Genomic analysis of emmer wheat: insights into domestication, dispersal and adaptation

Alice Iob

Genomic analysis of emmer wheat: insights into domestication, dispersal and adaptation

Alice Iob

DOCTORAL THESIS
to achieve the PhD degree in Genetics of the Universitat Autònoma de Barcelona, 2023

THESIS DIRECTOR
Dr. Laura R. Botigué

THESIS TUTOR
Dr. Sònia Casillas

UNIVERSITAT AUTÒNOMA DE BARCELONA
Departamento de Genética y de Microbiología

CENTRE FOR RESEARCH IN AGRICULTURAL GENOMICS
Plant and Animal Genomics program

Thesis submitted by Alice Iob in fulfilment of the requirements for the degree of Doctor by the Universitat Autonoma de Barcelona, under the supervision of Dr. Laura R. Botigué

Bellaterra, November 2023

Author

Alice Iob

Thesis supervisor

Laura R. Botigué

Cover designed by Elizabeth Anna Rovere Gaddi, November 2023

This work was funded by projects from the Spanish Ministerio de Ciencia e Innovación-Agencia Estatal de Investigación/Fondo Social Europeo, and from the Spanish Agencia Estatal de Investigación, through the "Severo Ochoa Programme for Centres of Excellence in R&D" with project references SEV-2015-0533 and CEX2019-000902-S. This work was also supported by the CERCA programme by the Generalitat de Catalunya.

Alice Iob was financially supported by a Formación del Personal Investigador (FPI) grant PRE2018-083529

Abstract

Emmer wheat, among the earliest plants domesticated in Southwest Asia, is the progenitor of the economically important durum and bread wheat. While it is acknowledged that domesticated emmer emerged from an admixture of wild wheat populations, details behind this process remain obscure. After domestication, emmer spread outside of Southwest Asia, adapting to a multitude of ecosystems. Open questions persist regarding the most likely route of dispersal to Africa and Asia, with various hypotheses being proposed. In this study, I use whole genome sequences from wild and domestic emmer wheat specimens to gain insights into these critical events and understand the role of wild ancestry in domestic populations. I analyse the population structure within the dataset and I delve into the wild ancestry of domestic landraces, estimating for the first time the contribution of each wild population to the domesticated ones. The results obtained, combined with archaeological evidence, provide a deeper and more detailed understanding of the domestication process. Around 9500 years ago, protodomestic emmer from the Southern Levant hybridized with its counterpart from the Northern Levant, giving rise to fully domesticated plants. Notably, I observe a higher proportion of wild ancestry from the Southern Levant in the population that dispersed southward and eastward, as opposed to the one that moved northward and westward. This discrepancy is attributed to post-domestication gene flow during dispersal approximately 6500 years ago, as confirmed by different tests. The inclusion of an ancient sample from Egypt in the analysis proves the antiquity of these events, suggesting that modern emmer from Ethiopia, Oman, and India descends from emmer that reached Egypt prior to 3000 years ago. This finding supports a dispersal route through Africa and the Arabian Peninsula before ultimately reaching India. Moreover, I explore the contribution of the until now unexplored Southern Levant ancestry to the functional fraction of the genome, identifying regions that are related to biotic and abiotic stress resistance. These findings underscore the potential of emmer wheat landraces in wheat improvement efforts and emphasize the vital significance of domestication studies in advancing our understanding of agricultural history and genetic diversity.

Table of contents

List of publications

This thesis is based on the work contained in the articles listed below:

1. **Iob, A.** & Botigué, L. (2023). "Genomic analysis of emmer wheat shows a complex history with two distinct domestic groups and evidence of differential hybridization with wild emmer from the western Fertile Crescent". *Vegetation History and Archaeobotany*, 32(5), 545-558.

2. **Iob, A.** & Botigué, L. (2023). "Wild emmer contribution in wheat domestication and adaptation to new environments". *Submitted.*

Other publications that are added in the Annex section:

I. **Iob, A**., Scott, M. F., Botigué L. "Ancient wheat genomes illuminate domestication, dispersal, and diversity" in "The wheat genome" (ed. Springer; in press)

II. **Iob, A**. & Botigué, L. (2022). "Crop archaeogenomics: A powerful resource in need of a well-defined regulation framework". Plants, People, Planet, 4(1), 44-50.

List of figures

List of tables

Acronyms

BP: Before Present

CWR: Crop Wild Relatives

DNW: Domestic North-West

DSE: Domestic South East

Dxy: Absolute Nucleotide Divergence

EPPNB: Early Pre-Pottery Neolithic B (10 700–10 200 cal BP)

Fst: Fixation Index

LD: Linkage Disequilibrium

LPPNB: Late Pre-Pottery Neolithic B (9 500 – 8 500 cal BP)

ML: Maximum Likelihood

MPPNB: Middle Pre-Pottery Neolithic B (10 200–9 500 cal BP)

NJ: Neighbor Joining

PI: Nucleotide Diversity (π)

PPNA: Pre-Pottery Neolithic A (11 700–10 700 cal BP)

QTL: Quantitative Trail Locus

WNL: Wild Northern Levant (synonym of W-NEFC)

WSL: Wild Southern Levant (synonym of W-WFC)

W-NEFC: Wild North-Eastern Fertile Crescent (synonym of WNL)

W-WFC: Wild Western Fertile Crescent (synonym of WSL)

XP-EHH: Cross Population Extended Homozygosity

Glossary

Adaptive introgression: refers to the process of hybridization in which the genes or alleles introduced from one population provide a selective advantage, allowing the receiving population to better adapt to specific environmental conditions.

Allopolyploidization: a biological process in which two different species interbreed, resulting in offspring with multiple complete sets of chromosomes from both parent species. This process leads to the formation of a new species.

Archaeological assemblage: a collection of artifacts, specimens, and other materials recovered from a specific archaeological site or context.

Crop improvement: refers to the process of enhancing the quality, yield, or resilience of crops through various agricultural techniques. In the context of this thesis it refers to a post-domestication selective process operated by Neolithic farmers, that led to the formation of new species and subspecies within the genus *Triticum*.

Domestication center: a geographic region where a particular plant or animal species was first domesticated.

Fertile Crescent: historical region in Southwest Asia, also known as the cradle of civilization, which includes parts of modern-day Syria, Lebanon, Jordan, Palestine, Israel, Iraq and south-eastern Turkey. In this thesis Fertile Crescent and Southwest Asia are often used as synonyms to indicate the broad region where emmer cultivation started.

Free threshing: refers to the ease with which seeds can be separated from the husks or chaff during the threshing process in agriculture. Seeds in free threshing wheat do not tightly adhere to outer glumes, making the separation of seeds relatively simple. The free-threshing phenotype is present in durum and bread wheat, not in emmer wheat.

Founder crops: A set of cereal and legumes that was considered the initial "package" of domestic plants in Southwest Asia. These species were considered to be the staple crops of Neolithic villages and include: emmer, einkorn, barley, lentils, peas and chickpeas. This idea was at the core of the single-origin theory of plant domestication.

Hitchhiking effect: The hitchhiking effect occurs when a specific genetic variant (allele) spreads in a population due to its association with another nearby advantageous allele. The allele "hitches a ride" with the neighboring beneficial variant during evolution, leading to both alleles becoming more prevalent in the population.

Hulled: seed or grain that is tightly enclosed within the glumes or hull. Hulled grains require additional processing to remove the hull before they can be consumed, unlike free-threshing grains. Emmer wheat has a hulled phenotype.

Monophyletic origin: indicates that a group of organisms shares a common ancestor and includes all of its descendants. In the context of this study it specifically refers to domestic populations deriving from a single wild population.

Reticulated origin: evolutionary process in which the hybridization between gene pools leads to a network-like pattern of relationships between populations rather than a strictly tree-like structure. The resulting genomes present complex patterns, characterized by diverse origin among distinct segments. In the context of this study it specifically refers to the origin of domestic crops as an admixture of diverse wild populations, in contrast to the monophyletic origin.

Selective sweep: an evolutionary process in which a specific allele becomes more common in a population due to positive selection. This increase in frequency can lead to a reduction or elimination of genetic variation in the surrounding genomic region (due to hitchhiking effect), creating a characteristic pattern in the genome.

Selfing: or self-pollination, is a reproductive process in which a plant fertilizes itself without the need for external pollen from another plant.

Southwest Asia: also known as Western Asia or the Middle East, is a geographical region comprising countries such as Iran, Iraq, Israel, Jordan, Lebanon, Syria, Turkey, and others. In this thesis Southwest Asia and Fertile Crescent are often used as synonyms to indicate the broad region where emmer cultivation started.

Subgenome: one of the constituent genomes in a polyploid organism. Each set of chromosomes in a polyploid species is often derived from different ancestral species and represents a subgenome.

Chapter 1

INTRODUCTION

1.1. The study of crop domestication

Domestication is a phenomenon that has long captivated the scientific interest of academics and continues to be a subject of enduring fascination and investigation.

While elaborating the theory of natural selection, Charles Darwin observed the great phenotypic variability of farm animals and how breeders were able to control it. He soon understood that domestication brings along phenotypic evolution, connected to human action and adaptation to a novel agro-ecosystem. He proposed that crop domestication might be thought of as a "giant experiment" to test the evolutionary hypothesis (Darwin, 1859).

In the early 20th century, crop domestication studies advanced significantly thanks to the innovative work of Nikolai Ivanovich Vavilov. Vavilov introduced the concept of "centers of domestication", identifying specific geographic regions for the origin of various crops (N. I. Vavilov et al. 1992), and the concept of "homologous series" (i.e. parallel variability of homologous characters in taxonomically near species) (N.I. Vavilov, 1922) shedding light on the vast diversity within crop species and aiding in their classification and understanding. His tireless efforts resulted in a sizable collection of germplasm, an invaluable repository of plant genetic resources. He was a pioneer in understanding the vital need to preserve crop diversity for future agricultural progress and global food security.

Since then, domestication studies have advanced significantly, highlighting the profound implications of this phenomenon in the Neolithic transition, a transformative chapter in the development of human civilization.

The shift from a hunter-gatherer subsistence economy to an agricultural one starting between 13000 and 10000 years ago can indeed be considered as one of the greatest innovations in human history (Diamond 2002). Food production arose independently in at least nine areas of the world: Fertile Crescent, China, Mesoamerica,

Andes/Amazonia, eastern North America, Sahel, Ethiopia, India and New Guinea, but more have been proposed (Diamond 2002; Larson et al. 2014; Spengler 2020), Figure 1.1. Thanks to the food surplus derived from domestication, man started forming sedentary societies and gave rise to modern human cultures (Purugganan and Fuller 2009).

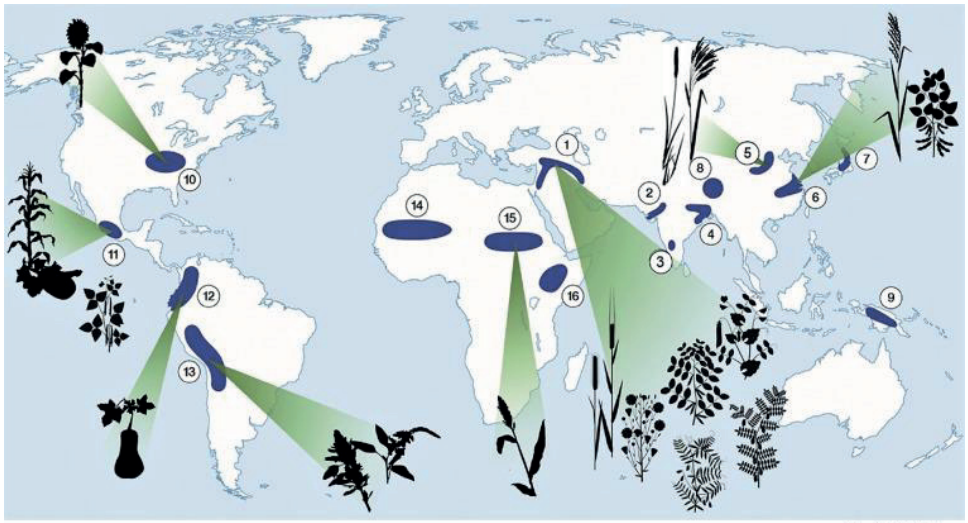


FIGURE 1.1: MAP OF THE REGIONS OR CENTERS OF DOMESTICATION GLOBALLY, AS PROPOSED AND DISCUSSED BY LARSON ET AL. 2014. RECOGNIZED REGIONS OF DOMESTICATION: 1) SOUTHWEST ASIA; 2) SAVANNAHS OF WEST INDIA; 3) SOUTH INDIA; 4) EAST INDIAN PLAINS; 5) NORTH CHINA PLAINS; 6) YANGTZE BASIN; 7) JAPANESE ISLANDS; 8) SOUTHERN HIMALAYA; 9) NEW GUINEA; 10) EASTERN NORTH AMERICAN PLAINS; 11) MESO-AMERICA; 12) LOWLANDS OF SOUTH AMERICA; 13) CENTRAL/SOUTH ANDES; 14) WEST AFRICA SAHEL; 15) EAST AFRICA SAVANNAH; 16) ETHIOPIAN PLATEAU. FIGURE FROM SPENGLER 2020.

Domestication is a complex phenomenon that can be described as feedback between human activities (e.g. harvesting, planting, and soil manipulation) and phenotypic changes in the plants (D. Q. Fuller and Lucas 2017). Through this process, domestic plants acquired traits that distinguish them from their wild ancestors, and that are advantageous in the human-mediated environment, while most likely unfavorable in the wild (Kantar et al. 2017; Purugganan and Fuller 2009), the so-called "domestication syndrome" (Charles 1859). The syndrome in plants includes a wide

variety of traits, that depend on the species and the use of the plant. Examples of such traits include a reduced ability to disperse seeds without human intervention (e.g. non shattering pods, non-brittle rachises), decreased physical and chemical defenses, less numerous unproductive side-shoots, reduction in seed dormancy, larger seeds, more predictable and synchronous germination, and in certain seed-propagated species, more numerous and bigger inflorescences (Larson et al. 2014). While plants adapted to the new human-modified environment, humans adapted to the new crop plants, developing cultivation and processing techniques, involving increased labor investment (D. Q. Fuller and Lucas 2017). Such mutualism can be seen as the result of niche-construction by both humans and plants (Zeder 2016).

In short, plant domestication consists in three fundamental steps: 1) wild harvesting; 2) unconscious selection modifying plant characteristics; and 3) conscious selection of plant material for specific locations and uses with the plant generally losing the ability to survive without human care (Kantar et al. 2017).

Early studies described domestication as a fast evolutionary process (Kohane and Parsons 1988), mainly human-driven (Innan and Kim 2004), that took place in rather restricted geographical areas (Abbo, Lev-Yadun, and Gopher 2010) and involved few populations of a small number of species, the so-called "founder crops" (Zohary 2013). Such scenario appeared to be supported by the relatively limited geographical ranges of wild relatives and their greater genetic diversity in contrast to domesticated species (Lev-Yadun, Simcha; Gopher, Avi; Abbo 2000).

However, thanks to the progress of archaeology and genetics, as well as techniques for the analysis of ancient DNA, in the last thirty years our understanding of domestication has changed significantly, and the scientific community converges today in the description of domestication as a co-evolutive process which is prolonged in time, geographically extended, genetically reticulated, and involving a wide variety of species, at least at its first stages (Allaby et al. 2017; Arranz-Otaegui et al. 2016, 2018; Crowther et al. 2018; Dorian Q. Fuller, Willcox, and Allaby 2011; Kantar et al. 2017; Purugganan 2019).

The history of emmer wheat domestication and dispersal is a good example of how discoveries in archaeology and genomics generate new questions and reframe our understanding of such complex phenomena.

1.2. The genus *Triticum* and its evolution

The genus *Triticum* (wheat) belongs to the Poaceae family and comprises six biological species of wild and domestic grasses(Dvořák 2001), table 1.1. All species are self-pollinating, with an outcrossing rate of ca 1% in field conditions (Dvořák 2001; Golenberg 1988).

TABLE 1.1: PLOIDY, DOMESTICATION STATUS, AND SPIKE CHARACTERISTICS OF *TRITICUM* SPECIES AND SUBSPECIES (DVOŘÁK 2001)

Ploidy	Status	Spike	Species	Subspecies
2×	wild	hulled, brittle	*T. monococcum*	*aegilopoides* (wild einkorn wheat)
	cult.	hulled, nonbrittle	*T. monococcum*	*monococcum* (cultivated einkorn wheat)
	wild	hulled, brittle	*T. urartu*	
4×	wild	hulled, brittle	*T. turgidum*	*dicoccoides* (wild emmer wheat)
	cult.	hulled, nonbrittle	*T. turgidum*	*dicoccon* (cultivated emmer wheat) *Paleocolchieum*
		naked, nonbrittle	*T. turgidum*	*durum* (durum), *turgidum* (pollard wheat), *turanicum* (Khorassan wheat), *polonicum* (Polish wheat), *carthlicum* (Persian wheat), *isphahanicum*
	wild	hulled, brittle	*T. timopheevii*	*armeniaeum* (syn. *araraticum*)
	cult.	hulled, nonbrittle	*T. timopheevii*	*timopheevii*
6×	cult.	hulled, brittle	*T. aestivum*	*macha*, *tibetanum* (Tibetan wheat)
		hulled, partially brittle	*T. aestivum*	*spelta* (spelt), *vavilovii*, *yunanense* (Yunan wheat)
		naked, nonbrittle	*T. aestivum*	*aestivum* (bread wheat), *compactum* (club wheat), *sphaerococcum* (Indian dwarf wheat), *petropavlovskyi* (Chinese rice wheat)
	cult.	hulled, nonbrittle	*T. zhukovskyi*	

Triticum species present inflorescences in the form of spikes. These are composed of spikelets attached to the rachis, each enclosed by two glumes at the base and containing 2-6 florets (Shitsukawa et al. 2009), figure 1.2.

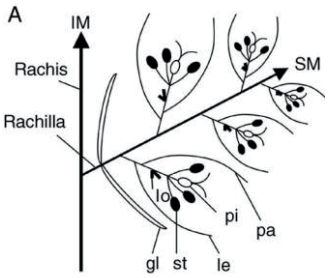


FIGURE 1.2: A SPIKELET OF HEXAPLOID WHEAT (*TRITICUM AESTIVUM*). THE SPIKELETS ARE ARRANGED AS TWO OPPOSITE ROWS OF LATERAL BRANCHES FROM THE MAIN AXIS (RACHIS). EACH SPIKELET IS COMPRISED OF FLORETS, JOINED AT THE AXIS (RACHILLA) ALTERNATELY ON OPPOSITE SIDES, ENCOMPASSED BY TWO GLUMES. EACH FLORET IS COMPRISED OF A LEMMA, A PALEA, TWO LODICULES, THREE STAMENS AND A PISTIL. ABBREVIATIONS: IM, INFLORESCENCE MERISTEM; SM, SPIKELET MERISTEM; GL, GLUME; LE, LEMMA; PA, PALEA; LO, LODICULE; ST, STAMEN; PI, PISTIL (SHITSUKAWA ET AL., 2009)

Two pivotal phenotypic traits within the spike structure have gained substantial attention due to their profound implications in both agricultural practices and evolutionary studies.

One of these characters is rachis brittleness, a trait of paramount importance as a discriminating factor in the study of wheat domestication. In wild species, the rachis is brittle, and spikelets disarticulate from it upon maturity, allowing seed dispersal in the surrounding area. On the other hand, in domestic species the rachis is non-brittle, and spikelets containing the mature grain remain attached to it, facilitating harvest, and determining dependence on human-mediated seed sowing for dispersal (Nave et al. 2019). For this reason, the non-brittle rachis is considered the quintessential domestication trait in wheat (Zohary 2013). This phenotype is controlled by two loci in chromosomes 3A (*TtBTR1-A*) and 3B (*TtBTR1-B*).

The second noteworthy character is threshability. In the spikelet, seeds can be toughly enclosed by the glumes, requiring mechanical work to extract them. Wheats with this ancestral character are called hulled, like emmer wheat and spelt. In free-threshing wheats, like durum and bread wheat, seeds are loosely encapsulated by glumes that fall apart during harvesting, hence minimizing the manual labor required for seed

collection by humans (Sharma et al. 2019; Simons et al. 2006). The presence of the free-threshing phenotype is not considered an indispensable hallmark of the domestication process, as domestic cereals can be hulled (like emmer and spelt). This trait emerged subsequent to domestication, and it is therefore regarded as a distinctive feature associated with crop improvement. The expression of the free-threshing phenotype is determined by the simultaneous occurrence of mutations within the non-brittle rachis loci on chromosomes 3A and 3B, as well as mutations in two quantitative trait loci (QTL) situated on chromosomes 2A (*Tg-A1*) and 2B (*Tg-B1*), and in the *Q* gene located on chromosome 5A. Over time, the production of free-threshing wheats has surpassed that of their hulled counterparts, establishing durum and bread wheat as the most prevalent wheats grown worldwide. These phenotypes and the underlying genes are reported in figure 1.3.

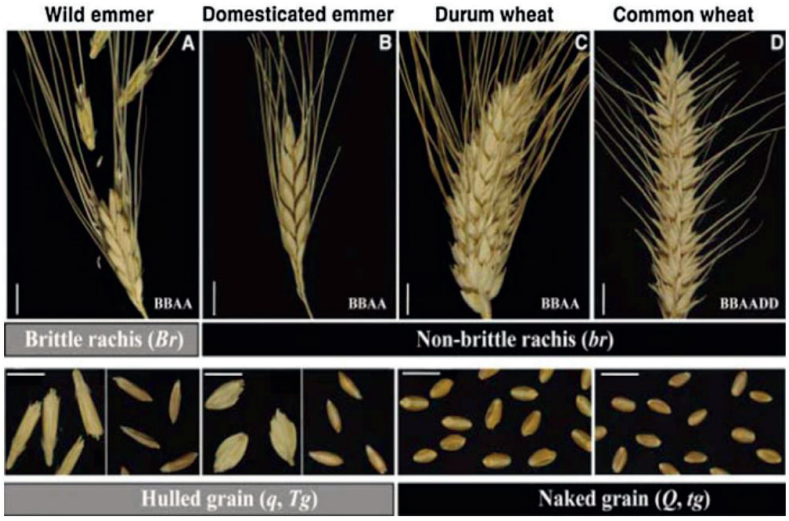


FIGURE 1.3: WHEAT SPIKES SHOWING BRITTLE RACHIS (A), NON-BRITTLE RACHIS (B TO D), HULLED GRAIN (A AND B), AND NAKED GRAIN (C AND D). A WILD EMMER WHEAT, B DOMESTICATED EMMER, C DURUM, AND D BREAD WHEAT. LETTERS AT THE LOWER RIGHT CORNER INDICATE THE GENOME FORMULA OF EACH TYPE OF WHEAT (J. H. PENG, SUN AND NEVO 2011)

The evolutionary history of the genus *Triticum* is very complex. It is characterized by the existence of four subgenomes, A, B, D and G, and three levels of ploidy: diploid, tetraploid and hexaploid (2x =14, 4x = 28 or 6x=42), as reported in table 1.1. The polyploid species are all derived by allopolyploidization events with weeds belonging to the closely related *Aegilopsis* genus (Golovnina et al. 2007; Goncharov 2011).

Diploid species (AA) comprise the wild *T. urartu* and *T. monococcum aegilopoides*, the latter having a domestic descendant *T. monococcum* (einkorn).

Tetraploid wild emmer (*T. turgidum dicoccoides*, BBAA) emerged from the allopolyploidization of *T. urartu* (AA) and *Ae. speltoides* (BB) around 0.5 million years ago (Haas, Schreiber, and Mascher 2018). Its domestic descendant is *T. turgidum dicoccon*, from which free-threshing tetraploid *T. turgidum durum* (also BBAA) derived as result of crop improvement under cultivation. Another wild tetraploid is *T. araraticum* (GGAA), which was domesticated into *T. timopheevi* (GGAA) (Dvořák 2001).

Finally, hexaploid, domestic *T. aestivum* (BBAADD) derived from the hybridization of a domestic tetraploid and *Ae. tauschii* (Y. Zhou et al. 2020).

The emergence of modern durum and bread wheat, the most grown wheats at a global level, can thus be summarized as the result of three processes: I) domestication of wild emmer wheat, associated with the loss of rachis brittleness; II) crop evolution (often also referred to as crop improvement under cultivation), which includes the emergence of the free-threshing phenotype and adaptation to new ecological niches; III) allopolyploidization between a free-threshing tetraploid with *A. tauschii*, giving rise to bread wheat. These steps are schematized in figure 1.4 (Iob, Scott, Botiguè 2023, see annex).

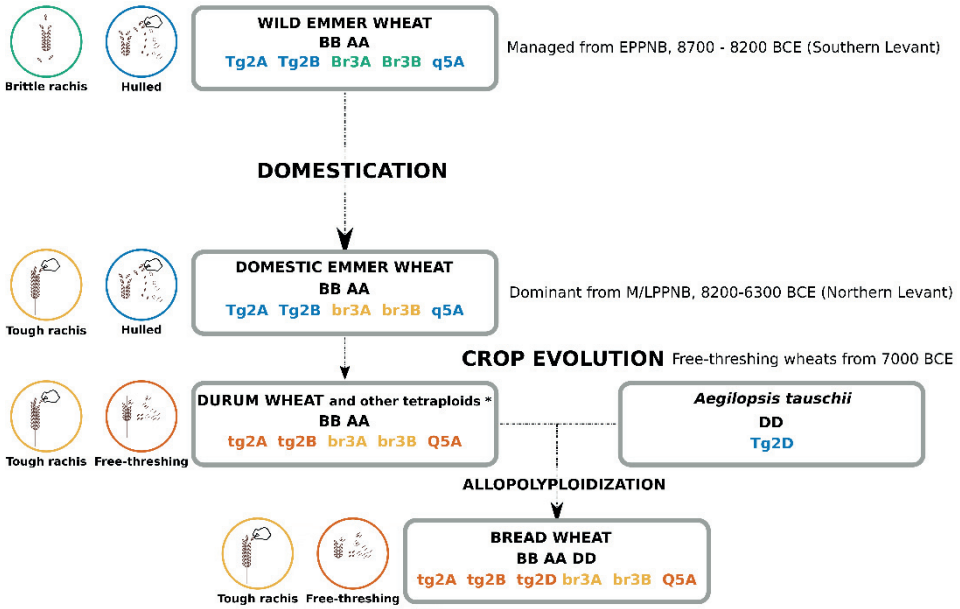


FIGURE 1.4: SCHEMATIC REPRESENTATION OF THE DOMESTICATION AND EVOLUTION OF THE MOST ECONOMICALLY IMPORTANT WHEATS TODAY, SHOWING IMPORTANT PHENOTYPES AND THE MUTATIONS THAT DETERMINE THEM. BASIC INFORMATION ABOUT THE APPEARANCE OF THE DIFFERENT WHEATS IN THE ARCHAEOLOGICAL RECORD IS GIVEN ON THE RIGHT. THE SMALL WHITE HAND REPRESENTS THE INVESTMENT OF HUMAN LABOR IN PROCESSING THE HARVEST (IOB, SCOTT, BOTIGUÉ, 2023, SEE ANNEX)

1.3. Emmer wheat domestication, a changing perspective

Emmer wheat is a tetraploid, hulled cereal. Wild emmer wheat grows in Southwest Asia, in the so-called Fertile Crescent (N. I. Vavilov et al. 1992), and is divided into two distinct populations living in different environments: one grows in cold and humid mountain slopes of Turkey, Iraq and Iran, identified as Northern Levant population, while the other, named Southern Levant population, grows in mild and dry areas of Israel, Jordan, Syria and Lebanon. Domestic emmer wheat first appeared

in the Fertile Crescent around 12 000 - 10 000 years before present (BP) (Maccaferri et al. 2019; Ozkan et al. 2002; Zohary 2013).

The very first evidence about emmer domestication came from archaeology and was based on the observation of the multiple morphotypes recovered at different archaeological sites. The loss of rachis brittleness (see section 1.2) leaves a recognizable sign at the base of the spikelet that is usually preserved in archaeobotanical remains. The spontaneous disarticulation of the spikelets in wild cereals leaves a smooth mark on the rachis (figure 1.5A), while the induced disarticulation of spikelets in domestic plants with tough rachis, leaves a rough scar (figure 1.5B) (Abbo et al. 2014; Nave et al. 2019; Snir et al. 2015). This phenotype, together with other evidence (e.g., distribution of wild relatives, radiocarbon dating, kernels dimensions), was used to determine the site of first appearance of domestic forms. (Abbo et al. 2014; Nave et al. 2019; Snir et al. 2015). Based on the proportion of wild (smooth scars) and domestic (rough scars) cereal remains at different sites, it was proposed that by 10 500 – 10 100 BP emmer, einkorn and barley "were intentionally sown and harvested" in Southwest Asia (Zohary 2013).

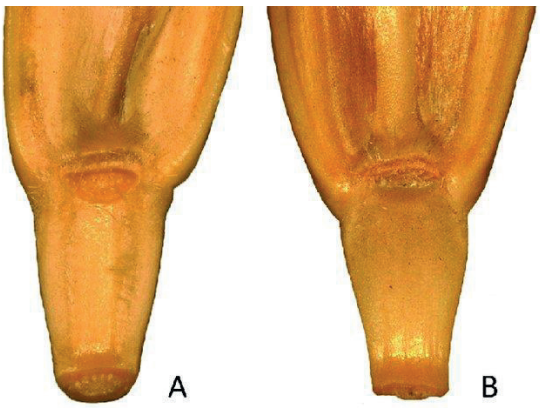


FIGURE 1.5: COMPARISON BETWEEN WILD AND DOMESTICATED EMMER WHEAT. A. WILD EMMER WITH SMOOTH ABSCISSION SCARS, B. DOMESTICATED EMMER WITH A ROUGH SURFACE OF THE UPPER SCAR AND A DAMAGED LOWER END OF THE INTERNODE (WEIDE 2015)

The earliest convincing evidence of domestic emmer wheat is attributed to the Pre-Pottery Neolithic B (PPNB, see table 1.2) sites of Çayönü (10 250 – 9 550 BP) and Cafer Höyük in Southern Turkey (Maeda et al. 2016; Zohary 2013).

TABLE 1.2: OVERVIEW OF TIME TRANSET BETWEEN EPIPALEOLITHIC AND CHALCOLITHIC IN SOUTHWEST ASIA, WITH THE NEOLITHIC PERIODS IN BOLD (ARRANZ-OTAEGUI AND ROE 2023)

Period	Ka cal BP	Subsistence
Late Epipal	15–11.7	Foraging
PPNA	11.7–10.7	Pre-domestication cultivation
EPPNB	10.7–10.2	Cultivation of domesticated species
MPPNB	10.2–9.5	Cultivation of domesticated species
LPPNB/C	9.5–8.5	Agriculture
Pottery Neolithic	8.5–6.5	Agriculture
Chalcolithic	6.5–5	Agriculture

In the early 2000's the popular "cradle of agriculture theory" (Lev-Yadun, Simcha; Gopher, Avi; Abbo 2000), identified South Easter Turkey and Northern Syria as the center of domestication of the "founder crops of agriculture" (Zohary 2013), including emmer wheat. Early genetic models, based on the similarity between domestic emmer and wild emmer from the Northern Levant, supported this theory (M. C. Luo et al. 2007; Ozkan et al. 2002, 2005), and pointed to the Karaca Dağ Mountain region (S-E Turkey) as the center of domestication of emmer wheat.

However, wild emmer wheat remains were recovered in Pre-Pottery Neolithic A (PPNA, see table 1.2) sites of the Southern Levant, but not in contemporary sites in the Northern Levant (Ozkan et al. 2011). Also, the proportion of domestic emmer remains slowly increases through time when broad regions are considered. For these reasons, a prolonged period of pre-domestication cultivation has been proposed

(Fuller, 2007), and some authors hypothesized that emmer cultivation spread from the South to the North of the Fertile Crescent during hundreds of years, leading to multiple domestication events (Feldman and Kislev 2007).

Over the past years, increasing evidence has challenged established paradigms on domestication, dramatically changing our understanding of such process. New archaeological findings and a re-evaluation of the extant record started questioning the role of the "founder crops" in the Neolithic transition (Arranz-Otaegui et al. 2018; Arranz-Otaegui and Roe 2023), the timing and pace of domestication (Allaby et al. 2017; D. Q. Fuller and Lucas 2017; Dorian Q Fuller, Asouti, and Purugganan 2012). At the same time, new genetic studies started questioning the monophyletic origin of domestic emmer, based on the levels of genetic diversity of Turkish wild emmer compared to emmer descendants (Jorgensen et al. 2017). Civáň et al., (2013) analyzed retrotransposons insertions and proposed a reticulated origin of emmer wheat, which would have arisen from a small but admixed wild population. A similar argument has been recently advocated by Oliveira et al., (2020) based on Genotyping By Sequencing (GBS) data, suggesting that wild emmer in the northern Fertile Crescent mixed with a pre-domesticated emmer population coming from the south of the region. Domestic emmer would be derived from this admixed population. Nave et al., (2019) investigated the origin of the main domestic haplotypes for the non-brittle rachis *(TtBTR1-A* and *TtBTR1-B*) phenotype, showing the contribution of the Southern Levant wild pool to at least one of the haplotypes for this domestication trait. The authors propose a two-step model for emmer domestication, in which mutations for domestication traits have appeared in different chromosomes at different times and possibly in different populations. This is in line with the observation that the domestic phenotype, which requires two independent recessive mutations, took millennia to be established (Avni et al. 2017; Dorian Q. Fuller et al. 2014). The results obtained by Wang et al., (2022) confirm that the donors of domestic emmer haplotypes and important domestication loci come from different wild lineages, covering a wide geographical area.

Nevertheless, the debate about how and when domestication took place is far from over, with some authors rejecting the autonomous-protracted model of domestication (Peleg, Abbo, and Gopher 2022), and others re-evaluating the monophyletic origin of domestic emmer (X. Zhao et al. 2023). Moreover, key uncertainties surround the contributions of different wild populations to the domestic gene pool and the dynamics leading to the fixation of domestic alleles.

1.4. Gene flow and dispersal

Despite the fact that the reticulated origin of emmer wheat needs further investigation, the role of introgression from the wild in wheat has been demonstrated by several studies e.g. (He et al. 2019; Keilwagen et al. 2022; Pont, Leroy, Seidel, and Tondelli 2019). Cheng et al. 2019, showed that even if the ancestry of tetraploid wheats points to the Northern Levant, important QTL in bread and durum wheat are possibly derived from different emmer wild populations, providing additional targets for selection and environmental adaptation thanks to gene flow during the spread of domesticated wheat (Cheng et al. 2019). Scott et al., (2019) analyzed one domestic ancient emmer specimen from Egypt, radiocarbon dated 3130 - 3000 BP, detecting signals of wild introgression from the Southern Levant, possibly occurring during cultivation before its introduction to Egypt, or during later interactions.

In general, big uncertainties surround the role played by wild emmer wheat from the Southern Levant in shaping the domestic populations. Specifically, conflicting evidence exists not only regarding its involvement in forming the domestic gene pool through reticulate origins but, according to ancient DNA studies (Scott et al. 2019), also in terms of the potential impact of post-domestication gene flow on the differentiation of domestic populations. This contribution could involve significant haplotypes that aid in adapting to new climatic conditions, known as adaptive introgression.

Domestic emmer spread from its area of origin to South Asia, Europe, North-East Africa, the Arabian Peninsula and the Indian subcontinent. At the beginning of the 9th millennium BP, domestic emmer spread to eastern Anatolia, northern Iraq and southwestern Iran. In Europe, emmer wheat accompanied the development of agriculture. In the South, it reached Egypt during the 7th millennium BP (Wendrich and Cappers 2005). It was introduced to Ethiopia and India around 5000 BP (Helbaek 1970; Zaharieva et al. 2010). It was the main crop of Babylon, ancient Egypt and Greece. Especially in Egypt, emmer was the only wheat species cultivated from the first settlements until Greco-Roman times (Nesbitt and Samuel 1996). While the diffusion of emmer wheat into Europe has been well studied in relation to the spread of agriculture (e.g., Coward et al., 2008), it remains unclear how emmer spread to the south and the east (Stevens et al. 2016). In particular, it's uncertain if the emmer wheat in Ethiopia came from Egypt or from the Iranian highlands and the Arabian Peninsula (Luo et al., 2007).

By the Bronze Age (ca 5 300 BP – 3 200 BP), emmer was replaced by its free-threshing descendants durum and bread wheat or barley in most regions. Nowadays emmer wheat can be considered a neglected crop: it is still cultivated only in few areas of the world, and it can still be considered as an important crop in India, Ethiopia and Yemen (Zaharieva et al. 2010).

Based on geographical and eco-morphological evidence, four subspecies of domestic emmer have been described (Dorofeev et al., 1979; Vavilov, 1926): *abyssinicum* Vav. (Abyssinian emmer), *asiaticum* Vav. (Eastern emmer), *dicoccum* (European emmer) and *maroccanum* Flaksb. (Moroccan emmer).

At a genetic level, four domestic gene pools have been identified: Mediterranean, Eastern Europe, Caucasus and Indian Ocean, all most closely related to the Northern Levant wild population (Avni et al. 2017). These groups reflect emmer wheat radial dispersal pattern outside the Fertile Crescent, with the Indian Ocean group being the most differentiated (Maccaferri et al. 2019).

1.5. The future of emmer wheat and its potential in wheat improvement

The domestic *Triticum* species have a huge economic and food security importance, as wheat is one of the most cultivated crops worldwide (FAO 2021). While einkorn and emmer wheat have nowadays a more marginal role in human nutrition and are considered "neglected crops", durum and bread wheat are used to make pasta and bakery products all over the world (Igrejas and Branlard 2020). Nevertheless, in recent years increasing attention has been given to "old crops" due to their nutritional characteristics (Bergamini et al. 2013; Campanaro et al. 2019; C. K. Khoury et al. 2014) and increasing research is demonstrating the great potential of traditional landraces and crop wild relatives to aid future crop improvement projects, especially in the context of climate change (Balla et al. 2022; Dwivedi et al. 2016; Saleh 2020).

Indeed, the current food system relies on a limited number of species that have been strongly selected especially for yield (Shewry, Pellny, and Lovegrove 2016). New studies have shown that the loss of genetic diversity during the process of domestication, the so-called domestication bottleneck, was not as strong as it was thought to be, at least for some crops (Allaby et al. 2022; Allaby, Ware, and Kistler 2019; Kilian et al. 2007; Russell et al. 2016). On the contrary, the greatest reduction in diversity of these crops is ascribable to modern breeding strategies (Louwaars 2018; Trucchi et al. 2021; Voss-Fels, Stahl, and Hickey 2019). One of the great successes of the Green Revolution is the creation of wheat varieties with heightened yields, remarkably responsive to agricultural inputs (Lopes et al. 2015). However, selection at specific loci reduces nucleotide diversity in the surrounding regions (selective sweeps or hitchhiking, Tanksley & McCouch, 1997), and the increased reliance on a limited spectrum of cultivars across most breeding programs has precipitated the depletion of finely adapted genetic diversity (Lopes et al. 2015). Such a reduction in diversity has a negative impact on the agricultural system, in the context of the climatic change the world is facing. In fact, elite cultivars lack the genetic diversity that is necessary for adapting and sustaining yields in shifting climatic conditions (Hufford,

Berny Mier Y Teran, and Gepts 2019; Labeyrie et al. 2021; Streit Krug et al. 2023), and rely extensively on substantial inputs of water, fertilizers, and pesticides. (Gonthier et al. 2014).

For this reason, increasing attention is given to crop wild relatives (CWR) and traditional landraces, as sources of genetic diversity (Bohra et al. 2022; Cortés and López-Hernández 2021; Marone et al. 2021). Wild relatives are naturally adapted to the environment they live in and did not undergo bottlenecks due to artificial (human-mediated) selection, as did domestic crops. For this reason, they represent an invaluable source of genetic diversity (Brozynska, Furtado, and Henry 2016; Tirnaz et al. 2022; H. Zhang et al. 2017). Traditional landraces are domestic and locally adapted varieties (Venkateswaran, Elangovan, and Sivaraj 2019). They differ from elite varieties that have been selectively improved by breeders for specific characteristics, (Dwivedi et al. 2016) and maintain high diversity between and within populations (Marone et al. 2021). While landraces are not as productive as elite cultivars, they are recognized for having high nutritional content (Dwivedi et al. 2016), and are often more resistant to biotic and abiotic stresses (Marone et al. 2021), thus maintaining yield stability and producing intermediate yield even under low-input agriculture (Lopes et al. 2015; Zeven 1998).

Emmer wheat is direct ancestor of bread and durum wheat, it has the same genome formula as durum and has contributed two genomes to bread wheat (J. Peng et al. 2013), figure 1.4. Several studies have reported reduced nucleotide diversity in domestic emmer compared to its wild counterpart (Avni et al. 2017; Jorgensen et al. 2017), a common feature of crops (Bai and Lindhout 2007; Karn, Gillman, and Flint-Garcia 2017; Kwak and Gepts 2009). Nevertheless, domestic emmer, which did not undergo improvement over the status of landraces, is substantially more diverse than its descendant's durum and bread wheat (Y. Zhou et al. 2020). The analysis of the durum wheat genome revealed widespread loss of diversity, consequence of selection and breeding (Maccaferri et al. 2019). A recent study showed that circa 19.6% of haplotypes found in domestic tetraploids are present in bread wheat landraces and

cultivars, and only 1.1% of haplotypes found in wild emmer wheat can be identified in bread wheat (Z. Wang et al. 2022).

Wild emmer presents agriculturally unfavorable traits, such as brittle rachis, non-free-threshing spikelet, few, small, and light spikes, small grains. Nevertheless, it harbors many useful traits related to grain quality (e.g. protein content) and tolerance to abiotic and biotic stress (e.g. salt, drought, stem rust). It also possesses genotypic variation for diverse agronomic and shoot morphology traits such as germination, biomass, earliness, nitrogen content, yield, short stature, and high tillering capacity (J. Peng et al. 2013; Rahman et al. 2023). A recent study (Balla et al. 2022) identified several maker-trait associations for heat resistance in wild emmer wheat, including favorable alleles that were absent or rare in the elite durum wheat germplasm.

In a similar way, domestic emmer can grow in adverse climatic conditions without use of pesticides and is resistant to several diseases (Saleh 2020). It can grow in soils with limited fertility (D'Antuono 1989), using low-input techniques (D'Antuono and Minelli 1997) and has highly competitive ability against weeds in comparison to bread wheat. Moreover, domestic emmer has higher fiber content than common wheat and is rich in protein, carbohydrates, minerals, and poor in fats (Dhanavath and Prasada Rao 2017). Its suitability for organic farming and the possible health benefits of its consumption have renewed its popularity (Pagnotta, Mondini, and Atallah 2005; Sanket 2023).

All this evidence indicates that both wild and domestic emmer hold great potential for durum and bread wheat breeding.

An important phase in plant breeding efforts is determining the origins of crops. It facilitates the discovery of distant relatives in the wild, closely related species, and novel genetic diversity. Such information, together with knowledge of patterns of genetic variation among different types, is essential to prevent genetic erosion and the ensuing loss of ecotypes and landraces (Venkateswaran et al., 2019). Exploring the genetic diversity and population structure of wild relatives and landraces is the first step towards the identification of alleles related to relevant agronomic traits, before

these can be introgressed into modern varieties by hybridization and other approaches in breeding programs (Lopes et al., 2015; Mourad et al., 2019).

1.6. Exploring large, polyploid genomes in self-pollinating species

The genomic analysis of emmer wheat is challenging for its big genome, its remarkably long chromosomes (> 800Mb), its ploidy and its self-pollinating reproductive strategy.

