITY_DN951_c0_g1_i73	Control_Stress_New	8.442899	1.13E-21	1.01E-17
TRINITY_DN2192_c0_g1_i10	Control_Stress_New	-5.42866	1.53E-21	1.25E-17
TRINITY_DN3059_c0_g1_i59	Control_Stress_New	4.503977	2.37E-21	1.77E-17
TRINITY_DN18960_c0_g3_i1	Control_Stress_New	-6.08397	2.69E-21	1.85E-17
TRINITY_DN1429_c3_g1_i1	Control_Stress_New	-7.61141	2.89E-21	1.85E-17
TRINITY_DN8908_c0_g1_i17	Control_Stress_New	4.778292	3.28E-21	1.88E-17
TRINITY_DN1091_c0_g1_i5	Control_Stress_New	-4.43185	3.36E-21	1.88E-17
TRINITY_DN951_c0_g1_i65	Control_Stress_Old	7.454535	2.19E-21	2.54E-17
TRINITY_DN8908_c0_g1_i16	Control_Stress_Old	4.173462	2.27E-21	2.54E-17
TRINITY_DN951_c0_g1_i73	Control_Stress_Old	7.777138	2.90E-21	2.78E-17
TRINITY_DN7239_c1_g1_i30	Control_Stress_Old	2.794789	3.11E-21	2.78E-17
TRINITY_DN18960_c0_g3_i1	Control_Stress_Old	-5.98287	4.15E-21	3.38E-17
TRINITY_DN21581_c0_g1_i3	Control_Stress_Old	-5.14731	6.66E-21	4.81E-17
TRINITY_DN537_c1_g1_i10	Control_Stress_Old	1.994875	7.31E-21	4.81E-17
TRINITY_DN8908_c0_g1_i14	Control_Stress_Old	4.092848	7.52E-21	4.81E-17
TRINITY_DN3971_c0_g1_i3	Control_Stress_New	1.39376	9.45E-21	4.98E-17
TRINITY_DN951_c0_g1_i41	Control_Stress_New	8.696689	1.04E-20	5.19E-17
TRINITY_DN8908_c0_g1_i35	Control_Stress_Old	4.30224	9.08E-21	5.28E-17
TRINITY_DN7239_c1_g1_i32	Control_Stress_Old	2.737324	9.42E-21	5.28E-17
TRINITY_DN5020_c1_g1_i7	Control_Stress_New	2.258454	1.30E-20	6.11E-17
TRINITY_DN16707_c1_g1_i4	Control_Stress_New	2.818234	1.39E-20	6.21E-17
TRINITY_DN2192_c0_g1_i10	Control_Stress_Old	-4.94184	1.27E-20	6.45E-17
TRINITY_DN1429_c3_g1_i1	Control_Stress_Old	-7.07309	1.30E-20	6.45E-17
TRINITY_DN1091_c0_g1_i5	Control_Stress_Old	-4.18774	1.38E-20	6.48E-17
TRINITY_DN39_c14_g1_i1	Control_Stress_Old	2.723138	1.45E-20	6.48E-17

TRINITY_DN3971_c0_g1_i23	Control_Stress_Old	1.483399	1.77E-20	7.41E-17
TRINITY_DN1894_c0_g1_i8	Control_Stress_Old	2.061879	1.82E-20	7.41E-17
TRINITY_DN3465_c0_g2_i2	Control_Stress_New	-14.3498	1.84E-20	7.54E-17
TRINITY_DN1894_c0_g1_i8	Control_Stress_New	2.063917	1.85E-20	7.54E-17
TRINITY_DN4434_c0_g1_i30	Control_Stress_Old	2.989593	2.25E-20	8.75E-17
TRINITY_DN5509_c0_g1_i2	Control_Stress_Old	-4.59362	2.42E-20	9.05E-17
TRINITY_DN3971_c0_g1_i21	Control_Stress_Old	1.513133	2.82E-20	1.01E-16
TRINITY_DN7239_c1_g1_i52	Control_Stress_Old	2.760609	3.23E-20	1.11E-16
TRINITY_DN2192_c0_g1_i2	Control_Stress_New	-5.05877	3.40E-20	1.26E-16
TRINITY_DN3059_c0_g1_i30	Control_Stress_New	4.472447	3.46E-20	1.26E-16
TRINITY_DN537_c1_g1_i10	Control_Stress_New	1.892362	3.51E-20	1.26E-16
TRINITY_DN7239_c1_g1_i48	Control_Stress_Old	2.778396	4.50E-20	1.49E-16
TRINITY_DN3897_c0_g1_i6	Control_Stress_Old	3.178948	4.97E-20	1.59E-16
TRINITY_DN951_c0_g1_i41	Control_Stress_Old	7.428851	6.07E-20	1.88E-16
TRINITY_DN1894_c0_g1_i6	Control_Stress_New	2.03628	5.52E-20	1.90E-16
TRINITY_DN1789_c0_g2_i14	Control_Stress_New	2.608387	5.95E-20	1.96E-16
TRINITY_DN3971_c0_g1_i21	Control_Stress_New	1.476039	6.13E-20	1.96E-16
TRINITY_DN3465_c0_g2_i2	Control_Stress_Old	-12.33	6.83E-20	2.02E-16