Emmer wheat has a genome size of approximately 10 Gb (Zhu et al. 2019), making it notably larger than the human genome by over threefold. This substantial genome size places emmer wheat in the category of cereals with relatively large genomes within the plant kingdom. For instance, when compared to other well-studied cereals, emmer wheat's genome size far exceeds that of maize (*Zea mays*), which measures approximately 2.4 Gb (Haberer et al. 2005). It also surpasses the genome sizes of rice (*Oryza sativa)* of approximately 400-500 Mb (Kurata, Nonomura, and Harushima 2002) and the model organism *Arabidopsis thaliana*, with a genome size of approximately 135 Mb (The Arabidopsis Genome Initiative 2000). Despite this, thanks to the advances of the last twenty years, genome assemblies are now available for several Triticeae species (Gao et al. 2023), but further improvements are necessary to effectively address the repetitive and low complexity regions of the genome, thereby minimizing the presence of non-assembled (Ns) sequences within the references.

The genomic analysis of such a big genome requires the use of large number of resources, starting from the sequencing investment to the computational power and the digital storage capacity. Most of the analyses must be conducted in High Performance Computers (HPCs) or clusters, and for a given analysis it is typically required to invest more time compared to what is needed for species with smaller genomes. This is exacerbated by specific features of the emmer wheat genome and by

self-pollination making the analyses more complex and time consuming. This complexity is further compounded by the fact that bioinformatic tools are seldom tailored to wheat's distinct characteristics.

Emmer wheat B and A subgenomes share over 95% sequence identity across coding regions. Additionally, these subgenomes comprise over 80% of repetitive DNA, primarily ascribable to transposable elements (Avni et al. 2017; Borrill, Adamski, and Uauy 2015; Krasileva et al. 2013). Common features of the wheat genome include gene duplication (e.g. Jia et al., 2023), gene gain and loss (Gao et al. 2023), pseudogene evolution (Wicker et al. 2011) and structural variation (e.g. R. De Oliveira et al., 2020).

Given the high homology between genomes, the quality of the reference assemblies and the reduced genetic diversity, short-sequence data may fail to map uniquely to the reference genome (Gardiner et al. 2019). For this reason, it is advisable to keep only reads mapping uniquely and to mask low complexity genomic regions. Similarly, structural variations can lead to mismapping issues. Reads from distinct genomic regions can map to the same location in the reference genome, yielding elevated mapping quality scores, hence passing quality filters. This may result in an inflated sample heterozygosity predicted by the genotype calling algorithm when there is a polymorphism in one of the duplications (see fig. 6 from Iob, Scott, Botigué in annex). A typical strategy to address this issue involves the elimination of variants that exhibit heterozygosity in multiple samples, since the predicted levels of heterozygosity at the population level in self-pollinating species are low (see below - e.g. Gardiner et al. 2019; He et al. 2019). Overall, despite the potential loss of valuable information, it is imperative to implement rigorous filtering methods when analyzing wheat sequencing data.

Moreover, emmer wheat is an almost completely self-pollinating species. Selfing has strong genetic consequences at the individual and population level. In self-pollinating species, homozygosity is increased due to the union of gametes from the same parent, while recombination is less efficient as cross-over happens between identical or nearly identical chromosomes. Consequently, Linkage Disequilibrium (LD) extends for large

regions of DNA and hitchhiking effects are exacerbated (Burgarella and Glémin 2017). Analyses conducted on durum wheat show that especially in centromeric regions, recombination blocks extend for hundreds of Megabases (Maccaferri et al. 2019).

All considered, selfing reduces effective population size (Ne) as it reduces the number of independent gametes for reproduction. This, coupled with the effects of strong LD, causes increased genetic drift and decreased gene flow, leading to strong population structure (Burgarella and Glémin 2017). These characteristics, visually summarized in figure 1.6, must be considered when conducting population genomic analysis on wheat species.

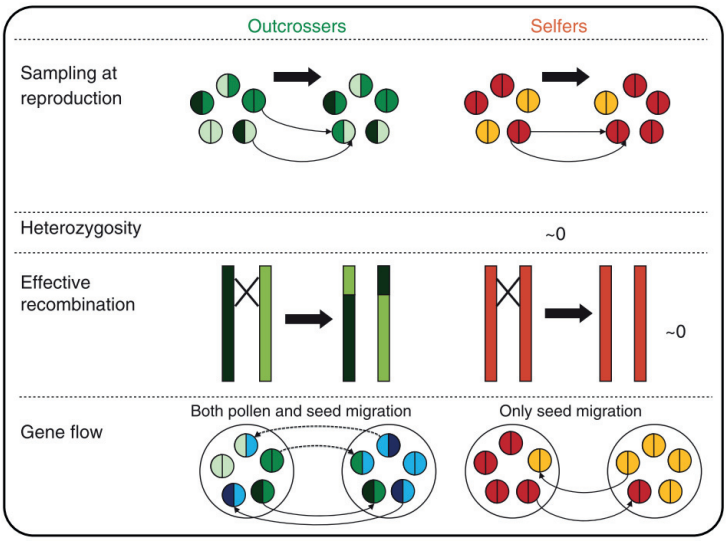


FIGURE 1.6: GENETIC EFFECTS OF SELF-FERTILIZATION COMPARED TO OUTCROSSING MATING (BURGARELLA AND GLÉMIN 2017)

Not only common standard measures of diversity depend on the variability between and within population, which are both altered by selfing (Burgarella and Glémin 2017), but also the presence of strong LD extremizes the signals of selection, as selective sweeps can extend for Mb. This makes it difficult to identify regions under

selection. Moreover, common population structure analyses are highly affected by LD (Price et al. 2008), and therefore strict filters must be applied when conducting analyses that assume independence between SNPs (e.g. DAPC). In addition, selfing species violate assumptions such as panmixia, which is at the basis of different methods. Correct interpretation of the results is guaranteed only if this is accounted for.

1.7. Population genetics analysis in wheat research

The field of domestication studies relies on a multitude of widely-adopted population genetic analysis methods. Within this section, I provide an overview of the foundational principles underpinning the primary methodologies employed in this thesis. Furthermore, I highlight the importance of employing various analytical approaches when investigating complex genomes. The integration of results between different methods is essential for deepening our understanding of crop evolution.

1.7.1. Genetic diversity, genetic distances and selection tests

The assessment of the levels of genetic diversity within and between populations is a fundamental step in genetic research. Nucleotide diversity (PI) is a common measure of variability within population and aids the identification of those populations and genomic regions that underwent reduction in diversity, due to phenomena such as selection, bottlenecks or genetic drift. However, as described above, diversity is lower in selfing species than in outcrossing organisms, and in domestic populations diversity is typically depleted compared to wild relatives. Such characteristics must not be overseen when comparing levels of diversity between populations, or calculating genetic distance between them. One of the most famous measures of genetic distance is Fst (Fixation Index), a relative measure of genetic differentiation that compares the genetic diversity between populations while considering the genetic diversity within each population. Given that Fst values depend on within-population genetic diversity,

peaks in Fst values are common when one population has low genetic diversity at a specific genomic region (Charlesworth 1998). A different measure of genetic distance is Dxy, which is an absolute measure of genetic divergence between populations. It measures the average number of nucleotide differences between two sequences from different populations (Nei and Li 1979) and is hence less dependent on within-population genetic diversity. For this reason, Dxy is a preferable choice when dealing with selfing species in the context of domestication studies.

The reduction in nucleotide diversity and the presence of extreme values of Fst between populations have often served as indicators of selection events. This reliance is grounded in the established observation that genetic diversity is depleted within a selective sweep (Stephan 2016). Traditionally, these statistics have been interpreted in conjunction with other measures of selection, such as Tajima's D. Tajima's D (Tajima 1989) is one of the most famous statistical methods in population genetics, used to detect deviation from neutral evolution. It is calculated at the population level as the difference between the mean number of pairwise differences and the number of segregating sites. Values of Tajima's D close to 0 indicate that the population is evolving neutrally, while positive values reflect an excess of intermediate-frequency alleles, which can indicate population bottleneck or balancing selection. Negative values suggest an excess of rare alleles, which can be indicative of population expansion or positive selection. However, the high levels of homozygosity and LD characteristic of the emmer genome can influence the calculation of Tajima's D by altering the expected patterns of genetic variation within a population. Inbreeding increases the frequency of homozygous individuals, leading to a decrease in the number of segregating sites within population and an increase in the number of low-frequency alleles. At the same time, when LD is high, alleles at distant loci are inherited together, reducing the number of unique combinations of alleles in the population (haplotype diversity).

Other methods use haplotype information for detecting selection. Positive selection can increase rapidly the frequency of a beneficial allele, resulting in an extension of the haplotype carrying it, a process called selective sweep. In other words, during a

selective sweep, the haplotype carrying the advantageous allele becomes more common, leading to an extended region of homozygosity around that allele. Therefore, the decay of haplotype homozygosity is expected to be slower for alleles that have undergone positive selection compared to those that have not. This is detected by EHH (Extended Haplotype Homozygosity) test, which measures the decay of haplotype homozygosity with distance from a core SNP. XP-EHH (Cross Population Extended Haplotype Homozygosity, Sabeti et al., 2007) is a cross-population extension of EHH, and it compares the haplotype homozygosity decay of two populations (one defined as target, the other as reference) allowing for the detection of selective sweeps that have occurred in one population but not the other. Extreme positive XP-EHH values indicate that the haplotypes in the target population are more extended than in the reference population, suggesting that a selective sweep has occurred. On the other hand, extremely negative values indicate the opposite trend (selection in the reference population). In the context of wheat population genetics analysis, where low recombination rates and high homozygosity impact selection statistics, the XP-EHH method offers a distinct advantage by discerning selection signals amidst shared genetic features across populations. Its unique efficacy lies in its capacity to pinpoint specific selection patterns within individual populations. To clarify, in instances where certain genomic regions exhibit diminished diversity due to inherent genetic characteristics of wheat rather than selective pressures, this uniformity should be mirrored in both populations. When applying the XP-EHH method to compare these populations, regions displaying analogous patterns of limited diversity in both groups are not, possibly erroneously, identified as under selection. Only regions demonstrating substantial disparities in homozygosity patterns between populations are indicated as under selection in one or the other population, ensuring the differentiation of genuine selective signals from shared genetic features. Consequently, this approach may miss signals of common selection present in both populations. However, it ensures the accurate identification of differences in selective pressures unique to each population.

1.7.2. Genetic clustering: dimension reduction and phylogenetic methods

In population genetics, the identification of distinct populations or genetic groups often relies on clustering individuals according to genetic similarity. This process involves the application of various methods, such as dimension reduction techniques and phylogenetic analyses.

PCA and DAPC are multivariate dimension reduction methods. PCA is one of the most used methods to investigate population structure and is based on mapping high-dimensional genetic data to a usually bi-dimensional space. The methods identify the major axes of variation of the data and project them into a subspace of equal or fewer dimensions (e.g. PC1 and PC2) (Alhusain and Hafez 2018). PCA summarizes the overall variability among individuals, including both the divergence between groups, and the variation within groups. DAPC is a two-step procedure that involves principal component analysis (PCA) followed by discriminant analysis (DA). In the first step, PCA is used to reduce the dimensionality of the data by identifying the major axes of variation. In the second step, DA is used to partition genetic variation in between-group and within-group components, and to identify the linear combinations of variables that best discriminate between groups of individuals while minimizing the variance within them (Jombart, Devillard, and Balloux 2010). For this reason, DAPC has enhanced power to identify groups within the dataset. Both methods are ideal for exploring the genetic structure of the dataset without any prior knowledge.

Phylogenetic trees are a powerful tool for understanding the evolutionary relationships among organisms and can be constructed using various methods based on genetic data. A phylogenetic tree represents a model of evolutionary history depicted by ancestor-descendant relationships between tree nodes and clustering of 'sister' organisms at a different level of relatedness. There are two classes of methods for constructing phylogenetic trees based on genetic data: distance-based methods and model-based methods. Distance-based methods, including the Unweighted Pair Group Method with Arithmetic means (UPGMA) and the Neighbor-Joining (NJ,

Saitou & Nei, 1987), are based on the matrix of pairwise genetic distances calculated between sequences. Model-based methods, like maximum parsimony (MP, Fitch, 1971), maximum likelihood (ML, Felsenstein, 1981), are based on the assumption that the sequences have evolved according to a specific model of evolution. MP infers the tree that requires the fewest number of evolutionary changes, while ML infers the tree that maximizes the probability of the observed data given a model of evolution. In ML methods, the likelihood of the tree is calculated, and its topology optimized until the best tree is found.

In this thesis I employ ML and NJ. While ML stands out as one of the most widely utilized techniques in phylogenetic reconstruction, NJ (distance-based) operates without the need for an underlying model. This characteristic ensures that the results are solely based on the available sequence data, mitigating the risk of biases induced by modelling assumptions. Moreover, it is a best practice in phylogenetic analysis to employ multiple methods and compare their results. This comparative approach ensures the robustness of the findings.

However, employing a tree-like model has inherent limitations in resolution due to its strict bifurcating nature, which fails to consider factors like gene flow and recombination. Introgression assumes a paramount significance in crop evolution. This importance arises not merely from the fact that many crops have evolved from multiple Crop Wild Relatives (CWR), but also due to human-mediated dispersal, which further complicates the dynamics of genetic exchange.

TreeMix (Pickrell and Pritchard 2012) is a ML phylogenetic method, which allows to infer population trees and identify patterns of gene flow among populations. In TreeMix a maximum likelihood tree of populations is first built (i.e. the nodes of the tree represent populations, not individuals). Then populations that are poor fits to the tree model are identified and migration events involving these populations are modeled with the aim of increasing the likelihood of the tree itself. TreeMix enables the inference of phylogenetic relationships between populations by incorporating genetic signals linked to gene flow events. This is achieved through the measurement

of shared genetic drift between populations, providing a detailed understanding of their evolutionary relationships and the extent of genetic exchange over time.

1.7.3. Ancestry inference: SNP- and haplotype-based methods

In population genetics, a fundamental pursuit involves unraveling ancestral relationships, descendance, and admixture events among populations. For this reason, numerous methodologies have been developed to address these intricate inquiries.

ADMIXTURE (Alexander, Novembre, and Lange 2009) is a maximum likelihood-based algorithm that models the probability of obtaining the observed SNP genotypes at the individual level using ancestry proportions and allele frequencies. Given *K* ancestral components, it estimates the proportions of each ancestry for every individual in the dataset, so that each individual has a proportion of ancestry from one or more components. It is a powerful tool for the identification of population structure and admixture events, as well as overall relationships between populations, but care should be taken in interpreting results when a selfing species is analyzed. Indeed, ADMIXTURE model assumes that individuals are unrelated (panmixia) and the high levels of homozygosity in selfing species may hinder the correct identification of ancestry proportions. Reduced genetic diversity within populations can lead to an overestimation of distinct clusters, potentially misrepresenting the true genetic structure. Additionally, selfing, causing increased homozygosity, can mask underlying genetic complexities within a population. Besides these limitations, ADMIXTURE have been successfully applied in the analysis of selfing species (e.g., Aranzana et al., 2010; Igolkina et al., 2023), but it remains pivotal to interpret the results in light of other genetic and historical evidence and not to over-interpret ADMIXTURE outcome (Daniel J. Lawson, van Dorp, and Falush 2018).

Another set of methods, including ChromoPainter, SourceFind and FastGlobetrotter use genetic data to infer haplotype sharing and admixture events. ChromoPainter (Daniel John Lawson et al. 2012) finds haplotypes in sequence data. It identifies

haplotypes and tries to reconstruct each individual's genetic makeup as a mosaic of diverse ancestral segments ("haplotype chunks"), each representing shared ancestry with specific individuals in the population. Essentially, it "paints" the genetic landscape of each individual, delineating the contributions from various ancestors within the population. The outcome is a matrix that quantifies the relatedness between individuals (co-ancestry matrix), used as input for SourceFind and FastGlobetrotter.

Based on ChromoPainter co-ancestry matrix, SourceFind (Chacón-Duque et al. 2018) uses a Bayesian model to determine ancestry proportions within populations. Given donor and target populations, it reconstructs the target population as a mixture of the donors, removing contributions that cannot be distinguished from background noise. Essentially, SourceFind outputs the ancestry proportion of each donor population in the target of interest.

The patterns of ancestry defined by ChromoPainter are used by Globetrotter (Hellenthal et al. 2014) to infer the time since admixture of donor populations in shaping the genome of target ones. Globetrotter constructs individual co-ancestry curves for pairs of donor populations. These curves illustrate the relationship between genetic distance (in cM) and the frequency with which haplotype chunks, separated by a certain distance, originate from each respective donor population. This methodology is grounded in the observation that after a single admixture event, inherited ancestry chunks from each source population follow an exponential size distribution. Consequently, the co-ancestry curves exhibit exponential decay, with the rate of decay across all curves directly reflecting the number of generations since the admixture event. A steeper decline in the curves indicates an older admixture event.

These methods have been mostly applied to human data and to my knowledge they have never been applied to plant data before. As mentioned above, the emmer wheat genome is characterized by long haplotypes, slow decay of LD, increased homozygosity, and very low outcrossing rates. All these aspects can potentially bias the results: it is essential to adjust the parameters of the analysis to account for the biology of the species of interest (see paper 2).

In summary, the intricate biological complexities and multifaceted history of emmer wheat's domestication, gene flow, and improvement necessitate a rigorous and comprehensive approach to data analysis. Employing a variety of methods is crucial in establishing a reliable and accurate interpretation of results.

1.8. A note on ancient DNA

(summary of the section "Analyzing degraded DNA from ancient polyploid wheat" – Iob, Scott, Botigué, in press, see annex)

Ancient DNA (aDNA) analysis revolutionized evolutionary genomics, by providing insights into the genetic diversity of past populations (Gutaker and Burbano 2017; Orlando et al. 2021; Der Sarkissian et al. 2014). Despite its tremendous potential, ancient DNA studies are challenged by inherent features like degradation and contamination. Consequently, specific methods have been developed in both sample preparation and subsequent analysis (reviewed by Orlando et al., 2021).

Typically, DNA in ancient samples is preserved in very small amounts, and it is highly fragmented, due to a biochemical process known as 'hydrolytic depurination,' which leads to the breakdown of the DNA 'backbone.' This degradation occurs at a faster pace in the presence of water and elevated temperatures (Lindahl 1993). Consequently, the local preservation conditions and environmental factors play pivotal roles in determining the yield and quality of DNA from different samples. Despite these challenges, successful DNA sequencing has been achieved from ancient plant tissues from tropical and warm climates (Fornaciari et al. 2018; Mascher et al. 2016; Ramos-Madrigal et al. 2016; Renner et al. 2019) and robust DNA preservation has been reported in plant remains found in desiccated and waterlogged conditions (Logan Kistler et al. 2020). Furthermore, the DNA sequence is altered after death, as a portion of cytosine residues undergoes deamination, transforming into uracil residues, which are read as thymines during sequencing (Briggs et al. 2007). This hydrolytic deamination predominantly occurs in the single-stranded overhangs of fragmented DNA molecules. Consequently, ancient DNA sequences exhibit a higher

frequency of C-to-T misincorporations at the 5' end of each fragment and, in double-stranded DNA, an increased occurrence of G-to-A at the 3' end following alignment. These characteristic degradation patterns serve as indicators of the ancient origin of the sample and are instrumental in ruling out contamination.

Contamination can arise from microbial decomposers infiltrating tissues post-mortem, and be therefore present within the sample before processing. This type of contamination cannot be avoided, but only assessed. The percentage of sequencing reads that can be confidently mapped to the reference genome of the targeted species is often used as an estimate of the overall contamination, although alternative methods are available (Peyrégne and Prüfer 2020). Furthermore, during sample processing even minor levels of contamination from contemporary sources can overwhelm the low amounts of the target DNA in the library (Renaud et al. 2019). Therefore, the extraction and manipulation of ancient DNA necessitate specialized facilities with protocols designed to minimize the introduction of modern DNA (Fulton 2012).

Standardized bioinformatic procedures have been established for processing fragmented and degraded DNA. These procedures involve the removal of damaged base pairs at the ends of DNA fragments (Jónsson et al. 2013) and the exclusion of transitions (SNPs where the two alleles are either C/T or G/A, including those possibly influenced by postmortem damage) to ensure robust analyses (Der Sarkissian et al. 2014). Moreover, established methodologies exist for aligning short-read data to reference genomes, and automated tools and pipelines are available for genotyping of ancient samples (Peltzer et al. 2016; Schubert et al. 2014).

Finally, investigations into plant archaeogenomics face more pronounced challenges, as the difficulties of ancient DNA analysis are compounded by the inherent difficulties of plant genomic analysis (Logan Kistler et al. 2020). Nevertheless, over the past decade, archaeogenomic studies have been conducted on several economically significant crops, with primary emphasis on the processes of domestication, crop dispersal, and subsequent adaptation (Orlando et al. 2021).

Notable examples include maize (Ramos-Madrigal et al. 2016), barley (Mascher et al. 2016; Palmer et al. 2009), cotton (Palmer et al. 2012), bean (Trucchi et al. 2021), sunflower (Wales et al. 2019), sorghum (Smith et al. 2019), watermelon (Renner et al. 2019), emmer wheat (Scott et al. 2019), grape (Ramos-Madrigal et al. 2019), and potato (Gutaker, Weiß, et al. 2019).

Chapter 2

OBJECTIVES

The primary aim of this thesis is to advance our understanding of the emmer wheat domestication, dispersal and adaptation to new environments through the genomic analysis of wild and domestic specimens. This study relies on a comprehensive dataset of samples from various geographical origins, alongside an ancient sample from Egypt. Special emphasis is placed on the contribution of wild relatives to the domestic gene pool and their role in the adaptation to new environments. Such knowledge is pivotal to set the basis for future wheat improvement efforts, as both wild relatives and traditional landraces such as emmer wheat represent invaluable sources of genetic variation.

To accomplish this overarching goal, four specific objectives are delineated:

Creation of Genomic Dataset: This objective involves the generation of a high-quality genomic dataset from sequence data, including the ancient sample. The process encompasses aligning sequence data to the reference genome, performing variant calling, and applying rigorous quality filtering to retain only highly reliable Single Nucleotide Polymorphisms (SNPs). Specific filtering techniques are employed in light of the specific characteristics of the wheat genome, and when including the ancient samples, to mitigate the effects of post-mortem damage.

Analysis of Population Structure: The second objective aims to investigate the population structure of modern wild specimens, domestic landraces, and their relationship with the ancient sample. Various population genetics methods, including DAPC, phylogenies, ADMIXTURE, and genetic diversity estimation (PI), are utilized to identify distinct populations, establish phylogenetic relationships, identify shared ancestry patterns, and estimate genetic diversity.

Role of wild populations in shaping domestic genomes: This objective focuses on understanding the contribution of the wild Northern and Southern Levant populations on the diversity of domestic landraces through pre-domestication hybridization and post-domestication gene flow. Methods such as ChromoPainter,

SourceFind, Globetrotter, Dxy, TreeMix and Patterson's D test (ABBA-BABA) are applied to evaluate wild ancestry proportions, estimate admixture time, calculate genetic distances between populations and identify gene flow. These analyses, in conjunction with archaeological evidence, provide insights into the dynamics of domestication and into dispersal routes from Southwest Asia to India.

Wild Southern Levant Population's Role in Adaptation: The final objective explores the potential role of the wild Southern Levant ancestry in adaptation. Annotated genes affected by high or moderate impact variants within regions closely related to the Southern Levant population are selected. These genes are scrutinized for overrepresentation of Gene Ontology (GO) terms to identify the Southern Levant population's impact on adapting to Southern climates. Selection is further tested using site- and haplotype-based methods, including Tajima's D and cross-population extended homozygosity (XP-EHH), providing an overview of regions under selection in domestic landraces.

Chapter 3

RESULTS

3.1. Genomic analysis of emmer wheat shows a complex history with two distinct domestic groups and evidence of differential hybridization with wild emmer from the western Fertile Crescent

Iob A., & Botigué, L.

Vegetation History and Archaeobotany. 2023

DOI: 10.1007/s00334-022-00898-7

This section is a verbatim reproduction from the following published paper:

Iob, A., & Botigué, L. (2023). "Genomic analysis of emmer wheat shows a complex history with two distinct domestic groups and evidence of differential hybridization with wild emmer from the western Fertile Crescent". *Vegetation History and Archaeobotany*, 32(5), 545-558.

NOTE: in this paper WSL (wild Southern Levant) is called W-WFC (wild Wester Fertile Crescent) and WNL (wild Northern Levant) is called W-NEFC (wild Northern and Eastern Fertile Crescent).

Abstract

Triticum turgidum ssp. *dicoccoides* (wild emmer wheat) was one of the first plants that gave rise to domestic wheat forms in southwest Asia. The details of the domestication of emmer and its early dispersal routes out of southwest Asia remain elusive, especially with regard to its dispersal to the east. In this study, we combine whole genome data from a selection of specimens of modern wild *T. turgidum* ssp. *dicoccoides* and domestic *T. turgidum* ssp. *dicoccon* (emmer wheats) with a previously published 3 000 years old sample, to explore the phylogenetic relationships between wild and domestic populations of emmer, and especially the early dispersal routes south and eastwards to Africa and Asia, respectively. Our data confirm a marked differentiation between landraces from Europe, the Caucasus and Iran, and those from Africa, the Arabian Peninsula and India, the first group being more closely related to wild emmer from the northern and eastern Fertile Crescent. Gene flow is detected between wild emmer from the western Fertile Crescent and the second group of domestic emmer. These results support a dispersal route from southwest Asia to Africa, the Arabian Peninsula and India. We also observe a lower genetic variability in the wild emmer from the northern and eastern compared with that of the western Fertile Crescent. It is possible that the ancestors of domestic emmer that spread into Egypt still remain to be surveyed and analyzed. Investigating the genetic content of ancient samples from Europe, the Caucasus or Iran would be very valuable to determine whether the two distinct types of germplasm arose during history or were already present during the early phases of dispersal.

Keywords Emmer wheat · Domestication · Dispersal · Genomics · Ancient DNA

Introduction

The domestication and spread of crop plants has aroused the curiosity of the scientific community for a century now, ever since the pioneering studies by Nikolai Vavilov, detailed in Studies on the origin of cultivated plants (N. Vavilov 1926). Technical innovations and methodological improvements have allowed a re-evaluation of old theories about plant domestication, in which evidence from both archaeobotany and genetics shows domestication as a geographically diffused and genetically varied phenomenon in southwest Asia (Allaby et al. 2017; Dorian Q. Fuller et al. 2014; Pankin et al. 2018), even if some of the details of this process at the species level remain elusive.

Triticum turgidum ssp. *dicoccoides* (Asch. & Graebn.) Thell. (wild emmer wheat) is a tetraploid mainly self-pollinating hulled cereal. Its range of distribution covers the Fertile Crescent, including Israel, Jordan, southwestern Syria, Lebanon (the western Fertile Crescent), and southeastern Turkey, northern Iraq and western Iran (the northern and eastern Fertile Crescent) (Ozkan et al. 2011; Zaharieva et al. 2010). It is unclear whether domestic emmer emerged in the western or the eastern and northern Fertile Crescent or independently in both regions. In the site of Ohalo II in the western Fertile Crescent (Kislev, Nadel, and Carmi 1992; Weiss et al. 2004) dated to ca. 23 000 – 21 000 years cal BP, as many as 36% of the emmer rachis remains carry the diagnostic scar associated with a non-shattering rachis and thus with domestication (Snir et al. 2015). Archaeological assemblages with 100% domestic emmer are first found dating to 10 600 – 10 200 years cal BP, from sites in the northern and eastern Fertile Crescent such as Çayönü, Turkey (Van Zeist and De Roller 2003).

Most of the studies of genetics and genomics have been based on modern material. Modern domestic landraces of emmer are closer to wild emmer accessions from the northern and eastern Fertile Crescent (Avni et al. 2017). Also, early genetic models based on the similarity between domestic and wild emmer from the northern and eastern Fertile Crescent suggested that southeastern Turkey and northern Syria were

the centers of emmer domestication (Lev-Yadun et al., 2000; Ozkan et al. 2002). More recently, the idea that domestic emmer originated from a single homogeneous wild population has been questioned (Jorgensen et al. 2017), with some authors proposing a model in which fully domestic emmer wheat emerged from a population in the northern and eastern Fertile Crescent that would carry the genetic background of multiple wild emmer populations, including those from the western Fertile Crescent (Civáň et al., 2013). This would be in agreement with other genetic studies, which find a greater level of similarity between domestic emmer populations and wild populations from the northern and eastern Fertile Crescent (H. R. Oliveira et al. 2020), but also a contribution from the western Fertile Crescent populations to important domestic haplotypes (Nave et al. 2019).

These genetic findings are compatible with archaeobotanical evidence of common finds of wild emmer in the western Fertile Crescent from the pre-pottery Neolithic A (PPNA) (11 600–10 700 cal BP) with increasing proportions of domestic types from the following periods, whereas in the northern and eastern Fertile Crescent other taxa were preferentially consumed. It was not until the middle and late pre-pottery Neolithic B (PPNB) (10 200 –8300 cal BP), that there was a change towards an increased management and consumption of domestic wheats in the northern and eastern Fertile Crescent, evident from findings from different archaeological sites (Arranz-Otaegui et al. 2016; Kabukcu et al. 2021). The Iranian site of Chogha Golan represents a good example of this, with its sequence of over 2200 years of plant management there. From the PPNA to mid Pre-pottery Neolithic B (MPPNB), only a small fraction of the remains from Chogha Golan is represented by wild emmer, while other cereals such as *Hordeum vulgare* (wild barley) are predominant. However, starting from 9800 cal BP, domestic emmer appears, soon outnumbering other large-seeded grasses (Riehl et al., 2013).

From southwest Asia, emmer spread towards Europe, central Asia and Africa. While dispersal of emmer into Europe has been well studied together with the spread of agriculture (for example, Coward et al. 2008), its dispersal to the south and east is still a matter of debate (Stevens et al. 2016). In Africa, the first settlements in Egypt date

to 7500–6650 cal BP (Wendrich and Cappers 2005). Around 5000 BP emmer reached Ethiopia (Helbaek 1970). Whether Ethiopian emmer is a descendant of Egyptian emmer or if it reached there through the Iranian highlands and the Arabian Peninsula is not known (M. C. Luo et al. 2007).

Triticum turgidum ssp. *dicoccum* (Schrank ex Schübl.) Thell. (fully domestic emmer), together with *T. monococcum* (einkorn), *Hordeum vulgare* (barley), *Pisum sativum* (pea), *Lens culinaris* (lentil) and *Linum usitatissimum* (flax) reached western Iran by 9000 BP (even though evidence of management of partially domestic emmer is known from Chogha Golan as far back as 9800 BP (Riehl et al., 2013)) northern India (the upper Punjab plain) by the first half of the 5th millennium BP and southern India during the 4th millennium BP. While emmer was important in India, it was not introduced into central Asia (north or east of Turkmenistan and to Afghanistan), where only *Triticum aestivum* (free-threshing wheat) is found in the archaeological record (Stevens et al. 2016).

Modern landraces of emmer from India were classified by Vavilov (1926) as *indostanicum* group within subspecies *abyssinicum* (Ethiopian emmer), suggesting that the dispersal of emmer into Ethiopia and India was somehow connected. Two routes have been proposed to explain this. One possibility is that the introduction into northwest India occurred via Iran and Afghanistan (Mani 2004) and then spread south into India. The second hypothesis proposes a first introduction to India by sea from the Arabian Peninsula, from a population characterized by low genetic diversity (Salunkhe et al. 2013). This theory would be compatible with known trade routes from the Red Sea across the Arabian Sea or the Indian Ocean which existed since Greek and Roman times (M. C. Luo et al. 2007). Both possibilities are compatible with trade routes connecting India through the coast of Iran with the Arabian Peninsula and from there to Ethiopia.

One limitation of the genetic studies is that they are all based on modern data, and it is not possible to determine whether modern landraces are representative of those present at the time of early dispersal of cultivated emmer. The publication of the

genomic sequence of an ancient sample of emmer from Egypt dated to 3130–3000 BP (Scott et al. 2019) re-opened interesting questions about its dispersal. This sample resembled present-day landraces from India, the Arabian Peninsula and Turkey. Interestingly, this study showed signals of genetic introgression from the western Fertile Crescent, which possibly occurred when emmer was cultivated, but before its introduction to Egypt, or during later interactions, indicating a connection between early emmer dispersal eastwards across the Iranian plateau and into the Indus valley and also to the southwest into the Nile valley. However, due to the lack of examples of emmer from Africa in the dataset, it was not possible to find out about these early emmer trade routes from this study.

Here we re-analyze this ancient sample in the context of a more extended modern genomic dataset with representation of emmer landraces from Ethiopia, to elucidate the phylogenetic relationships between domestic and wild emmer populations, and with the ancient sample, with the ultimate goal to bring insight into the routes of dispersal of emmer (Fig. 1).

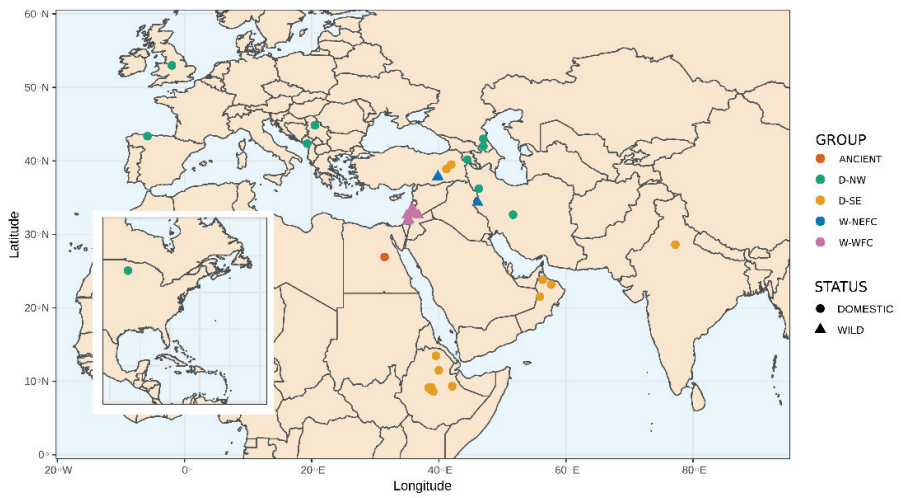


FIGURE 1: MAP SHOWING THE DISTRIBUTION OF SITES OF THE SAMPLES ANALYSED IN THIS STUDY, COLOURED ACCORDING TO THE GENETIC CLUSTERING IDENTIFIED IN THE ANALYSES. WWFC WILD WESTERN FERTILE CRESCENT, WNEFC WILD NORTHERN AND EASTERN FERTILE CRESCENT, DNW DOMESTIC NORTHWEST ROUTE OF DISPERSAL, DSE DOMESTIC SOUTHEAST ROUTE OF DISPERSAL. SAMPLES FROM THE WNEFC POPULATION COME FROM ONLY TWO SAMPLING SITES (7 SAMPLES, 2 SITES). SAMPLE WLBN3 FROM LEBANON IS NOT REPRESENTED DUE TO LACK OF GEOGRAPHICAL INFORMATION

Materials and methods

Samples

Main dataset

Previously published whole genome sequence data from 57 emmer accessions were downloaded from https://bigd.big.ac.cn/search/?dbId=gsa&q=CRA001951 (Y. Zhou et al. 2020). This dataset comprises 28 wild samples and 29 domestic samples covering Europe, western Asia and the Horn of Africa.

Extended dataset

For some analyses, the main dataset is extended to include an ancient sample from Egypt, 14C dated to 3130–3000 cal BP (Scott et al. 2019), and four domestic landraces from Turkey, Oman and India, sequenced by exome capture (Avni et al. 2017). Details for the dataset can be found in ESM Table S1, and the geographical distribution of the samples is shown in Fig. 1.

Alignment and variant calling

Main dataset

We aligned all the samples to the durum wheat genome (*Triticum durum*), using the program bwa v. 0.7.17 (H. Li and Durbin 2009). For each wheat specimen, this process takes all the sequenced DNA fragments (reads) and finds its coordinates (chromosome, position) in the reference genome. We then sorted and indexed the resulting *.bam files using Samtools 1.9 (Danecek et al. 2021). Variant calling (the process of identification of genetic differences between the sequenced specimen and the reference genome) was done with the genome analysis toolkit GATK v. 4.1.6 (Van der Auwera and O'Connor 2020) and Picard v. 2.22.3 (Institute 2019), retaining only biallelic SNPs (single nucleotide polymorphisms, which are changes to a single DNA

base, hereafter referred to as "variants"), and hard-filtering the dataset following Zhou et al. (2020) to keep only those genetic variants that were identified with high confidence. We further removed variants (SNPs) with more than 10% missing information or that were found in one sample only using VCFtools v. 0.1.16 (Danecek et al. 2011) with the commands -max-missing 0.1 and -mac 3. The total number of variants in this dataset was 66 million (66,097,433 SNPs). We further filtered our dataset for linkage disequilibrium using plink v. 1.9 whole genome association analysis toolset (Purcell et al. 2007), allowing a maximum r^2 value of 0.1 calculated in 50 kb windows with a step size of 10 kb, reducing the dataset to 4.5 million variants (4,444,631 SNPs). Increasing the value of r^2 to 0.4 gave the same results in both principal component analysis (PCA) and discriminant analysis of principle components (DAPC). For this reason, we used the filtering for r^2 0.1 in all analyses.

Extended dataset

The modern samples from the extended dataset were also aligned to the durum wheat reference genome and processed in the same way as the main dataset. The data from the ancient samples were processed as in Scott et al. (2019) to adequately account for the characteristics of degraded DNA. Variants were re-called on the whole dataset as previously described and sites present in both datasets were kept, resulting in 3.6 million (3,689,770 SNPs) variants. After filtering for linkage disequilibrium (LD, r^2 0.1 as above) and removing transitions to avoid errors from postmortem damage, the number of variants was ca. 400,000 (433,973 SNPs).

Outgroup

Some of the analyses required the comparison of the genetic data observed in emmer wheat with that of an outgroup, a more distantly related group that can serve as a reference when determining the relationships within the ingroup. Since wild emmer is the result of an ancient hybridization event between *Triticum urartu* (A genome) and *Aegilops speltoides* (B genome), it is not possible to find a tetraploid outgroup in nature. We circumvented this by following Scott et al. (2019) and used publicly available genomic data of the donors of the A and B genomes of emmer as an outgroup. We

downloaded genomic sequences of *A. speltoides* (SAMEA2342530, European Nucleotide Archive) representing the B subgenome and *T. urartu* (sample A082 from Zhou et al. 2020, https://bigd.big.ac.cn/searc h/?dbId=gsa&q=CRA001951), representing the A subgenome. Paired-end reads were aligned to the durum reference genome. We retained only reads mapping to the A subgenome for *T. urartu* and the B subgenome for *A. speltoides*. Retaining only these reads allowed us to avoid mis-mapping biases from reads mapping to the wrong homologue, an issue that is common in wheat studies, due to the high levels of homology between subgenomes.

After re-calling variants in the whole dataset, we kept only the sites that were polymorphic in the emmer dataset. This led to the identification of 56 million variants (56,200,433 SNPs), which after LD filtering reduced to 4 million variants (4,078,964 SNPs).

All the analyses used the dataset which had been filtered for LD, except the calculation of the nucleotide diversity.

Population structure of emmer

PCA

In order to get a general overview of the relationships between our samples we did a Principal Components Analysis (PCA) on the main dataset, using plink v. 1.9 and plotted the results in R v. 4.1.0 (R Core Team 2021). This analysis led to the identification of two outliers (samples ISR5 and SRB3, ESM Fig. S1). In the absence of additional information about these samples, we decided to exclude them from further analyses.

DAPC

Since emmer wheat is a self-fertilizing, highly inbred taxon, we also performed a discriminant analysis of principal components (DAPC) on the main dataset to better discriminate between groups. DAPC identifies clusters by minimizing the differences

within groups while maximizing the differences between groups. This was done with the R package Adegenet, in R v. 4.1.0 on variants that were at least 25,000 bases apart (VCFtools v. 0.1.16 command—thin 25 000 will take one variant every 25,000 bases) resulting in 319,331 variants. In DAPC the number of retained principle components (PCs) is critical, as retaining too many of them compared with the sample size could lead to over-fitting and a subsequent distortion of the results. For this reason, we performed the cross validation using the xvalDapc function, and retained the first ten principle components (ESM Fig. S2).

ADMIXTURE

This is a maximum likelihood based unsupervised clustering algorithm that estimates the proportions of an established number of ancestries for each specimen in the dataset (Alexander et al., 2009). Maximum likelihood methods estimate the most probable model given the observed data. Given a certain number of ancestral components, K, the individuals can be represented as a mixture of such components. In order to determine the best number of K for the main dataset, we used cross validation error analysis in ADMIXTURE, with values of K from 1 to 10. The best values for K proved to be 3 (cross validation error 0.58) and 4 (cross validation error 0.58), even if very close values were obtained for the values 2, 5 and 6 for K (cross validation error 0.61–0.63). All analyses were done using ADMIXTURE v. 1.3.0 and results plotted using R v 4.1.0.

Phylogenetic analysis

In order to discover the phylogenetic relationships between the samples in our dataset, we first converted the format of the dataset from vcf to relaxed Phylip using the script vcf2phylip.py, downloaded from https://github.com/edgardomortiz/vcf2phylip, and used the resulting phylip file as input for RaxML (randomized axelerated maximum likelihood, Stamatakis 2014) and MEGA XI (molecular evolutionary genetics analysis, Kumar et al. 2018). We constructed a phylogenetic tree using maximum likelihood with RaxML v. 8.2.12 and our main dataset. We ran a RaxML rapid bootstrap analysis (option -f a), searching

for the best tree out of 20 runs (option -# 20). For the extended dataset, since it included an ancient sample, and in order to avoid introducing any bias due to model selection, we used neighbor joining clustering to construct a phylogenetic tree, based on p-distance with 100 bootstrap replicates using MEGA XI; bootstrapping is a re-sampling method for assessing the reliability of the results. We kept only transversions (variants that entail a change from a purine (Adenine, Guanine) to a pyrimidine (Cytosine, Thymine) or vice versa) in order to eliminate the potential misincorporations from Cytosine to Thymine and from Guanine to Adenine related to ancient DNA (aDNA) damage. The resulting phylogenetic tree topology (which is the branching structure of the tree, indicating the patterns of relatedness among taxa) was mainly consistent with that from the main dataset based on a maximum likelihood approach, which was further verified by constructing a new tree for the main dataset samples with neighbor-joining (ESM Fig. S3).

Genetic diversity

The amount of genetic diversity in each group was calculated as nucleotide diversity (PI) using VCFtools v. 0.1.16. For groups with few samples it is possible that the genetic variability within the group is not representative of that of the real population. Since the northern and eastern Fertile Crescent group consisted of only seven samples compared to other group sizes of > 13, bootstrapping was used to enable comparison between groups, in which four random samples were extracted from each group to calculate nucleotide diversity (PI), and the process was repeated ten times. The averaged value of PI from all the subsets within a population is taken as its value for the whole group.

Analysis of the southeastern dispersal route and gene flow

We analyzed the genetic make-up of populations of domestic emmer that dispersed to the south towards Africa and the Arabian Peninsula, and also to the east towards Asia and India, in an attempt to detect hybridization events, especially with the wild western Fertile Crescent (WWFC) population.

D statistic

In order to detect gene flow between different populations, we calculated Patterson's D (ABBA-BABA test) (Green et al. 2010) between the two wild groups (WNEFC for northern and eastern Fertile Crescent and WWFC for western Fertile Crescent) and the two domestic groups, DNW for the northwestern route to Europe, the Caucasus, Balkans and Iran and DSE for the southeastern route to Ethiopia and India in the main dataset. This analysis is based on the fact that, given a known phylogeny with four populations (((P1, P2) P3) OUTGROUP) represented as BBAA, in which A is the ancestral condition (allele) and B the derived one, by analyzing different genomic regions one can obtain a certain number of phylogenies which do not fit, such as ABBA and BABA (P2 and P3 sharing the derived allele and P1 and P3 sharing the derived allele), due to incomplete lineage sorting. This is random, and in the absence of gene flow between populations, the number of ABBAs and BABAs should be equal or not significantly different and the D statistic should be zero. On the other hand, an excess of ABBAs or BABAs and the resulting deviation of D from 0 is a sign of gene flow between the two populations. The statistic was computed using Dsuite v. 0.4 r41 (Malinsky et al., 2021). Significant results are defined by an absolute Z-score for relationship to the average larger than 3.

TreeMix

TreeMix v. 1.13 (Pickrell and Pritchard 2012) builds a phylogenetic tree using maximum likelihood and allowing for gene flow between populations. We used this on the modern dataset after converting the *.vcf data to TreeMix input format, using the script vcf2treemix.sh, downloaded from https://github.com/speciationgenomics/scripts/blob/master/vcf2t reemix.sh, and used the output file as input for TreeMix. We tested for 0 to 3 migration edges, which represent events of migration (gene flow) from one population to another (represented as arrows in the output) (-m 0 to 3). Applying bootstrap validation with blocks of 500 SNPs (-bootstrap -k 500), we found that the only meaningful trees were the ones with 0 or 1 migration edges (more edges show gene flow from the outgroup). We also used TreeMix on the extended dataset, this time without the sample size

correction (applying -noss) and, as in the tree using neighbor joining, we kept transversions only and filtered for linkage disequilibrium (LD). The results were plotted in R using the script plotting funcs.R, which is included in TreeMix.

Results

Re-processing the whole genome sequence data

The samples published by Zhou et al. (2020) were aligned to the bread wheat reference genome. In order to be able to capture as much genetic variation as possible, we re-processed them to align them to the durum wheat reference genome (Maccaferri et al. 2019). The geographical distribution of the samples is shown in Fig. 1, with the samples colored according to the results from discriminant analysis of principal components (DAPC).

After re-processing this genomic sequenced data, we had a high-quality dataset of 66 million variants (66,097,433 SNPs) for 55 samples. After filtering for linkage disequilibrium (LD) the dataset was reduced to 4.4 million variants (4,444,631 SNPs). For those analyses that required an outgroup, we kept only the sites that were polymorphic in the emmer dataset. This led to the identification of 56 million variants (56,200,433 SNPs), which after LD filtering reduced to 4 million (4,078,964 SNPs) variants. For the analysis of the southeastern migration route, adding the data from exome capture and the ancient samples resulted in a dataset of 3.7 million (3,689,770 SNPs) variants, which after filtering for LD and removing transitions to avoid errors related to postmortem damage resulted in ca. 400,000 (433,973 SNPs) variants. For these variants the mean depth of the whole genome sequenced (WGS) samples (including outgroup) calculated with VCFtools is 5.2×, that is, on average each of the positions where these variants are found, is covered 5.2 times by the sequenced DNA fragments. We called variants on the exome capture samples and on the ancient sample only on sites known to be polymorphic in the main dataset, allowing for a

minimum depth of 1×. The covered sites for the ancient sample have a mean depth of 1.47×.

Population structure

To get a general overview of the relationships between modern emmer landraces, we used principal component analysis (PCA) and discriminant analysis of principal components (DAPC). Both these methods are used to infer population structure without previous knowledge, by determining the number of observed clusters in the dataset, based on principal components, which are the ones that explain most of the genetic variability in the dataset. In DAPC the differences between groups are maximized, while the differences within them are minimized. These methods are ideal for measuring the degree of differentiation between the samples when using genetic information only.

The resulting clustering is used for finding out about the possible causes of these observed patterns, for example, geography. The PCA led to the identification of four groups of samples, differentiating wild from domestic and grouping samples according to geographical patterns (ESM Fig. S1). Two outliers were detected, most probably due to inaccurate passport information, the meta-information such as origin, species, data of collection, that has been associated with this accession. After excluding these outliers, the samples were grouped into clusters for further analysis. Wild emmer samples from the western Fertile Crescent are shortened to WWFC and those from the northern and eastern Fertile Crescent to WNEFC in the following text. Domestic samples from Ethiopia and Oman are called DSE as in domestic southeast, to refer to the southeast route, whereas samples from Europe, the Caucasus, Balkans and Iran are called DNW as in domestic northwest, since they broadly represent the northwest route of emmer dispersal. We discuss the intriguing grouping of the domestic Iranian landraces in the DNW group, below.

This sample clustering was confirmed by DAPC (Fig. 2A). Interestingly, not all domestic samples are equally close to the WNEFC group, which would be expected

if there had been a single domestication and dispersal event. DNW landraces appear closer to the WNEFC group in the PCA and are on the same x-axis in the DAPC, whereas the DSE specimens are more distant from the WNEFC cluster in both analyses.

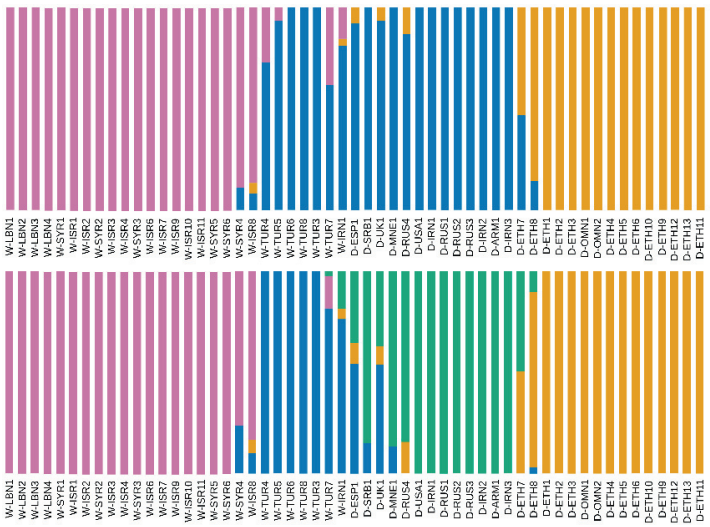
To investigate the population structure in more detail and identify shared genetic components between samples, we applied ADMIXTURE clustering analysis, which models shared ancestries between individuals. Values of K (representing the number of ancestries) between 2 and 6 were tested (ESM Fig. S4), and cross-validation errors determined that modelling three or four ancestries provided a best fit with the data. With a K value of 2, the dataset is divided into two ancestries, one characterizing all WWFC samples (shown in pink) and the other all domestic samples (blue), with WNEFC emmer specimens having a varying proportion of the two ancestries. When an additional ancestral component is allowed (K = 3), the domestic DSE group is assigned to this new component (yellow), the WWFC group keeps its ancestry (pink) and the WNEFC and DNW groups are largely represented by the other ancestral component (blue) (Fig. 2b). It is interesting to note that, in agreement with the PCA results, allowing a third ancestry results in the differentiation between DSE and the DNW-WNEFC group, which now appears homogeneous. At K = 4, the WNEFC and DNW groups are further differentiated, with domestic specimens from DNW being largely assigned to a new ancestral component (green) (Fig. 2B). Very low levels of admixture are found between the wild populations, and only a few samples from DNW appear to be admixed between WNEFC, DNW and DSE components, perhaps as a result of the under-representation of the Mediterranean landraces. Overall, these results are in agreement with the DAPC results.

We finally further explored the genetic affinities between wild and domestic modern emmer specimens by constructing a phylogenetic tree using maximum likelihood (Fig. 2C). Overall, the specimens are grouped into clades with the same grouping, such as that when using PCA and ADMIXTURE. The first node in the tree divides WWFC from all other populations, while the WNEFC samples cluster together and are an outgroup to all the domestic samples (with the exception of the only wild Iranian

emmer specimen, that appears as the closest relative to all other domestic emmers). Within the domestic cluster, samples maintain the DNW and DSE groupings, even though samples from Spain and the UK appear as outliers to the DNW clade, not following a clear geographical pattern. The DSE cluster is subdivided in three subclades, the most divergent being the one with the samples from Oman. In the WWFC cluster, on the other hand, the subclades mirror the geographical origins of the samples. Judging by their scattered position within the WWFC group, emmer from Syria seems to be the most diverse. This pattern was replicated with the neighbor-joining tree (ESM Fig. S3).

A.

B.



C.

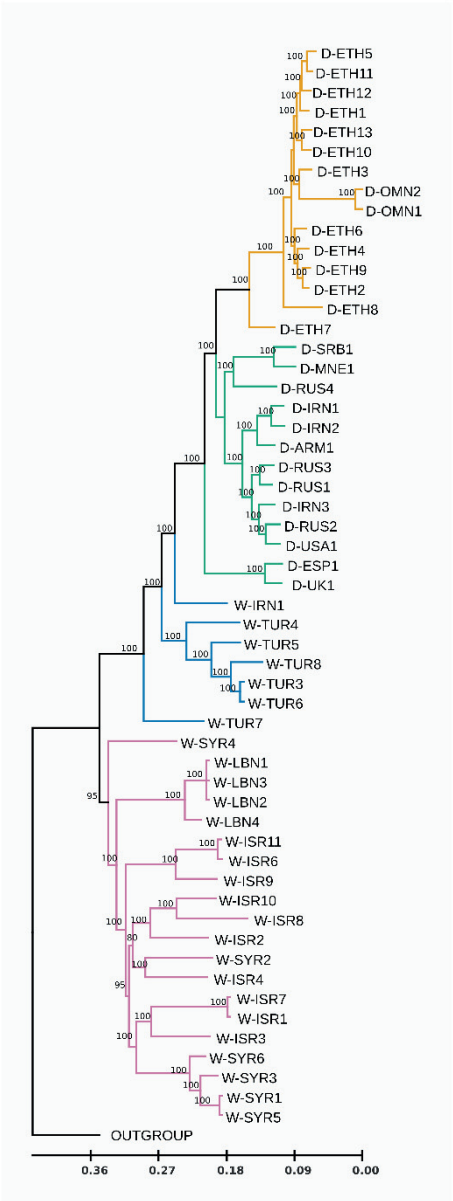


FIGURE 2: A. DISCRIMINANT ANALYSIS OF PRINCIPAL COMPONENTS (**DAPC**). GROUPS SHOWN BY THEIR COLOURS ARE ARRANGED ACCORDING TO THE PCA RESULTS AND CONFIRMED BY THIS ANALYSIS. **B ADMIXTURE** ANALYSIS FOR THE BEST VALUES OF K. UPPER PANEL, K = 3; LOWER PANEL, K = 4. **C MAXIMUM LIKELIHOOD TREE,** MODERN WHOLE GENOME DATA ONLY. THE TREE WAS CREATED WITH RAXML, USING A FAST SEARCH TO FIND THE BEST TREE OUT OF 20 RUNS. BOOTSTRAP VALUES ARE REPORTED AT THE TOP OF THE NODES

Overall these results confirm that WWFC is the most differentiated population within the dataset and they reinforce the observation of the genetic differences between the DNW and DSE landraces, as well as their varying affinity to the WNEFC samples. It is unclear why the DSE modern landraces are so different from this wild population. These results do not suggest that domestic emmer emerged in the north-eastern and western Fertile Crescent independently, since the DSE samples are not closer to WWFC. A wide range of modern wheat landraces were studied to investigate the allelic variability in the genes responsible for a non-shattering rachis (*TtBTR1-A* and *TtBTR1-B*) (Nave et al. 2019). The authors found no diversity within the domestic samples, suggesting that the fully domestic phenotype, to which wild populations WWFC and WNEFC probably contributed, had a single origin. The observed results could be explained if the ancestors of the modern DSE group had experienced a strong founder effect, the loss of genetic variability that occurs when a new population is established by a very small number of individuals from a larger population, possibly during the early dispersal of the group. In such a situation, the low genetic diversity within the group would have increased its genetic differences from other groups, deleting the genetic signature linking this group to the WNEFC emmer wheat ancestor, a process called genetic drift.

In order to determine whether genetic drift caused this differentiation in the DSE group, we calculated nucleotide diversity (Fig. 3). As expected, the wild populations show higher levels of nucleotide diversity, even if, intriguingly, the WWFC group shows levels of diversity (PI 0.19) almost twice those of WNEFC (PI 0.10). Domestic groups show lower levels of nucleotide diversity than WWFC and WNEFC, but similar levels of genetic diversity among themselves (DSE PI 0.06, DNW 0.07). As the genetic diversity of the DSE group is comparable to that of DNW, this evidence does not suggest a strong genetic bottleneck, a sharp reduction in the size of the population of the DSE ancestor, pointing to other possible explanations for its differentiation.

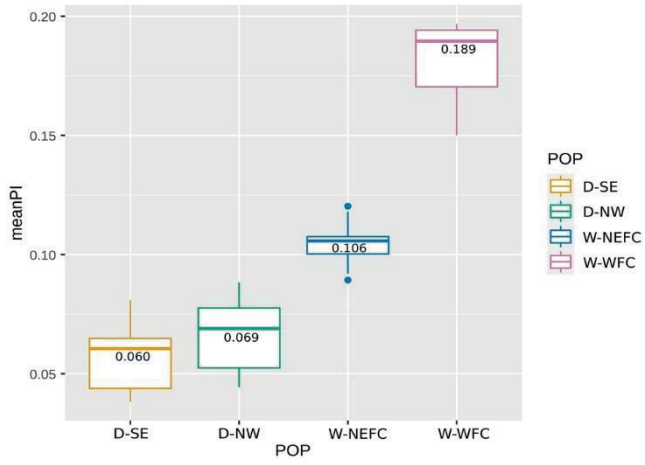


FIGURE 3: NUCLEOTIDE DIVERSITY (PI) IN DIFFERENT POPULATIONS, SHOWING THAT THE WWFC POPULATION IS THE MOST DIVERSE, WHILE WNEFC IS LESS SO. DOMESTIC POPULATIONS, DNW AND DSE, HAVE SIMILAR LEVELS OF DIVERSITY

Gene flow

Another explanation for the distinctiveness of the DSE group could lie in a different level of contribution of the WWFC populations to this group. This could have occurred during the domestication process or the early dispersal of the domestic emmer into Africa. In order to investigate if the domestic landraces from the DSE group show signs of admixture with wild specimens from the western Fertile Crescent, we calculated Patterson's D statistic (ABBA–BABA) (Green et al. 2010) for the deviation from the expected ratio of allele sharing between WWFC and the two domestic groups, plus the outgroup, using the phylogenetic tree: (((DNW, DSE) WWFC) OUTGROUP). The results show indeed an excess of allele sharing between WWFC and DSE (tree (((DNW; DSE) WWFC) OUTGROUP) D = 0.106, Z score = 3.15), compatible with gene flow from the wild population into the domestic one (Fig. 4 and ESM Table S2). Replacing the WWFC population by the WNEFC one yielded no evidence of gene flow between any of the domestic groups (topology

(((DSE; DNW) WNEFC) OUTGROUP) D = 0.074, Z score = 2.03), consistent with both domestic populations having more affinities with WNEFC than with WWFC.

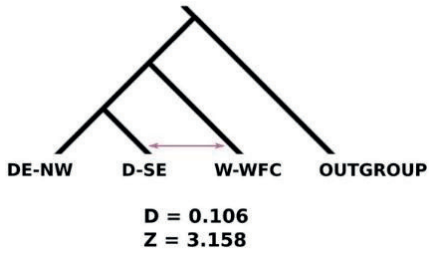


FIGURE 4: SCHEMATIC REPRESENTATION OF PATTERSON'S D (ABBA-BABA) STATISTICAL TEST RESULTS; THE ARROW SHOWS THE DIRECTION OF GENE FLOW BETWEEN WWFC AND DSE

In light of the evidence of hybridization observed between the WWFC and DSE groups, we constructed a maximum likelihood phylogenetic tree with TreeMix, which models gene flow by introducing edges of migration between populations (Fig. 5). As expected, under no migration, TreeMix constructed the same tree topology previously obtained. However, the plot of the residuals showed that the tree does not properly fit the data (ESM Fig. S5a). When allowing for an edge of migration, gene flow is identified from the WWFC population branch to the DSE leaf, giving a tree with increased likelihood (ln(likelihood) = − 4477.06 vs. ln(likelihood) = − 224.897 for 0 and 1 migration events, respectively) and lower standard error, ± 43.4 vs. ± 15.2 for 0 and 1 migrations events, respectively (ESM Fig. S5a, b), confirming the Patterson's D statistic results.

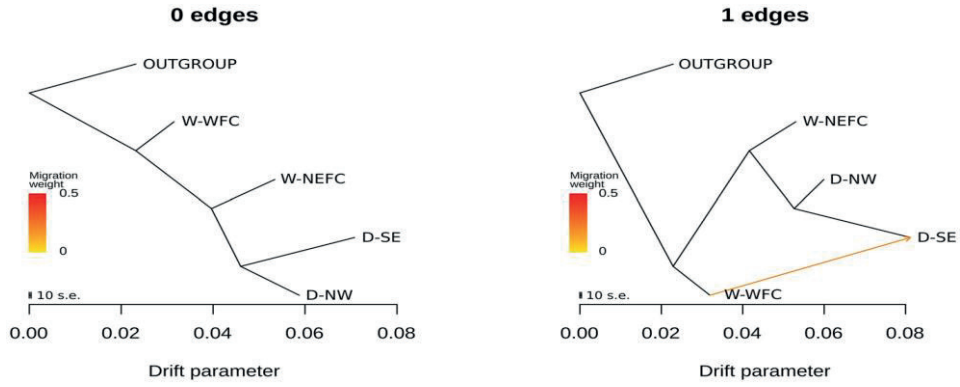


FIGURE 5: TREEMIX. MAXIMUM LIKELIHOOD ANALYSIS OF THE MODERN WHOLE GENOME DATASET ALLOWING FOR 0 (LEFT PANEL) AND 1 (RIGHT PANEL) MIGRATION EVENTS, SHOWING GENE FLOW BETWEEN WWFCSL AND DSE (ARROW). THE COLOUR OF THE ARROW IS PROPORTIONAL TO THE AMOUNT OF MIGRATION

In order to better characterize the nature of the DSE group, we included publicly available data on emmer from Turkey, Oman and India (referred to as the Indian Ocean group) that also differed from domestic samples from Europe and the Caucasus (Avni et al. 2017). We note that the inclusion of modern landraces from Turkey could potentially differentiate between whether hybridization happened during the domestication process or during the early dispersals from the northern and eastern Fertile Crescent southwards. We also included the genomic data of a 3,000 year old ancient Egyptian emmer sample (Scott et al. 2019) to determine whether the genetic make-up of this domestic group is modern or if it already existed in the past. We first constructed a neighbor joining tree to study how the specimens from this extended dataset were related to those from the main one. The extended dataset does not change the arrangement of the tree obtained with respect to the main one (Fig. 6). Interestingly, the samples D-TUR1 and D-TUR2 from Turkey cluster together with the samples from Ethiopia, while D-OMN3 from Oman and D-IND1 from India cluster with the other samples from Oman. The ancient sample UC10164 from Egypt is placed as an outgroup to the DSE group and shows a quite long branch, which could be due to the low coverage of the sample or a slightly different genetic make-up.

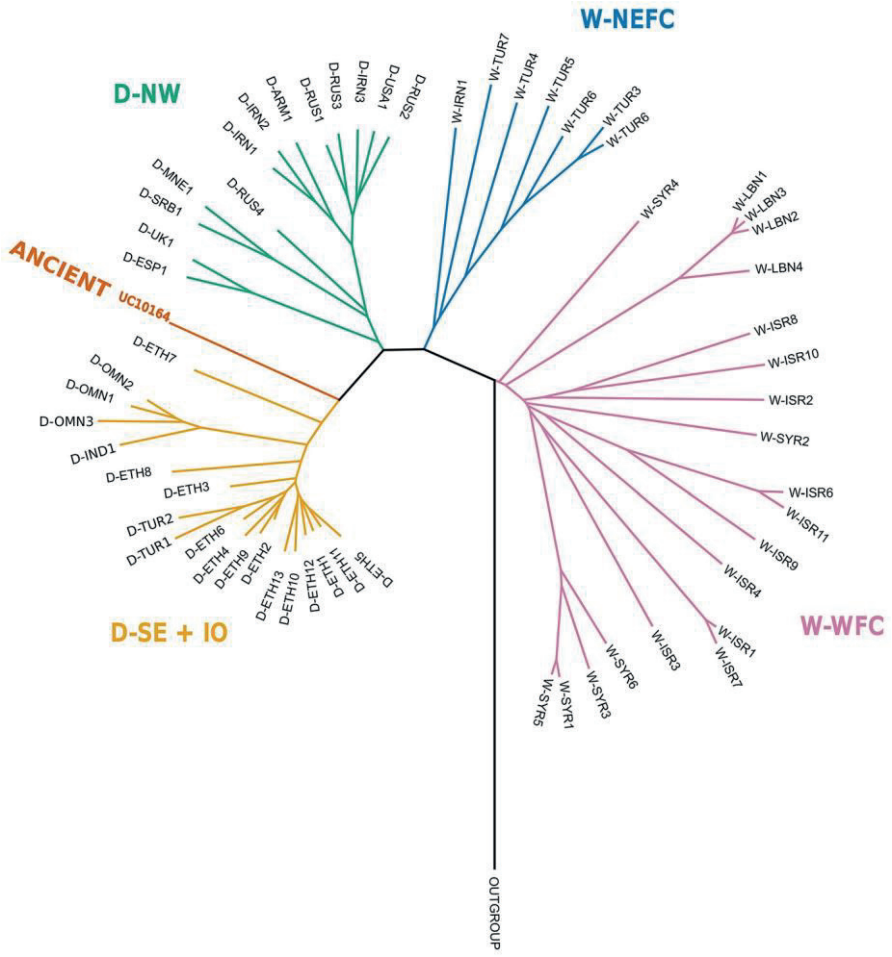


FIGURE 6: NEIGHBOR JOINING PHYLOGENETIC TREE OF THE EXTENDED DATASET, INCLUDING THE INDIAN OCEAN SAMPLES AND THE ANCIENT EGYPTIAN SAMPLE. BOOTSTRAP 100 REPLICATES, ALL NODES HAVE BOOTSTRAP VALUES ARE ABOVE 90.

We next constructed a tree that allowed for gene flow between its branches (Fig. 7). The new emmer specimens from India, Oman and Turkey were merged into the DSE group (now DSE + IO). One edge of migration in the tree continued to support gene flow from WWFC, this time to the branch of the common ancestor of the ancient sample and the DSE + IO group, indicating some mixture in both of them. The likelihood of a tree with one edge of migration is higher than one with no edges, while the Standard Error (SE) is lower (ESM Fig. S5c-d). The (ln(likelihood) with 0 migration events = -4.62704, SE = ± 8.4; ln(likelihood) with 1 migration event = 159.731, SE = ± 3.4).

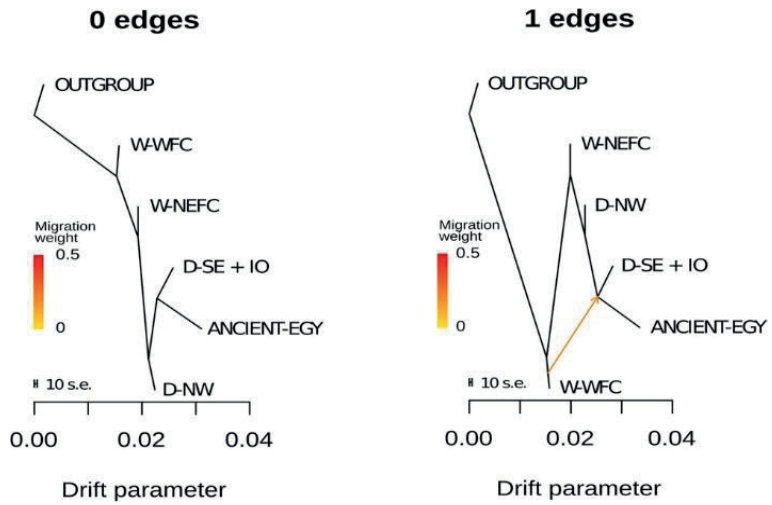


FIGURE 7: TREEMIX MAXIMUM LIKELIHOOD ANALYSIS OF THE ENTIRE DATASET INCLUDING THE INDIAN OCEAN AND THE ANCIENT SAMPLE, ALLOWING FOR 0 (LEFT) AND 1 (RIGHT) MIGRATION EVENTS, SHOWING GENE FLOW BETWEEN WWFC AND THE NODE TO MODERN DSE + IO AND THE ANCIENT SAMPLE (ARROW). THE COLOUR OF THE ARROW IS PROPORTIONAL TO THE AMOUNT OF MIGRATION

Discussion

There are many aspects surrounding the emergence of domestic emmer and its dispersal that even after decades of study remain elusive. While archaeobotanical evidence shows that the earliest domestic emmer appeared in sites in the western Fertile Crescent, the earliest assemblages with only domestic emmer are found in the northern and eastern Fertile Crescent. Genomic data has always supported the view that modern domestic emmer is clearly closer to emmer from the northern and eastern Fertile Crescent. Regarding its dispersal, it is not clear how and when emmer spread east and south. How it arrived into Ethiopia (whether from Egypt or through the Arabian Peninsula) or into India (considering that only free-threshing wheats are found in central Asia) is so far unknown.

Our analysis of a comprehensive collection of modern emmer landraces from the southern and eastern route together with an ancient specimen from Egypt, provides some new insights into these events, but also raises new questions. It is now clear that modern emmer landraces can be clearly differentiated into two groups, at least in our dataset, in which landraces from other regions such as the Mediterranean area are under-represented. These groups can be assigned to two geographical regions, loosely reflecting the proposed dispersal routes. One includes Europe and the Caucasus, and the other Africa and the Arabian Peninsula. Notably, modern landraces from Iran are grouped with the ones from Europe and the Caucasus, which is even more puzzling in light of the fact that modern landraces from India are grouped with those from the Arabian Peninsula and Africa. The similarity between the ancient Egyptian sample and this DSE + IO group from Africa, the Arabian Peninsula and south Asia demonstrates that the genetic make-up of these modern landraces has not changed much, at least over the past 3,000 years. The clustering of the landraces from India and Oman is in agreement with known trade routes connecting southern Asia and Africa (Cuny and Mouton 2009). However, with the current data it is not possible to know whether modern Indian landraces are descendants of emmer which came there

during early dispersal or whether they arrived from the Arabian Peninsula at a later stage.

The differentiation of these landraces from those from Iran poses a perhaps bigger mystery. Indeed, Iran and particularly the surroundings of the Zagros mountains experienced the same change towards the consumption of increasing proportions of domestic emmer during the Middle and Late PPNB, at the expense of other cereals (Riehl et al. 2013), as did other sites in the northern and eastern Fertile Crescent, reflecting similar patterns of early agriculture. Without ancient samples from Europe, the Caucasus or Iran, it is not possible to know if the genetic make-up of the DNW group is ancient or more recent. If this group represents an ancient genetic structure and assuming that the Iranian landraces which have been studied reflect the diversity of the whole country, this would show that domestic emmer soon replaced other wheats in the most eastern parts of the Fertile Crescent, reaching Iran around 9000 BP. It then spread north and west, but without reaching south Asia through Iran. *Triticum aestivum* (free-threshing wheat) would have replaced *T. turgidum* ssp. *dicoccum* (hulled emmer wheat) in this route, causing the dispersal of *T. aestivum* through the inner Asian mountain corridor (X. Zhou et al. 2020), eventually reaching China (Stevens et al. 2016). In parallel, domestic emmer would also have first spread south to Egypt, then to Ethiopia, through the Arabian Peninsula and finally from there to India, where its first evidence dates to 4700–4500 BP (Stevens et al. 2016). Even though the genomic data of our dataset supports this route, it must be noted that this route would have needed a second adaptation, to high altitudes, during the dispersal from Egypt to the high plains of Ethiopia. Introduction of emmer into India would have occurred either through the sea routes, or through Iran but without introduction of the germplasm to this area. Another option is that it was introduced into Iran, and that two different types of emmer were cultivated in this region in the past, but only one persisted into modern times. In a third scenario, modern landraces from the DSE + IO group might exist in Iran but have not been studied. Finally, we cannot exclude the possibility that the genetic make-up of the DNW group, including the modern Iranian samples, is the product of a much recent dispersal into these areas, perhaps

during the last millennium. The study of ancient samples from this region would be invaluable to investigate these different scenarios.

The clear genetic differences between the DNW and DSE groups raise another important question. In this paper we have ruled out that this differentiation is driven by drastic founder effects, since the genetic diversity in the two groups is similar. If this differentiation reflects a past genetic structure, this would suggest that early dispersal of domestic emmer occurred from two different germplasm sources, opening the possibility that the full domestic phenotype was present in markedly different populations before its dispersal. This would be compatible with a diffused and protracted domestication process that may be reflected by the abundance of domestic emmer remains in archaeological sites throughout southwest Asia over thousands of years. Also, analysis of the genes diagnostic for domestication, the *TtBTR* loci, show that all modern domestic landraces carry the same haplotype (the same variants in a sequence of DNA) (Nave et al. 2019), as do the samples studied in this dataset. The immediate wild ancestor of one of the domestic haplotypes is almost ubiquitous among wild emmer, both from the north-eastern and the western Fertile Crescent, so, even if a single origin is the simplest hypothesis, it is nonetheless possible that the haplotype arose independently in different populations. A more careful examination of the genetic composition around this locus could help test this hypothesis. However, if this were the case, the key question that would remain to be answered is how did the same fully domestic phenotype emerge in different wheat germplasms?

Another interesting aspect that we have found in this study is the gene flow from WWFC to the ancestor of the ancient Egyptian sample and the DSE + IO group. The presence of two modern landraces from Turkey within the DSE + IO group raises the possibility that gene flow from wild emmer from the western Fertile Crescent did not occur during early dispersal to the south, but rather before then. These results, together with archaeological evidence, point to an early contribution from the western Fertile Crescent and possibly even from the first domesticated emmer to at least part of the modern gene pool. Also, the clustering of these modern

landraces from Turkey with those from Ethiopia may suggest the existence of different wheat germplasms before their dispersal. However, if this was so, we would expect the Turkish samples to be sister clades to the DSE + IO group on the phylogenetic tree, and not the ancient Egyptian sample. However, the low genomic coverage obtained with the sequencing of the ancient sample could explain its position in the tree. On the other hand, we cannot rule out a recent re-introduction of DSE + IO germplasm to Turkey, perhaps during the last millennium. If gene flow did occur during emmer dispersal to the south, the wild component of the wheat from the western Fertile Crescent would have perhaps provided a basis for its adaptation to hot and dry environments, very different to the climate of the mountain slopes in Turkey (Ozkan et al. 2011). In any case, more ancient and modern samples from these regions would be needed to accurately understand this matter.

Overall, our results reveal that modern landraces from Ethiopia, the Arabian Peninsula and Africa have a genetic structure that is at least 3000 years old, very similar to emmer from Pharaonic Egypt, and most probably similar to the genetic component of the first domestic emmer that spread from southwest Asia to Egypt. While it is not possible to establish whether the modern Indian landraces are descendants of the first emmer that arrived into the region, their similarity with modern landraces from Oman supports a theory by which domestic emmer spread from southwest Asia into Egypt, from there to Ethiopia, crossing the Arabian Peninsula and eventually reaching India, perhaps through sea trade. Our data suggest that the emmer that arrived in Iran was probably from a different germplasm than the one that was introduced into India. While the analysis of the gene responsible for the brittle rachis trait supports a single origin for all domestic emmer landraces, we do find two very different germplasms in modern emmer landraces. Ancient data of the DNW germplasm would be needed to confirm how old is the genetic structure of these landraces.

Finally, some of the results of this study show the need for further analyses. The low genetic variability of the WNEFC groups highlights the importance of adopting a more meaningful sampling strategy that shows whether this characteristic is real, or an artefact if perhaps this group is under-represented in the number of samples from

the area. It remains to be confirmed whether gene flow between wild emmer from the western Fertile Crescent and domestic emmer of the DSE + IO group occurred during the early part of the dispersal of emmer or before then. More information on these processes would allow testing whether domestic emmer from the DSE + IO group could have emerged from a now extinct wild population in the northern and eastern Fertile Crescent and hybridized with wild emmer from the western Fertile Crescent when it dispersed towards the south. Moreover, several studies, according to the archaeological evidence, indicate the origin for the domestic gene pool from an admixture of wild populations, followed by a process of this ancestral group becoming more wild type or feral, that explains the high similarity of the WNEFC population to the domesticated emmer gene pool. Considering our dataset, this would mean that the extant feral population was originally derived from the DNW, not the DSE population. However, our results are insufficient to make any speculation about this, and further analyses are needed to shed light on this possible sequence of events.

Supplementary Information The online version contains supplementary material available at https://d oi.org/1 0 .1007 /s 00334-022-0 0898- 7.

Author contributions LB planned and designed the research. AI conducted the research. AI and LB wrote the manuscript.

Funding A.I. is a FPI fellow (PRE2018-083529). L. B. is a Ramón y Cajal Fellow. (RYC2018-024770-I) both fellowships funded by the Ministerio de Ciencia e Innovación—Agencia Estatal de Investigación/ Fondo Social Europeo. We acknowledge financial support from the Spanish Agencia Estatal de Investigación (Ministry of Science and Innovation-State Research Agency) (AEI), through the "Severo Ochoa Programme for Centres of Excellence in R&D" SEV-2015-0533 and CEX2019-000902-S. This work was also supported by the CERCA programme by the Generalitat de Catalunya.

3.2. Wild emmer contribution in wheat domestication and adaptation to new environments

Iob A., & Botigué, L.

[1].Centre for Research in Agricultural Genomics (CRAG), Cerdanyola del Vallès, 08193 Barcelona, Spain

Manuscript submitted.

Pre-print available in bioRxiv:

DOI: https://doi.org/10.1101/2023.10.25.563983

This chapter is a verbatim reproduction from the last version of the manuscript.

Abstract

Wheat is a staple crop, and its production is heavily threatened by climate change and soil erosion. Breeding programmes commonly rely on the wild progenitor, wild emmer wheat, *Triticum turgidum* subsp. *dicoccoides*, to find markers associated with traits conferring higher resistance to biotic and abiotic stresses, albeit how its genetic diversity contributed into domestication and adaptation to agricultural systems has never been studied. We explore the population structure and the influence of wild emmer populations from both the Northern and Southern Levant during the domestication process. Additionally, we examine their potential contribution to facilitating the adaptation and dispersal of domestic landraces to new environments. We quantify the genomic proportion of wild Southern Levant ancestry in two different domestic germplasms, including landraces from Europe, Caucasus, Africa and Asia. We obtain direct evidence that as much as 26% of the genome has Southern Levant ancestry in the population from Europe, and up to 40% in the population from Africa and Asia. We also estimate the time since admixture of the two wild populations to produce the domestic forms, obtaining two dates, one matching the domestication in Southwest Asia (ca 9500 BP) and the other matching the dispersal towards Africa (ca 6500 BP). We also inquire about the possible adaptive role of wild emmer from the Southern Levant into domestication and dispersal and find an overrepresentation of genes associated with resistance to biotic stress and drought. Overall, our work provides more information on the origins of domestic wheat and highlights the potential of modern domestic landraces of emmer wheat in the study of the genetic basis of resilience. Modeling wheat genome evolution under different demographic scenarios is needed to confirm the observed signals of positive selection and facilitate the use of emmer landraces in future wheat breeding programmes.

Introduction

Over the past decades human food supply has become increasingly dependent on the cultivation of a few elite cultivars from a handful of species, prioritizing yield and caloric content while often overlooking other essential traits. This breeding process has led to a decrease in genetic diversity (Colin K. Khoury et al. 2022), including variation conferring resistance to biotic and abiotic stresses, making modern cultivars heavily dependent on intensive inputs of water, fertilizers and pesticides. Consequently, agriculture is one of the major contributors to ecological degradation at the global scale, encompassing biodiversity loss (Gonthier et al. 2014), soil erosion and climate change. At the same time, climate change threatens the resilience and stability of our food system (Streit Krug et al. 2023) and great concerns arise from the decrease in productivity and increase in vulnerability of agriculture. In the past years global agricultural productivity has decreased (Arora 2019; Ortiz-Bobea et al. 2021), in part due to the lack of sufficient diversity in elite cultivars to adapt and maintain yields amidst climate change (Hufford, Berny Mier Y Teran, and Gepts 2019; Labeyrie et al. 2021; Streit Krug et al. 2023). Furthermore, predictions of greenhouse gas (GHG) emissions under the current food system would prevent the achievement of the targeted 1.5ºC or 2ºC degrees limits, even if all non-food emissions were to rapidly decrease (Clark et al. 2020).

The scientific community has turned to crop wild relatives (CWR) and traditional landraces to face these challenges. They constitute reservoirs of genetic diversity, harboring variation that allows them to adapt to diverse climatic conditions, and to grow with little input. (Cortés and López-Hernández 2021; Marone et al. 2021). To incorporate desirable variants into high-performing cultivars, it is essential to conduct phenotyping of CWR and traditional landraces, while simultaneously preserving and characterizing their genetic diversity. At the same time, understanding genome evolution during domestication and dispersal to new environments is essential to identify genomic regions under positive selection, find polymorphisms influencing traits of agronomic interest and ultimately contribute to the development of new

breeding strategies after functional validation (Venkateswaran, Elangovan, and Sivaraj 2019).

Wheat is one of the most widely cultivated crops on earth (FAO 2021), and it is seriously threatened by climate change. Each degree-Celsius increase in global mean temperature is predicted to reduce global yields by 6.0% (C. Zhao et al. 2017), in part due to the extremely low genetic variability of modern cultivars. Recent investigations have revealed that only a small fraction of the extensive diversity present in its ancestor wild emmer (*Triticum turgidum* subsp. *dicoccoides*) is retained within its domestic tetraploid descendants, and only 1.1% in bread wheat cultivars (Z. Wang et al. 2022). Domesticated emmer wheat, the ancestor of bread wheat, is known to be one of the most ancient cultivated cereals and is characterized by qualities such as high fiber content, resistance to biotic and abiotic stresses and ability to grow in adverse environmental conditions (Lucas et al. 2017; Saleh 2020). Despite the exciting potential of emmer wheat to contribute to wheat improvement (Mohammadi et al. 2021; Zaharieva et al. 2010; Zohary 2013), the evolution from wild to domestic emmer forms and the differentiation of domestic emmer are still largely unexplored topics.

Wild emmer wheat only grows in Southwest Asia (the Fertile Crescent) and is formed by two distinct populations living in different environments: one in the Northern Levant, mountainous, cold and humid, and the other in the Southern Levant, close to the Mediterranean coast with milder and dryer weather (Ozkan et al. 2011). The origins of domestic wheat remain uncertain. Several studies have shown that domestic emmer is genetically closer to wild emmer form the Northern Levant, as evidenced by phylogenetic and genome-wide analyses (Avni et al. 2017; M. C. Luo et al. 2007; Ozkan et al. 2002), but multiple studies have demonstrated that wild emmer from Southern Levant also contributed to the domestic gene pool (Cheng et al. 2019; Iob and Botigué 2022; Pont, Leroy, Seidel, and Tondelli 2019).

The model currently accepted distinguishes between a first phase of intensive exploitation and increasing management of wild or proto-domestic emmer starting in the Southern Levant, followed by dispersal and hybridization in Northern Levant

leading to the emergence of the fully domestic phenotype, centuries later (Arranz-Otaegui et al. 2016; Civáň, Ivaničová, and Brown 2013; H. R. Oliveira et al. 2020; Z. Wang et al. 2022). Fully domestic emmer would then spread out of the Fertile Crescent, adapting to new environments (Avni et al. 2017; Maccaferri et al. 2019; Zaharieva et al. 2010). This model aligns with evidence supporting the contribution of the wild emmer from the Southern Levant population at least one haplotype associated with the non-brittle rachis (Nave et al. 2019). The loss of rachis brittleness is considered the quintessential trait in cereal domestication, as it disrupts the plant's natural seed dispersal mechanism. It is determined by two recessive mutations, one in chromosome 3A (*TtBTR1-A*) and the other in chromosome 3B (*TtBTR1-B*). According to this study, the domestic allele in chromosome 3A could be derived from the Northern Levant population, and the one in chromosome 3B from the Southern Levant one. However, a third scenario in which both alleles originated in Southern Levant could not be ruled out.

Despite these advancements, several key questions are still unanswered, including the timing and composition of wild admixture, as well as the extent of the Southern Levant's contribution to the domesticated emmer gene pool. Additionally, some researchers continue to consider the genetic similarities between domestic emmer and its wild relatives from the Northern Levant as support for the formerly popular hypothesis that domestic emmer descended monophyletically from this wild population (X. Zhao et al. 2023).

Less attention has been given to the process of adaptation during the expansion of crops from their centers of origin (Janzen, Wang, and Hufford 2019). Domestic emmer wheat spread from the Fertile Crescent to the Balkans, Western Europe, the Mediterranean basin, Eastern Africa, and eventually reached India adapting along the way to a wide range of different ecosystems (Maccaferri et al. 2019; Zaharieva et al. 2010). Adaptation can occur *de novo,* on standing variation or, when introduced through gene flow from a wild population, through a process known as *adaptive introgression*. Adaptive introgression has been proposed to have enabled changes in

flowering time in flax during its expansion into Central Europe (Gutaker, Zaidem, et al. 2019) or in fish living at high altitude (Qian et al. 2023).

Previous studies on the population structure of emmer wheat (Iob and Botigué 2022), including wild and domestic specimens, identified two differentiated domestic germplasms, one representing the northwestern route of dispersal (DNW) into the Caucasus and Europe and another reflecting the southeastern route of dispersal (DSE) into Africa and Asia. Furthermore, evidence of extensive gene flow from wild emmer from Southern Levant into the DSE population was detected, explaining the differentiation between the two domestic germplasms. The uneven contribution from the wild Southern Levant population can be explained by two plausible scenarios: genetic drift following a reticulated domestication event, or hybridization during domestic wheat dispersal into Africa.

In this study, we aim to quantify the contribution of the wild emmer population from Southern Levant in the domestic pool, model the time since admixture and determine whether gene flow was linked to positive selection. Using haplotype-based techniques and selection statistics, we aim to gain insights into adaptive traits in wild emmer wheat and landraces, which can lead to the identification of advantageous alleles in front of the challenges of climate change in wheat cultivation. The identification of genomic regions that received one or the other influx from wild Southern Levant population, can aid understanding domestication mechanisms as well as adaptation.

Results

Earlier research on the population structure of wild and domestic emmer wheat revealed that domestic emmer germplasm is generally more closely related to wild emmer samples from the Northern Levant. However, there was evidence of gene flow from the wild emmer population in the Southern Levant to the domestic germplasm from Africa and India (Iob and Botigué 2022; Scott et al. 2019). To further characterize the contribution of the genetically diverse wild emmer from the Southern

Levant to the domestic populations, we analyzed the genetic variability of an emmer collection (Y. Zhou et al. 2020) containing specimens from Europe, the Caucasus and Africa (Figure 1A and S. Table 1).

A.

FIGURE 1: 1A. DISTRIBUTION OF SITES OF THE SAMPLES ANALYSED IN THIS STUDY, COLORED ACCORDING TO THE GENETIC CLUSTERING IDENTIFIED IN THE ANALYSES. WSL: WILD SOUTHERN LEVANT, WNL: WILD NORTHERN LEVANT, DNW: DOMESTIC NORTHWEST ROUTE OF DISPERSAL, DSE: DOMESTIC SOUTHEAST ROUTE OF DISPERSAL. MAP WAS CREATED IN DATABASIN (HTTPS://DATABASIN.ORG/). **1B**: DISCRIMINANT ANALYSIS OF PRINCIPAL COMPONENTS (**DAPC**) RETAINING 5 PC AND 2 DA. THE RELATIONSHIPS BETWEEN SAMPLES ARE THE SAME AS IN IOB & BOTIGUÉ 2022

B.