TRINITY_DN7239_c1_g1_i22	Control_Stress_Old	2.748656	6.98E-20	2.02E-16
TRINITY_DN9724_c45_g1_i1	Control_Stress_New	-8.41618	6.96E-20	2.15E-16
TRINITY_DN2690_c0_g1_i5	Control_Stress_Old	-11.3387	8.20E-20	2.30E-16
TRINITY_DN1559_c0_g1_i137	Control_Stress_Old	2.935453	8.59E-20	2.33E-16
TRINITY_DN1559_c0_g1_i180	Control_Stress_Old	3.215062	9.23E-20	2.43E-16
TRINITY_DN21206_c0_g1_i2	Control_Stress_Old	-3.17299	1.04E-19	2.66E-16
TRINITY_DN142_c1_g1_i1	Control_Stress_Old	-9.51535	1.08E-19	2.70E-16
TRINITY_DN3971_c0_g1_i23	Control_Stress_New	1.405484	9.35E-20	2.79E-16
TRINITY_DN21206_c0_g1_i2	Control_Stress_New	-3.16731	1.03E-19	2.91E-16
TRINITY_DN3465_c0_g2_i1	Control_Stress_New	-12.7898	1.04E-19	2.91E-16
TRINITY_DN7239_c1_g1_i30	Control_Stress_New	2.466178	1.12E-19	3.05E-16
TRINITY_DN49642_c0_g1_i7	Control_Stress_New	-4.54092	1.16E-19	3.05E-16
TRINITY_DN8908_c0_g1_i17	Control_Stress_Old	4.138674	1.27E-19	3.08E-16
TRINITY_DN6193_c0_g1_i6	Control_Stress_New	-3.71876	1.27E-19	3.17E-16
TRINITY_DN16707_c1_g1_i13	Control_Stress_New	2.770154	1.27E-19	3.17E-16
TRINITY_DN3059_c0_g1_i59	Control_Stress_Old	3.857375	1.35E-19	3.18E-16
TRINITY_DN6193_c0_g1_i23	Control_Stress_New	-3.42639	1.42E-19	3.35E-16
TRINITY_DN5020_c1_g1_i5	Control_Stress_New	2.145436	1.42E-19	3.35E-16
TRINITY_DN2192_c0_g1_i2	Control_Stress_Old	-4.7373	1.48E-19	3.40E-16
TRINITY_DN16707_c1_g1_i2	Control_Stress_New	2.814041	1.51E-19	3.46E-16
TRINITY_DN3059_c0_g1_i91	Control_Stress_New	4.522146	1.58E-19	3.53E-16
TRINITY_DN5509_c0_g1_i2	Control_Stress_New	-4.20585	1.62E-19	3.53E-16
TRINITY_DN9724_c45_g1_i1	Control_Stress_Old	-8.04763	1.73E-19	3.86E-16
TRINITY_DN3059_c0_g1_i79	Control_Stress_New	4.546911	1.89E-19	3.99E-16
TRINITY_DN49642_c0_g1_i2	Control_Stress_New	-4.6764	1.93E-19	3.99E-16

TRINITY_DN49642_c0_g1_i13	Control_Stress_New	-4.71325	1.96E-19	3.99E-16
TRINITY_DN49642_c0_g1_i7	Control_Stress_Old	-4.50159	1.89E-19	4.12E-16
TRINITY_DN3465_c0_g2_i1	Control_Stress_Old	-11.871	2.05E-19	4.38E-16
TRINITY_DN21581_c0_g1_i3	Control_Stress_New	-4.39065	2.35E-19	4.67E-16
TRINITY_DN1559_c0_g1_i31	Control_Stress_Old	2.903299	2.36E-19	4.83E-16
TRINITY_DN1894_c0_g1_i6	Control_Stress_Old	1.937425	2.41E-19	4.83E-16
TRINITY_DN1175_c0_g1_i67	Control_Stress_Old	2.721406	2.43E-19	4.83E-16
TRINITY_DN3897_c0_g1_i6	Control_Stress_New	3.008659	2.52E-19	4.91E-16
TRINITY_DN3971_c0_g1_i5	Control_Stress_Old	1.457329	2.56E-19	4.99E-16
TRINITY_DN4434_c0_g1_i27	Control_Stress_Old	2.708808	2.71E-19	5.17E-16
TRINITY_DN2690_c0_g1_i5	Control_Stress_New	-10.4998	2.73E-19	5.20E-16
TRINITY_DN7239_c1_g1_i32	Control_Stress_New	2.431276	2.81E-19	5.25E-16
TRINITY_DN44_c0_g2_i3	Control_Stress_Old	-2.69112	2.98E-19	5.56E-16
TRINITY_DN6193_c0_g1_i15	Control_Stress_New	-3.16319	3.08E-19	5.63E-16
TRINITY_DN5509_c0_g1_i7	Control_Stress_New	-2.32273	3.26E-19	5.82E-16
TRINITY_DN1894_c0_g1_i2	Control_Stress_New	2.137142	3.31E-19	5.82E-16
TRINITY_DN1275_c0_g1_i7	Control_Stress_New	-2.34508	3.99E-19	6.74E-16