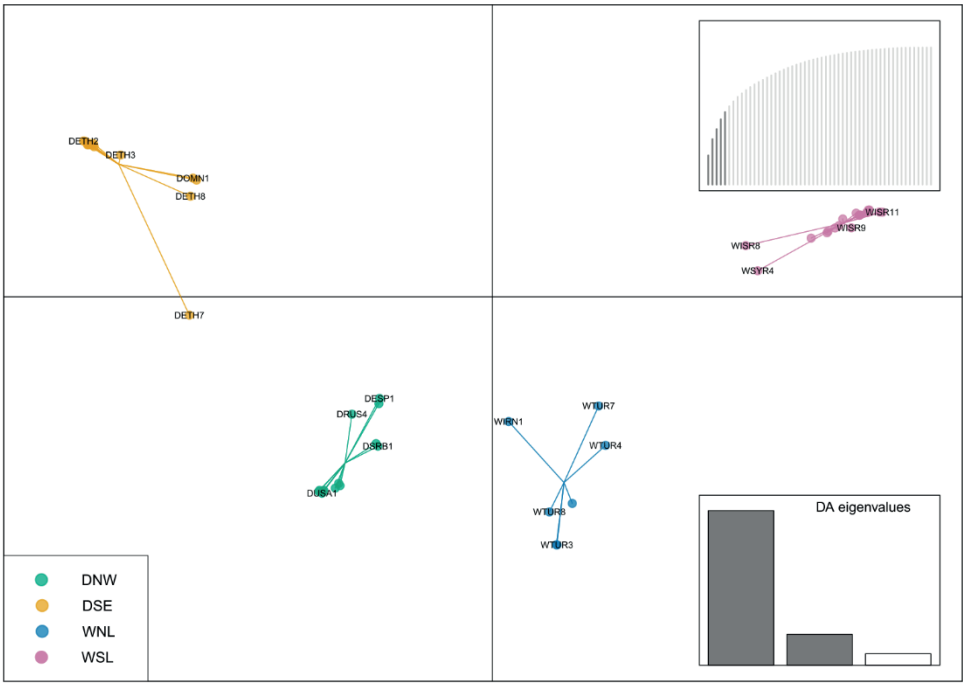


FIGURE 1 (CONTINUES)

We applied a strict filtering approach to remove all potential artifacts from our dataset (see Methods). Specifically, after quality filtering we removed all heterozygote sites to eliminate artificial polymorphisms arising from the misalignment of repetitive regions, which reduced the number of SNPs by almost a half (from 66M to ca 38M, representing the "unmasked dataset"), and we further masked low complexity and repetitive regions based on the reference genome. This led to the retention of ca 10M SNPs representing the "masked dataset". Given that this approach drastically reduced the number of SNPs for analysis, we tested the effect of the filtering by performing a Discriminant Analysis of Principal Components (DAPC), which confirms previously identified patterns. Wild samples from Southern Levant (Israel, Syria and Lebanon) cluster together and we henceforth refer to them as WSL population, while wild samples from Northern Levant (Turkey and Iran) form another cluster that we refer

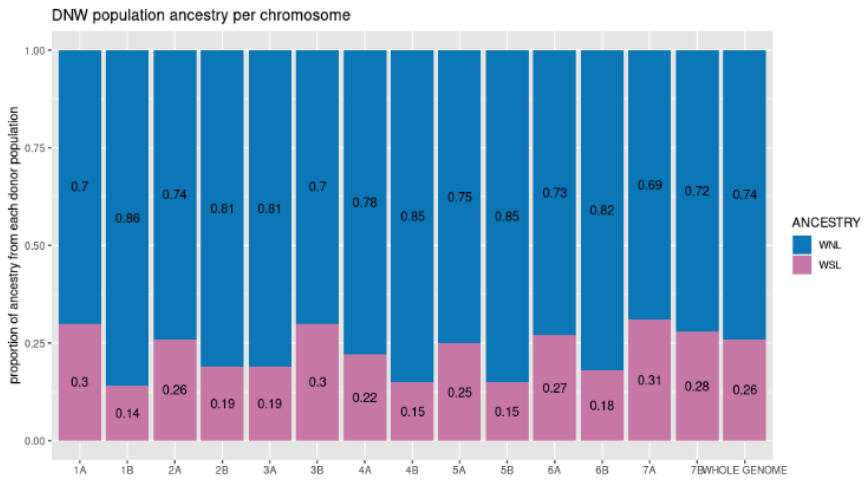
to as WNL. Domestic samples are distributed in two clusters, one containing specimens from Ethiopia and Oman, what we refer to as the DSE population, and another containing samples from Europe, the Caucasus, the Balkans and Iran, forming what we refer to as the DNW population. As previously assessed, the WNL population is closer to the domestic populations than WSL, and closest to the DNW group, while DSE looks more differentiated (Figure 1B). These results, that follow those obtained in Iob and Botigué 2022, show that the removal of heterozygotes and the masking of low complexity regions do not affect the relationships between populations within the dataset, while it removes potential unreliable signals.

Wild ancestry of domestic populations

The use of unsupervised clustering algorithms to investigate the structure of the emmer wheat populations does not allow to model the contribution of each wild population in the domestic germplasm. The high degree of self-pollination in wheat combined with the low recombination rates in pericentromeric regions are translated in low levels of genome diversity. Such reduced diversity is interpreted as little admixture proportions at the genome level by these algorithms, as previously observed (Iob and Botigué 2022). We used SourceFind, a haplotype-based method, to assess the extent of each wild population's contribution to the two domestic populations, DNW and DSE.

Variability in both domestic populations arises mostly from the WNL group, in line with the literature (Avni et al. 2017; Oliveira et al. 2020). Nevertheless, we find that WSL not only contributed to the DSE domestic germplasm, but also to the genetic architecture of the DNW germplasm, obtaining direct genomic evidence of the influence of wild emmer from Southern Levant into the whole domestic pool. The estimated proportion of ancestry coming from WSL amounts to 26% for DNW and reaches as much as 40% of the whole genome contribution for DSE (Figures 2A and 2B).

A.



B.

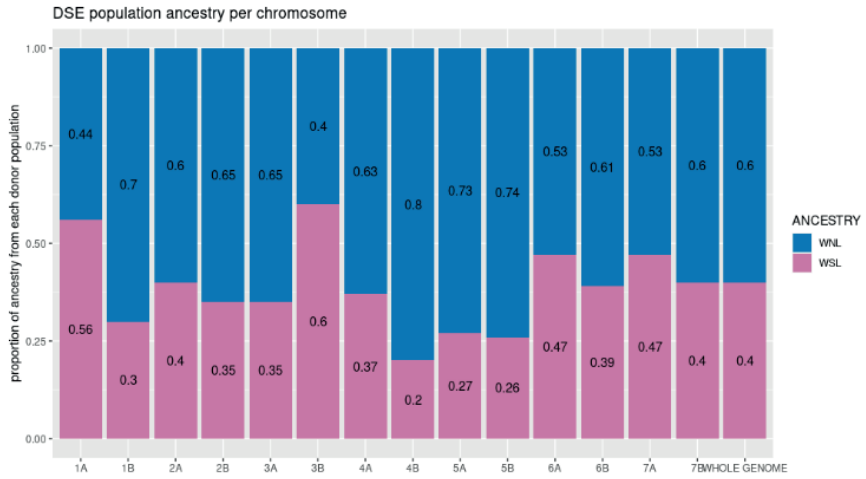


FIGURE 2: WILD ANCESTRY IN DOMESTIC POPULATIONS AS ESTIMATED BY SOURCEFIND. RESULTS ARE REPORTED FOR EACH CHROMOSOME AND FOR THE WHOLE GENOME. **2A**: DNW; **2B**: DSE.

Results at the chromosome level show variable proportions of ancestry between chromosomes, which is expected by the combination of the effects of drift and differences in recombination rate along the chromosome. However, patterns of variation between chromosomes are shared by the two domestic populations. For instance, chromosomes 1B, 4B, 5A and 5B show the lowest levels of WSL ancestry

in both populations, ranging from 14% to 26%, while chromosomes 1A and 3B, on the other hand, have the highest input from WSL in both populations, ranging from 30% to 60%. Interestingly, the copy of the *TtBTR1* gene responsible for the non-brittle rachis located in chromosome 3B has southern Levant ancestry (Nave et al. 2019), and out of the two gene copies is the one displaying the largest selective sweep (Scott et al. 2019), which is in line with the overall high proportion of Southern Levant ancestry in this chromosome. The presence of shared patterns of ancestry between the two domestic populations indicates that the contribution of WSL pre-dates the split between DSE and DNW.

We used the identified ancestry patterns to model the time since admixture between the two wild populations giving rise to the first domestic forms using FastGlobetrotter (Figure 3). When we considered DNW as the target population, we got one admixture event between WNL and WSL, dated 9537 years ago (90% CI 4900 – 13 900 years ago). This estimate is remarkably in line with the first archaeobotanical findings of fully domestic emmer (around 10 000 years ago) (Arranz-Otaegui et al. 2016). The majority contributing source is WNL, with 80% of ancestry, while WSL is the minority contributing source, with 20% ancestry. Interestingly, when we considered DSE as target population, the inferred date since admixture is 6,485 years ago (90% CI: 1700 – 10 505 years ago), with WNL contributing 39% of the ancestry and WSL contributing 61%. This admixture date is compatible with the early dispersal of domestic emmer into the south, reaching Egypt between 7500 and 6500 years ago (Scott et al. 2019; Zaharieva et al. 2010).

These results support the hypothesis that the excess of WSL ancestry in modern DSE landraces comes from a hybridization event during the dispersal of early domestic forms into Africa.

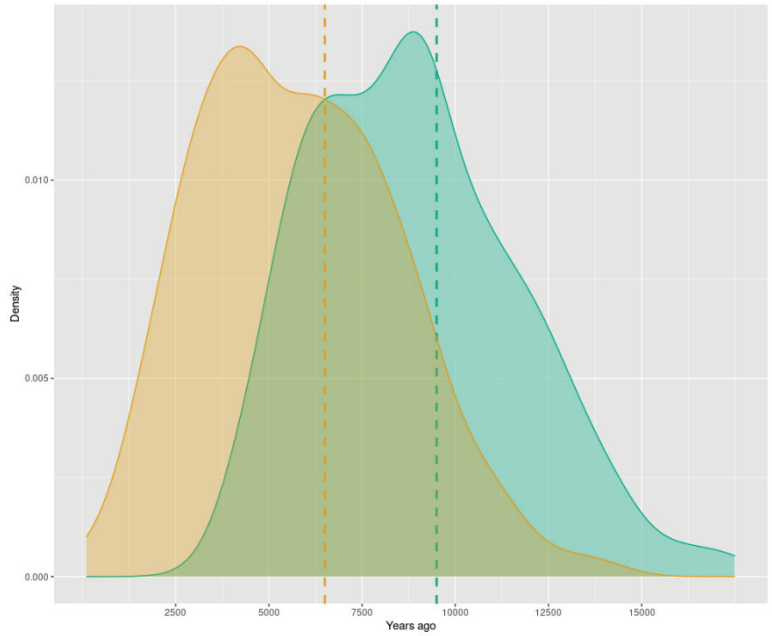


FIGURE 3: FIGURE 3: DENSITY PLOT FOR THE ADMIXTURE DATES ESTIMATES AFTER 500 BOOTSTRAP ITERATIONS OF GLOBETROTTER. THE X-AXIS SHOWS THE DATE SINCE ADMIXTURE (YEARS AGO). THE GREEN CURVE REPRESENTS DNW, THE YELLOW CURVE REPRESENTS DSE.

Once we studied the wild ancestry proportions at the chromosome level, we aimed to study the contribution of wild population into the domestic pool at a finer scale. To do so, we calculated the absolute nucleotide divergence (Dxy) between populations along genomic intervals of 2Mb. As Dxy is not dependent on within-population diversity, it is better suited for the study of inbred populations. Averaging the values to get an estimate at the whole genome level, we obtained: Dxy WSL-DSE = 0.231, SD=0.066; Dxy WSL-DNW = 0.245, SD=0.066; Dxy WNL-DSE = 0.170, SD=0.093; Dxy WNL-DNW=0.150, SD= 0.081, confirming known relationships between the populations. Next, we compared the closeness of each domestic population to the two wild populations by plotting the joint distribution of Dxy scores between each domestic and the two wild populations (Figure 4A and 4B). In line with the haplotype-based results, most of the windows show that the two domestic populations have lower Dxy values (hence lower differentiation) with WNL than with WSL. However, differences can be observed between the two domestic populations.

The DSE has an excess of WSL ancestry compared to DNW (1912 and 1015 out of 9881 windows showing lower Dxy values between DSE-WSL and DNW-WSL, respectively), Figure 4. This represents almost a two-fold increase, the same proportion that SourceFind estimated. On the contrary, DNW has more windows that show a high differentiation with WSL and low differentiation with WNL.

A.

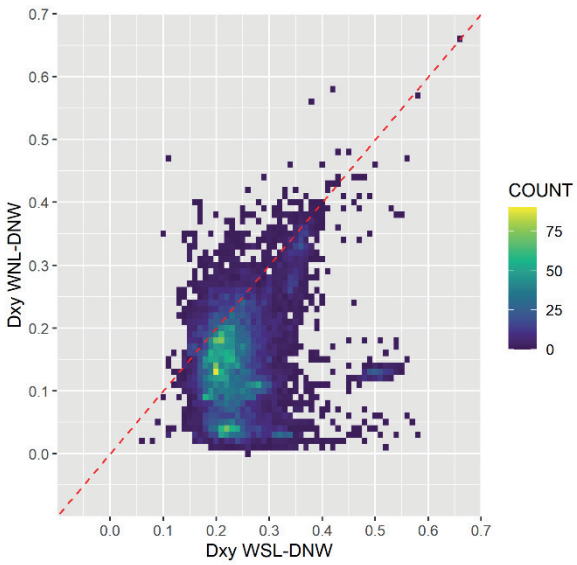


FIGURE 4: DXY HEATMAP. X AXIS SHOWS VALUES OF DXY WSL-DOMESTIC AND THE Y AXIS SHOWS VALUES OF DXY WNL-DOMESTIC. EACH SQUARE REPRESENTS A COUPLE OF VALUES OF DXY FOR 2MB GENOMIC WINDOWS, INDICATING THE RELATIONSHIP OF SUCH WINDOWS TO ONE AND THE OTHER WILD POPULATION. THE COLOR DEPENDS ON THE NUMBER OF OCCURRENCES OF SUCH DXY VALUE COUPLES. THE DIAGONAL RED LINE INDICATES EQUIDISTANCE TO THE TWO WILD POPULATIONS; VALUES BELOW SUCH LINE INDICATE CLOSER RELATEDNESS TO THE WNL POPULATION, WHILE VALUES ABOVE THE DIAGONAL LINE INDICATE CLOSER RELATEDNESS TO THE WSL POPULATION. **4A: DNW; 4B: DSE**, SHOWING AN EXCESS OF WINDOWS ABOVE THE DIAGONAL.

B.

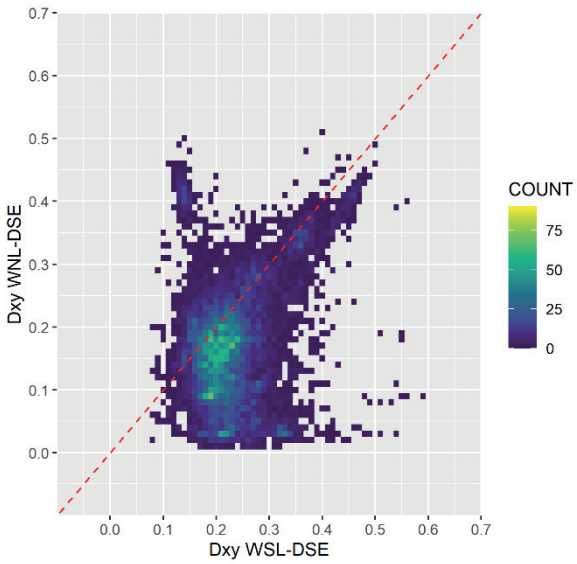


FIGURE 4: (CONTINUES)

Identification of shared WSL contribution to the domestic gene pool

In light of the results obtained, we propose a two-phase contribution of WSL to the domestic pool. An initial phase in which WSL hybridized with WNL (probably in the Northern Levant region in light of archaeobotanical evidence) to generate the domestic wheat, and a second phase in which WSL hybridized with the already domestic wheat spreading into Africa.

The legacy of wild populations from the Southern Levant in modern emmer wheat landraces has only been studied with regards to the evolution of the brittle rachis trait, but it has never been investigated at the whole genome level. We first focused on the influence of WSL in the domestication process by identifying regions of the genome where the two domestics show low differentiation with WSL and are genetically similar between them. For all those windows showing lower Dxy values between domestics and WSL than between domestic and WNL, we selected those that had a Dxy score between the two domestics smaller than 0.05. This pattern of

differentiation reflects the contribution of the WSL into the domestic pool that has remained relatively unchanged in the two domestic germplasms since the domestication process.

This led to the identification of 184 2-Mb, overlapping windows (Table S2), corresponding to a total of 279Mb of Southern Levant origin that remain similar in the two domestic populations. Notably, the windows containing the *Tt-BTR1-B* locus controlling rachis brittleness on chromosome 3B (of putative Southern Levant origin in the domestic forms) were included (96-98 Mb in chromosome 3B). A similar pattern was observed when genetic variability in the domestics was compared with that of WSL. XP-EHH score was in the top 5 and top 1 percentile for DSE and DNW, respectively in the 94-96Mb window, but not in the 96-98Mb. These results reflect the importance of using overlapping windows and support the idea that using a genetic distance statistic such as Dxy it is possible to retrieve the signal of positive selection. On the other hand, the window containing the *TtBTR1-A* locus on chromosome 3A shows lower distance to the WNL population for both domestics (Dxy WSL-domestics 0,231; Dxy WNL-domestics 0,173, Dxy DSE-DNW 0,007). These results support one of the two-step domestication scenarios hypothesized by Nave et al. 2019, namely an independent origin of the domestic *TtBTR1* genes in north and south Levant and subsequent hybridization that would generate the fully domestic phenotype.

Within the 184 overlapping windows, we found 170 SNPs with a predicted high impact on the biological function of the protein products in 85 genes, and over 4500 (4535) with moderate impact on the biological function in 829 genes (Table S3). Three genes carrying high impact variants drive the statistical overrepresentation in *systemic acquired resistance*, (GO:0009627, adjusted P-value 0,0212), a mechanism of induced defense occurring in the distal parts of the plant following localized infection and conferring protection against a broad spectrum of microorganisms (Durrant and Dong 2004). Interestingly, the canonical Ensembl transcript of one of these genes has an NPR3 domain, where the high impact polymorphism is located. It is possible that two transcripts from different genes have been fused, one of them coding for an

NPR3 protein. Only NPR1 proteins have been annotated in this reference genome, so it is likely that this fused transcript contains an NPR3 protein. The other two genes encode for Lipid Transfer Proteins that have also been associated with plant immunity (Finkina et al. 2016).

The genes affected by moderate and high impact variants taken together show enrichment in galactoside 2-alpha-L-fucosyltransferase activity (GO:0008107, adjusted P-value 0,0237), alpha-(1,2)-fucosyltransferase activity (GO:0031127, adjusted P-value 0,00237), and binding (GO:0005488, adjusted P-value 0,0010), molecular functions known to be related to the synthesis of cell wall matrix and hence involved in several biological processes (Reiter 2002; Tryfona et al. 2014). Fucosyl transferases are known to play a role in cell wall biosynthesis in cereals (Hazen et al. 2003) and their activity has been linked to both salt sensitivity (Tryfona et al. 2014) and immunity (L. Zhang et al. 2019).

Introgression from WSL to DSE

We next focused on the unique contribution of WSL to the DSE population, which we hypothesize is the consequence of the second hybridization phase during the early dispersal of the first domestic wheat forms out of the Fertile Crescent. We identified those regions of the emmer wheat genome in which not only DSE is closer to WSL than to WNL but also the differentiation between the two domestic populations is high. From the calculation of Dxy as described above, we selected only those windows in which Dxy WSL-DSE is smaller than Dxy WNL-DSE and Dxy DSE-DNW is bigger than 0.33, based on patterns of Dxy (top 5% values).

The analysis revealed 247 2-Mb overlapping windows, corresponding to 271Mb in total, that show contribution from WSL to DSE only (Table S4). Within the 247 windows, we found 311 SNPs that have high (16 SNPs) or moderate impact on protein products of 186 genes (Table S5). Such genes showed overrepresentation for molecular functions, biological processes and cellular components mainly related to

cell cycle, metabolism, cellular organization, membrane and organelles, as reported in table 1.

TABLE 1: OVERREPRESENTATION TEST RESULTS FOR GENES RELATED TO INTROGRESSION FROM WSL TO DSE.

SOURCE	TERM NAME	TERM ID	ADJUSTED P-VALUE
GO:MF	molecular_function	GO:0003674	0,0021
GO:BP	cell cycle G1/S phase transition	GO:0044843	0,0011
GO:BP	G1/S transition of mitotic cell cycle	GO:0000082	0,0011
GO:BP	metabolic process	GO:0008152	0,0045
GO:BP	biological_process	GO:0008150	0,0082
GO:BP	primary metabolic process	GO:0044238	0,0162
GO:BP	organic substance metabolic process	GO:0071704	0,0172
GO:BP	intracellular transport	GO:0046907	0,0455
GO:BP	establishment of localization in cell	GO:0051649	0,0488
GO:CC	membrane coat	GO:0030117	0,0001
GO:CC	coated membrane	GO:0048475	0,0001
GO:CC	cytoplasm	GO:0005737	0,0010
GO:CC	intracellular anatomical structure	GO:0005622	0,0028
GO:CC	membrane protein complex	GO:0098796	0,0067
GO:CC	cellular_component	GO:0005575	0,0080
GO:CC	cellular anatomical entity	GO:0110165	0,0173
GO:CC	chloroplast	GO:0009507	0,0319
GO:CC	Intracellular membrane-bounded organelle	GO:0043231	0,0334
GO:CC	membrane-bounded organelle	GO:0043227	0,0339
GO:CC	intracellular organelle	GO:0043229	0,0354
GO:CC	organelle	GO:0043226	0,0358
GO:CC	plastid	GO:0009536	0,0396

Forty percent of these windows (n=99) are found in a contiguous stretch in chromosome 4B, between 189 and 378 Mb (Figure 5A). The low recombination rate in this area, 8.6 * 10^{-9} cM/bp, based on Maccaferri 2019, explains the length of the region. Within this region the two domestic populations are quite divergent (Dxy=0.4356, SD=0.0236), and DNW is clearly closer to WNL, as shown by Dxy values (Dxy WNL-DNW=0.1288, Dxy WSL-DNW=0.5027) (Fig. 5B), evidencing that this region has different wild ancestries in the two domestic populations. These results are surprising, since chromosome 4B is among the chromosomes with fewest WSL ancestry in both domestic populations (15% in DNW and 20% in DSE)

according to SourceFind. This slight difference of 5% in the estimated contribution of WSL between the two domestics rules out that the signal in chromosome 4B is related with the second hybridization phase. By examining Dxy values for DNW and DSE with the two wild populations we could corroborate that for most of the chromosome not only were both domestics closer to WNL than WSL, but that Dxy values were low, ranging between 0 and 0.1. Rather than adaptive introgression, these results support an evolutionary constraint to preserve ancestry from WNL during the domestication process and that only the centromeric region in DSE accumulated WSL ancestry. This hypothesis is further supported by the fact that according to SourceFind, only chromosome 4B and chromosome 5A have similar amounts of WSL ancestry in the two domestics, while other chromosomes display between 11 and 20% increase in WSL ancestry in DSE compared to DNW. Notably, chromosome 5A harbors the Q gene, known for its many effects on different phenotypes associated with early crop improvement (Simons et al. 2006).

The WSL ancestry stretch in DSE could have evolved through the effect of drift or selection. Since our goal was to assess the role of wild wheat from the Southern Levant in the domestic pool, we next investigated this region for signals of selection. We calculated Tajima's D scores and found that both domestic populations have negative Tajima's D values, especially DNW (Figs 5A and 5B). On the other hand, DSE shows lower nucleotide diversity (PI=0.0333, SD=0.0036) than DNW (PI=0.1012, SD=0.0109). Both negative Tajima's D and reduced nucleotide diversity are potential indicators of a selective sweep. However, this excess of rare alleles (Tajima's $D < 0$) can also be explained by the combination of inbreeding and low recombination rate of pericentromeric regions in wheat. For this reason, we decided to compute the XP-EHH statistic in DNW vs. DSE for the identification of a selection signal within this region. XP-EHH is a haplotype-based statistic that allows for the identification of recent selective sweeps. XP-EHH implicitly accounts for the recombination rate within the genome and is also less sensitive to population structure, as it detects selection by comparing haplotypes between two populations

(Eydivandi et al. 2021). XP-EHH results (Fig 5C) show top values for DSE within this region, indicating strong evidence of a selective sweep in this population.

A.

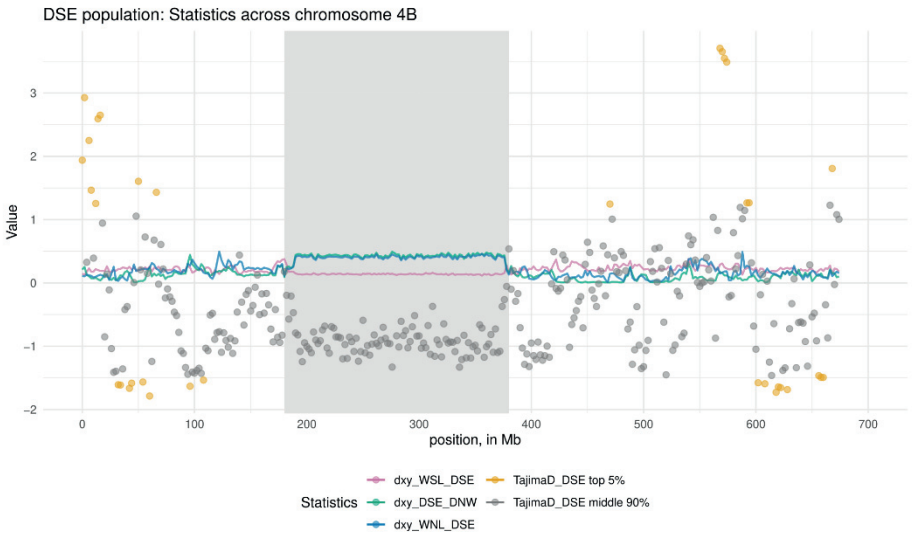
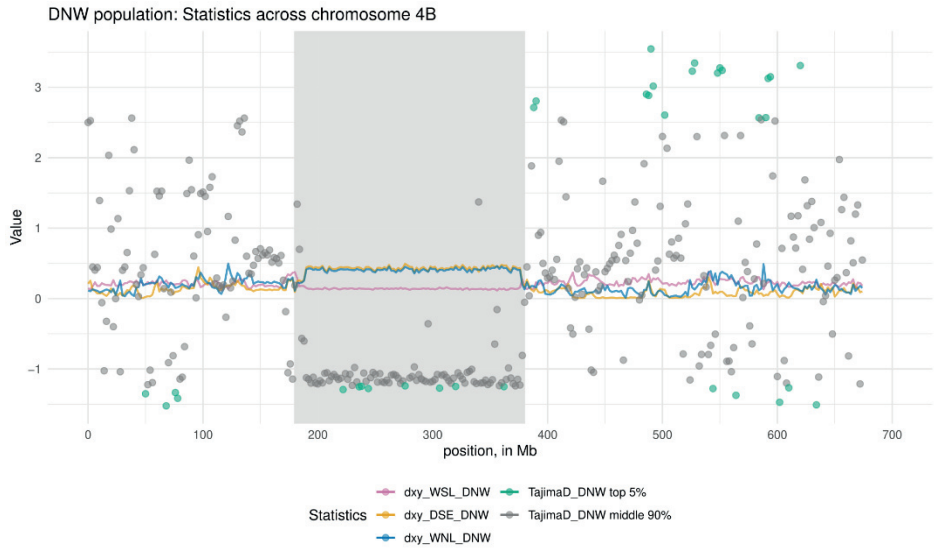


FIGURE 5: STATISTICS ACROSS CHROMOSOME 4B. TAJIMA'S D AND XPEHH ARE CALCULATED IN 2MB NON-OVERLAPPING WINDOWS, WHILE DXY IN 2MB OVERLAPPING WINDOWS, WITH A STEP OF 1MB**. 5A: TAIJMA'S D AND DXY FOR DSE POPULATION**. YELLOW POINTS SHOW TOP 5% TAJIMA'S D VALUES; **5B: TAIJMA'S D AND DXY FOR DNW POPULATION**. GREEN POINTS SHOW THE TOP 5% TAJIMA'S D VALUES; **5C: XP-EHH DSE-DNW**. YELLOW POINTS SHOW TOP 5% VALUES FOR SELECTION IN DSE, GREEN POINTS SHOW TOP 5% VALUES FOR SELECTION IN DNW

B.



C.

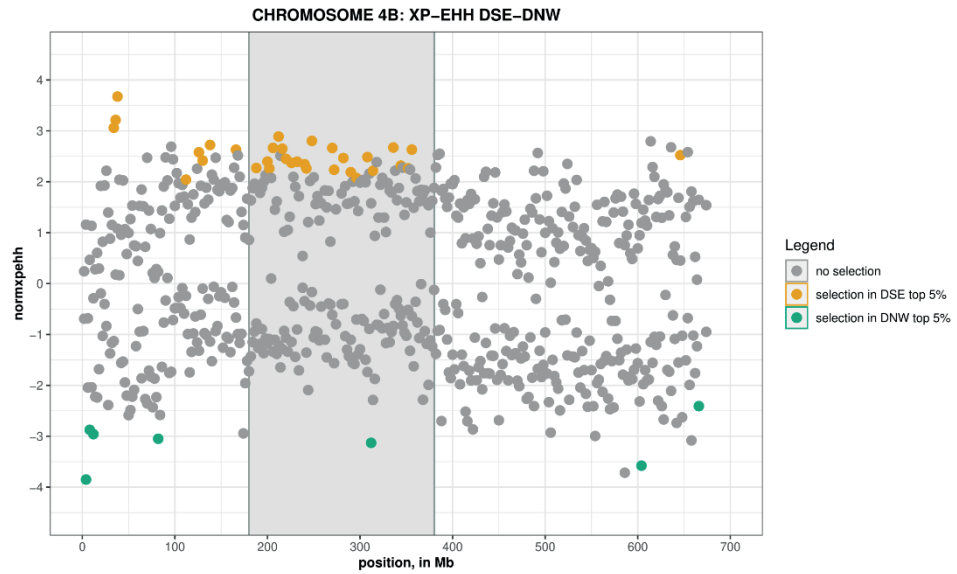


FIGURE 5: (CONTINUES)

We found 71 genes affected by moderate or high impact variants in this region. These genes show enrichment in organic substance metabolic process (GO:077170, adjusted p-value 0,038), chloroplast (GO:0009507, adjusted p-value 0,001), plastid (GO:0009536, adjusted p-value 0,001), intracellular membrane bounded organelle (GO:0043231, adjusted p-value 0,032) and membrane-bounded organelle (GO:0043227, adjusted p-value 0,033). Plant metabolism plays a crucial role in acclimation and survival under stress conditions (Fraire-Velazquez and Emmanuel 2013), and the chloroplast is considered a metabolic center with key role in adaptation to heat stress (Q. L. Wang et al. 2018).

Overall, these results suggest that the WSL population may have played a role in adaptation of domestic wheat to the African milder and dryer climates.

In addition to the region in chromosome 4B, the differentiation-based filtering also led to the identification of a smaller stretch of 2-Mb windows in chromosome 6A, between 177 and 208 Mb. Windows within this region also produce negative Tajima's D scores. Within this region, the XP-EHH values don't reach the top 5% for DSE, but the values are higher than the chromosome average (average within this region =1.48; chromosome average = 1.2).

These results are compatible with a selective process in wheat from DSE in genomic regions with a WSL genetic ancestry. Moreover, these results highlight once again the necessity of applying a combination of different methods when analyzing introgression and selection.

Discussion

Emmer wheat, nowadays considered as a "neglected" crop, is cultivated only in a few areas of the world such as India, Yemen or Ethiopia, where it is consumed in moderate amounts (Zaharieva et al. 2010). Its wild progenitor, wild emmer wheat, has been used as a genetic resource for QTL mapping and potential wheat improvement.

However, the generation of wild and domestic wheat hybrids carry many unfavorable traits, making the task towards wheat improvement slow and arduous (Engels and Thormann 2020). On the contrary, modern landraces of domestic emmer carry many traits of interest for wheat improvement without the burden of undesirable wild traits (Swarup et al. 2021). At the genome level, domestic emmer has been often investigated in the context of genetic variation in durum and bread wheat, while its individual genetic characteristics have received substantially less attention.

Our analysis based on genetic differentiation and polymorphisms with a probable high impact on the biological function reveal that several regions of the genome with a most likely ancestry from Southern Levant may have been selected both during domestication and during the early dispersal of the domestic forms into Africa. Interestingly, our results highlight the potential of emmer wheat landraces in wheat improvement. When focusing on selection during the domestication process, we found a significant overrepresentation in genes involved in systemic acquired resistance (SAR, GO:0009627) in the regions of the genome with shared WSL ancestry in the two domestic populations. Such enrichment holds great scientific interest, as plant resistance to herbivory and pathogens is a primary phenotype found predominantly in wild ancestral species (Carmona, Lajeunesse, and Johnson 2011; Chaudhary 2013; Y. H. Chen, Gols, and Benrey 2015), and typically depleted in modern wheat cultivars. Wild plants, constantly exposed to diverse pathogens, rely on inherent genetic resistance for fitness and survival in natural habitats. However, in cultivated environments, the use of agronomic practices and chemical interventions gradually diminished the need for natural pathogen immunity in cultivated plants (Singh and van der Knaap 2022). These results support the hypothesis that alleles increasing biological resistance coming from wild wheat from Southern Levant would have been selected and conserved in traditional emmer landraces but lost in elite cultivars of modern bread and durum wheat, as reflected by efforts to introduce resistance genes from their CWR (Hajjar and Hodgkin 2007).

Haplotype-based analysis carried out with sourceFind allowed us not only to quantify the amount of wild Southern Levant ancestry in domestic populations but also detect

the post-domestication hybridization event in domestic emmer landraces from Africa and the Arabian Peninsula, previously detected at a lower resolution. Remarkably, genes containing polymorphisms with an estimated high and moderate impact in the biological function were overrepresented in processes that have been associated to drought stress and tolerance in maize by more than one study (Jiang et al. 2014; Xu et al. 2014). Among the overrepresented cellular components, it is worth highlighting organelles and plastids. A recent study on drought resistance in wheat (Lv et al. 2020) suggest that that cellular organizations play a crucial role in drought stress, highlighting the susceptibility of the cytoplasm, peroxisome, and chloroplast to drought stress treatment during early developmental stages in wheat. The authors also point out that drought stress significantly affects various processes, including, among others, the cell cycle. Chloroplast overrepresentation largely stems from the long stretch in chromosome 4B. Incidentally, recent studies have shown that the chloroplast is strongly affected by heat stress (Akter and Rafiqul Islam 2017) and it is known to undergo metabolic reprogramming in response to it (Hu, Ding, and Zhu 2020; Q. L. Wang et al. 2018). Moreover, one of the genes in this region that is involved in all categories showing enrichment, is *TRITD4BV1G106550*, which codes for Photosystem II Psb27 protein, a protein that transiently binds to PSII assembly intermediates before a fully functional PSII is formed (Liu et al. 2011). It has been shown that such protein is involved in response to photodamage of PSII in *Arabidopsis* (H. Chen et al. 2006) and a recent study showed that Psb27 participates in light energy dissipation to allow correct maturation of PSII (Johnson et al. 2022).

Overall, our results strongly suggest that there has been positive selection on regions of the genome of WSL ancestry in domestic wheat spreading towards Africa. This selection was likely linked to light, heat and drought stresses, and therefore adaptation to hotter and dryer climates. Modern emmer from these regions should be further investigated to validate the effect of these polymorphisms on heat and drought resistance.

More studies are needed to confirm that these sites are under positive selection indeed, both at the functional level and by comparing wheat genome evolution under

drift and under adaptation. The combination of long chromosomes and self-pollination increases the effect of drift, but at the same time under certain conditions can facilitate the fixation of positive alleles. Looking for evidence of differential selective pressure in the two domestics, we did find overrepresentation of biological processes like transmembrane transport activity in DNW, and genes associated to 1,4-alpha-glucan branching enzyme activity (GO:0003844) in DSE, known to play a fundamental role in starch biosynthesis (Bahaji et al. 2014; Tetlow and Emes 2014). However, genetic variability between the two domestics is substantially smaller than between the domestics and wild wheat from Southern Levant, the focus of this manuscript, and other methodological approaches are needed to confirm this potential evidence of selection.

Refined insights into wheat domestication and dispersal

With regards to the historical significance, our results not only support a reticulated model of wheat domestication but also quantify the contribution of each wild population in the domestic pool, also at the chromosome level. We find that the proportion of WSL ancestry in the two domestics is correlated for most of the chromosomes, pointing to an initial phase of admixture that pre-dates the split of the domestic gene pool into the two populations. The proportions of each wild ancestry that we find suggest that the two proto-domestic populations hybridized in the Northern Levant, generating the fully domestic phenotype. This evidence is compatible with the presence of fully domestic archaeological assemblages in the NL starting from Middle-Late Pre-Pottery Neolithic B (MLPPB) and previous hypotheses (H. R. Oliveira et al. 2020). This genetic model also reconciles with archaeological models that indicate domestication as a long and patchy process, involving different wild populations and human communities (Asouti and Fuller 2013; Dorian Q. Fuller, Denham, and Allaby 2023; Mithen, Richardson, and Finlayson 2023).

Further evidence of the reticulated domestication is provided with our analyses on genetic differentiation and the *TtBTR* loci, which confirm a dual ancestry in chromosomes 3A and 3B. If we incorporate archaeobotanical evidence and previous results, the most plausible scenario is that the domestic allele would have emerged first in chromosome 3B in the Southern Levant, since it is the region where wild emmer was first managed during Pre-Pottery Neolithic A (PPNA) and the first findings of partially domestic wheat occur. During this period incipient agricultural communities in the Northern Levant were focused on other species, based on the archaeobotanical remains identified (Arranz-Otaegui et al. 2016). The second haplotype (on chromosome 3A) would have emerged later, when emmer cultivation increased in the Northern Levant. Admixture between the two proto-domestics carrying each of the two haplotypes would have led to the appearance of the fully domestic phenotype.

We also used haplotype information from the whole genome to estimate the time since admixture between the two wild populations to model a domestication scenario. The date obtained for DNW, 9500 years ago, is very close to the appearance of fully domestic assemblages in the archaeological record. For DSE, the time since the last major admixture event is around 6500 years ago, matching the Southern dispersal of emmer wheat to Africa (Özkan et al. 2011). Based on these results in combination with the other findings, we hypothesize that the additional 20% of WSL ancestry in DSE compared to DNW reflects a second wave of hybridization between the domestic forms and wild wheat from Southern Levant. Even if FastGlobetrotter indicates a single event of admixture, our results are limited by having only two source populations, which reduces the power of the model to distinguish between one or multiple hybridizations. Furthermore, emmer wheat has long chromosomes that rarely recombine around the pericentromeric regions forming long haplotypes.

In light of this, we propose a model of wheat dispersal towards Africa where hybridization occurred not between the wild populations in Northern and Southern Levant, but between the fully domestic forms spreading southwards from the Northern Levant and either wild populations from Southern Levant or, most likely,

proto-domestic forms in the area. SourceFind and Dxy results support that the hybridization occurred predominantly on the Northern Levant genetic background, as reflected by the higher proportion of WNL-derived ancestry in both populations. A schematic representation of the main events leading to the appearance and diversification of the domestic landraces can be found in Figure 6.

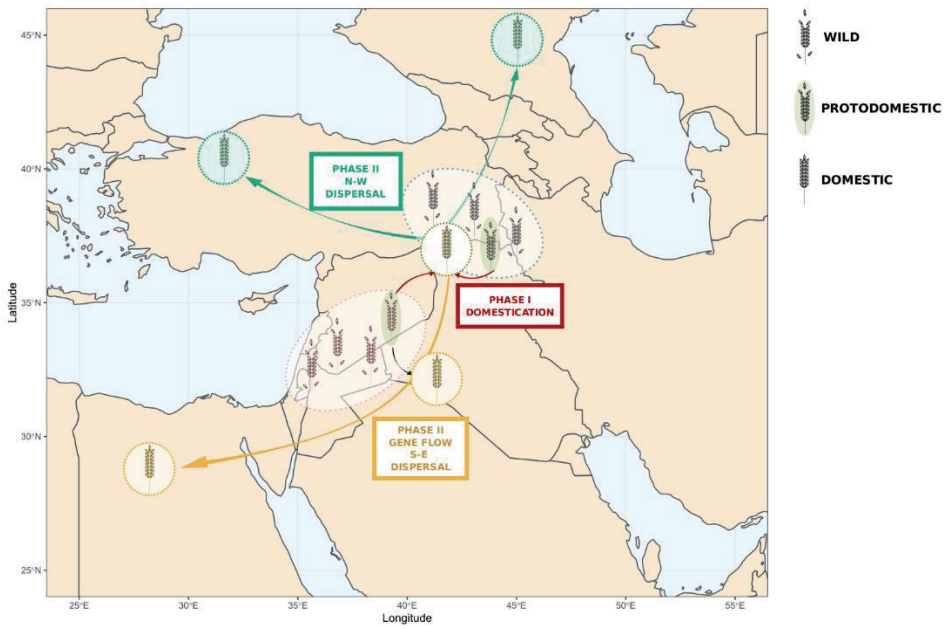


FIGURE 6: SCHEMATIC REPRESENTATION OF THE EVENTS LEADING TO THE APPEARANCE AND DIVERSIFICATION OF DOMESTIC EMMER. WILD, PROTODOMESTIC AND DOMESTIC STATUS ARE REPRESENTED THROUGH VARYING DEGREES OF RACHIS BRITTLENESS. THE PINK AND BLUE DOTTED CIRCLES REPRESENT THE WILD AND PROTODOMESTIC SOUTHERN AND NORTHERN LEVANT POPULATIONS, RESPECTIVELY. THE GREEN AND YELLOW CIRCLES REPRESENT THE DOMESTIC NW AND SE POPULATIONS, RESPECTIVELY. IN THE FIRST PHASE, PROTODOMESTIC INDIVIDUALS FROM THE SL AND THE NL HYBRIDIZE (RED ARROWS) IN THE NORTH OF SW-ASIA (OR NORTHERN LEVANT), LEADING TO THE APPEARANCE OF THE DOMESTIC PHENOTYPE. IN A SECOND PHASE, DOMESTIC EMMER THAT DISPERSED TO THE NORTH AND WEST OF SW-ASIA DIFFERENTIATED INTO CONTEMPORARY DNW POPULATION (GREEN ARROWS). DOMESTIC EMMER THAT DISPERSED TO THE SOUTH HYBRIDIZED WITH WILD OR PROTODOMESTIC EMMER FROM THE SL (BLACK ARROW), GIVING RAISE TO THE CONTEMPORARY DSE POPULATION, THAT REACHED AFRICA AND INDIA (YELLOW ARROW)

Conclusions

Our genomic analysis of traditional emmer landraces and their wild relatives has unveiled new details of the domestication process. More importantly, our combined survey of ancestry and of genomic regions that show evidence of positive selection has revealed several aspects to take into consideration. First, studying different germplasms is crucial to have a comprehensive catalogue of genetic variations. This is true not only for traditional landraces, in this case with the inclusion of accessions from Ethiopia and Oman, but also Crop Wild Relatives. Generating a wheat pangenome would allow to better characterize genetic variability and include structural variation, giving another level of resolution of ancestral wheat genome evolution. Second, the lack of a good, parametrized demographic model for wheat does not allow us to distinguish between the effect of drift from that of adaptation. Despite this, our conservative approach has unveiled many regions that are candidates for positive selection and that have been previously associated with resistance to biotic and abiotic stresses. Efforts to understand wheat genome evolution under domestication and dispersal are essential to increase our ability to distinguish the genomic footprints of adaptation. Finally, our results show that not only CWR are a desirable tool for breeding improvement. Our understanding of how WSL contributed to the adaptation of domestic wheat suggests that these domestic landraces can serve as valuable resources for breeding, as they already possess a multitude of desirable traits.

Materials and methods

In this study we analyze 55 emmer wheat samples, representative of two wild and two domestic populations: 7 are from Wild Northern Levant (WNL), 20 are from Wild Southern Levant (WSL), 15 are from the domestic southeastern route of dispersal (DSE) and 13 from the northwestern route of dispersal (DNW). Passport information is available in Supplementary Table 1 (Table S1). Samples were processed as in Iob

and Botigué (2022), with further filtering applied. Briefly, reads were aligned to the durum reference genome (*Triticum turgidum* subsp. *durum,* Maccaferri et al., 2019) and variants were called in the whole dataset using GATK v. 4.1.6 (Van der Auwera, GA O'Connor 2020). Hard filters were applied after genotype calling as in (Y. Zhou et al. 2020), and 66M SNPs were kept. Emmer wheat is a highly homogeneous and repetitive genome, and as such is prone to read misalignment that translates into inflated (false) heterozygosity (Bukowski et al. 2018; H. Li and Wren 2014). To circumvent this problem, we applied a very conservative approach and removed all sites that showed at least one heterozygote genotype. We refer to the resulting dataset as "unmasked dataset", comprising ca 38 million (38 099 555) homozygote SNPs.

For all analyses involving the calculation of genetic differentiation and selection statistics, we masked SNPs located in low complexity and repetitive regions based on the masked version of the reference genome. To do so, we utilized the "generate_masked_ranges.py" script (Cook, D.E., GitHub), which translated the coordinates of masked regions into a bed file. Using this bed file, we excluded SNPs within these regions from the VCF file and ensured no missing information using the vcftools v.0.1.16 (Danecek et al. 2021) options "-exclude-bed" and "-max-missing 1." This filtered dataset, referred to as the "masked dataset", comprises over 10 million (10,398,941) SNPs.

Additional filters specific to each analysis are discussed in their respective methodological sections.

To test the goodness of our strict filtering approach, we performed a DAPC (Discriminant Analysis of Principal Components) on the masked dataset. First, we filtered it for linkage disequilibrium, using plink v. 1.9 (Purcell et al. 2007), allowing a maximum r^2 value of 0.1 calculated in 50 kb windows with a step size of 10 kb, reducing the dataset to ca 750 000 (748 812) variants. Then, we used the Adegenet package (R 4.1.0, R Core Team, 2021) to perform the DAPC, retaining 5 principal components (PC, based on the result of the xvalDapc function) and 2 discriminant analysis eigenvalues (DA).

Wild Ancestry of domestic populations

In this study, we employed Chromopainter v.2 (Daniel John Lawson et al. 2012), SourceFind v.2 (Chacón-Duque et al. 2018) and FastGlobetrotter (fastGT) (Wangkumhang, Greenfield, and Hellenthal 2022), to estimate ancestry proportions and admixture times for a specific target population using source populations. We used this strategy to model the admixture process of the two domestic populations (DSE and DNW) separately, using the two wild populations (WSL and WNL) as sources. We followed the protocol described in (Hellenthal et al. 2014), and we utilized perl 5.26.0 and R 4.1.0 for its implementation. The unmasked dataset was phased and imputed using ShapeIT v.2 (Delaneau et al. 2014), applying the recombination rate of different chromosomic regions as published in Maccaferri et al., (2019). To avoid introduction of biases due to lack of heterozygote sites, we "haploydized" the dataset by removing one allele for each genotype, and ran all subsequent steps accordingly.

We first ran ChromoPainter to infer the proportion and number of haplotypes shared between individuals, using all samples as possible recipients but only wild samples as donors.

We performed an initial run to estimate switch rate parameters (-n) and global mutation parameters (-M) using ten iterations of expectation maximization per chromosome, and then we averaged the resulting values across all chromosomes, weighting by the number of SNPs per chromosome. The resulting fixed values were used to run ChromoPainter for all chromosomes and the final ancestry matrices (i.e., *.chuncklengths.out and *.chunckcounts.out files) were summed across chromosomes. We then ran ChromoPainter a second time, painting each domestic target using wild sources as donors only (disallowing "self-copying" from other members of the same population), to generate painting sample files (i.e. *samples.out files). The resulting copy vector files from the first ChromoPainter run, and the

painting sample files from the second ChromoPainter run were used as input files for both SourceFind and FastGlobetrotter. We ran SourceFind using default parameters to infer the admixing proportion of WNL and WSL for both DSE and DNW.

In the fastGT runs (one for each domestic population), the null.ind parameter was set to 1, as recommended, to account for decay in linkage disequilibrium that may not be due to authentic admixture signals. We set the haploid option with "haploid.ind 1", changed the curve range from 30 to 15 with "cr 15" to account for the slow decay of linkage in the emmer wheat genome, and the generation time to 1 year. We performed 500 bootstrap replicates and calculated the 90% confidence interval from the distributions of the replicates. FastGT assumes an outcrossing rate of 100%, but emmer wheat is a mostly self-pollinating species. For this reason, we adjusted the admixture times as follows: considering that emmer wheat has a selfing rate of around 99% (Dvořák 2001; Golenberg 1988), we expect an outcrossing event with a probability of 1%, or, in other words, 1 every 100 generations. Hence, we divided the estimate times since admixture by the outcrossing rate ($T_{real} = T_{inferred} /(1/100)$).

Genetic differentiation

We calculated Dxy (absolute nucleotide divergence) to measure genetic distance between populations. We opted for Dxy over Fst because Dxy is unaffected by within-population diversity (Henderson and Brelsford 2020). In self-pollinating species, diversity is generally lower than in outcrossing ones, and domestic populations tend to have reduced diversity compared to their wild relatives. These characteristics are crucial when comparing diversity levels and genetic distances between populations. While Fst relies on within-population genetic diversity and can yield high values when one population has low diversity in a specific genomic region, Dxy provides an absolute measure of genetic divergence between populations, making it a better choice for self-pollinating species in domestication studies. For the calculation of Dxy we used the masked dataset and a window size of 2 Mb and a step of 1Mb and applied the scripts from (Martin, S., GitHub): we converted vcf to "geno" format using parseVCF.py and then calculated Dxy using popgenWindows.py

considering only windows with at least 100 good sites (options -w 2000000, -s 1000000, -m 100). We rounded the values of Dxy to two digits and we plotted the number of windows pairs with the same Dxy couples of values using ggplot2 (R 4.1.0).

Positive selection

We applied two statistics to the masked dataset to identify regions of the genome with patterns of genetic variability compatible with positive selection. Within each population, we used Tajima's D (Tajima 1989) based on nucleotide diversity and number of segregating sites. Tajima's D was calculated in 2-Mb non-overlapping windows using vcftools -TajimaD 2000000. We also used XP-EHH (cross-population extended haplotype homozygosity), a haplotype-based method, which allows the identification of regions under selection by comparing two different populations. XP-EHH was calculated using Selscan v.2 (Szpiech 2022; Szpiech and Hernandez 2014) --xpehh with options --trunc-ok to include calculations at the boundaries of chromosomes, --max-extend 0 for no distance restriction in the calculation of EHH, and default –cutoff 0.5 for EHH decay stopping condition. Then, we searched for 2Mb windows with extreme values using Selscan function --norm with parameters --crit-percent 0.05 --bp-win --winsize 2000000. Per-chromosome distribution of Tajima's D and XP-EHH(DSE-DNW) values can be found in Figures S1 and S2.

Investigating the biological effect of the regions putatively under selection

Windows of interest based on patterns of genetic differentiation were further studied to focus on changes in the biological function and investigate their possible underlying biological cause. Vcftools –bed command was used to extract the windows of interest. The impact on the biological function of polymorphisms falling in the coding regions within these windows was estimated using Ensembl's Variant Effect Predictor (VEP). Overrepresentation of certain biological pathways was then tested by performing a Gene Ontology enrichment analysis. Genes containing High and Moderate impact variants were tested for statistical overrepresentation in g:Profiler (Raudvere et al.

2019) using all known genes annotated in the Durum Reference genome as background and a significance g:SCS threshold of 0.05 (default).

Acknowledgments

A.I. has been supported by an FPI fellowship (PRE2018-083529). L. B. is a Ramón y Cajal Fellow. (RYC2018-024770-I) both fellowships funded by the Ministerio de Ciencia e Innovación—Agencia Estatal de Investigación/Fondo Social Europeo. We acknowledge financial support from the Spanish Agencia Estatal de Investigación (Ministry of Science and Innovation-State Research Agency) (AEI), through the "Severo Ochoa Programme for Centres of Excellence in R&D" SEV-2015-0533 and CEX2019-000902-S. This work was also supported by the CERCA programme by the Generalitat de Catalunya. The Authors thank Cristóbal Uauy for his valuable comments on the manuscript.

Chapter 4

DISCUSSION

The study of domestication encompasses a spectrum of disciplines across social and life sciences, including genetics, archaeology, anthropology, evolutionary biology, ecology and agricultural sciences (Larson et al. 2014). In the past decades, our understanding of the domestication process has shifted from a model of rapid, localized change driven by human action, to a model of slow, dispersed change involving complex interactions between humans, plants, animals, and the environment (Brown et al. 2009; Purugganan 2019). Technical advancements in both archaeology and genetics have enabled the accumulation of extensive datasets and evidence, hitherto inaccessible, fostering a comprehensive understanding of the domestication process. Beyond the historical relevance, the study of domestication is a necessary step to understand the transition from wild to domesticated plants, identify genomic regions under selection and contribute to the improvement of agriculture and the well-being of human populations.

This thesis represents a comprehensive population genomic analysis of wild and domestic emmer wheat germplasm from different geographic origins. Taken together, the results obtained bring new knowledge into the domestication, dispersal and adaptation of emmer wheat. The study encompasses four main aspects: the population structure of emmer wheat, the contribution of wild populations to the domestic genomes, the dispersal of domestic landraces and the role of wild ancestry in adaptation to new environments. Special focus is given to the role of the wild emmer population from the Southern Levant and its contribution to the domestic populations.

To address the research questions outlined in this thesis, I employed a diverse array of techniques. Initially, clustering methods and phylogenetic analyses were utilized to reconstruct the population structure of the dataset. These methods primarily examine genomic variation in the form of independent variables (unlinked SNPs) and allow to get an overview of the relationships between samples, a necessary step to define populations and to unveil the relationships between them. Additionally, I applied techniques for detecting gene flow between populations, facilitating the identification of inter-population interactions. In a subsequent phase, I delved into genomic

variability using haplotypes, hence considering the linkage between individual SNPs. This approach was essential for discerning ancestry proportions and detecting signals of selection. Furthermore, the integration of contemporary data with evidence from an ancient sample was undertaken. This integration was fundamental to confirm the antiquity of genetic signatures found in modern landraces, and to validate potential dispersal routes outside the domestication area.

The findings will be discussed in the context of existing knowledge, integrating evidence from various methods and disciplines.

4.1. Exploring emmer wheat diversity

The genetic relationships between wild and domestic wheat populations have been object of study for several years now, but most of the attention was focused on the relationships between emmer and its economically important domestic descendants (durum and bread wheat, e.g., Pont et al., 2019; Wang et al., 2014; Zhao et al., 2023; Zhou et al., 2020). Previous research has explored emmer wheat populations; however, comprehensive studies utilizing whole genome sequences are scarce. Therefore, this study aims to fill this gap by specifically focusing on the relationships between emmer wheat populations, employing whole genome sequences.

The initial step in population genetic analysis involves identifying genetic clusters or distinct populations within the dataset and determining the underlying factors. This essential step was undertaken in the first study (paper 1, section 3.1) presented in this thesis. The results provide an overview of the relationships among wild and domestic emmer populations, confirming evidence that was obtained using other types of data, such as restriction fragment length polymorphisms RFLP (M. C. Luo et al. 2007; Ozkan et al. 2002), and exome capture (Avni et al. 2017), among others.

Employing DAPC and various phylogenetic approaches based on Maximum Likelihood and genetic distance, I identified a close relationship between the Wild

Northern Levant (WNL) and domestic emmer populations, while the Wild Southern Levant (WSL) population appears more distantly related (paper 1, figures 2A, 2C). Domestic samples are clearly divided into two groups reflecting dispersal outside the center of origin (North-West, DNW and South-East, DSE).

As described in the introduction, this closer relationship between wild emmer from the Northern Levant and domestic emmer populations have been considered as evidence for the monophyletic Northern Levant origin of the domestic gene pool in the past (e.g., Luo et al., 2007; Ozkan et al., 2002, 2005; Salamini et al., 2004). Such conclusion was considered particularly reliable at the end of the 90s and at the beginning of the 2000s, when the model of domestication as a fast and geographically restricted phenomenon was still accepted (Hillman and Davies 1992; Innan and Kim 2004).

While these methods offer valuable insights into population relationships, the study of crop domestication requires a deeper analysis of the genetic diversity, to allow the dissection of local ancestries along the genome, going beyond a tree-like model of evolution. Indeed, classical phylogenetic methods, while valuable, struggle to reflect the complex dynamics of domestication, which likely involve gene flow between plant lineages. The current models of domestication indeed describe it as a slow and protracted phenomenon, characterized by intricate interactions between plant lineages and human groups (Allaby et al. 2022). Furthermore, self-pollination in emmer wheat leads to significant genetic drift, resulting in a strong population structure that complicates the identification of underlying ancestries. For example, the ADMIXTURE analysis (paper 1, figure 2B) did not detect any genetic contribution from the WSL to the DSE population, a finding supported by various other methods employed in this study (paper 1, figures 4, 5; paper 2, figures 2B, 4B).

Considering these limitations, in the second study presented in this thesis I employ various methods to specifically investigate the presence of local wild ancestries within the genomes of domestic populations.

4.2. Unraveling the ancestry of domestic populations

The current state of the art in wheat domestication studies support a reticulated origin of the domestic gene pool, which is derived from the admixture of different wild populations. However, while the contribution of the WSL population to the domestic gene pool has been identified by several studies (e.g., Cheng et al. 2019; He et al. 2019; Pont, Leroy, Seidel, Tondelli, et al. 2019), it has never been quantified so far, and many questions on the emergence of the first domestics remain open, giving rise to multiple plausible models (Civáň et al. 2013; Nave et al. 2019; H. R. Oliveira et al. 2020).

For this reason, in the second paper presented in this thesis, I examined and quantified the wild ancestry within domestic populations, by applying both haplotype-based and distance-based methods. Transitioning from the analysis of unlinked SNPs to the analysis of haplotypes was crucial in this context. Haplotypes offer information on the linkage between genetic variants, allowing for the dissection of ancestry patterns along the genome.

The analysis led to the identification of WSL ancestry in both domestic populations. For the first time, the contribution of this wild population to the domestic gene pool is measured, with a proxy of the ancestry proportion given by sourceFind results (paper 2, section 3.2, figures 2A-B). Both domestic populations share most of the genome with the WNL population, in accordance to the overall genomic relationships between populations identified using unlinked SNPs and clustering techniques (paper 1, figures 2A, 2C). Nevertheless, as much as 26% of the DNW genome and up to 40% of the DSE genome has WSL ancestry. Such pattern is further confirmed by the genetic distance analysis, in which most 2-Mb windows exhibit closer relation to the WNL population, but a subset of windows, more numerous in DSE than in DNW, is closer to WSL (paper 2, figures 4A-B). Despite varying levels of WSL ancestry between the two domestic populations, shared patterns emerged. Specifically, chromosomes with lower (1B, 4B, 5A, and 5B) and higher (1A, 3B, 7A) amounts of WSL ancestry are consistent across both populations. In a scenario of domestication

from the NL population and of post-domestication gene flow from the SL to the domestic populations, one would expect different WSL-ancestry patterns along the chromosomes of different populations. My results, on the other hand, indicate that the shared amount of WSL ancestry in DNW and DSE must pre-date the split of the domestic landraces, and confirm therefore the reticulated origin of the domestic emmer gene pool, providing new details on the extent of the WSL contribution to domestication.

The reticulated origin of the domestic gene pool is reflected by the ancestry of the non-brittle rachis loci. By analyzing the genetic distances between wild and domestic populations, I have observed that while the window containing the *TtBTR1-B* gene is more closely related to the WSL population (Dxy WSL-domestics 0,183; Dxy WNL-domestics 0,250; Dxy DSE-DNW 0,006; Dxy WSL-WNL 0,338), the window surrounding the *TtBTR1-A* copy of the gene is substantially more closely related to the WNL population (Dxy WSL-domestics 0,231; Dxy WNL-domestics 0,173, Dxy DSE-DNW 0,007; Dxy WSL-WNL 0,264). This result proves the different wild origin of the two domestic haplotypes, and supports one of the scenarios proposed by Nave et al., (2019), in which the two haplotypes emerged in different protodomestic populations that hybridized, leading to the fully domestic phenotype.

These findings on the other hand challenge the model proposed by Civan et al., 2013, which suggested both wild emmer from the Northern Levant and domestic populations descended from protodomestic forms coming from the Southern Levant. Instead, this research highlights a distinct narrative (paper 2, figure 6): an initial phase of wild emmer management and cultivation occurred in the Southern Levant, leading to the emergence of protodomestic forms (*TtBTR1-B* mutants). Subsequently, these forms were introduced into the Northern Levant and hybridized with protodomestics derived from the WNL population *(TtBTR1-A* mutants). This aligns with the evidence from the archaeological record, indicating that the exploitation and cultivation of emmer wheat started in the Southern Levant before than in the Northern Levant. Indeed, archaeological evidence shows that wild cereals were mainly exploited in the Southern-Central Levant and the Euphrates during PPNA (11 700–10 700 cal BP),

while in the Northern Levant, legumes and other wild plants were preferentially consumed. During Early Pre-Pottery Neolithic B (EPPNB, 10 700–10 200 cal BP), wild cereals continued to be exploited in the Southern Levant with isolated evidence of cultivation (domestic-type cereal chaff), while in the Northern Levant subsistence was still based on other plants. It is only during the Middle- Late Pre-Pottery Neolithic B (M/L PPNB, 10 200 – 8 500 cal BP), that domesticated-type cereal species are found across southwest Asia, especially in Turkey, Anatolia and the Zagros mountains in Iran, and start outnumbering other plant species (Arranz-Otaegui et al. 2016; Arranz-Otaegui and Roe 2023). The appearance of domestic emmer in the Northern Levant is remarkably reflected by the estimated time of wild admixture I obtained for the DNW population, approximately 9500 years ago (paper 2, figure 3).

Overall, these findings not only validate the model of domestication proposed by Oliveira et al., 2020, but also provide valuable new insights into the dynamics of emmer domestication. The results in this thesis go beyond mere confirmation by quantifying the contribution of the WSL population to the domestic gene pool. Furthermore, they verify the reticulated origin of the *Tt-BTR* haplotypes and provide essential estimates regarding the time since admixture between the wild (protodomestic) lineages.

The reticulated origin of crops and domestic animals has been proposed for other species, such as barley (Pankin et al. 2018), goat (Daly et al. 2018) and rice (Choi et al. 2017). These discoveries give support to the emerging model of domestication as a landscape process (Allaby et al. 2022), figure 4.1. Such framework pictures domestication as the result of slow exploitation and management processes that involved large populations over wide areas with pervasive cultural connections, pre-dating the onset of cultivation practices. In other words, the process that led to domestication was driven by weak selection pressures for thousands of years, that allowed for the accumulation and interchange of domestication alleles, ultimately resulting in the establishment of the domestic phenotypes. In this scenario, human exchange between different groups of hunter-gatherers and incipient agricultural communities would have played a fundamental (yet unconscious) role.

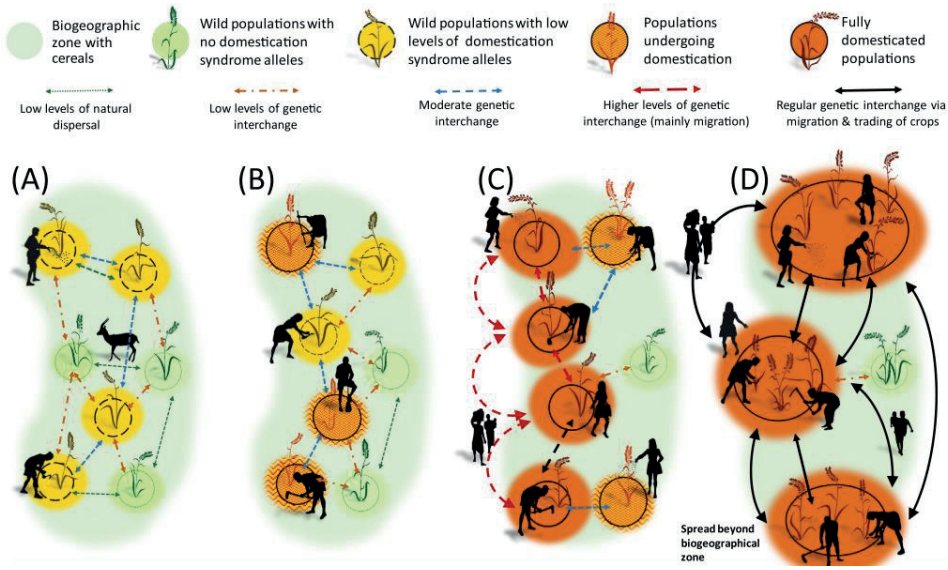


FIGURE 4.1: DOMESTICATION AS A LANDSCAPE PROCESS. INTERACTION BETWEEN THE CHANGING DISTRIBUTION OF DOMESTICATION SYNDROME ALLELES WITHIN A CEREAL POPULATION (A) UNDER LOW-LEVEL EXPLOITATION AND MANAGEMENT, (B) AT BEGINNINGS OF DOMESTICATION (PRE-DOMESTICATION CULTIVATION), (C) AT THE END OF DOMESTICATION (PRE-DOMESTICATION CULTIVATION), AND (D) IN FULLY DOMESTICATED POPULATIONS. THE ARROWS INDICATE GENETIC TRANSFER BETWEEN CEREAL POPULATIONS, IN WHICH INCREASING TRANSFER OF DOMESTICATION SYNDROME ALLELES BETWEEN POPULATIONS THROUGH TIME TAKES PLACE PREDOMINATELY THROUGH HUMAN MOVEMENT AND EXCHANGE (ALLABY ET AL. 2022).

This scenario aligns with the emmer domestication model outlined in this study. The consumption and exploitation of wild emmer wheat initiated in the Southern Levant during the Epipaleolithic and PPNA periods, evolving into cultivation during the EPPNB. Subsequently, emmer cultivation spread from the South to the North of the Levant, facilitated by sustained human interaction and mobility over an extended temporal span. Ultimately, the fully domestic phenotype emerged in the M/LPPNB period, thanks to hybridization of different protodomestic populations. This sequence of events not only underscores the gradual process of wheat domestication but also highlights the crucial role of human exchange and mobility in the diffusion of agricultural practices and in the appearance of fully domestic wheat.

Finally, with regard to the possibility that the current WNL population from Turkey represents a feral descendant of protodomestic emmer, as proposed by Civan et al., and Oliveira 2020, more research is needed to investigate this possibility. In paper 1 I have identified reduced genetic diversity in the WNL population compared to the WSL, which could align with the feralization scenario. Nevertheless, the WNL population presents almost twice as the nucleotide diversity found in the domestic landraces (paper 1, figure 3). The lower level of diversity compared to WSL could be explained by the distribution of wild emmer in the Northern Levant, which is found in small and isolated strands (Harlan and Zohary 1966), giving less chances for gene flow between groups. Also, another explanation could be found in the different sample size of the two populations within this study (20 vs 7), and in the limited geographic distribution of the analyzed WNL samples. Besides these limitations, the possibility of the current WNL population from Turkey being a feral descendant of protodomestic emmer raises intriguing questions. For a protodomestic to become feral, it would need to "revert" at least one domestication allele mutation, a process whose likelihood and mechanisms remain uncertain. Furthermore, considering that cultivation began in the Southern Levant region and gene flow persisted there even after domestication (WSL to DSE, see next section), one would expect a similar presence of feral populations in the Southern Levant area. However, the absence of evidence of such feral populations challenges this hypothesis. This discrepancy raises important questions about the conditions necessary for a population to become feral and the potential barriers to this process.

4.3. Dispersal of DSE and post domestication gene flow

The analyses in paper 1 (figures 2A-C), demonstrate the striking distinctiveness of the DSE population from the DNW one. In this thesis, I carry out a first attempt to reconstruct the ancient patterns that determined it. Such reconstruction is made possible considering complementary evidence from paper 1 and paper 2, as I will discuss in this section.

In paper 1, I compared genetic diversity in the different populations and excluded that the distinctiveness of DSE was derived by strong bottleneck or founder effects during dispersal. On the contrary, despite covering an overall smaller geographical range of provenance, this population does not show a considerable reduction in diversity when compared to DNW (paper 1, figure 3). On the other hand, I gathered evidence of gene flow from the WSL population to the DSE one, using both Patterson's D test and TreeMix (paper 1, figures 4 and 5). This allowed to determine that the differences between domestics are ascribable to uneven gene flow from the WSL after domestication. This is revealed at a higher resolution by the results obtained in paper 2, in which the ancestry proportions and the genetic distance analysis confirmed the higher levels of WSL descendance in the DSE genome (paper 2, figures 2A-B and 4A-B). While a certain amount of WSL ancestry is shared by the domestic populations, a fraction of it is found exclusively in the DSE population, and must therefore be derived from a process that involved this population only. I have indeed identified several genomic windows in which DNW is closer to WNL but DSE is closer to WSL. Remarkably, the estimated time since admixture computed using FastGT (ca 6500 years ago; paper 2, figure 3) aligns with a post-domestication gene flow scenario. It is known that domestic emmer wheat was introduced to Egypt between 7500 and 6500 years ago (Nesbitt and Samuel 1996; Zaharieva et al. 2010). As previously mentioned, the fully domestic phenotype first emerged in the North of the Fertile Crescent. Consequently, the initial domestic plants must have traveled through the Southern Levant to reach Egypt, allowing for further hybridization along the way. This admixture phase gains support when considering the over 2000-year timeframe between the emergence of the first domesticated plants, estimated around 9500 years ago, and their arrival in Egypt. As these domesticated plants migrated southward, it is highly likely that they coexisted with their wild and protodomestic counterparts. Given that the success of domesticated plants was primarily determined by their fitness in the human-mediated environment (Dorian Q. Fuller et al. 2014), rather than conscious human selection, it is improbable that the initial domestic fields exclusively contained fully domestic emmer. This coexistence of domestic, protodomestic, and wild plants significantly amplified opportunities for genetic

exchange. The genome of the DSE population was therefore shaped by the first hybridization phase during domestication (9500 years ago) and a second hybridization event with a population in the southern Levant during dispersal.

The FastGT analysis, however, was not able to retrieve the double wave of admixture in the DSE population, and estimated WSL to be the majority contributing source in the inferred admixture event (paper 2, figure 3). The analysis is constrained by the presence of only two wild populations as sources, the unusual long haplotypes and extremely low recombination rate of the emmer genome. In this context, accurately distinguishing between chromosomal segments of varying sizes originating from either ancestry is challenging. It is plausible that haplotypes resulting from different historical events (i.e. having different lengths) have "merged", creating extended haplotypes resembling those stemming from a singular, more recent event. Besides this technical constraint, another limitation arises from the inherent complexities of utilizing contemporary populations to reconstruct intricate historical events. Indeed, the described hybridization event occurred not between modern WSL and WNL (the populations used as donors in the analysis), but rather involved early domestic emmer spreading to the South, which had already acquired some WSL ancestry, and wild, or more likely protodomestic emmer in the Southern Levant. This can explain the inferred higher proportion of WSL contribution to this event. Considering these limits and the evidence from all the other analyses conducted in this study, together with archaeobotanical evidence, I conclude that the estimated time since admixture in DSE reflects a post-domestication gene flow event during dispersal.

To overcome the identified limitations, future research efforts could benefit from high coverage data and methodological advancements. Obtaining high coverage sequence data, would enable the retention of a greater number of variants, facilitating the exploration of a larger fraction of the genome. Additionally, exploring alternative computational methods or refining existing algorithms to account for the peculiarities of the emmer genome could enhance the accuracy of ancestry estimation.

Despite the constraints imposed by these limitations, it is noteworthy that the analysis yielded a time since admixture that aligns with documented events, such as the dispersal of domestic emmer to Egypt.

4.3.1. The contribution of aDNA to the study of gene flow and dispersal

The inclusion of one ancient sample from Egypt, along with samples from India and Oman, proved that the results obtained from modern data are accurate reflections of ancient genetic patterns. The ancient sample was collected in the archaeological site of Hememiah North Spur in Egypt (Brunton 1928), and is radiocarbon dated 3130-3000 years BP. It is therefore a representative of the emmer wheat that was cultivated in the region during the New Kingdom's Late Ramesside period, twentieth Dynasty. This sample was previously analyzed (Scott et al. 2019) together with exome data from diverse emmer populations (Avni et al. 2017). The study revealed the closer relationship between this sample and modern accessions from Oman, Turkey and India. Moreover, it shows signals of introgression from the WSL population. The authors concluded that eastward and southward dispersals of domesticated emmer are connected. However, due to small sample size of the "Indian Ocean" (Oman, Turkey, India) cluster and the lack of samples from Ethiopia, it was not possible to infer the most probable route of dispersal to India.

According to archaeological evidence, cereal agriculture arrived in the Iranian Plateau and into the Indus valley by 8000 BP and into the Nile Valley around 6500 BP (Dorian Q Fuller 2006). By 5000 years BP emmer had reached Ethiopia and India (Zaharieva et al. 2010). Two routes are possible: emmer dispersed to the South (Egypt) and to the East (Iranian plateau). Then, from the east it was introduced to Oman and Ethiopia ("clockwise" dispersal). The other possible route starts with the introduction to Egypt followed by dispersal to Ethiopia, Oman and finally India ("anticlockwise" dispersal).

In paper 1 I have extended my dataset to include the ancient sample and the four above mentioned modern samples. This expansion enabled me to contextualize the ancient sample within a more comprehensive representation of emmer populations that migrated both southward and eastward.

The ancient and modern samples clustered together with accession from the DSE population (paper 1, figure), indicating that modern landraces from Ethiopia, Oman and India are close relatives of the descendants of the ancient emmer cultivated in Egypt. Moreover, gene flow from the WSL population to both this ancient samples and the DSE population was confirmed by the results obtained in this study (treeMix, paper 1, figure 7). This evidence is highly revealing as it validates two fundamental discoveries. First, the signal of gene flow from the WSL is consistent in both modern and ancient specimens from the DSE population indicating that this phenomenon must have happened before 3000 years ago. This aligns with the model of post-domestication hybridization during southern dispersal, around 6500 years ago. Second, the fact that the emmer that was introduced into both Ethiopia and India is related to this ancient sample connects the early dispersal to Egypt with the later introduction in those countries. These results, taken together, rule out a possible route of dispersal of the DSE population through Iran to reach India, while they corroborate the other proposed route of dispersal, from the Fertile Crescent to Egypt, reaching Ethiopia, Oman and finally India. This route of dispersal finds support in the evidence of early maritime crop transfers between eastern Africa and South Asia, suggesting interactions across the Indian Ocean or potentially along the southern coast of Arabia as early as 5000-4000 years ago (Dorian Q. Fuller and Boivin 2009).

Despite the minimal sample size, and the limited amount of DNA that could be retrieved from the ancient seed (coverage 0.48X, Scott et al., 2019), these findings highlight the great resource that aDNA represents in genetic and domestication studies.

In conclusion, the findings of this study substantiate a refined model of emmer wheat dispersal. According to this model, one domesticated population migrated northward

and westward from the Fertile Crescent, ultimately giving rise to the modern DNW group observed in regions spanning Europe, the Caucasus, and the Balkans. Another population of domesticated emmer dispersed southward, where it underwent a subsequent hybridization event with wild or protodomestic types in the Southern Levant. The descendants of this southern dispersion constitute the contemporary DSE group, encompassing landraces from Ethiopia, Oman, and India (paper 2, figure 6).

4.4. The role of WSL ancestry in domestication and adaptation

Throughout this thesis, I have evaluated the impact of the WSL population on the domestic gene pool, both during and after the domestication process. Particularly, in paper 2, I have estimated the extent of WSL ancestry within domestic populations. This involved distinguishing between the WSL contribution to the domestication process (shared among domestic populations) and the gene flow occurring after domestication into the DSE population. As the final stage of this study, I endeavored to ascertain whether genes encoded in regions of WSL ancestry are associated with specific biological processes, cellular components, or molecular functions.

With respect to domestication, the genomic regions exhibiting WSL ancestry in both domestic populations, as indicated by Dxy patterns, demonstrate an overrepresentation of genes associated with systemic acquired resistance and molecular functions related to cell wall matrix synthesis, pivotal in various biological processes (Reiter 2002; Tryfona et al. 2014), and immunity (L. Zhang et al. 2019). Notably, both wild and domestic emmer are known to harbor many useful traits related to biotic stress tolerances and resistance to fungal diseases (e.g. powdery mildew, leaf rust, stem rust, and many more) (J. Peng et al. 2013; Rahman et al. 2023; Saleh 2020). On the other hand, efforts to introduce resistance genes from wild cereals

into wheat varieties underpin the critical need to enhance this aspect of modern cultivars (Kumar et al. 2022; Mapuranga et al. 2022).

Given the shared WSL ancestry of these regions in both domestic populations, it would be useful to identify allelic variants that are not shared between them. The DSE and the DNW groups are indeed found in different areas of the world, growing in different climatic conditions and possibly exposed to different pathogens. This would allow to discern between shared (conserved) immune mechanisms and post-dispersal adaptations. Even if more research is needed, these results suggest that emmer landraces, capable of growing without extensive use of pesticides, may serve as a genetic resource for enhancing disease resistance in wheat cultivars.

Examining WSL ancestry unique to the DSE population, I delved into genomic regions closely related to WSL in DSE but not in DNW. Analysis of these regions revealed an overrepresentation of categories primarily linked to the cell cycle, metabolism, cellular organization, membranes, and organelles. Particularly intriguing was the abundance of genes related to organelles. Cellular organization plays a pivotal role in drought stress response (Lv et al. 2020), with cytoplasm, peroxisomes, and chloroplasts highly susceptible to drought stress during early developmental stages in wheat. Additionally, recent studies have emphasized the susceptibility of chloroplasts to heat stress (Akter and Rafiqul Islam 2017; Hu, Ding, and Zhu 2020; Q. L. Wang et al. 2018). A specific region on chromosome 4B housing genes associated with chloroplasts exhibited strong signals of selection in the DSE population (paper 2, figure 5C), indicating that this chromosomal segment with WSL ancestry was favored in populations inhabiting Southern climates. These findings imply that the contribution of WSL to the DSE gene pool possibly facilitated adaptation to hot and arid environments during the dispersal process.

The presence of advantageous variants inherited from WSL relatives within emmer landraces suggests the potential for integrating these beneficial traits into wheat varieties. This approach bypasses the challenges associated with breeding cultivars

with wild plants, which often retain traits considered unfavorable in agriculture, such as small spikes, low yield, and brittle rachises.

While these findings necessitate further validation through additional research and offer only a preliminary glimpse into potentially selected regions with intriguing traits, the reported results emphasize the significance of domestication studies. They lay a crucial foundation for future improvement initiatives and breeding efforts. Understanding the intricate interplay between ancient genetic adaptations and contemporary agricultural challenges is pivotal, guiding us towards more informed and effective strategies for enhancing crop resilience and productivity.

4.5. Future perspectives

4.5.1. Future challenges and open questions

Expanding the dataset to include more samples from underrepresented geographical regions is a fundamental step towards a deeper comprehension of emmer wheat's genetic diversity and migratory routes.

The main limit of the dataset is the small number of samples of the WNL population, spanning a narrow geographical range. This "double" (sample size and geographical) underrepresentation affects various statistical analyses, requiring for example the adoption of a re-sampling strategy in the calculation of genetic diversity (see paper 1) and hindering inferences about potential evolutionary scenarios (see discussion). Extending the dataset with more samples covering the Northern Levant geographical range, and especially Iran, is fundamental to address questions like: *Is this wild population way less diverse than the WSL one? Is there any substructure within this population? And what are the relationships between wild types from Turkey, Iran and domestic populations?*

Conversely, the WSL population is better represented and comprises samples from different countries. In this study, I have identified some level of genetic structure between them (see paper 1), and more targeted sampling strategy would allow unravel

possible domestication dynamics. *Is there any WSL subpopulation that shares higher level of ancestry with the domestics?*

Regarding the eastward expansion of emmer wheat, it is essential to obtain a more robust representation of populations from Iran and India. This will allow to get detailed insights into the proposed dispersal route. Key questions linger: *Was the DSE population ever introduced to Iran? Do accessions from different origins within India all belong to the DSE population?*

It would also be important to extend the dataset to include samples from the Mediterranean basin. *What are the relationships between Mediterranean emmer and the identified domestic populations? Do Mediterranean emmer share the same WSL introgression signatures as the DSE population?*

While retrieving DNA from Neolithic archaeological seeds in Southwest Asia proves exceptionally challenging due to unfavorable preservation conditions (see annex), exploring ancient samples from diverse dispersal routes in Europe, Africa, and Asia and from different historical periods (e.g. Neolithic, Roman, Medieval) promises to shed light on ancient wheat populations and their connections to contemporary landraces. The analysis of diverse ancient samples in terms of geographical and historical origin would allow us to answer questions like: *How different was ancient European or Asian emmer from modern one? Can we explore ancient, potentially useful genetic variation that has been lost in modern landraces?*

Moreover, investing in high coverage sequencing and long-read technologies will not only enhance the quality of the data but also allow the exploration of previously inaccessible genomic regions. High coverage sequencing will allow a thorough analysis of haplotypes and the validation of heterozygous sites, eliminating the need for stringent filtering approaches. Meanwhile, long-read sequencing, would allow reliable alignment of complex genetic regions, aiding the identification of structural variants and transposable elements and allowing to assess the role that this variation has on wheat genome evolution under domestication.

Finally, regions with WSL ancestry and regions under differential selection in the domestic emmer population deserve a deeper study. The observed signatures of selection identified in this thesis should be examined in a larger wheat dataset, including durum and bread wheat landraces and elite cultivars, to verify whether adaptive traits have been lost in recent times. Subsequent functional validation in model organisms would be necessary to assess the phenotypic effect of these alleles.

4.5.2. Advancing Genomic Tools and Evolutionary Models for Tetraploid Wheat Studies

On a broader scale, studies on tetraploid wheat would dramatically benefit from improved reference genomes. Characterizing low complexity regions in both emmer and durum reference genomes will facilitate more accurate alignments and SNP identification. Likewise, efforts must be made to improve gene annotation by carrying out multi-omics studies, which will enable the identification of important alleles in coding portions of the genome, as well as non-coding variants playing an essential role in the regulation of agronomically important traits.

Moreover, adapting genomic tools for lengthy and repetitive genomes is pivotal. Genomic tools are seldom tailored for the analysis of genomes that are very long and repetitive. As an example, the widely-used variant calling toolkit GATK (v.4.1.6) cannot handle long chromosomes, requiring custom adaptations such as the creation of scripts to "cut and paste" them. These adaptations not only extend the processing time but also introduce a potential for errors. The adaptation of genomic tools to big, repetitive genomes will not only enhance the efficiency of data processing but also ensure the accuracy of variant detection, providing a solid foundation for downstream analyses.

Finally, the development of demographic models calibrated for the wheat genome would revolutionize our understanding of its evolution. Simulating the combined effects of long chromosomes, self-pollination, and domestication on population size

(Ne) over time is essential. It will allow to develop models which will serve as invaluable tools for hypothesis testing and scenario exploration. Thanks to such a tool, researchers will gain deeper insights into the intricate interplay between selection and environmental factors that have shaped the emmer wheat genome.

Chapter 5

CONCLUSIONS

As a concluding remark, this work highlights the importance of interdisciplinarity in research. The integration of evidence from diverse disciplines such as archaeology and genomics is fundamental and complementary in the study of crop domestication and dispersal. Insights from various fields have proven essential, offering valuable glimpses into our agricultural past. Ancient DNA analysis, in particular, emerges as a powerful tool, allowing us to analyze past genetic diversity.

Despite the limits imposed by the dataset and the inherent complexities of the wheat genome, this study has successfully reconstructed the ancestry of domestic populations and brought new insight into the dynamics of the domestication process, dispersal and adaptation of emmer wheat.

1. Insights into Population Structure

This research provides detailed insights into the population structure of emmer wheat, uncovering distinct genetic clusters within wild and domestic populations. Two wild populations have been identified, one from the north and the other from the south of the Fertile Crescent. Domestic populations reflect dispersal routes outside of Southwest Asia (south-east and north-west) and are more closely related to the wild Northern Levant population at a genome-wide level.

2. Reticulated Origin

This study provides evidence supporting the reticulated origin of domestic emmer wheat. Specifically, the research quantifies the contribution of the Wild Northern and Southern Levant (WNL and WSL) populations to domestic genomes and provides a refined model of domestication, in which protodomestic plants derived from different wild populations hybridized around 9500 years ago in the North of the Fertile Crescent.

3. Dispersal Patterns and Post-Domestication Gene Flow

This study sheds light on the dispersal patterns of emmer wheat, highlighting the distinctiveness of different domestic populations. It identifies uneven gene flow from the WSL population after domestication in the DSE population. This post-domestication gene flow aligns with historical events such as the spread of domestic emmer to Egypt (6500 years ago) and subsequent southward dispersal, providing valuable insights into the ancient movements of agricultural societies and their crops.

4. The contribution of aDNA

The inclusion of ancient DNA in this study has proven instrumental in reconstructing historical dispersal patterns and gene flow events. By integrating genetic information from ancient emmer wheat, it was possible to verify the ancient origin of hybridization events, that happened before 3000 years ago, and validate possible dispersal routes to the south and to the east.

5. Applications

This research underscores the significance of domestication studies for future crop improvement. The identification in domestic landraces of regions having WSL ancestry associated with disease resistance and environmental adaptation lays the foundations for future research and opens the doors to targeted breeding programs. By harnessing the genetic diversity inherent in wild strands and domestic landraces, researchers and can work towards developing resilient crop varieties capable of withstanding changing environmental conditions and evolving pathogens.

Chapter 6

BIBLIOGRAPHY

Abbo, Shahal et al. 2014. "Plant Domestication versus Crop Evolution: A Conceptual Framework for Cereals and Grain Legumes." *Trends in Plant Science* 19(6): 351–60. http://dx.doi.org/10.1016/j.tplants.2013.12.002.

Abbo, Shahal, Simcha Lev-Yadun, and Avi Gopher. 2010. "Agricultural Origins: Centers and Noncenters; A Near Eastern Reappraisal." *Critical Reviews in Plant Sciences* 29(5): 317–28. https://doi.org/10.1080/07352689.2010.502823.

Akter, Nurunnaher, and M Rafiqul Islam. 2017. "Heat Stress Effects and Management in Wheat. A Review." *Agronomy for Sustainable Development* 37(5): 37. https://doi.org/10.1007/s13593-017-0443-9.

Alexander, David H., John Novembre, and Kenneth Lange. 2009. "Fast Model-Based Estimation of Ancestry in Unrelated Individuals." *Genome Research* 19(9): 1655–64.

Alhusain, Luluah, and Alaaeldin M. Hafez. 2018. "Nonparametric Approaches for Population Structure Analysis." *Human Genomics* 12(1): 1–12.

Allaby, Robin G. et al. 2017. "Geographic Mosaics and Changing Rates of Cereal Domestication." *Philosophical Transactions of the Royal Society B: Biological Sciences* 372(1735).

Allaby, Robin G., Monica Banerjee, and Terence A. Brown. 1999. "Evolution of the High Molecular Weight Glutenin Loci of the A, B, D, and G Genomes of Wheat." *Genome* 42(2): 296–307. https://doi.org/10.1139/g98-114.

Allaby, Robin G., Chris J. Stevens, Logan Kistler, and Dorian Q. Fuller. 2022. "Emerging Evidence of Plant Domestication as a Landscape-Level Process." *Trends in Ecology and Evolution* 37(3): 268–79. https://doi.org/10.1016/j.tree.2021.11.002.

Allaby, Robin G., Roselyn L. Ware, and Logan Kistler. 2019. "A Re-Evaluation of the Domestication Bottleneck from Archaeogenomic Evidence." *Evolutionary Applications* 12(1): 29–37.

Alonge, Michael et al. 2020. "Chromosome-Scale Assembly of the Bread Wheat Genome Reveals Thousands of Additional Gene Copies." *Genetics*: genetics.303501.2020.

Appels, Rudi et al. 2018. "Shifting the Limits in Wheat Research and Breeding Using a Fully Annotated Reference Genome." *Science* 361(6403).

Aranzana, Maria J., El Kadri Abbassi, Werner Howad, and Pere Arús. 2010. "Genetic Variation, Population Structure and Linkage Disequilibrium in Peach Commercial Varieties." *BMC Genetics* 11.

Arora, Naveen Kumar. 2019. "Impact of Climate Change on Agriculture Production and Its Sustainable Solutions." *Environmental Sustainability* 2(2): 95–96. https://doi.org/10.1007/s42398-019-00078-w.

Arranz-Otaegui, Amaia et al. 2016. "Regional Diversity on the Timing for the Initial Appearance of Cereal Cultivation and Domestication in Southwest Asia." *Proceedings of the National Academy of Sciences of the United States of America* 113(49): 14001–6.

Arranz-Otaegui, Amaia, Lara González Carretero, Joe Roe, and Tobias Richter. 2018. "'Founder Crops' v. Wild Plants: Assessing the Plant-Based Diet of the Last Hunter-Gatherers in Southwest Asia." *Quaternary Science Reviews* 186: 263–83.

Arranz-Otaegui, Amaia, and Joe Roe. 2023. "Revisiting the Concept of the 'Neolithic

Founder Crops' in Southwest Asia." *Vegetation History and Archaeobotany*. https://doi.org/10.1007/s00334-023-00917-1.

Asouti, Eleni, and Dorian Q Fuller. 2013. "A Contextual Approach to the Emergence of Agriculture in Southwest Asia: Reconstructing Early Neolithic Plant-Food Production." *Current Anthropology* 54(3): 299–345. https://doi.org/10.1086/670679.

Van der Auwera, GA O'Connor, BD. 2020. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (1st Edition)*.

Ávila-Arcos, María C. et al. 2011. "Application and Comparison of Large-Scale Solution-Based DNA Capture-Enrichment Methods on Ancient DNA." *Scientific Reports* 1.

Avni, Raz et al. 2017. "Wild Emmer Genome Architecture and Diversity Elucidate Wheat Evolution and Domestication." *Science* 97(July): 93–97.

Bahaji, Abdellatif et al. 2014. "Starch Biosynthesis, Its Regulation and Biotechnological Approaches to Improve Crop Yields." *Biotechnology Advances* 32(1):87–106. https://www.sciencedirect.com/science/article/pii/S0734975013001122.

Bai, Yuling, and Pim Lindhout. 2007. "Domestication and Breeding of Tomatoes: What Have We Gained and What Can We Gain in the Future?" *Annals of botany* 100(5): 1085–94.

Balla, Mohammed Yousif et al. 2022. "Exploiting Wild Emmer Wheat Diversity to Improve Wheat A and B Genomes in Breeding for Heat Stress Adaptation." *Frontiers in Plant Science* 13(July).

Bergamini, Nadia, Stefano Padulosi, S Bala Ravi, and Nirmala Yenagi. 2013. "Minor Millets in India: A Neglected Crop Goes Mainstream." *Diversifying food and diets: Using agricultural biodiversity to improve nutrition and health*: 313–25.

Bilgic, Hatice et al. 2016. "Ancient DNA from 8400 Year-Old Çatalhöyük Wheat: Implications for the Origin of Neolithic Agriculture." *Plos One* 11(3): e0151974.

Blatter, Robert H.E., Stefanie Jacomet, and Angela Schlumbaum. 2002. "Spelt-Specific Alleles in HMW Glutenin Genes from Modern and Historical European Spelt (Triticum Spelta L.)." *Theoretical and Applied Genetics* 104(2–3): 329–37.

Bohra, Abhishek et al. 2022. "Reap the Crop Wild Relatives for Breeding Future Crops." *Trends in Biotechnology* 40(4): 412–31. https://doi.org/10.1016/j.tibtech.2021.08.009.

Borrill, Philippa, Nikolai Adamski, and Cristobal Uauy. 2015. "Genomics as the Key to Unlocking the Polyploid Potential of Wheat." *New Phytologist* 208(4): 1008–22.

Botigue, Laura R et al. 2017. "Ancient European Dog Genomes Reveal Continuity since the Early Neolithic." *Nature Communications* (May).

Briggs, A. W. et al. 2010. "Removal of Deaminated Cytosines and Detection of in Vivo Methylation in Ancient DNA." *Nucleic acids research* 38.

Briggs, A. W., U. Stenzel, P. L. F. Johnson, and S. Pääbo. 2007. "Patterns of Damage in Genomic DNA Sequences from a Neandertal." *Proceedings of the National Academy of Sciences* 104: 14616–14621.

Brown, Terence A, Martin K Jones, Wayne Powell, and Robin G Allaby. 2009. "The

Complex Origins of Domesticated Crops in the Fertile Crescent." *Trends in Ecology & Evolution* 24(2): 103–9. https://www.sciencedirect.com/science/article/pii/S0169534708003479.

Brozynska, Marta, Agnelo Furtado, and Robert J. Henry. 2016. "Genomics of Crop Wild Relatives: Expanding the Gene Pool for Crop Improvement." *Plant Biotechnology Journal* 14(4): 1070–85.

Brunton, Guy. 1928. "The Badarian Civilisation and Predynastic Remains near Badari." *(No Title)*.

Bukowski, Robert et al. 2018. "Construction of the Third-Generation Zea Mays Haplotype Map." *GigaScience* 7(4): 1–12.

Burgarella, Concetta, and Sylvain Glémin. 2017. "Population Genetics and Genome Evolution of Selfing Species." *eLS*: 1–8.

Campanaro, Ausilia et al. 2019. "DNA Barcoding to Promote Social Awareness and Identity of Neglected, Underutilized Plant Species Having Valuable Nutritional Properties." *Food Research International* 115: 1–9. https://www.sciencedirect.com/science/article/pii/S0963996918305696.

Carmona, Diego, Marc J. Lajeunesse, and Marc T.J. Johnson. 2011. "Plant Traits That Predict Resistance to Herbivores." *Functional Ecology* 25(2): 358–67.

Chacón-Duque, Juan Camilo et al. 2018. "Latin Americans Show Wide-Spread Converso Ancestry and Imprint of Local Native Ancestry on Physical Appearance." *Nature Communications* 9(1).

Charles, Darwin. 1859. "On the Origin of Species by Means of Natural Selection."

Charlesworth, B. 1998. "Measures of Divergence between Populations and the Effect of Forces That Reduce Variability." *Molecular Biology and Evolution* 15(5): 538–43. https://doi.org/10.1093/oxfordjournals.molbev.a025953.

Chaudhary, Bhupendra. 2013. "Plant Domestication and Resistance to Herbivory." *International Journal of Plant Genomics* 2013.

Chen, Hua et al. 2006. "A Psb27 Homologue in Arabidopsis Thaliana Is Required for Efficient Repair of Photodamaged Photosystem II." *Plant Molecular Biology* 61(4): 567–75. https://doi.org/10.1007/s11103-006-0031-x.

Chen, Yolanda H, Rieta Gols, and Betty Benrey. 2015. "Crop Domestication and Its Impact on Naturally Selected Trophic Interactions." *Annual Review of Entomology* 60(1): 35–58. https://doi.org/10.1146/annurev-ento-010814-020601.

Cheng, Hong et al. 2019. "Frequent Intra- and Inter-Species Introgression Shapes the Landscape of Genetic Variation in Bread Wheat." *Genome Biology* 20(1): 1–16.

Choi, Jae Young et al. 2017. "The Rice Paradox: Multiple Origins but Single Domestication in Asian Rice." *Molecular Biology and Evolution* 34(4): 969–79. https://doi.org/10.1093/molbev/msx049.

Civáň, Peter, Zuzana Ivaničová, and Terence A. Brown. 2013. "Reticulated Origin of Domesticated Emmer Wheat Supports a Dynamic Model for the Emergence of Agriculture in the Fertile Crescent." *PLoS ONE* 8(11): 1–11.

Clark, Michael A et al. 2020. "Global Food System Emissions Could Preclude Achieving the 1.5° and 2°C Climate Change Targets." *Science* 370(6517): 705–8. https://doi.org/10.1126/science.aba7357.

Cook, Daniel E. "Generate Masked Ranges." https://gist.github.com/danielecook/cfaa5c359d99bcad3200.

Cortés, Andrés J., and Felipe López-Hernández. 2021. "Harnessing Crop Wild Diversity for Climate Change Adaptation." *Genes* 12(5): NA.

Coward, Fiona et al. 2008. "The Spread of Neolithic Plant Economies from the Near East to Northwest Europe : A Phylogenetic Analysis." 35: 42–56.

Crowther, Alison, Mary E Prendergast, Dorian Q Fuller, and Nicole Boivin. 2018. "Subsistence Mosaics , Forager-Farmer Interactions , and the Transition to Food Production in Eastern Africa." *Quaternary International* 489: 101–20. https://doi.org/10.1016/j.quaint.2017.01.014.

Cuny, J., and M. Mouton. 2008. "La Transition Vers La Période Sassanide Dans La Péninsule d'Oman: Chronologie et Modes de Peuplement." In *L'Arabie à La Veille de l'Islam: Bilan Clinique*, De Boccard.

Czajkowska, Beata I. et al. 2020. "Ancient DNA Typing Indicates That the 'New' Glume Wheat of Early Eurasian Agriculture Is a Cultivated Member of the Triticum Timopheevii Group." *Journal of Archaeological Science* 123(October): 105258. https://doi.org/10.1016/j.jas.2020.105258.

D'Antuono, L.F., and M. Minelli. 1997. "Yield and Yield Components Analysis of Emmer Wheat (Triticum Dicocum Schubler) Landraces from Italy." In *Proceedings of the 3rd International Triticeae Symposium*, Aleppo Syria, 393–404.

D'Antuono, L F. 1989. "Il Farro: Areali Di Coltivazione, Caratteristiche Agronomiche, Utilizzazione e Prospettive Colturali." *L'Informatore agrario* 45(24): 49–57.

Daly, Kevin. G. et al. 2018. "Ancient Goat Genomes Reveal Mosaic Domestication in the Fertile Crescent." *Science , 85–88* 361: 24–27.

Danecek, Petr et al. 2011. "The Variant Call Format and VCFtools." *Bioinformatics*.

Danecek, Petr et al. 2021. "Twelve Years of SAMtools and BCFtools." *GigaScience* 10(2): 1–4.

Delaneau, Olivier et al. 2014. "Integrating Sequence and Array Data to Create an Improved 1000 Genomes Project Haplotype Reference Panel." *Nature Communications* 5: 1–9.

Dhanavath, Srinu, and U. J.S. Prasada Rao. 2017. "Nutritional and Nutraceutical Properties of Triticum Dicoccum Wheat and Its Health Benefits: An Overview." *Journal of Food Science* 82(10): 2243–50.

Diamond, Jared. 2002. "Evolution, Consequences and Future of Plant and Animal Domestication." *Nature* 418(6898): 700–707. http://www.nature.com/doifinder/10.1038/nature01019.

di Donato, Antimo, Edgardo Filippone, Maria R. Ercolano, and Luigi Frusciante. 2018. "Genome Sequencing of Ancient Plant Remains: Findings, Uses and Potential Applications for the Study and Improvement of Modern Crops." *Frontiers in Plant Science* 9(April).

DOROFEEV, V.F. et al. 1979. *Wheat. Origin, Classification, Cultivation, and Use. (English Translation)*. Pullman, WA: Washington State University Press.

Durrant, W. E., and X. Dong. 2004. "SYSTEMIC ACQUIRED RESISTANCE." *Annual Review of Phytopathology* 42(1): 185–209. https://doi.org/10.1146/annurev.phyto.42.040803.140421.

Dvořák, J. 2001. "Triticum Species (Wheat)." *Encyclopedia of Genetics*: 2060–68.

Dwivedi, Sangam L. et al. 2016. "Landrace Germplasm for Improving Yield and

Abiotic Stress Adaptation." *Trends in Plant Science* 21(1): 31–42. http://dx.doi.org/10.1016/j.tplants.2015.10.012.

Engels, Johannes M M, and Imke Thormann. 2020. "Main Challenges and Actions Needed to Improve Conservation and Sustainable Use of Our Crop Wild Relatives." *Plants (Basel, Switzerland)* 9(8).

Estrada, Oscar, James Breen, Stephen M. Richards, and Alan Cooper. 2018. "Ancient Plant DNA in the Genomic Era." *Nature Plants* 4(7): 394–96. http://dx.doi.org/10.1038/s41477-018-0187-9.

Eydivandi, Sirous, Mahmoud Amiri Roudbar, Mohammad Osman Karimi, and Goutam Sahana. 2021. "Genomic Scans for Selective Sweeps through Haplotype Homozygosity and Allelic Fixation in 14 Indigenous Sheep Breeds from Middle East and South Asia." *Scientific Reports* 11(1): 2834. https://doi.org/10.1038/s41598-021-82625-2.

FAO. 2021. "Agricultural Production Statistics 2000-2021." *FAOSTAT ANALYTICAL BRIEF 60*. https://www.fao.org/3/cc3751en/cc3751en.pdf.

Feldman, M., and E. M. Kislev. 2007. "Domestication of Emmer Wheat and Evolution of Free-Threshing Tetraploid Wheat." *Isr J Plant Sci* 55: 207–221.

Felsenstein, J. 1981. "Evolutionary Trees from DNA Sequences: A Maximum Likelihood Approach." *Journal of molecular evolution* 17(6): 368–76.

Fernández, E. et al. 2013. "DNA Analysis in Charred Grains of Naked Wheat from Several Archaeological Sites in Spain." *Journal of Archaeological Science*: 659–670.

Filipovic, Dragana. 2014. "Southwest Asian Founder- and Other Crops at Neolithic Sites in Serbia." *Bulgarian e-Journal of Archaeology* 4(2): 229–47.

Finkina, E I, D N Melnikova, I V Bogdanov, and T V Ovchinnikova. 2016. "Lipid Transfer Proteins As Components of the Plant Innate Immune System: Structure, Functions, and Applications." *Acta naturae* 8(2): 47–61.

Fitch, Walter M. 1971. "Toward Defining the Course of Evolution: Minimum Change for a Specific Tree Topology." *Systematic Zoology* 20(4): 406–16. http://www.jstor.org/stable/2412116.

Da Fonseca, Rute R. et al. 2015. "The Origin and Evolution of Maize in the Southwestern United States." *Nature Plants* 1(January): 1–5.

Fornaciari, R. et al. 2018. "Panicum Spikelets from the Early Holocene Takarkori Rockshelter (SW Libya): Archaeo-Molecular and -Botanical Investigations." *Plant Biosystems* 152(1): 1–13. http://dx.doi.org/10.1080/11263504.2016.1244117.

Fraire-Velazquez, Saul, and Victor Emmanuel. 2013. "Abiotic Stress in Plants and Metabolic Responses." *Abiotic Stress - Plant Responses and Applications in Agriculture*.

Frantz, Laurent A. F. et al. 2019. "Ancient Pigs Reveal a Near-Complete Genomic Turnover Following Their Introduction to Europe." *Pnas* 116.

Fuller, D. Q., and L. Lucas. 2017. "Adapting Crops, Landscapes, and Food Choices: Patterns in the Dispersal of Domesticated Plants across Eurasia." In *Human Dispersal and Species Movement: From Prehistory to the Present*, , 304–31.

Fuller, Dorian Q. 2007. "Contrasting Patterns in Crop Domestication and Domestication Rates: Recent Archaeobotanical Insights from the Old World." *Annals of Botany* 100(5): 903–24.

Fuller, Dorian Q., et al. 2014. "Convergent Evolution and Parallelism in Plant

Domestication Revealed by an Expanding Archaeological Record." *Proceedings of the National Academy of Sciences of the United States of America* 111(17): 6147–52.

Fuller, Dorian Q., Robin G. Allaby, and Chris Stevens. 2010. "Domestication as Innovation: The Entanglement of Techniques, Technology and Chance in the Domestication of Cereal Crops." *World Archaeology* 42(1): 13–28.

Fuller, Dorian Q., and Nicole Boivin. 2009. "Crops, Cattle and Commensals across the Indian Ocean." *Études Océan Indien* (42–43): 13–46.

Fuller, Dorian Q., Tim Denham, and Robin Allaby. 2023. "Plant Domestication and Agricultural Ecologies." *Current Biology* 33. https://doi.org/10.1016/j.cub.2023.04.038.

Fuller, Dorian Q., George Willcox, and Robin G. Allaby. 2011. "Cultivation and Domestication Had Multiple Origins: Arguments against the Core Area Hypothesis for the Origins of Agriculture in the Near East." *World Archaeology* 43(4): 628–52.

Fuller, Dorian Q. 2006. "Agricultural Origins and Frontiers in South Asia : A Working Synthesis." : 1–86.

Fuller, Dorian Q, Eleni Asouti, and Michael D Purugganan. 2012. "Cultivation as Slow Evolutionary Entanglement : Comparative Data on Rate and Sequence of Domestication." : 131–45.

Fulton, T. L. 2012. "Setting Up an Ancient DNA Laboratory." In *Shapiro, B. and Hofreiter, M. (Eds) Ancient DNA: Methods and Protocols.*, , 1–11.

Gao, Zhaoxu et al. 2023. "Triticeae Crop Genome Biology: An Endless Frontier." *Frontiers in Plant Science* 14(July): 1–14.

Gardiner, Laura Jayne et al. 2019. "Integrating Genomic Resources to Present Full Gene and Putative Promoter Capture Probe Sets for Bread Wheat." *GigaScience* 8(4): 1–13.

Gaunitz, Charleen et al. 2018. "Ancient Genomes Revisit the Ancestry of Domestic and Przewalski ' s Horses." (April): 1–5.

Golenberg, E. M. 1988. "Outcrossing Rates and Their Relationship to Phenology in Triticum Dicoccoides." *Theoretical and Applied Genetics* 75(6): 937–44. http://link.springer.com/10.1007/BF00258057.

Golovnina, K. A. et al. 2007. "Molecular Phylogeny of the Genus Triticum L." *Plant Systematics and Evolution* 264(3–4): 195–216.

Goncharov, Nikolay P. 2011. "Genus Triticum L. Taxonomy: The Present and the Future." *Plant Systematics and Evolution* 295(1): 1–11.

Gonthier, David J et al. 2014. "Biodiversity Conservation in Agriculture Requires a Multi-Scale Approach." *Proceedings. Biological sciences* 281(1791): 20141358.

Green, R. E. et al. 2010. "A Draft Sequence of the Neandertal Genome." *Science* 328: 710–722.

Günther, Torsten, and Carl Nettelblad. 2019. "The Presence and Impact of Reference Bias on Population Genomic Studies of Prehistoric Human Populations." *PLoS Genetics* 15(7): 1–20.

Gutaker, Rafal M., and Hernán A. Burbano. 2017. "Reinforcing Plant Evolutionary Genomics Using Ancient DNA." *Current Opinion in Plant Biology* 36: 38–45.

Gutaker, Rafal M, Maricris Zaidem, et al. 2019. "Flax Latitudinal Adaptation at LuTFL1 Altered Architecture and Promoted Fiber Production." *Scientific Reports*

9(1): 976. https://doi.org/10.1038/s41598-018-37086-5.

Gutaker, Rafal M, Clemens L Weiß, et al. 2019. "The Origins and Adaptation of European Potatoes Reconstructed from Historical Genomes." *Nature Ecology & Evolution* 3(July). http://dx.doi.org/10.1038/s41559-019-0921-3.

Haas, Matthew, Mona Schreiber, and Martin Mascher. 2018. "Domestication and Crop Evolution of Wheat and Barley: Genes, Genomics, and Future Directions." *Journal of Integrative Plant Biology* XXXX(Xxxx).

Haberer, Georg et al. 2005. "Structure and Architecture of the Maize Genome." *Plant physiology* 139(4): 1612–24.

Hajjar, Reem, and Toby Hodgkin. 2007. "The Use of Wild Relatives in Crop Improvement: A Survey of Developments over the Last 20 Years." *Euphytica* 156(1): 1–13. https://doi.org/10.1007/s10681-007-9363-0.

Harlan, Jack R, and Daniel Zohary. 1966. "Distribution of Wild Wheats and Barley." *Science* 153: 1074–79.

Hazen, Samuel P et al. 2003. "Quantitative Trait Loci and Comparative Genomics of Cereal Cell Wall Composition." *Plant Physiology* 132(1): 263–71. https://doi.org/10.1104/pp.103.020016.

He, Fei et al. 2019. "Exome Sequencing Highlights the Role of Wild-Relative Introgression in Shaping the Adaptive Landscape of the Wheat Genome." *Nature Genetics* 51(5): 896–904. http://dx.doi.org/10.1038/s41588-019-0382-2.

Helbaek, HH. 1970. "The Plant Husbandry of Hac¸ilar." In *Mel- Laart J (Ed) Excavations at Hac¸ilar*, , 189–244.

Hellenthal, Garrett et al. 2014. "A Genetic Atlas of Human Admixture History." 343(February): 747–51.

Henderson, Elisa C., and Alan Brelsford. 2020. "Genomic Differentiation across the Speciation Continuum in Three Hummingbird Species Pairs." *BMC Evolutionary Biology* 20(1): 1–11.

Higuchi, R. et al. 1984. "DNA Sequences from the Quagga, an Extinct Member of the Horse Family." *Nature* 312: 282–84.

Hillman, Gordon C, and M Stuart Davies. 1992. "Domestication Rate in Wild Wheats and Barley under Primitive Cultivation: Preminary Results and Archaeological Implications of Field Measurements of Selection Coefficient." *Monographie du CRA* (6): 113–58.

Hofreiter, Michael et al. 2015. "The Future of Ancient DNA: Technical Advances and Conceptual Shifts." *BioEssays* 37(3): 284–93.

Hu, Shanshan, Yanfei Ding, and Cheng Zhu. 2020. "Sensitivity and Responses of Chloroplasts to Heat Stress in Plants." *Frontiers in Plant Science* 11(April): 1–11.

Hufford, Matthew B., Jorge C. Berny Mier Y Teran, and Paul Gepts. 2019. "Crop Biodiversity: An Unfinished Magnum Opus of Nature." *Annual Review of Plant Biology* 70: 727–51.

Igolkina, Anna A. et al. 2023. "Historical Routes for Diversification of Domesticated Chickpea Inferred from Landrace Genomics." *Molecular Biology and Evolution* 40(6).

Igrejas, Gilberto, and Gérard Branlard. 2020. "The Importance of Wheat." *Wheat quality for improving processing and human health*: 1–7.

Innan, Hideki, and Yuseob Kim. 2004. "Pattern of Polymorphism after Strong

Artificial Selection in a Domestication Event." *Proceedings of the National Academy of Sciences* 101(29): 10667–72.

Institute, Broad. 2019. "Picard Toolkit." https://broadinstitute.github.io/picard/.

Iob, Alice, and Laura Botigué. 2022. "Genomic Analysis of Emmer Wheat Shows a Complex History with Two Distinct Domestic Groups and Evidence of Differential Hybridization with Wild Emmer from the Western Fertile Crescent." *Vegetation History and Archaeobotany* (0123456789). https://doi.org/10.1007/s00334-022-00898-7.

Janzen, Garrett M., Li Wang, and Matthew B. Hufford. 2019. "The Extent of Adaptive Wild Introgression in Crops." *New Phytologist* 221(3): 1279–88.

Jia, Yong et al. 2023. "Comparative Gene Retention Analysis in Barley, Wild Emmer, and Bread Wheat Pangenome Lines Reveals Factors Affecting Gene Retention Following Gene Duplication." *BMC Biology* 21(1): 25. https://doi.org/10.1186/s12915-022-01503-z.

Jiang, Yun Feng et al. 2014. "Genome-Wide Quantitative Trait Locus Mapping Identifies Multiple Major Loci for Brittle Rachis and Threshability in Tibetan Semi-Wild Wheat (Triticum Aestivum Ssp. Tibetanum Shao)." *PLoS ONE* 9(12): 1–15.

Johnson, Virginia M et al. 2022. "Psb27, a Photosystem II Assembly Protein, Enables Quenching of Excess Light Energy during Its Participation in the PSII Lifecycle." *Photosynthesis Research* 152(3): 297–304. https://doi.org/10.1007/s11120-021-00895-3.

Jombart, Thibaut, Sébastien Devillard, and François Balloux. 2010. "Discriminant Analysis of Principal Components: A New Method for the Analysis of Genetically Structured Populations." *BMC Genetics* 11(1): 94. https://doi.org/10.1186/1471-2156-11-94.

Jones, Glynis, Soultana Valamoti, and Michael Charles. 2000. "Early Crop Diversity: A 'New' Glume Wheat from Northern Greece." *Vegetation History and Archaeobotany* 9(3): 133–46.

Jonsson, H. et al. 2014. "Ancient Genomics." *Philosophical Transactions of the Royal Society B: Biological Sciences* 370.

Jónsson, Hákon et al. 2013. "MapDamage2.0: Fast Approximate Bayesian Estimates of Ancient DNA Damage Parameters." *Bioinformatics* 29(13): 1682–84.

Jordan, Katherine W. et al. 2015. "A Haplotype Map of Allohexaploid Wheat Reveals Distinct Patterns of Selection on Homoeologous Genomes." *Genome Biology* 16(1): 1–18.

Jorgensen, Chad et al. 2017. "A High-Density Genetic Map of Wild Emmer Wheat from the Karaca Dağ Region Provides New Evidence on the Structure and Evolution of Wheat Chromosomes." *Frontiers in Plant Science* 8(October): 1–13.

Kabukcu, Ceren et al. 2021. "Pathways to Plant Domestication in Southeast Anatolia Based on New Data from Aceramic Neolithic Gusir Höyük." *Scientific Reports* 11(1): 1–15. https://doi.org/10.1038/s41598-021-81757-9.

Kantar, Michael B. et al. 2017. "The Genetics and Genomics of Plant Domestication." *BioScience* XX(X): 1–12. http://academic.oup.com/bioscience/article/doi/10.1093/biosci/bix114/4371394/The-Genetics-and-Genomics-of-Plant-Domestication.

Kaplan, Lawrence, Mary B. Smith, and Lesley Ann Sneddon. 1992. "Cereal Grain Phytoliths of Southwest Asia and Europe." In *Phytolith Systematics in Archaeological and Museum Science*, eds. G. Rapp and S.C. Mulholland. Springer US, 149–74. https://doi.org/10.1007/978-1-4899-1155-1_8.

Karn, Avinash, Jason D Gillman, and Sherry A Flint-Garcia. 2017. "Genetic Analysis of Teosinte Alleles for Kernel Composition Traits in Maize." *G3 (Bethesda, Md.)* 7(4): 1157–64.

Keilwagen, Jens et al. 2022. "Detecting Major Introgressions in Wheat and Their Putative Origins Using Coverage Analysis." *Scientific Reports* 12(1): 1–11. https://doi.org/10.1038/s41598-022-05865-w.

Key, Felix M et al. 2020. "Emergence of Human-Adapted Salmonella Enterica Is Linked to the Neolithization Process." *Nature Ecology & Evolution* 4(March). http://dx.doi.org/10.1038/s41559-020-1106-9.

Khoury, C. K. et al. 2014. "Increasing Homogeneity in Global Food Supplies and the Implications for Food Security." *Proceedings of the National Academy of Sciences* 111(11): 4001–6.

Khoury, Colin K. et al. 2022. "Crop Genetic Erosion: Understanding and Responding to Loss of Crop Diversity." *New Phytologist* 233(1): 84–118.

Kilian, Benjamin et al. 2007. "Molecular Diversity at 18 Loci in 321 Wild and 92 Domesticate Lines Reveal No Reduction of Nucleotide Diversity during Triticum Monococcum (Einkorn) Domestication: Implications for the Origin of Agriculture." *Molecular Biology and Evolution* 24(12): 2657–68.

Kislev, ME, D Nadel, and I Carmi. 1992. "Epipalaeolithic (19,000 Bp) Cereal and Fruit Diet at Ohalu II, Sea of Galilee, Israel." *Review of Palaeobotany and Palynology* 73: 161–66.

Kistler, L. et al. 2017. "A New Model for Ancient DNA Decay Based on Paleogenomic Meta-Analysis." *Nucleic acids research* 45: 6310–6320.

Kistler, Logan et al. 2020. "Ancient Plant Genomics in Archaeology, Herbaria, and the Environment." *Annual Review of Plant Biology* 71: 605–29.

Kohane, M J, and P A Parsons. 1988. "Domestication: Evolutionary Change under Stress." *Evolutionary Biology* 23: 31–48.

Korneliussen, T. S. Albrechtsen, A., and R. Nielsen. 2014. "ANGSD: Analysis of Next Generation Sequencing Data." *BMC Bioinformatics* 15: 1–13.

Krasileva, Ksenia V et al. 2013. "Separating Homeologs by Phasing in the Tetraploid Wheat Transcriptome." *Genome Biology* 14(6): 1–19.

Krause, Johannes et al. 2010. "The Complete Mitochondrial DNA Genome of an Unknown Hominin from Southern Siberia." 464(April): 894–97.

Kumar, Kuldeep et al. 2022. "An Update on Resistance Genes and Their Use in the Development of Leaf Rust Resistant Cultivars in Wheat ." *Frontiers in Genetics* 13. https://www.frontiersin.org/articles/10.3389/fgene.2022.816057.

Kurata, Nori, Ken-Ichi Nonomura, and Yoshiaki Harushima. 2002. "Rice Genome Organization: The Centromere and Genome Interactions." *Annals of botany* 90(4): 427–35.

Kwak, Myounghai, and Paul Gepts. 2009. "Structure of Genetic Diversity in the Two Major Gene Pools of Common Bean (Phaseolus Vulgaris L., Fabaceae)." *Theoretical and Applied Genetics* 118(5): 979–92.

Labeyrie, Vanesse et al. 2021. "The Role of Crop Diversity in Climate Change Adaptation: Insights from Local Observations to Inform Decision Making in Agriculture." *Current Opinion in Environmental Sustainability* 51: 15–23.

Lacan, Marie et al. 2011. "Ancient DNA Reveals Male Diffusion through the Neolithic Mediterranean Route."

Larson, G. et al. 2014. "Current Perspectives and the Future of Domestication Studies." *Proceedings of the National Academy of Sciences* 111(17): 6139–46. http://www.pnas.org/cgi/doi/10.1073/pnas.1323964111.

Lawson, Daniel J., Lucy van Dorp, and Daniel Falush. 2018. "A Tutorial on How Not to Over-Interpret STRUCTURE and ADMIXTURE Bar Plots." *Nature Communications* 9(1): 1–11. http://dx.doi.org/10.1038/s41467-018-05257-7.

Lawson, Daniel John, Garrett Hellenthal, Simon Myers, and Daniel Falush. 2012. "Inference of Population Structure Using Dense Haplotype Data." *PLoS Genetics* 8(1): 11–17.

Leathlobhair, Máire Ní, Angela R Perri, Evan K Irving-pease, and Kelsey E Witt. 2018. "The Evolutionary History of Dogs in the Americas." 85(July): 81–85.

Lev-Yadun, Simcha; Gopher, Avi; Abbo, Shalal. 2000. "The Cradle of Agriculture." *Science* 288(June).

Li, Chunxiang et al. 2011. "Ancient DNA Analysis of Desiccated Wheat Grains Excavated from a Bronze Age Cemetery in Xinjiang." *Journal of Archaeological Science* 38(1): 115–19. http://dx.doi.org/10.1016/j.jas.2010.08.016.

Li, Heng, and Richard Durbin. 2009. "Fast and Accurate Short Read Alignment with Burrows – Wheeler Transform." 25(14): 1754–60.

Li, Heng, and Jonathan Wren. 2014. "Toward Better Understanding of Artifacts in Variant Calling from High-Coverage Samples." *Bioinformatics* 30(20): 2843–51.

Librado, P., Khan, N., Fages, A. et al. 2021. "The Origins and Spread of Domestic Horses from the Western Eurasian Steppes." *Nature* 598.

Lindahl, T. 1993. "Instability and Decay of the Primary Structure of DNA." *Nature* 362: 709–15.

Ling, Hong Qing et al. 2018. "Genome Sequence of the Progenitor of Wheat A Subgenome Triticum Urartu." *Nature* 557(7705): 424–28.

Link, V. et al. 2017. "ATLAS: Analysis Tools for Low-Depth and Ancient Samples." *BioRxiv* 33(16): 1–7.

Liu, Haijun et al. 2011. "Psb27, a Transiently Associated Protein, Binds to the Chlorophyll Binding Protein CP43 in Photosystem II Assembly Intermediates." *Proceedings of the National Academy of Sciences of the United States of America* 108(45): 18536–41.

Lopes, Marta S. et al. 2015. "Exploiting Genetic Diversity from Landraces in Wheat Breeding for Adaptation to Climate Change." *Journal of Experimental Botany* 66(12): 3477–86.

Louwaars, Niels P. 2018. "Plant Breeding and Diversity: A Troubled Relationship?" *Euphytica* 214(7): 1–9. https://doi.org/10.1007/s10681-018-2192-5.

Lucas, Stuart J., Ayten Salantur, Selami Yazar, and Hikmet Budak. 2017. "High-Throughput SNP Genotyping of Modern and Wild Emmer Wheat for Yield and Root Morphology Using a Combined Association and Linkage Analysis." *Functional & Integrative Genomics* 17(6): 667–85.

http://link.springer.com/10.1007/s10142-017-0563-y.

Luo, M. C. et al. 2007. "The Structure of Wild and Domesticated Emmer Wheat Populations, Gene Flow between Them, and the Site of Emmer Domestication." *Theoretical and Applied Genetics* 114(6): 947–59.

Luo, Ming Cheng et al. 2017. "Genome Sequence of the Progenitor of the Wheat D Genome Aegilops Tauschii." *Nature* 551(7681): 498–502. http://dx.doi.org/10.1038/nature24486.

Lv, Liangjie et al. 2020. "Gene Co-Expression Network Analysis to Identify Critical Modules and Candidate Genes of Drought-Resistance in Wheat." *PLOS ONE* 15(8): e0236186. https://doi.org/10.1371/journal.pone.0236186.

Maccaferri, Marco et al. 2019. "Durum Wheat Genome Highlights Past Domestication Signatures and Future Improvement Targets." *Nature Genetics* 51(5): 885–95. http://dx.doi.org/10.1038/s41588-019-0381-3.

Maeda, Osamu et al. 2016. "Narrowing the Harvest: Increasing Sickle Investment and the Rise of Domesticated Cereal Agriculture in the Fertile Crescent." *Quaternary Science Reviews* 145: 226–37. http://dx.doi.org/10.1016/j.quascirev.2016.05.032.

Malinsky, Milan, Michael Matschiner, and Hannes Svardal. 2021. "Dsuite - Fast D-Statistics and Related Admixture Evidence from VCF Files." *Molecular Ecology Resources* 21(2): 584–95.

Mani, B. R. 2004. "Further Evidence on Kashmir Neolithic in the Light of Recent Excavations at Kanishkapura." *Journal of interdisciplinary studies in history and archaeology* 1.1: 137–42.

Mapuranga, Johannes et al. 2022. "Harnessing Genetic Resistance to Rusts in Wheat and Integrated Rust Management Methods to Develop More Durable Resistant Cultivars ." *Frontiers in Plant Science* 13. https://www.frontiersin.org/articles/10.3389/fpls.2022.951095.

Marciniak, Stephanie, and George H Perry. 2017. "Harnessing Ancient Genomes to Study the History of Human Adaptation." *Nature Publishing Group* 18(11): 659–74. http://dx.doi.org/10.1038/nrg.2017.65.

Marone, Daniela et al. 2021. "Importance of Landraces in Cereal Breeding for Stress Tolerance." *Plants* 10(7).

Martin, Simon. "Genomics General." https://github.com/simonhmartin/genomics_general.

Mascher, Martin et al. 2016. "Genomic Analysis of 6,000-Year-Old Cultivated Grain Illuminates the Domestication History of Barley." *Nature Genetics* 48(9): 1089–93. http://dx.doi.org/10.1038/ng.3611.

McCarthy, S. et al. 2016. "A Reference Panel of 64,976 Haplotypes for Genotype Imputation." *Nature Genetics* 48: 1279–83.

Mithen, Steven, Amy Richardson, and Bill Finlayson. 2023. "The Flow of Ideas: Shared Symbolism during the Neolithic Emergence in Southwest Asia: WF16 and Göbekli Tepe." *Antiquity* (October 2022): 1–21.

Mohammadi, Majid, Aghafakhr Mirlohi, Mohammad Mahdi Majidi, and Esmaeil Soleimani Kartalaei. 2021. "Emmer Wheat as a Source for Trait Improvement in Durum Wheat: A Study of General and Specific Combining Ability." *Euphytica* 217(4): 1–20. https://doi.org/10.1007/s10681-021-02796-x.

Morozova, Irina et al. 2016. "Toward High-Resolution Population Genomics Using

Archaeological Samples." *DNA Research* 23(4): 295–310.
N.I. Vavilov. "The Law of Homologous Series in Variation." *Journal of Genetics* 1922: 47–89.
Nave, Moran et al. 2019. "Wheat Domestication in Light of Haplotype Analyses of the Brittle Rachis 1 Genes (BTR1-A and BTR1-B)." *Plant Science* 285(May): 193–99.
Nei, M., and W. H. Li. 1979. "Mathematical Model for Studying Genetic Variation in Terms of Restriction Endonucleases." *Proceedings of the National Academy of Sciences of the United States of America* 76(10): 5269–73.
Nesbitt, M., and D. Samuel. 1996. "From Staple Crop to Extinction? The Archaeology and History of the Hulled Wheat." In *Hulled Wheats, Promoting the Conservation and Used of Underutilized and Neglected Crops.*, eds. S. Padulosi, K. Hammer, and J. Heller. Rome, 40–99.
Nielsen, Rasmus et al. 2017. "Tracing the Peopling of the World through Genomics." *Nature* 541.
Oliveira, Hugo R. et al. 2020. "Multiregional Origins of the Domesticated Tetraploid Wheats." *PLoS ONE* 15(1): 1–20. http://dx.doi.org/10.1371/journal.pone.0227148.
De Oliveira, Romain et al. 2020. "Structural Variations Affecting Genes and Transposable Elements of Chromosome 3B in Wheats." *Frontiers in Genetics* 11(August): 1–15.
Orlando, Ludovic et al. 2021. "Ancient DNA Analysis." *Nature Reviews Genetics*: 1–26.
Ortiz-Bobea, Ariel et al. 2021. "Anthropogenic Climate Change Has Slowed Global Agricultural Productivity Growth." *Nature Climate Change* 11(4): 306–12. http://dx.doi.org/10.1038/s41558-021-01000-1.
Özbaşaran, M., G. Duru, M. C. Stiner, and U. Esin. 2018. *The Early Settlement at Aşıklı Höyük: Essays in Honor of Ufuk Esin*. Ege Yayınları.
Ozkan, H. et al. 2005. "A Reconsideration of the Domestication Geography of Tetraploid Wheats." *Theoretical and Applied Genetics* 110: 1052– 1060.
Ozkan, H. et al. 2011. "Geographic Distribution and Domestication of Wild Emmer Wheat (Triticum Dicoccoides)." *Genetic Resources and Crop Evolution* 58: 11–53.
Ozkan, H., A. Brandolini, R. Schafer-Pregl, and F. Salamini. 2002. "AFLP Analysis of a Collection of Tetraploid Wheat Indicated the Origin of Emmer and Hard Wheat Domestication in South- Eastern Turkey." *Molecular Biology and Evolution* 19(10): 1797–1801. https://doi.org/10.1093/oxfordjournals.molbev.a004002.
Özkan, Hakan et al. 2011. "Geographic Distribution and Domestication of Wild Emmer Wheat (Triticum Dicoccoides)." *Genetic Resources and Crop Evolution* 58(1): 11–53.
Pagnotta, Mario Augusto, Linda Mondini, and Maroun Fandy Atallah. 2005. "Morphological and Molecular Characterization of Italian Emmer Wheat Accessions." *Euphytica* 146(1): 29–37. https://doi.org/10.1007/s10681-005-8607-0.
Palmer, Sarah A. et al. 2009. "Archaeogenetic Evidence of Ancient Nubian Barley Evolution from Six to Two-Row Indicates Local Adaptation." *PLoS ONE* 4(7): 2–8.
Palmer, Sarah A. et al. 2012. "Archaeogenomic Evidence of Punctuated Genome

Evolution in Gossypium." *Molecular Biology and Evolution* 29(8): 2031–38.
Pankin, Artem, Janine Altmüller, Christian Becker, and Maria von Korff. 2018. "Targeted Resequencing Reveals Genomic Signatures of Barley Domestication." *New Phytologist* 218(3): 1247–59.
Parducci, Laura et al. 2017. "Ancient Plant DNA in Lake Sediments." *New Phytologist* 214(3): 924–42.
Pedersen, Jakob Skou et al. 2014. "Genome-Wide Nucleosome Map and Cytosine Methylation Levels of an Ancient Human Genome." : 454–66.
Peleg, Zvi, Shahal Abbo, and Avi Gopher. 2022. "When Half Is More than the Whole: Wheat Domestication Syndrome Reconsidered." *Evolutionary Applications* 15(12): 2002–9.
Peltzer, Alexander et al. 2016. "EAGER: Efficient Ancient Genome Reconstruction." *Genome biology* 17: 60. http://dx.doi.org/10.1186/s13059-016-0918-z.
Peng, J., D. Sun, Y. Peng, and E. Nevo. 2013. "Gene Discovery in Triticum Dicoccoides, the Direct Progenitor of Cultivated Wheats." *Cereal Research Communications* 41(1): 1–22.
Peng, Junhua H., Dongfa Sun, and Eviatar Nevo. 2011. "Domestication Evolution, Genetics and Genomics in Wheat." *Molecular Breeding* 28(3): 281–301.
Peyrégne, S., and K. Prüfer. 2020. "Present-Day DNA Contamination in Ancient DNA Datasets." *BioEssays* 42(1–11).
Pickrell, Joseph K., and Jonathan K. Pritchard. 2012. "Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data." *PLoS Genetics* 8(11).
Pont, Caroline, Stefanie Wagner, Antoine Kremer, Ludovic Orlando, et al. 2019. "Paleogenomics: Reconstruction of Plant Evolutionary Trajectories from Modern and Ancient DNA." *Genome Biology* 20(1): 29. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1627-1.
Pont, Caroline, Thibault Leroy, Michael Seidel, and Alessandro Tondelli. 2019. "Tracing the Ancestry of Modern Bread Wheats." *Nature Genetics* 51: 905–11.
Price, Alkes L et al. 2008. "Long-Range LD Can Confound Genome Scans in Admixed Populations." *American journal of human genetics* 83(1): 132–39.
Prüfer, Kay, and Matthias Meyer. 2015. "Comment on 'Late Pleistocene Human Skeleton and MtDNA Link Paleoamericans and Modern Native Americans.'" *Science* 347(6224): 835a.
Przelomska, Natalia A.S., Chelsey G. Armstrong, and Logan Kistler. 2020. "Ancient Plant DNA as a Window Into the Cultural Heritage and Biodiversity of Our Food System." *Frontiers in Ecology and Evolution* 8(March): 1–8.
Przewieslik-Allen, Alexandra M. et al. 2021. "The Role of Gene Flow and Chromosomal Instability in Shaping the Bread Wheat Genome." *Nature Plants* 7(2): 172–83. http://dx.doi.org/10.1038/s41477-020-00845-2.
Purcell, S et al. 2007. "PLINK: A Toolset for Whole-Genome Association and Population-Based Linkage Analysis." *American Journal of Human Genetics* 81.
Purugganan, Michael D. 2019. "Evolutionary Insights into the Nature of Plant Domestication." *Current Biology* 29(14): R705–14. https://doi.org/10.1016/j.cub.2019.05.053.
Purugganan, Michael D., and Dorian Q. Fuller. 2009. "The Nature of Selection during

Plant Domestication." *Nature* 457(7231): 843–48. http://www.nature.com/doifinder/10.1038/nature07895.

Qian, Yuting et al. 2023. "The Role of Introgression During the Radiation of Endemic Fishes Adapted to Living at Extreme Altitudes in the Tibetan Plateau." *Molecular Biology and Evolution* 40(6): msad129. https://doi.org/10.1093/molbev/msad129.

R Core Team. 2021. "R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria." https://www.r-project.org/.

Rahman, Shanjida et al. 2023. "Characterizing Agronomic and Shoot Morphological Diversity across 263 Wild Emmer Wheat Accessions." *Agriculture (Switzerland)* 13(4).

Ramos-Madrigal, Jazmín et al. 2016. "Genome Sequence of a 5,310-Year-Old Maize Cob Provides Insights into the Early Stages of Maize Domestication." *Current Biology* 26(23): 3195–3201.

Ramos-Madrigal, Jazmín et al. 2019. "Palaeogenomic Insights into the Origins of French Grapevine Diversity." *Nature Plants* 5(6): 595–603. http://dx.doi.org/10.1038/s41477-019-0437-5.

Raudvere, Uku et al. 2019. "G:Profiler: A Web Server for Functional Enrichment Analysis and Conversions of Gene Lists (2019 Update)." *Nucleic Acids Research* 47(W1): W191–98.

Reich, David et al. 2010. "Genetic History of an Archaic Hominin Group from Denisova Cave in Siberia."

Reiter, Wolf-Dieter. 2002. "Biosynthesis and Properties of the Plant Cell Wall." *Current Opinion in Plant Biology* 5(6): 536–42. https://www.sciencedirect.com/science/article/pii/S1369526602003060.

Renaud, G., M. Schubert, S. Sawyer, and L. Orlando. 2019. "Authentication and Assessment of Contamination in Ancient DNA." In *Shapiro, B. et Al. Ancient DNA: Methods and Protocols*, Springer US, 163–194.

Renner, S. S. et al. 2019. "A 3500-Year-Old Leaf from a Pharaonic Tomb Reveals That New Kingdom Egyptians Were Cultivating Domesticated Watermelon." *bioRxiv*.

Riehl, Simone, Mohsen Zeidi, and Nicholas J. Conard. 2013. "Emergence of Agriculture in the Foothills of the Zagros Mountains of Iran." *Science* 341(6141): 65–67.

Rohland, Nadin et al. 2015. "Partial Uracil – DNA – Glycosylase Treatment for Screening of Ancient DNA." *Philosophical Transactions of the Royal Society B: Biological Sciences* 370(1660).

Russell, Joanne et al. 2016. "Exome Sequencing of Geographically Diverse Barley Landraces and Wild Relatives Gives Insights into Environmental Adaptation." *Nature Genetics* 48(9): 1024–30.

Sabeti, Pardis C. et al. 2007. "Genome-Wide Detection and Characterization of Positive Selection in Human Populations." *Nature* 449(7164): 913–18. http://www.nature.com/articles/nature06250.

Saitou, Naruya, and Masatoshi Nei. 1987. "The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees." *Molecular biology and evolution*

4(4): 406–25.
Salamini, F et al. 2004. "Comment on" AFLP Data and the Origins of Domesticated Crops"." *Genome* 47(3): 615–20.
Saleh, Maysoun M. 2020. "Stress Breeding of Neglected Tetraploid Primitive Wheat (Triticum Dicoccum, Triticum Carthlicum and Triticum Polonicum)." *Current Botany* 11: 99–110.
Salunkhe, Arvindkumar et al. 2013. "Molecular Genetic Diversity Analysis in Emmer Wheat (Triticum Dicoccon Schrank) from India." *Genetic Resources and Crop Evolution*: 165–74.
Sanket, Jain. 2023. "The Indian Farmers Battling Climate Change With 10,000-Year-Old Emmer Wheat." *foodunfolded*. https://www.foodunfolded.com/article/the-indian-farmers-battling-climate-change-with-10000-year-old-emmer-wheat.
Der Sarkissian, C. et al. 2014. "Ancient Genomics." *Philosophical Transactions of the Royal Society B: Biological Sciences* 370(1660): 20130387–20130387. http://rstb.royalsocietypublishing.org/cgi/doi/10.1098/rstb.2013.0387.
Schubert, Mikkel et al. 2014. "Characterization of Ancient and Modern Genomes by SNP Detection and Phylogenomic and Metagenomic Analysis Using PALEOMIX." *Nature Protocols* 9(5): 1056–82.
Scott, Michael F. et al. 2019. "A 3,000-Year-Old Egyptian Emmer Wheat Genome Reveals Dispersal and Domestication History." *Nature Plants* 5(11): 1120–28. http://www.nature.com/articles/s41477-019-0534-5.
Sharma, Jyoti S. et al. 2019. "Genetic Analysis of Threshability and Other Spike Traits in the Evolution of Cultivated Emmer to Fully Domesticated Durum Wheat." *Molecular Genetics and Genomics* 294(3): 757–71. http://dx.doi.org/10.1007/s00438-019-01544-0.
Shewry, Peter R., Till K. Pellny, and Alison Lovegrove. 2016. "Is Modern Wheat Bad for Health?" *Nature Plants* 2(7).
Shitsukawa, Naoki, Hiroko Kinjo, Shigeo Takumi, and Koji Murai. 2009. "Heterochronic Development of the Floret Meristem Determines Grain Number per Spikelet in Diploid, Tetraploid and Hexaploid Wheats." *Annals of Botany* 104(2): 243–51.
Simons, Kristin J. et al. 2006. "Molecular Characterization of the Major Wheat Domestication Gene Q." *Genetics* 172(1): 547–55.
Singh, Jugpreet, and Esther van der Knaap. 2022. "Unintended Consequences of Plant Domestication." *Plant and Cell Physiology* 63(11): 1573–83. https://doi.org/10.1093/pcp/pcac083.
Smith, Oliver et al. 2019. "A Domestication History of Dynamic Adaptation and Genomic Deterioration in Sorghum." *Nature Plants* 5(4): 369–79. http://dx.doi.org/10.1038/s41477-019-0397-9.
Snir, Ainit et al. 2015. "The Origin of Cultivation and Proto-Weeds, Long before Neolithic Farming." *PLoS ONE* 10(7): 1–12.
Spengler, Robert N III. 2020. "Anthropogenic Seed Dispersal : Rethinking the Origins of Plant Domestication." *Trends in Plant Science* 25(4): 340–48. https://doi.org/10.1016/j.tplants.2020.01.005.
Spyrou, Maria A, Kirsten I Bos, Alexander Herbig, and Johannes Krause. 2019. "Ancient Pathogen Genomics as an Emerging Tool for Infectious Disease

Research." *Nature Reviews Genetics* 20(JUNE). http://dx.doi.org/10.1038/s41576-019-0119-1.

Stephan, Wolfgang. 2016. "Signatures of Positive Selection: From Selective Sweeps at Individual Loci to Subtle Allele Frequency Changes in Polygenic Adaptation." *Molecular Ecology* 25(1): 79–88.

Stevens, Chris J et al. 2016. "Between China and South Asia : A Middle Asian Corridor of Crop Dispersal and Agricultural Innovation in the Bronze Age." : 31–34.

Streit Krug, Aubrey, Emily B. M. Drummond, David L. Van Tassel, and Emily J. Warschefsky. 2023. "The next Era of Crop Domestication Starts Now." *Proceedings of the National Academy of Sciences* 120. https://www.pnas.org/doi/epdf/10.1073/pnas.2205769120.

Swarts, Kelly et al. 2017. "Genomic Estimation of Complex Traits Reveals Ancient Maize Adaptation to Temperate North America." 515(August): 512–15.

Swarup, Shilpa et al. 2021. "Genetic Diversity Is Indispensable for Plant Breeding to Improve Crops." *Crop Science* 61(2): 839–52.

Szpiech, Zachary A. 2022. "Selscan 2.0: Scanning for Sweeps in Unphased Data." *bioRxiv*: 2021.10.22.465497.

Szpiech, Zachary A., and Ryan D. Hernandez. 2014. "Selscan: An Efficient Multithreaded Program to Perform EHH-Based Scans for Positive Selection." *Molecular Biology and Evolution* 31(10): 2824–27.

Tajima, F. 1989. "Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism." *Genetics* 123(3): 585–95.

Tanksley, Steven D, and Susan R McCouch. 1997. "Seed Banks and Molecular Maps: Unlocking Genetic Potential from the Wild." *Science* 277(5329): 1063–66. https://doi.org/10.1126/science.277.5329.1063.

Tetlow, Ian J, and Michael J Emes. 2014. "A Review of Starch-Branching Enzymes and Their Role in Amylopectin Biosynthesis." *IUBMB Life* 66(8): 546–58. https://doi.org/10.1002/iub.1297.

The Arabidopsis Genome Initiative. 2000. "Analysis of the Genome Sequence of the Flowering Plant Arabidopsis Thaliana." *Nature* 408(6814): 796–815. https://doi.org/10.1038/35048692.

Tirnaz, Soodeh et al. 2022. "Application of Crop Wild Relatives in Modern Breeding: An Overview of Resources, Experimental and Computational Methodologies." *Frontiers in Plant Science* 13(November): 1–17.

Trucchi, Emiliano et al. 2021. "Ancient Genomes Reveal Early Andean Farmers Selected Common Beans While Preserving Diversity." *Nature Plants* 7(2): 123–28.

Tryfona, Theodora et al. 2014. "Characterisation of FUT4 and FUT6 α-(1 → 2)-Fucosyltransferases Reveals That Absence of Root Arabinogalactan Fucosylation Increases Arabidopsis Root Growth Salt Sensitivity." *PloS one* 9(3): e93291.

Ulaş, Burhan, and Girolamo Fiorentino. 2021. "Recent Attestations of ' New ' Glume Wheat in Turkey : A Reassessment of Its Role in the Reconstruction of Neolithic Agriculture." *Vegetation History and Archaeobotany* 30(5): 685–701. https://doi.org/10.1007/s00334-020-00807-w.

Vavilov, N. I. 1926. *The Origin, Variation, Immunity and Breeding of Cultivated Plants (K. S.*

Chester Trans.). New York, NY: Ronald Press.

Vavilov, N.I., M. I. Vavylov, N. Í. Vavílov, and V. F. Dorofeev. 1992. *Origin and Geography of Cultivated Plants*. Cambridge: Cambridge University of Press.

Vavilov, NI. 1926. *Studies on the Origin of Cultivated Plants*. Institut Botanique Applique´ et d'Amelioration des Plantes.

Venkateswaran, Kamala, M Elangovan, and N Sivaraj. 2019. "Chapter 2 - Origin, Domestication and Diffusion of Sorghum Bicolor." In *Woodhead Publishing Series in Food Science, Technology and Nutrition*, eds. C Aruna, K.B.R.S. Visarada, B Venkatesh Bhat, and Vilas A B T - Breeding Sorghum for Diverse End Uses Tonapi. Woodhead Publishing, 15–31. https://www.sciencedirect.com/science/article/pii/B9780081018798000024.

Verdugo, Marta Pereira et al. 2019. "Ancient Cattle Genomics, Origins, and Rapid Turnover in the Fertile Crescent." 176(July): 173–76.

Voss-Fels, Kai P., Andreas Stahl, and Lee T. Hickey. 2019. "Q&A: Modern Crop Breeding for Future Food Security 07 Agricultural and Veterinary Sciences 0703 Crop and Pasture Production 06 Biological Sciences 0607 Plant Biology 06 Biological Sciences 0604 Genetics." *BMC Biology* 17(1): 1–7.

Wagenaar, E. B. 1966. "Studies on the Genome Constitution of Triticum Timopheevi Zhuk . II. The T . Timopheevi Complex and Its Origin." *Society for the Study of Evolution* 20(2): 150–64. https://www.jstor.org/stable/2406569.

Wales, Nathan et al. 2019. "Ancient DNA Reveals the Timing and Persistence of Organellar Genetic Bottlenecks over 3000 Years of Sunflower Domestication and Improvement." *Evolutionary Applications* 12: 1–16. http://doi.wiley.com/10.1111/eva.12594.

Walkowiak, Sean et al. 2020. "Multiple Wheat Genomes Reveal Global Variation in Modern Breeding." *Nature* 588(7837): 277–83. http://dx.doi.org/10.1038/s41586-020-2961-x.

Wang, Qing Long, Juan Hua Chen, Ning Yu He, and Fang Qing Guo. 2018. "Metabolic Reprogramming in Chloroplasts under Heat Stress in Plants." *International Journal of Molecular Sciences* 19(3): 9–11.

Wang, Shichen et al. 2014. "Characterization of Polyploid Wheat Genomic Diversity Using a High-Density 90 000 Single Nucleotide Polymorphism Array." *Plant Biotechnology Journal* 12(6): 787–96.

Wang, Zihao et al. 2022. "Dispersed Emergence and Protracted Domestication of Polyploid Wheat Uncovered by Mosaic Ancestral Haploblock Inference." *Nature Communications* 13(2022): 1–14.

Wangkumhang, Pongsakorn, Matthew Greenfield, and Garrett Hellenthal. 2022. "An Efficient Method to Identify, Date, and Describe Admixture Events Using Haplotype Information." *Genome Research* 32(8): 1553–64.

Warinner, Christina et al. 2014. "Pathogens and Host Immunity in the Ancient Human Oral Cavity." *Nature Publishing Group* 46(4).

Weiss, Ehud, Mordechai E Kislev, Orit Simchoni, and Dani Nadel. 2004. "SMALL-GRAINED WILD GRASSES AS STAPLE FOOD AT THE 23 000-YEAR-OLD SITE OF OHALO II, ISRAEL." *Economic Botany* 58.

Wendrich, W. Z., and R. T. J. Cappers. 2005. "Egypt's Earliest Granaries: Evidence from the Fayum." *Egyptian Archaeology* 27: 12–15.

Weyrich, Laura S., Keith Dobney, and Alan Cooper. 2015. "Ancient DNA Analysis of Dental Calculus." *Journal of Human Evolution* 79: 119–24. http://dx.doi.org/10.1016/j.jhevol.2014.06.018.

Wicker, Thomas et al. 2011. "Frequent Gene Movement and Pseudogene Evolution Is Common to the Large and Complex Genomes of Wheat, Barley, and Their Relatives." *Plant Cell* 23(5): 1706–18.

Wu, Xiyan et al. 2019. "Phylogenetic and Population Structural Inference from Genomic Ancestry Maintained in Present-Day Common Wheat Chinese Landraces." *Plant Journal* 99(2): 201–15.

Xu, Jie et al. 2014. "Identification of Candidate Genes for Drought Tolerance by Whole-Genome Resequencing in Maize." *BMC Plant Biology* 14(1): 83. https://doi.org/10.1186/1471-2229-14-83.

Yang, Melinda A et al. 2020. "Ancient DNA Indicates Human Population Shifts and Admixture in Northern and Southern China." *Science* 288(July): 282–88.

Yoshida, K. et al. 2013. "The Rise and Fall of the Phytophthora Infestans Lineage That Triggered the Irish Potato Famine." *Elife* 2: 1–25.

Zaharieva, Maria et al. 2010. "Cultivated Emmer Wheat (Triticum Dicoccon Schrank), an Old Crop with Promising Future: A Review." *Genetic Resources and Crop Evolution* 57(6): 937–62.

Zeder, Melinda A. 2016. "Domestication as a Model System for Niche Construction Theory." *Evolutionary Ecology* 30(2): 325–48.

VAN ZEIST, W., and G. J. DE ROLLER. 2003. "The Çayönü Archaeobotanical Record." *Reports on archaeobotanical studies in the Old World*: 143–66.

Zeven, Anton C. 1998. "Landraces: A Review of Definitions and Classifications." *Euphytica* 104: 127–39.

Zhang, Hengyou et al. 2017. "Back into the Wild—Apply Untapped Genetic Diversity of Wild Relatives for Crop Improvement." *Evolutionary Applications* 10(1): 5–24.

Zhang, Li et al. 2019. "An Important Role of L-Fucose Biosynthesis and Protein Fucosylation Genes in Arabidopsis Immunity." *The New phytologist* 222(2): 981–94.

Zhao, Chuang et al. 2017. "Temperature Increase Reduces Global Yields of Major Crops in Four Independent Estimates." *Proceedings of the National Academy of Sciences of the United States of America* 114(35): 9326–31.

Zhao, Xuebo et al. 2023. "Population Genomics Unravels the Holocene History of Bread Wheat and Its Relatives." *Nature Plants* 9(March).

Zhou, Xinying et al. 2020. "5,200-Year-Old Cereal Grains from the Eastern Altai Mountains Redate the Trans-Eurasian Crop Exchange." *Nature Plants* 6(2): 78–87. http://dx.doi.org/10.1038/s41477-019-0581-y.

Zhou, Yao et al. 2020. "Triticum Population Sequencing Provides Insights into Wheat Adaptation." *Nature Genetics* 52(12): 1412–22. http://dx.doi.org/10.1038/s41588-020-00722-w.

Zhu, T. et al. 2019. "Improved Genome Sequence of Wild Emmer Wheat Zavitan with the Aid of Optical Maps." *G3: Genes, Genomes, Genetics* 9(3): 619–24.

Zohary, D. 2013. "Domestication of Crop Plants." *Encyclopedia of Biodiversity: Second Edition* (March 2018): 657–64.

Chapter 7

ANNEX

7.1. Supplementary material paper 1: "Genomic analysis of emmer wheat shows a complex history with two distinct domestic groups and evidence of differential hybridization with wild emmer from the western Fertile Crescent"

Table S1: passport information for all the samples analyzed in the study, including accession number and sample name from the reference publication. The field "group" is assigned based on the DAPC analyses.

SAMPLE	ACCESSION	COUNTRY	LATITUDE	LONGITUDE	PUBLICATION	OLD SAMPLE	GROUP	STATUS
W-ISR1	PI 466970	Israel	32.60	35.00	Zhou et al., 2020	B025	W-SL	WILD
W-ISR10	PI 466986	Israel	31.80	35.03	Zhou et al., 2020	B047	W-SL	WILD
W-ISR11	PI 471012	Israel	32.97	35.53	Zhou et al., 2020	B048	W-SL	WILD
W-ISR2	PI 471062	Israel	31.95	35.33	Zhou et al., 2020	B026	W-SL	WILD
W-ISR3	PI 466959	Israel	32.60	35.00	Zhou et al., 2020	B028	W-SL	WILD
W-ISR4	PI 428107	Israel	32.97	35.53	Zhou et al., 2020	B029	W-SL	WILD
W-ISR6	PI 470998	Israel	32.97	35.53	Zhou et al., 2020	B036	W-SL	WILD
W-ISR7	PI 466980	Israel	32.60	35.00	Zhou et al., 2020	B039	W-SL	WILD
W-ISR8	PI 466987	Israel	31.80	35.03	Zhou et al., 2020	B040	W-SL	WILD
W-ISR9	PI 470991	Israel	32.97	35.53	Zhou et al., 2020	B044	W-SL	WILD
W-LBN1	PI 428138	Lebanon	33.52	35.87	Zhou et al., 2020	B031	W-SL	WILD
W-LBN2	PI 428127	Lebanon	33.52	35.87	Zhou et al., 2020	B034	W-SL	WILD
W-LBN3	TRI 11505	Lebanon	NA	NA	Zhou et al., 2020	B035	W-SL	WILD
W-LBN4	PI 538702	Lebanon	33.50	35.84	Zhou et al., 2020	B042	W-SL	WILD
W-SYR1	PI 466933	Syria	32.99	35.69	Zhou et al., 2020	B024	W-SL	WILD
W-SYR2	PI 487254	Syria	32.87	36.03	Zhou et al., 2020	B027	W-SL	WILD
W-SYR3	PI 466943	Syria	32.99	35.69	Zhou et al., 2020	B030	W-SL	WILD
W-SYR4	PI 487260	Syria	32.65	36.79	Zhou et al., 2020	B037	W-SL	WILD
W-SYR5	PI 466934	Syria	32.99	35.69	Zhou et al., 2020	B049	W-SL	WILD
W-SYR6	PI 466941	Syria	32.99	35.69	Zhou et al., 2020	B052	W-SL	WILD
W-IRN1	PI 428016	Iran	34.37	46.10	Zhou et al., 2020	B023	W-NL	WILD
W-TUR3	PI 428041	Turkey	37.88	39.87	Zhou et al., 2020	B033	W-NL	WILD
W-TUR4	PI 428071	Turkey	37.87	39.88	Zhou et al., 2020	B038	W-NL	WILD
W-TUR5	PI 538630	Turkey	37.88	39.87	Zhou et al., 2020	B041	W-NL	WILD
W-TUR6	PI 428029	Turkey	37.88	39.87	Zhou et al., 2020	B045	W-NL	WILD
W-TUR6	PI 428020	Turkey	37.88	39.87	Zhou et al., 2020	B050	W-NL	WILD
W-TUR7	PI 428082	Turkey	37.78	39.77	Zhou et al., 2020	B046	W-NL	WILD
D-ETH1	PI 479958	Ethiopia	9.20	38.60	Zhou et al., 2020	B063	D-SE	DOMESTIC
D-ETH10	CItr 7966	Ethiopia	11.50	40.00	Zhou et al., 2020	B077	D-SE	DOMESTIC
D-ETH11	PI 479968	Ethiopia	9.20	38.60	Zhou et al., 2020	B079	D-SE	DOMESTIC
D-ETH12	PI 480465	Ethiopia	9.12	38.38	Zhou et al., 2020	B082	D-SE	DOMESTIC
D-ETH13	PI 58788	Ethiopia	8.60	39.12	Zhou et al., 2020	B088	D-SE	DOMESTIC
D-ETH2	PI 197492	Ethiopia	9.32	42.12	Zhou et al., 2020	B065	D-SE	DOMESTIC
D-ETH3	CItr 14824	Ethiopia	13.48	39.55	Zhou et al., 2020	B066	D-SE	DOMESTIC
D-ETH4	PI 197488	Ethiopia	9.32	42.12	Zhou et al., 2020	B069	D-SE	DOMESTIC
D-ETH5	PI 534275	Ethiopia	9.17	38.82	Zhou et al., 2020	B070	D-SE	DOMESTIC
D-ETH6	CItr 14916	Ethiopia	8.87	38.78	Zhou et al., 2020	B071	D-SE	DOMESTIC
D-ETH7	PI 197481	Ethiopia	9.32	42.12	Zhou et al., 2020	B073	D-SE	DOMESTIC
D-ETH8	PI 196905	Ethiopia	8.73	38.98	Zhou et al., 2020	B074	D-SE	DOMESTIC
D-ETH9	PI 197484	Ethiopia	9.32	42.12	Zhou et al., 2020	B075	D-SE	DOMESTIC
D-IND1	PI 322232	India	28.61	77.21	Avni et al., 2017	DE2	D-SE	DOMESTIC
D-OMN1	PI 532305	Oman	23.83	56.33	Zhou et al., 2020	B067	D-SE	DOMESTIC
D-OMN2	PI 532306	Oman	23.17	57.67	Zhou et al., 2020	B068	D-SE	DOMESTIC
D-OMN3	PI 532302	Oman	21.51	55.92	Avni et al., 2017	DE1	D-SE	DOMESTIC
D-TUR1	PI 319868	Turkey	38.90	41.19	Avni et al., 2017	DE3	D-SE	DOMESTIC
D-TUR2	PI 319869	Turkey	39.51	41.95	Avni et al., 2017	DE4	D-SE	DOMESTIC
D-ARM1	PI 499973	Armenia	40.18	44.50	Zhou et al., 2020	B090	D-NW	DOMESTIC
D-ESP1	PI 355484	Spain	43.37	-5.83	Zhou et al., 2020	B064	D-NW	DOMESTIC
D-IRN1	PI 626391	Iran	32.68	51.68	Zhou et al., 2020	B081	D-NW	DOMESTIC
D-IRN2	PI 624908	Iran	36.23	46.27	Zhou et al., 2020	B089	D-NW	DOMESTIC
D-IRN3	PI 624904	Iran	36.23	46.27	Zhou et al., 2020	B091	D-NW	DOMESTIC
D-MNE1	PI 362696	Montenegro	42.35	19.30	Zhou et al., 2020	B083	D-NW	DOMESTIC
D-RUS1	PI 94668	Russia	42.00	47.00	Zhou et al., 2020	B084	D-NW	DOMESTIC
D-RUS2	PI 254190	Russia	42.00	47.00	Zhou et al., 2020	B085	D-NW	DOMESTIC
D-RUS3	PI 349045	Russia	43.00	47.00	Zhou et al., 2020	B086	D-NW	DOMESTIC
D-RUS4	PI 254191	Russia	42.00	47.00	Zhou et al., 2020	B087	D-NW	DOMESTIC
D-SRB1	PI 254192	Serbia	44.83	20.50	Zhou et al., 2020	B072	D-NW	DOMESTIC
D-UK1	PI 266842	United Kingdo	53.00	-2.00	Zhou et al., 2020	B076	D-NW	DOMESTIC
D-USA1	CItr 3686	America	46.00	-94.00	Zhou et al., 2020	B078	D-NW	DOMESTIC
UC10164	UC10164	Egypt - Heme	26.92	31.47	Scott et al., 2019	UC10164	ANCIENT	DOMESTIC

Table S2: Patterson's D (ABBA-BABA), all tested triplets. Significant results are shown in bold. As per D-suite, the populations are always ordered to infer gene flow between P2 and P3 (D>0)

P1	P2	P3	Dstatistic	Z-score	p-value
W-NL	D-SE	W-SL	0.05362	1.5436	0.12268
D-NW	W-NL	W-SL	**0.05825**	**4.4092**	**1.04E-05**
D-NW	D-SE	W-SL	**0.10672**	**3.1587**	**0.0015847**
D-SE	D-NW	W-NL	0.07405	2.0327	0.0420823

Fig S1: Principal Component Analysis of the modern (whole genome) dataset, prior to group determination: two outliers are identified and excluded from further analysis, while major groups are identified. The first two PCs explain a good amount of diversity in the dataset (15.3 % PC1 and 9.94% PC2). In line with previous studies, PC1 clearly divides Southern Levant from all domestic samples, while the Northern Levant wild group appears to be closer to them. On the other hand, PC2 divides the dataset in three groups: African domestic (Ethiopia and Oman), wild emmer from Southern Levant (W-SL), and a third group comprised by wild emmer from Northern Levant (W-NL) and domestic samples from Iran, Turkey, Europe, Caucasus and Balkans.

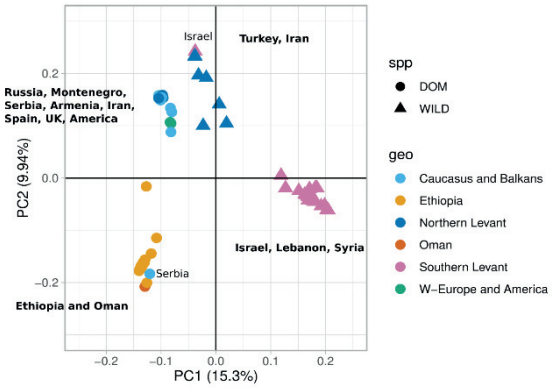


Fig S2: xValDAPC showing the best number of PCs to retain for DAPC analysis, based on the highest proportion of successful outcome prediction (0-1).

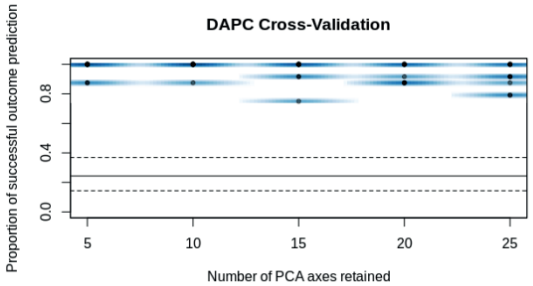


Fig S3: Admixture plots from K=2 to K=6, modern whole genome dataset

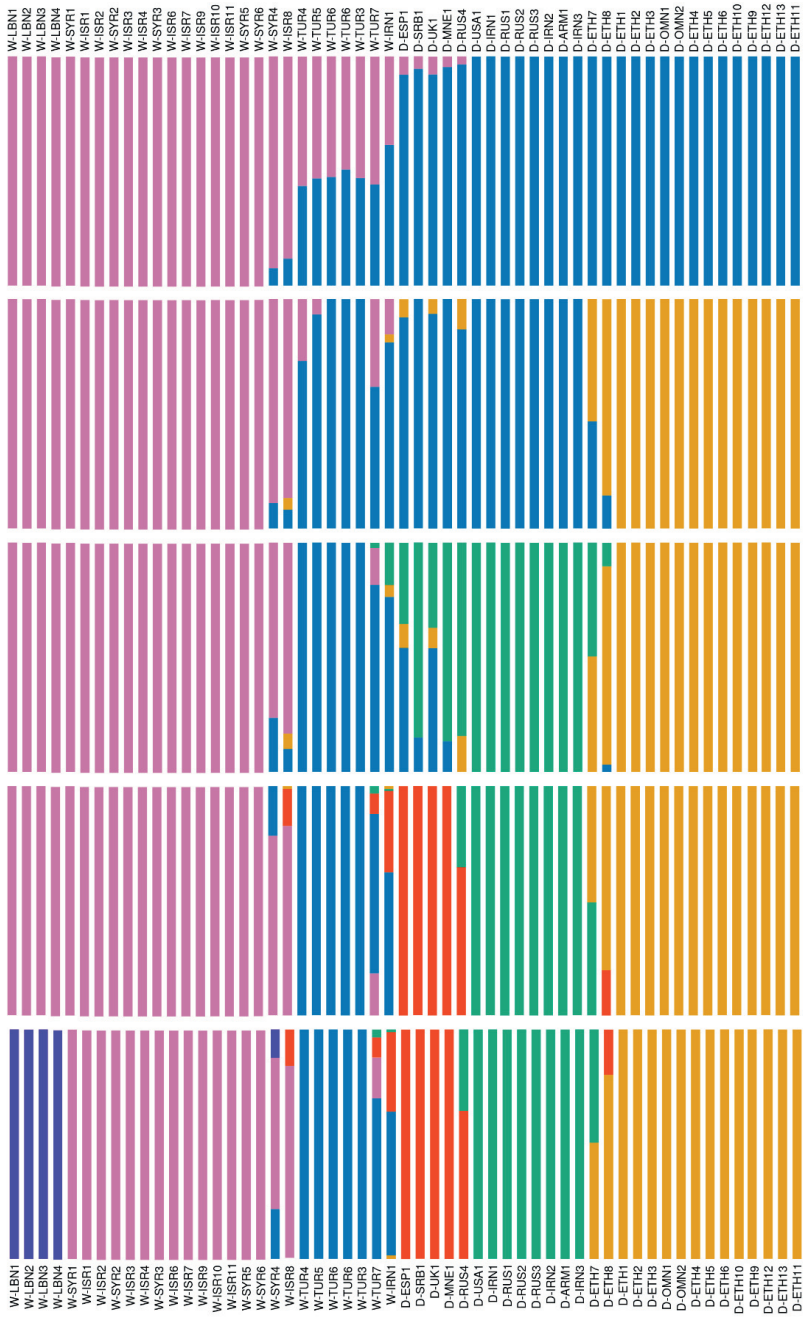


Fig S4: Neighbor Joining analysis with 100 bootstrap replicates of the modern (whole genome) dataset. Bootstrap values are not shown (all nodes have bootstrap more than 90).

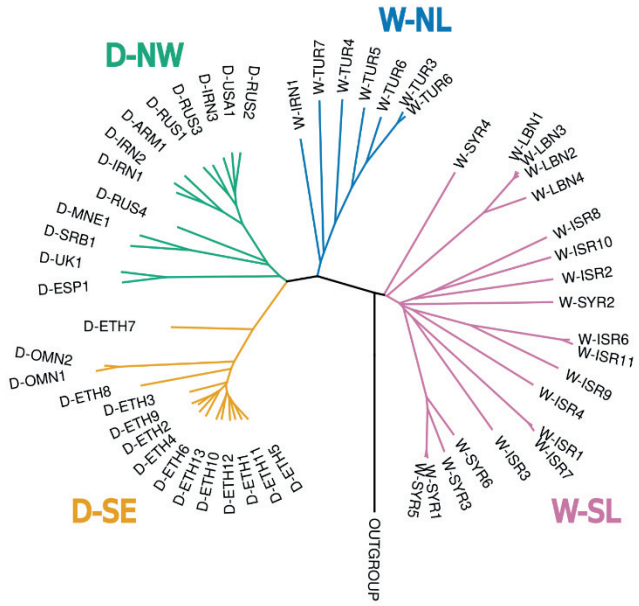
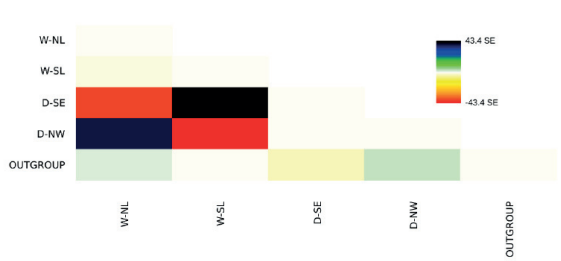
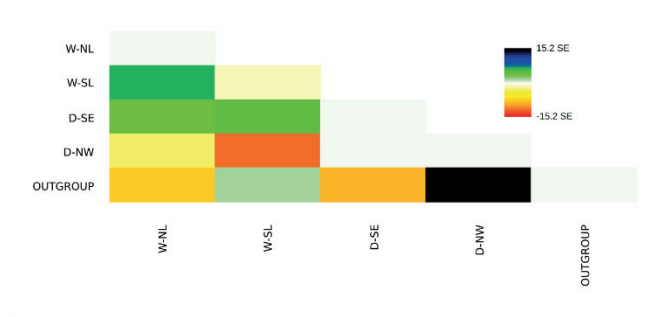


Fig S5 (a-d): TreeMix residuals. The residual covariance between pairs of populations is divided by the average standard error across pairs. Residuals above 0 (green to black) represent populations that are more closely related to each other in the data than in the best-fit tree and thus candidate for admixture events. Values below 0 (yellow to red) represent populations that appear to be more closely related in the best fit-tree than in the data. (SE= standard error).

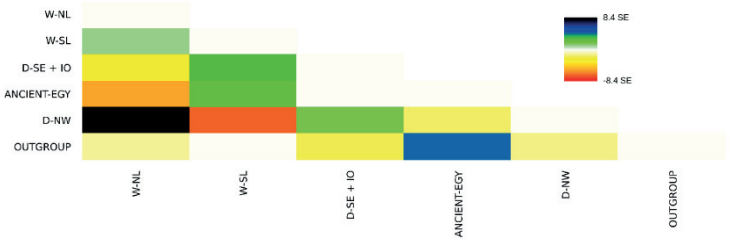
5a) Main (whole genome) dataset 0 edges of migration



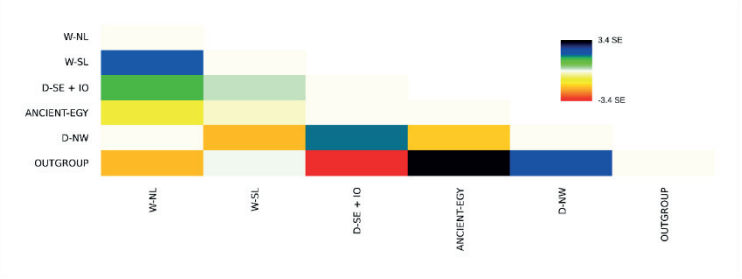
5b) Main (whole genome) dataset with 1 edge of migration



5c) Extended dataset (whole genome + exome capture + ancient) with 0 edges of migration



5d) Extended dataset (wgs + exome capture + ancient) with 1 edge of migration



7.2. Supplementary material for paper 2: "Wild emmer contribution in wheat domestication and adaptation to new environments"

S. Table 1: passport information of the samples included in the study: SEE SUPP. TABLE 1 IN SECTION 7.1 – only samples from Zhou et al., 2020

S. Table 2: Contribution of WSL to domestication; genomic windows showing closer relationship to WSL than to WNL for both domestic populations, and close relatedness between domestics, according to Dxy values

scaffold	start	end	sites	Dxy WSL-DSE	Dxy WSL-DNW	Dxy WNL-DSE	Dxy WNL-DNW	Dxy DSE-DNW
1A	22000001	24000000	3536	0.1738	0.1943	0.1773	0.1997	0.0467
1A	338000001	340000000	2111	0.1897	0.1921	0.2117	0.2123	0.0242
1A	339000001	341000000	2121	0.1944	0.1976	0.2182	0.2204	0.0257
1A	340000001	342000000	2104	0.1906	0.1977	0.2014	0.2072	0.0241
1A	393000001	395000000	2354	0.2094	0.2072	0.278	0.2761	0.0276
1A	394000001	396000000	2608	0.2107	0.2156	0.2429	0.2473	0.0285
1B	57000001	59000000	4216	0.167	0.1667	0.1744	0.1742	0.0076
1B	58000001	60000000	3719	0.1354	0.133	0.2577	0.2554	0.0092
1B	59000001	61000000	4043	0.1686	0.1672	0.2716	0.2702	0.0081
1B	80000001	82000000	2871	0.1938	0.2033	0.229	0.2386	0.032
1B	81000001	83000000	2920	0.1829	0.1908	0.2289	0.2404	0.0335
1B	82000001	84000000	2459	0.1699	0.1772	0.1902	0.2024	0.0413
1B	88000001	90000000	3645	0.1649	0.1636	0.2854	0.2835	0.0178
1B	89000001	91000000	3911	0.1974	0.1987	0.2331	0.2355	0.0246
1B	90000001	92000000	5292	0.2038	0.2125	0.2319	0.2393	0.0357
1B	91000001	93000000	6809	0.1887	0.2048	0.2486	0.2626	0.0436
1B	92000001	94000000	4566	0.19	0.2107	0.2297	0.2509	0.0491
1B	95000001	97000000	523	0.227	0.2327	0.2341	0.2369	0.0264
1B	100000001	102000000	3525	0.1408	0.1399	0.1502	0.1504	0.019

1B	101000001	103000000	2256	0.1423	0.1408	0.1541	0.153	0.0192
1B	208000001	210000000	1095	0.1314	0.1451	0.1687	0.184	0.0498
1B	224000001	226000000	792	0.1574	0.1627	0.183	0.1913	0.0482
1B	225000001	227000000	841	0.1646	0.164	0.2051	0.2049	0.0437
1B	242000001	244000000	537	0.1452	0.1428	0.2069	0.205	0.0463
1B	247000001	249000000	849	0.1437	0.1575	0.164	0.179	0.0475
1B	248000001	250000000	789	0.1485	0.1574	0.1739	0.1831	0.0478
2A	93000001	95000000	3006	0.2107	0.2088	0.2341	0.2328	0.0153
2A	94000001	96000000	2061	0.1928	0.206	0.2456	0.2576	0.0338
2A	399000001	401000000	111	0.0955	0.0969	0.3475	0.3489	0.0014
2A	745000001	747000000	2078	0.201	0.2034	0.2212	0.2236	0.0079
2A	746000001	748000000	2382	0.1867	0.1934	0.2228	0.2262	0.0139
2A	747000001	749000000	2165	0.1746	0.1812	0.2152	0.219	0.021
2A	748000001	750000000	1761	0.1851	0.1853	0.1884	0.1896	0.0226
2B	84000001	86000000	3991	0.2281	0.2279	0.2609	0.2606	0.0081
2B	85000001	87000000	3068	0.1872	0.1759	0.2328	0.2055	0.0413
2B	186000001	188000000	2652	0.198	0.2058	0.2182	0.2288	0.0346
2B	192000001	194000000	3669	0.1881	0.1996	0.1963	0.2088	0.0472
2B	557000001	559000000	3002	0.1723	0.1704	0.2055	0.194	0.0371
2B	558000001	560000000	2664	0.1553	0.1541	0.2691	0.2683	0.0164
2B	559000001	561000000	2332	0.1848	0.183	0.2699	0.2674	0.0203
2B	560000001	562000000	2522	0.2316	0.2359	0.2658	0.2587	0.0417
2B	591000001	593000000	3438	0.1886	0.1874	0.2084	0.2085	0.0227
3A	492000001	494000000	3029	0.197	0.2147	0.2162	0.2173	0.0497
3A	595000001	597000000	3484	0.2045	0.2022	0.2086	0.2061	0.0128
3A	596000001	598000000	3260	0.2205	0.2171	0.2374	0.2339	0.0184
3B	83000001	85000000	2819	0.1976	0.1943	0.2113	0.2079	0.0067
3B	84000001	86000000	2978	0.2111	0.2083	0.236	0.2332	0.0051
3B	85000001	87000000	3391	0.2165	0.2157	0.2187	0.2179	0.0065
3B	86000001	88000000	2930	0.2205	0.2183	0.2281	0.2258	0.0097
3B	87000001	89000000	2311	0.2194	0.2157	0.2411	0.2371	0.01

3B	88000001	90000000	2257	0.2071	0.2034	0.2416	0.2378	0.0082
3B	89000001	91000000	1795	0.2016	0.2073	0.2484	0.2373	0.0215
3B	96000001	98000000	2240	0.1837	0.1832	0.249	0.2485	0.0063
3B	97000001	99000000	2381	0.1926	0.2053	0.2561	0.2513	0.0232
3B	102000001	104000000	3191	0.153	0.1608	0.3615	0.3517	0.0299
3B	103000001	105000000	2895	0.1987	0.1977	0.3776	0.3764	0.0094
3B	104000001	106000000	2285	0.2319	0.2381	0.2991	0.305	0.0209
3B	105000001	107000000	2077	0.1803	0.1874	0.2467	0.2533	0.0451
3B	107000001	109000000	3163	0.1933	0.1977	0.2005	0.2066	0.0498
3B	112000001	114000000	2438	0.1564	0.1821	0.2709	0.2908	0.0465
3B	113000001	115000000	3185	0.1962	0.222	0.2139	0.2432	0.0475
3B	118000001	120000000	2458	0.1097	0.1434	0.3283	0.347	0.0485
3B	119000001	121000000	2850	0.1347	0.1539	0.296	0.3152	0.0436
3B	121000001	123000000	3307	0.197	0.2087	0.2537	0.2584	0.0323
3B	122000001	124000000	3287	0.1902	0.1994	0.2347	0.2368	0.0265
3B	123000001	125000000	1965	0.1724	0.1885	0.1909	0.199	0.036
3B	641000001	643000000	2518	0.2287	0.2476	0.2398	0.2619	0.0457
3B	671000001	673000000	1533	0.166	0.1703	0.3673	0.3693	0.0137
3B	672000001	674000000	2712	0.1678	0.1762	0.242	0.2426	0.0282
3B	819000001	821000000	2116	0.1629	0.1695	0.2428	0.2559	0.0312
3B	820000001	822000000	1730	0.1778	0.185	0.2293	0.2388	0.0304
4A	41000001	43000000	3357	0.1845	0.1879	0.2476	0.2528	0.0387
4A	42000001	44000000	2739	0.1991	0.1931	0.2635	0.2592	0.0319
4A	188000001	190000000	834	0.3249	0.3205	0.3434	0.3392	0.0426
4A	189000001	191000000	845	0.3413	0.336	0.3435	0.3382	0.0446
4A	209000001	211000000	858	0.3709	0.3599	0.383	0.3719	0.0439
4A	214000001	216000000	819	0.337	0.3347	0.3398	0.3367	0.0403
4A	215000001	217000000	973	0.3532	0.3503	0.3627	0.3595	0.0422
4A	235000001	237000000	1081	0.36	0.3589	0.3889	0.3879	0.0424
4A	238000001	240000000	714	0.3665	0.3651	0.3756	0.374	0.0378
4A	254000001	256000000	961	0.3621	0.3602	0.3727	0.3704	0.0437

4A	255000001	257000000	948	0.3821	0.3773	0.3836	0.3787	0.04
4A	281000001	283000000	243	0.3383	0.3305	0.367	0.3592	0.0417
4A	285000001	287000000	622	0.3659	0.3546	0.368	0.3566	0.0487
4A	286000001	288000000	566	0.368	0.351	0.3839	0.3667	0.0453
4A	294000001	296000000	928	0.3606	0.3612	0.3633	0.3639	0.0341
4A	295000001	297000000	761	0.3481	0.3476	0.3518	0.3509	0.041
4A	305000001	307000000	1246	0.3643	0.3553	0.3871	0.3779	0.0426
4A	324000001	326000000	836	0.3372	0.3348	0.3474	0.345	0.0437
4A	339000001	341000000	1413	0.3473	0.3389	0.3487	0.3403	0.0461
4A	343000001	345000000	953	0.3452	0.3388	0.3523	0.3459	0.0408
4A	344000001	346000000	1067	0.3464	0.3368	0.3481	0.3386	0.0486
4A	349000001	351000000	892	0.3577	0.3527	0.3839	0.3789	0.0415
4A	350000001	352000000	826	0.3538	0.3499	0.3599	0.3553	0.0496
4A	368000001	370000000	989	0.3326	0.3316	0.3454	0.3443	0.0425
4A	375000001	377000000	1189	0.3629	0.3635	0.3692	0.3704	0.0408
4A	378000001	380000000	834	0.3502	0.3505	0.3596	0.3602	0.0421
4A	384000001	386000000	841	0.3707	0.376	0.3803	0.3859	0.0333
4A	385000001	387000000	695	0.3683	0.3601	0.3707	0.3627	0.0395
4A	389000001	391000000	1110	0.3601	0.3588	0.371	0.3693	0.0398
4A	390000001	392000000	931	0.3545	0.3556	0.3643	0.3641	0.0388
4A	403000001	405000000	990	0.3419	0.336	0.3477	0.3418	0.0396
4A	404000001	406000000	1100	0.3451	0.3438	0.3472	0.346	0.0402
4A	424000001	426000000	1047	0.3506	0.3406	0.3524	0.3417	0.0434
4A	432000001	434000000	904	0.3373	0.3405	0.3393	0.343	0.0315
4A	433000001	435000000	975	0.3223	0.3248	0.3624	0.3662	0.0349
4A	437000001	439000000	915	0.332	0.3263	0.3334	0.3278	0.0346
4A	444000001	446000000	1472	0.2332	0.2307	0.2435	0.2409	0.0281
4A	445000001	447000000	1651	0.2392	0.2377	0.256	0.2545	0.0278
4A	450000001	452000000	1393	0.235	0.2335	0.2407	0.2389	0.0237
4A	477000001	479000000	1375	0.1864	0.1748	0.3966	0.3843	0.0414
4A	478000001	480000000	1543	0.1706	0.1629	0.4088	0.4005	0.0348

4A	479000001	481000000	1414	0.1836	0.1771	0.331	0.3241	0.0334
4A	480000001	482000000	1266	0.1877	0.1854	0.3007	0.2974	0.0327
4A	481000001	483000000	1188	0.1978	0.1954	0.2588	0.2552	0.0328
4A	482000001	484000000	1379	0.2259	0.2191	0.2604	0.253	0.0271
4A	521000001	523000000	2174	0.1724	0.1823	0.2768	0.2756	0.0303
4B	45000001	47000000	3723	0.2155	0.2205	0.2537	0.2615	0.0242
4B	46000001	48000000	3634	0.2461	0.2421	0.2907	0.2868	0.0104
4B	49000001	51000000	4461	0.1631	0.1612	0.2122	0.2092	0.0164
4B	50000001	52000000	4012	0.1956	0.1921	0.2051	0.2022	0.0237
4B	54000001	56000000	3454	0.2046	0.1949	0.2247	0.2197	0.031
4B	524000001	526000000	2834	0.2219	0.226	0.2633	0.2598	0.0217
4B	525000001	527000000	3501	0.205	0.209	0.2147	0.213	0.029
4B	555000001	557000000	2070	0.2475	0.2333	0.3765	0.3665	0.0375
4B	556000001	558000000	2590	0.2596	0.2437	0.3443	0.3373	0.0393
4B	557000001	559000000	2380	0.2882	0.2711	0.2962	0.2917	0.0431
4B	558000001	560000000	1862	0.2692	0.2562	0.2872	0.2841	0.0415
4B	559000001	561000000	1458	0.2665	0.2532	0.2996	0.2954	0.0409
4B	560000001	562000000	1497	0.2911	0.2735	0.3338	0.3272	0.0445
4B	562000001	564000000	3072	0.2611	0.2494	0.2654	0.255	0.0344
4B	563000001	565000000	4123	0.2188	0.216	0.2279	0.2304	0.0241
4B	588000001	590000000	2560	0.1315	0.1124	0.49	0.4686	0.0384
4B	589000001	591000000	2635	0.1763	0.1645	0.351	0.3354	0.0436
4B	593000001	595000000	1104	0.2901	0.3109	0.3608	0.3829	0.0477
4B	609000001	611000000	1988	0.1643	0.1611	0.1921	0.19	0.0147
4B	610000001	612000000	2582	0.1661	0.161	0.2277	0.2246	0.0129
4B	611000001	613000000	3404	0.1582	0.1555	0.2368	0.2357	0.0107
4B	612000001	614000000	3272	0.1689	0.1825	0.1782	0.1907	0.0399
5A	40000001	42000000	2898	0.2037	0.2047	0.2096	0.2098	0.0205
5A	41000001	43000000	2392	0.1992	0.1997	0.2548	0.2545	0.0149
5A	42000001	44000000	4201	0.1756	0.1739	0.2017	0.2011	0.0276
5A	44000001	46000000	2747	0.1509	0.1525	0.1705	0.1747	0.0415

5A	45000001	47000000	3561	0.1649	0.152	0.179	0.1689	0.0483
5A	67000001	69000000	3773	0.2446	0.2308	0.2951	0.2715	0.0479
5A	68000001	70000000	2011	0.2411	0.2274	0.284	0.26	0.0487
5A	75000001	77000000	1967	0.2417	0.2183	0.39	0.3655	0.0419
5A	76000001	78000000	3151	0.1901	0.1727	0.4062	0.388	0.0348
5A	77000001	79000000	3229	0.1736	0.1683	0.406	0.4002	0.0213
5A	78000001	80000000	2934	0.203	0.2073	0.2845	0.2885	0.0102
5A	108000001	110000000	1402	0.3001	0.3033	0.3118	0.3156	0.0212
5A	109000001	111000000	1674	0.3096	0.3142	0.3457	0.3508	0.0151
5A	454000001	456000000	4323	0.1655	0.1861	0.1981	0.2212	0.0473
5A	455000001	457000000	3119	0.1882	0.1897	0.2232	0.2265	0.0261
5A	659000001	661000000	3056	0.1795	0.1796	0.1996	0.2007	0.0107
5B	27000001	29000000	3409	0.2046	0.1979	0.2254	0.213	0.0469
5B	28000001	30000000	4310	0.2224	0.2169	0.2458	0.2406	0.0147
5B	89000001	91000000	1798	0.2111	0.209	0.2609	0.2588	0.0195
5B	90000001	92000000	2201	0.2223	0.2194	0.2663	0.2633	0.0168
5B	91000001	93000000	2083	0.2386	0.2344	0.2699	0.2656	0.0193
5B	273000001	275000000	1467	0.1585	0.1532	0.2334	0.228	0.0254
5B	274000001	276000000	1631	0.1634	0.1612	0.2099	0.2076	0.0197
5B	402000001	404000000	1786	0.2181	0.2168	0.3139	0.3123	0.0152
5B	403000001	405000000	2204	0.2094	0.2082	0.3311	0.3288	0.0215
5B	652000001	654000000	2797	0.1825	0.1875	0.2055	0.2042	0.0358
6A	63000001	65000000	3287	0.2397	0.2377	0.2677	0.2657	0.01
6A	64000001	66000000	2568	0.2362	0.2346	0.2798	0.2782	0.0066
6A	65000001	67000000	2440	0.218	0.217	0.2473	0.2464	0.0091
6A	66000001	68000000	2854	0.2239	0.2239	0.249	0.2491	0.0111
6B	91000001	93000000	1908	0.1198	0.1068	0.1554	0.1475	0.0261
6B	92000001	94000000	2103	0.1289	0.117	0.1338	0.1271	0.0253
6B	164000001	166000000	2930	0.2109	0.2094	0.2221	0.2214	0.0232
7A	578000001	580000000	2064	0.1868	0.1742	0.2233	0.2128	0.0482
7A	660000001	662000000	2325	0.1604	0.1633	0.2352	0.2374	0.0193

7A	661000001	663000000	3362	0.1573	0.1681	0.1853	0.1944	0.0281
7A	678000001	680000000	2680	0.1706	0.16	0.3439	0.338	0.0222
7A	679000001	681000000	3567	0.1732	0.1749	0.18	0.1821	0.0243
7A	687000001	689000000	2550	0.1883	0.1908	0.1909	0.199	0.0494
7A	688000001	690000000	1790	0.2021	0.1964	0.2208	0.224	0.0198
7B	85000001	87000000	3374	0.1771	0.1973	0.2259	0.2422	0.0486
7B	156000001	158000000	1916	0.1643	0.1489	0.1746	0.1622	0.0359
7B	384000001	386000000	1418	0.1444	0.1513	0.1965	0.2013	0.0493
7B	546000001	548000000	1993	0.2145	0.2032	0.2705	0.2644	0.0341
7B	685000001	687000000	2303	0.2128	0.2031	0.2345	0.233	0.0461

S. Table 3: list of genes contained in the windows from S. Table 2; genes highlighted in bold are affected by high impact variants

Gene stable ID	*Gene description*	*Chrom*	*Gene start (bp)*	*Gene end (bp)*
TRITD1Av1G010440	DUF674 family protein	1A	22263590	22264321
TRITD1Av1G010510	Ankyrin repeat-containing protein	1A	22460834	22466344
TRITD1Av1G010600	UDP galactose transporter-related protein	1A	22852317	22859292
TRITD1Av1G010690	MADS-box transcription factor-like protein	1A	22929041	22933171
TRITD1Av1G010700	F-box family protein	1A	22944522	22946180
TRITD1Av1G124860	PHD finger family protein	1A	338100397	338108233
TRITD1Av1G125140	3-hydroxyisobutyryl-CoA hydrolase-like protein	1A	338797312	338805187
TRITD1Av1G125250	PP2A regulatory subunit TAP46	1A	338986994	338987767
TRITD1Av1G125470	Carbohydrate kinase, putative, expressed	1A	339477451	339483753
TRITD1Av1G125480	D-mannose binding lectin protein with Apple-like carbohydrate-binding domain-containing protein G	1A	339484873	339486241
TRITD1Av1G125490	F-box/RNI-like superfamily protein TE?	1A	339485042	339485842
TRITD1Av1G125910	Mitochondrial carrier protein	1A	340974330	340976220
TRITD1Av1G126070	60S ribosomal export protein NMD3	1A	341284289	341288975
TRITD1Av1G126080	Nuclear control of ATPase protein 2	1A	341290671	341297944
TRITD1Av1G146420	Histone H3	1A	393699589	393701201
TRITD1Av1G146450	Histone H3	1A	393711368	393713943
TRITD1Av1G146660	Pathogen-related protein G	1A	394266862	394267208
TRITD1Av1G146930	Leucine-rich repeat receptor kinase-like protein	1A	394940840	394948088
TRITD1Av1G146950	Receptor-like protein kinase	1A	394948541	394950540
TRITD1Av1G146960	Elongation factor 4 G	1A	394951901	394955702
TRITD1Av1G147010	Phytochrome A-associated F-box protein	1A	394996702	394997814

TRITD1Av1G147120	splicing factor-like protein	1A	395555916	395559707
TRITD1Av1G147150	Chaperone protein dnaJ, putative	1A	395561765	395562865
TRITD1Av1G147230	Transducin/WD40 repeat-like superfamily protein G	1A	395693510	395701445
TRITD1Bv1G022440	50S ribosomal protein L22 G	1B	57137849	57138381
TRITD1Bv1G022650	MADS-box transcription factor family protein	1B	57288866	57289402
TRITD1Bv1G022770	Glycosyltransferase G	1B	57489873	57494137
TRITD1Bv1G022910	4'-phosphopantetheinyl transferase	1B	57975903	57980360
TRITD1Bv1G022920	Protease inhibitor/seed storage/lipid transfer protein family protein	1B	57989579	58681446
TRITD1Bv1G022930	Protease inhibitor/seed storage/lipid transfer protein family protein G	1B	58410729	58551860
TRITD1Bv1G022940	Protease inhibitor/seed storage/lipid transfer protein family protein	1B	58600561	58601048
TRITD1Bv1G022980	DNL-type zinc finger protein	1B	58743011	58744528
TRITD1Bv1G022990	Ubiquitin-conjugating enzyme E2	1B	58746022	58748328
TRITD1Bv1G023010	NBS-LRR-like resistance protein	1B	58750482	58753325
TRITD1Bv1G023390	Glycosyltransferases	1B	60246087	60251114
TRITD1Bv1G023410		1B	60287861	60288772
TRITD1Bv1G023480	Photosystem II reaction center protein H	1B	60386147	60386377
TRITD1Bv1G023550	Ankyrin repeat protein	1B	60475689	60476960
TRITD1Bv1G023750	Peroxidase G	1B	60703725	60705378
TRITD1Bv1G023770	Histone H4	1B	60833295	60838156
TRITD1Bv1G030300	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	1B	80294941	80295237
TRITD1Bv1G030320	Chaperone protein dnaJ	1B	80300847	80306029
TRITD1Bv1G030710	AT5G11810-like protein G	1B	81760426	81761583
TRITD1Bv1G030900	WRKY transcription factor	1B	82473590	82475847
TRITD1Bv1G030970	Aminotransferase-related family protein	1B	82703303	82705602
TRITD1Bv1G032950	RNA polymerase II-associated protein 3	1B	88666579	88670684
TRITD1Bv1G033000	Kinase, putative	1B	88739908	88743012
TRITD1Bv1G033010	Leucine-rich repeat receptor-like protein kinase family protein	1B	88745613	88747025
TRITD1Bv1G033110	transmembrane protein, putative (DUF247)	1B	88936441	88938018
TRITD1Bv1G033120	Beta-carotene isomerase d27, chloroplastic	1B	88938536	89007514
TRITD1Bv1G033180	Pentatricopeptide repeat-containing protein	1B	89032323	89034317
TRITD1Bv1G033420	Peroxidase	1B	89599753	89607725
TRITD1Bv1G033430	Peroxidase	1B	89614649	89615810
TRITD1Bv1G033440	Peroxidase	1B	89620241	89621442
TRITD1Bv1G033580	F2P16.20 protein, putative isoform 1	1B	90031758	90035856
TRITD1Bv1G033600	Protein phosphatase 2C-like protein	1B	90037083	90039174
TRITD1Bv1G033820	Ankyrin repeat-containing protein	1B	90694791	90760524
TRITD1Bv1G033830	transmembrane protein, putative (DUF594)	1B	90695723	90697831
TRITD1Bv1G034090	Dentin sialophosphoprotein-related, putative isoform 1 G	1B	91527631	91529040
TRITD1Bv1G034300	Regulator of chromosome condensation (RCC1) family protein	1B	92357548	92367555
TRITD1Bv1G034590	MYB transcription factor	1B	93544573	93545811

TRITD1Bv1G034650	O-fucosyltransferase family protein	1B	93627654	93632690
TRITD1Bv1G035340	Pentatricopeptide repeat-containing protein	1B	95111904	95113817
TRITD1Bv1G036690	Peroxidase	1B	100116336	100117452
TRITD1Bv1G036760	Lecithin-cholesterol acyltransferase-like 1	1B	100243077	100244378
TRITD1Bv1G036890	Glutathione S-transferase	1B	100469953	100470919
TRITD1Bv1G036920	Lecithin-cholesterol acyltransferase-like 1	1B	100478096	100479610
TRITD1Bv1G037300	Myb-like protein	1B	101582391	101597265
TRITD1Bv1G037500	Myb-like transcription factor	1B	102318877	102325802
TRITD1Bv1G072950	Tetraspanin family protein	1B	208452635	208477660
TRITD1Bv1G077600	Protein transport protein Sec61 subunit gamma	1B	225910725	225911018
TRITD1Bv1G084140	Autoinducer 2 import ATP-binding protein LsrA	1B	248893728	248894084
TRITD1Bv1G084180	Werner Syndrome-like exonuclease	1B	249108626	249108976
TRITD2Av1G042800	Pyrrolidone-carboxylate peptidase	2A	93085221	93086870
TRITD2Av1G042910	Peroxidase	2A	93232956	93233915
TRITD2Av1G042920	Peroxidase	2A	93236596	93239057
TRITD2Av1G042930	Peroxidase	2A	93271315	93272384
TRITD2Av1G042950	Glucose-6-phosphate 1-dehydrogenase G	2A	93299355	93299603
TRITD2Av1G042960	Peroxidase	2A	93352036	93359640
TRITD2Av1G042980	Peroxidase	2A	93367330	93368981
TRITD2Av1G043020	Peroxidase	2A	93519820	93520257
TRITD2Av1G043180	Phosphate carrier protein, mitochondrial	2A	94099305	94099637
TRITD2Av1G043240	Peroxidase	2A	94211740	94215254
TRITD2Av1G043500	Leucine--tRNA ligase	2A	94953246	94961150
TRITD2Av1G043580	Phosphoserine phosphatase	2A	95042901	95045768
TRITD2Av1G043650	Kinesin-like protein	2A	95133908	95146519
TRITD2Av1G043700	Serine/threonine-protein kinase haspin	2A	95215880	95221883
TRITD2Av1G280320	Pentatricopeptide repeat-containing protein	2A	745026562	745262259
TRITD2Av1G280380	Cellulose synthase-like protein	2A	745267141	745269194
TRITD2Av1G280400	Pentatricopeptide repeat-containing protein	2A	745275012	745325467
TRITD2Av1G280490	Wound-responsive family protein	2A	745493379	745493862
TRITD2Av1G280780	Serine/threonine-protein kinase	2A	746159448	746163780
TRITD2Av1G280820	F-box family protein	2A	746229712	746230877
TRITD2Av1G280860	Cytochrome P450	2A	746298416	746299936
TRITD2Av1G280950	Invertase inhibitor	2A	746583954	746584484
TRITD2Av1G280970	Heme oxygenase 1	2A	746616997	746622417
TRITD2Av1G281000	Serine/threonine-protein kinase	2A	746663591	746665923
TRITD2Av1G281040	Serine/threonine-protein kinase	2A	746728010	746732486
TRITD2Av1G281150	Cytochrome P450	2A	746947252	746948931
TRITD2Av1G281190	indeterminate(ID)-domain 11	2A	746993085	746999189
TRITD2Av1G281200	Serine/threonine-protein kinase	2A	747009262	747010870
TRITD2Av1G281220	Serine/threonine-protein kinase	2A	747024766	747025764

TRITD2Av1G281230	Copalyl diphosphate synthase	2A	747066836	747073986
TRITD2Av1G281280	Ubiquitin carboxyl-terminal hydrolase 2	2A	747146892	747175966
TRITD2Av1G281290	Ubiquitin carboxyl-terminal hydrolase 2	2A	747179495	747181714
TRITD2Av1G281300	FBD / Leucine Rich Repeat domains containing protein G	2A	747188293	747188544
TRITD2Av1G281430	Non-specific serine/threonine protein kinase	2A	747297985	747299283
TRITD2Av1G281440	Sulfotransferase	2A	747358533	747359709
TRITD2Av1G281450	Serine/threonine-protein kinase	2A	747362868	747366490
TRITD2Av1G281460	Serine/threonine-protein kinase	2A	747368931	747374845
TRITD2Av1G281470	Serine/threonine-protein kinase	2A	747373416	747448431
TRITD2Av1G281660	Serine/threonine-protein kinase	2A	747850469	747852040
TRITD2Av1G281680	Superoxide dismutase	2A	747913884	747918634
TRITD2Av1G281750	Sulfotransferase	2A	748004479	748005507
TRITD2Av1G281770	Serine/threonine-protein kinase	2A	748087675	748089115
TRITD2Av1G281780	Serine/threonine-protein kinase	2A	748089310	748090701
TRITD2Av1G281840	Serine/threonine-protein kinase	2A	748147304	748148629
TRITD2Av1G281860	basic helix-loop-helix (bHLH) DNA-binding superfamily protein	2A	748267711	748270103
TRITD2Av1G281930	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	2A	748371056	748372502
TRITD2Av1G282020	Heavy metal transport/detoxification superfamily protein	2A	748443578	748484279
TRITD2Av1G282170	Regulatory protein recX	2A	748691289	748693174
TRITD2Av1G282200	Phosphatidylinositol 4-kinase gamma-like protein	2A	748712457	748714355
TRITD2Av1G282430	Exocyst complex component, putative	2A	749305164	749306651
TRITD2Av1G282440	Exocyst complex component, putative	2A	749383899	749385392
TRITD2Av1G282480	Exocyst complex component, putative	2A	749393360	749394952
TRITD2Av1G282590	Exocyst complex component, putative	2A	749515144	749516370
TRITD2Av1G282610	Triacylglycerol lipase 2, putative G	2A	749521574	749524665
TRITD2Av1G282640	Blue copper protein	2A	749592536	749593974
TRITD2Av1G282680	Blue copper protein	2A	749658879	749659394
TRITD2Av1G282690	Blue copper protein	2A	749666839	749667354
TRITD2Av1G282740	FAD-binding Berberine family protein	2A	749716112	749717707
TRITD2Av1G282840	Acyl-protein thioesterase 1	2A	749865146	749867705
TRITD2Av1G282850	Acyl-protein thioesterase 1	2A	749868122	749869884
TRITD2Bv1G034660	NAC domain-containing protein	2B	83650221	84083682
TRITD2Bv1G034930	Stress responsive protein	2B	85485409	85487268
TRITD2Bv1G034950	Stress responsive protein	2B	85637941	85639667
TRITD2Bv1G035010	NAC domain protein	2B	85747668	85749357
TRITD2Bv1G035220	Pro-resilin, putative G	2B	86316705	86320161
TRITD2Bv1G035390	RNA-binding domain CCCH-type zinc finger protein TE?	2B	86883400	86886195
TRITD2Bv1G072150	F-box family protein	2B	186411035	186411586
TRITD2Bv1G072160	Zinc finger family protein	2B	186419498	186420892
TRITD2Bv1G072190	Myosin heavy chain-like protein G	2B	186672887	186675499

TRITD2Bv1G072410	O-fucosyltransferase family protein G	2B	187448400	187457922
TRITD2Bv1G072480	S phase cyclin A-associated protein in the endoplasmic reticulum	2B	187539412	187549539
TRITD2Bv1G072490	F-box domain containing protein TE?	2B	187552531	187553073
TRITD2Bv1G072520	DNA helicase	2B	187563440	187578259
TRITD2Bv1G072540	EID1-like F-box protein 2	2B	187690729	187691475
TRITD2Bv1G072640	DNA repair protein-like protein	2B	187956803	187964463
TRITD2Bv1G072650	tRNA guanosine-2'-O-methyltransferase	2B	187968469	187971995
TRITD2Bv1G072680	Salt overly sensitive 1 G	2B	187976653	187977039
TRITD2Bv1G073940	Eukaryotic aspartyl protease family protein	2B	192673515	192683611
TRITD2Bv1G073970	Short-chain dehydrogenase/reductase	2B	192870458	192871510
TRITD2Bv1G073980	Hyccin G	2B	192904709	192905981
TRITD2Bv1G074000	Gibberellin regulated protein	2B	192954239	192961523
TRITD2Bv1G074150	glutamyl-tRNA (Gln) amidotransferase subunit A (DUF620)	2B	193316915	193318869
TRITD2Bv1G074420	GPI-anchored adhesin-like protein, putative (DUF3741) G	2B	193846316	193847150
TRITD2Bv1G188130	Protein kinase family protein	2B	557049328	557051926
TRITD2Bv1G188140	Protein kinase family protein	2B	557060115	557062658
TRITD2Bv1G188160	Kinase family protein	2B	557112380	557113440
TRITD2Bv1G188210	Hexosyltransferase G	2B	557277823	557283289
TRITD2Bv1G188380	MYB-related protein	2B	557708354	557709844
TRITD2Bv1G188410	Histone-lysine N-methyltransferase	2B	557743697	557747679
TRITD2Bv1G188470	Protodermal factor 1 G	2B	558258884	558260722
TRITD2Bv1G188540	NAC domain protein,	2B	558687739	558690001
TRITD2Bv1G188590	L-allo-threonine aldolase	2B	559109795	559111781
TRITD2Bv1G188650	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein G	2B	559218356	559219906
TRITD2Bv1G189010	Calmodulin-binding family protein, putative, expressed G	2B	559868742	559872370
TRITD2Bv1G189040	Ribosomal protein L11 methyltransferase G	2B	559950738	559951340
TRITD2Bv1G189060	DNA helicase INO80-like protein TE?	2B	560181903	560182550
TRITD2Bv1G189110	CC-NBS-LRR family disease resistance protein	2B	560256887	560260520
TRITD2Bv1G189180	Beta-glucosidase	2B	560723172	560728433
TRITD2Bv1G189200	Beta-glucosidase, putative	2B	560826893	560834452
TRITD2Bv1G189340	Protein BREAST CANCER SUSCEPTIBILITY 1-like protein	2B	561452213	561457205
TRITD2Bv1G189350	Actin-related protein 2/3 complex subunit 2	2B	561457854	561460054
TRITD2Bv1G189380	Fatty acid hydroxylase superfamily protein	2B	561469552	561474041
TRITD2Bv1G189410	Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B, putative isoform 1 G	2B	561613092	561615480
TRITD2Bv1G189430	E3 ubiquitin-protein ligase SDIR1 G	2B	561714208	561718407
TRITD2Bv1G189450	Transporter	2B	561769664	561772356
TRITD2Bv1G189460	Caleosin	2B	561796934	561798389
TRITD2Bv1G189540	Vacuolar-processing enzyme	2B	561889138	561891676
TRITD2Bv1G198590	Homeobox-leucine zipper protein	2B	591288052	591293185
TRITD2Bv1G198720	Heat Stress Transcription Factor family protein	2B	591760019	591761025

TRITD2Bv1G198800	3-hexulose-6-phosphate isomerase	2B	591938371	591938976
TRITD2Bv1G198890	Bax inhibitor 1 G	2B	592426518	592428291
TRITD2Bv1G198930	WD-repeat protein, putative	2B	592476030	592478588
TRITD3Av1G176500	Protein DETOXIFICATION	3A	492204114	492209315
TRITD3Av1G176640	transmembrane protein, putative (Protein of unknown function, DUF599)	3A	492387489	492388331
TRITD3Av1G176860	Glycerate kinase	3A	492994629	492998205
TRITD3Av1G176970	Glycosylphosphatidylinositol anchor attachment 1 protein G	3A	493265520	493269156
TRITD3Av1G177110	Glutamate synthase, putative	3A	493574543	493584567
TRITD3Av1G177240	Elongation factor 4 G	3A	493830782	493831201
TRITD3Av1G177250	Splicing factor 3a, subunit 3	3A	493834382	493838051
TRITD3Av1G177310	O-acyltransferase WSD1	3A	493874539	493877656
TRITD3Av1G216140	p-loop ntpase domain-containing protein lpa1-like 2	3A	595286648	595290756
TRITD3Av1G216170	Zinc finger protein, putative TE?	3A	595304836	595305450
TRITD3Av1G216350	Protein UPSTREAM OF FLC	3A	595825192	595833718
TRITD3Av1G216430	ATP-dependent zinc metalloprotease FtsH	3A	596044713	596047527
TRITD3Av1G216570	Nucleosome assembly protein 1-like 1	3A	596475878	596478089
TRITD3Av1G216670	Pentatricopeptide repeat-containing protein, putative	3A	596649471	596650847
TRITD3Av1G216690	Histone H2B	3A	596718058	596718468
TRITD3Av1G216940	Histone H2B	3A	597299560	597309615
TRITD3Av1G217220	HSP20-like chaperones superfamily protein	3A	597895215	597896789
TRITD3Av1G217230	Ankyrin repeat family protein, putative	3A	597899266	597902843
TRITD3Av1G217260	Wuschel-related homeobox protein	3A	597915150	597916973
TRITD3Av1G217280	Wuschel-related homeobox protein	3A	597987122	597990043
TRITD3Bv1G031790	MYB transcription factor	3B	83063102	83064048
TRITD3Bv1G031860	Cytochrome P450 family protein	3B	83280443	83282063
TRITD3Bv1G031870	Cellulose synthase, putative	3B	83296170	83301934
TRITD3Bv1G031940	Cytochrome P450 family protein	3B	83438839	83440483
TRITD3Bv1G031950	Receptor-like kinase LIP2	3B	83445643	83448226
TRITD3Bv1G032110	TPR repeat protein	3B	83753358	83754029
TRITD3Bv1G032120	Receptor protein kinase, putative	3B	83755461	83760216
TRITD3Bv1G032190	O-fucosyltransferase family protein	3B	83805766	83808687
TRITD3Bv1G032200	WD repeat domain phosphoinositide-interacting protein	3B	83829827	83833727
TRITD3Bv1G032410	F-box family protein	3B	84757176	84759343
TRITD3Bv1G032510	import inner membrane translocase subunit G	3B	84988357	84992056
TRITD3Bv1G032570	Amino acid transporter	3B	85169562	85176729
TRITD3Bv1G032630	F-box protein	3B	85295488	85296681
TRITD3Bv1G032730	F-box protein	3B	85649478	85650659
TRITD3Bv1G032770	Sialyltransferase-like protein 2	3B	85789275	85789625
TRITD3Bv1G032800	Kinase interacting (KIP1-like) family protein	3B	85804638	85806576
TRITD3Bv1G032830	PP2A regulatory subunit TAP46	3B	85929885	85931195
TRITD3Bv1G032870	Sn1-specific diacylglycerol lipase alpha	3B	85991795	86001985

TRITD3Bv1G032910	Pentatricopeptide repeat-containing protein	3B	86029443	86031317
TRITD3Bv1G033050	p0028E10.10 protein G	3B	86193627	86200344
TRITD3Bv1G033200	Nucleotide/sugar transporter family protein G	3B	86541101	86544274
TRITD3Bv1G033210	Mini-chromosome maintenance complex-binding protein	3B	86545470	86549189
TRITD3Bv1G033370	Phosphatidylinositol N-acetylglucosaminyltransferase subunit H	3B	86736658	86738306
TRITD3Bv1G033420	Alpha-ketoglutarate-dependent dioxygenase alkB-like protein 6	3B	86747334	86749691
TRITD3Bv1G033470	Nuclease	3B	86904663	86907097
TRITD3Bv1G033490	Lecithin-cholesterol acyltransferase-like 1	3B	86987421	86988829
TRITD3Bv1G033740	Carboxymethylenebutenolidase-like protein	3B	87376089	87382944
TRITD3Bv1G033800	Eukaryotic aspartyl protease family protein	3B	87408874	87410244
TRITD3Bv1G033870	Splicing factor U2AF 50 kDa subunit G	3B	87582937	87583734
TRITD3Bv1G033880	Guanine-nucleotide-exchange protein	3B	87584127	87588758
TRITD3Bv1G033910	Avr9/Cf-9 rapidly elicited protein G	3B	87619665	87619880
TRITD3Bv1G034470	ethylene responsive element binding factor 2	3B	89444925	89458252
TRITD3Bv1G034480	Breast cancer type 2 susceptibility protein-like protein	3B	89463365	89474954
TRITD3Bv1G034550	Calcium-dependent protein kinase	3B	89688329	89698912
TRITD3Bv1G034620	External alternative NAD(P)H-ubiquinone oxidoreductase B2, mitochondrial	3B	89760237	89761405
TRITD3Bv1G034800	Receptor kinase	3B	90027745	90031230
TRITD3Bv1G034940	Serine carboxypeptidase, putative	3B	90263412	90287229
TRITD3Bv1G034960	Protein DETOXIFICATION	3B	90289548	90292293
TRITD3Bv1G036920	Histone H3	3B	96167917	96169565
TRITD3Bv1G036940	Leucine-rich repeat receptor-like protein kinase family protein	3B	96240126	96240761
TRITD3Bv1G036970	Lectin receptor kinase	3B	96246525	96247748
TRITD3Bv1G037040	receptor lectin kinase	3B	96275988	96277674
TRITD3Bv1G037100	Chlorophyllase family protein, expressed	3B	96348346	96349468
TRITD3Bv1G037200	receptor lectin kinase	3B	96536081	96541797
TRITD3Bv1G037240	Pyruvate decarboxylase	3B	96767958	97279827
TRITD3Bv1G037280	DNA-binding storekeeper protein-related transcriptional regulator	3B	96835026	96836012
TRITD3Bv1G037290	Amino acid transporter, putative	3B	97271894	97279005
TRITD3Bv1G037340	Basic helix-loop-helix transcription factor	3B	97312369	97313496
TRITD3Bv1G037350	Sugar transporter, putative	3B	97319360	97320816
TRITD3Bv1G037470	RING/U-box superfamily protein	3B	97627464	97631699
TRITD3Bv1G037620	Fasciclin-like arabinogalactan protein	3B	98029254	98030039
TRITD3Bv1G037670	Ribonuclease P protein subunit p14 G	3B	98162142	98166916
TRITD3Bv1G037730	transcription factor-like protein	3B	98316727	98317575
TRITD3Bv1G037770	ATP-dependent RNA helicase p62	3B	98602824	98606916
TRITD3Bv1G038880	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family	3B	102015889	102016776
TRITD3Bv1G038950	Serine/threonine-protein kinase PBS1 G	3B	102304118	102316292
TRITD3Bv1G039230	Cytochrome p450	3B	103224901	103225239
TRITD3Bv1G039450	Ankyrin repeat-containing protein	3B	103967873	103971576

TRITD3Bv1G039470	Receptor-like protein kinase	3B	103989618	104002396
TRITD3Bv1G039500	Protein IQ-DOMAIN 1	3B	104227692	104230129
TRITD3Bv1G039510	Myb-like transcription factor family protein	3B	104236173	104241510
TRITD3Bv1G039670	Always early, putative	3B	104891465	104897376
TRITD3Bv1G039700	23S rRNA (Uracil-5-)-methyltransferase	3B	104907153	104911172
TRITD3Bv1G039760	p-loop containing nucleoside triphosphate hydrolases superfamily protein, putative	3B	105224110	105232190
TRITD3Bv1G040000	Dof zinc finger protein	3B	106314457	106315167
TRITD3Bv1G040180	cDNA clone:J013058P10, full insert sequence G	3B	106837392	106842671
TRITD3Bv1G040190	U3 small nucleolar RNA-associated protein 25	3B	106901928	106903195
TRITD3Bv1G040640	Pollen-specific protein SF21	3B	108103711	108106228
TRITD3Bv1G040910	Zinc finger CCCH domain-containing protein	3B	108934952	108936115
TRITD3Bv1G041840		3B	112046063	112047334
TRITD3Bv1G041880	Mitochondrial carrier family protein	3B	112135004	112155499
TRITD3Bv1G041900	Rhomboid-like protein	3B	112233768	112234367
TRITD3Bv1G042000	Replication factor-A carboxy-terminal domain protein G	3B	112374420	112375156
TRITD3Bv1G042220	Transportin-1 G	3B	113054086	113059544
TRITD3Bv1G042340	Phosphotransferase	3B	113255013	113258363
TRITD3Bv1G042470	CRS2-associated factor 1	3B	113617455	113619268
TRITD3Bv1G042480	CRS2-associated factor 1, chloroplastic G	3B	113625559	113626284
TRITD3Bv1G042580	Malic enzyme	3B	113743798	113749135
TRITD3Bv1G042680	Werner Syndrome-like exonuclease	3B	113990928	113993933
TRITD3Bv1G042690	Werner Syndrome-like exonuclease	3B	114007311	114008277
TRITD3Bv1G042700	Werner Syndrome-like exonuclease	3B	114026469	114026819
TRITD3Bv1G042710	Ribosomal protein S12 methylthiotransferase RimO	3B	114030232	114031727
TRITD3Bv1G042780	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	3B	114101359	114105058
TRITD3Bv1G042940	Cytokinin oxidase/dehydrogenase	3B	114730559	114732700
TRITD3Bv1G043950	ATP-dependent RNA helicase, putative	3B	118285722	118293296
TRITD3Bv1G044050	Protein TWIN LOV 1	3B	118454592	118458322
TRITD3Bv1G044070	WRKY transcription factor	3B	118543710	118544378
TRITD3Bv1G044210	WRKY family transcription factor	3B	118862683	118864671
TRITD3Bv1G044220	Wd40-repeat-like protein	3B	118877776	118884148
TRITD3Bv1G044280	Heat shock protein	3B	119062894	119063382
TRITD3Bv1G044290	Heat shock protein	3B	119094677	119422259
TRITD3Bv1G044320	Heat shock protein	3B	119877827	119878306
TRITD3Bv1G044330	17. class II heat shock protein	3B	119882730	119883476
TRITD3Bv1G044340	Heat shock protein	3B	120062680	120063219
TRITD3Bv1G044370	17. class II heat shock protein	3B	120091228	120091716
TRITD3Bv1G044380	ATP-dependent RNA helicase	3B	120094185	120099605
TRITD3Bv1G044560	Homeobox protein	3B	120673031	120683400
TRITD3Bv1G044630	Cytochrome c oxidase subunit 1	3B	120725948	120726211
TRITD3Bv1G044650	FACT complex subunit SSRP1	3B	120773307	120780430

TRITD3Bv1G044810	FAD-binding Berberine family protein	3B	121311317	121312903
TRITD3Bv1G044820	Glutamine--tRNA ligase	3B	121314741	121320218
TRITD3Bv1G044850	3'-N-debenzoyl-2'-deoxytaxol N-benzoyltransferase	3B	121460414	121463330
TRITD3Bv1G044890	Extracellular ribonuclease	3B	121661281	121663640
TRITD3Bv1G044960	regulatory particle non-ATPase subunit 5B	3B	121795467	121796441
TRITD3Bv1G044970	Calcium/calmodulin-dependent 3',5'-cyclic nucleotide phosphodiesterase 1B G	3B	121799571	121800110
TRITD3Bv1G045000	Cysteine/Histidine-rich C1 domain family protein	3B	121829376	121831149
TRITD3Bv1G045070	cDNA clone:001-038-D11, full insert sequence G	3B	121978272	121978808
TRITD3Bv1G045080	F-box protein	3B	121979464	121981392
TRITD3Bv1G045110	Histone-lysine N-methyltransferase	3B	122129332	122131831
TRITD3Bv1G045200	Alkaline alpha-galactosidase seed imbibition protein	3B	122363453	122366785
TRITD3Bv1G045330	Cytochrome P450 family protein, expressed	3B	122662011	122663782
TRITD3Bv1G045340	P53/DNA damage-regulated protein G	3B	122691033	122691936
TRITD3Bv1G045350	Metacaspase-1	3B	122692424	122693747
TRITD3Bv1G045380	Cytochrome P450 family protein	3B	122726442	122728109
TRITD3Bv1G045420	Type I inositol 1,4,5-trisphosphate 5-phosphatase 1	3B	122808594	122811302
TRITD3Bv1G045450	Zinc finger family protein	3B	122999853	123000766
TRITD3Bv1G045960	Choline/ethanolamine kinase	3B	124173577	124175535
TRITD3Bv1G045980	spatacsin carboxy-terminus protein	3B	124192176	124195694
TRITD3Bv1G045990	spatacsin carboxy-terminus protein	3B	124206560	124211288
TRITD3Bv1G046010	Werner Syndrome-like exonuclease	3B	124217494	124218096
TRITD3Bv1G046020	Phosphoribulokinase / Uridine kinase family G	3B	124327072	124328422
TRITD3Bv1G046120	Gigantea-like protein	3B	124530113	124536664
TRITD3Bv1G046220	Sequence-specific DNA binding transcription factor	3B	124738464	124740092
TRITD3Bv1G209020	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	3B	640834958	641091073
TRITD3Bv1G209120	F-box protein	3B	641494881	641499540
TRITD3Bv1G209140	Zinc finger protein	3B	641542913	641544940
TRITD3Bv1G209260	BZIP transcription factor	3B	642143053	642144584
TRITD3Bv1G209270	BZIP transcription factor	3B	642187859	642955038
TRITD3Bv1G209350	Protein ABSCISIC ACID-INSENSITIVE 5 G	3B	642964091	642964557
TRITD3Bv1G209360	ABA response element binding factor G	3B	642966452	642967302
TRITD3Bv1G219910	Sentrin-specific protease	3B	671016807	671019221
TRITD3Bv1G219920	DUF1685 family protein	3B	671025612	671091338
TRITD3Bv1G220060	Phthiocerol synthesis polyketide synthase type I PpsE	3B	671237227	671237880
TRITD3Bv1G220140	Protein EFR3-like protein G	3B	671451826	671458007
TRITD3Bv1G220470	Leucine-rich repeat receptor-like protein kinase	3B	672270386	672275210
TRITD3Bv1G220550	Sister chromatid cohesion 1 protein	3B	672654992	672665092
TRITD3Bv1G220660	Formin-like protein	3B	672803685	672806786
TRITD3Bv1G220690	Pentatricopeptide repeat protein	3B	672821835	672824390
TRITD3Bv1G220750	Ribonuclease	3B	673012197	673014643
TRITD3Bv1G220790	Ribonuclease	3B	673091153	673093520

TRITD3Bv1G220800	Glycine--tRNA ligase beta subunit G	3B	673095054	673097399
TRITD3Bv1G220820	alpha/beta-Hydrolases superfamily protein G	3B	673134642	673135658
TRITD3Bv1G220830	NADH dehydrogenase subunit 9	3B	673154002	673154223
TRITD3Bv1G220960	Ribosomal protein S3	3B	673286285	673287129
TRITD3Bv1G220970	ATP synthase subunit alpha	3B	673294508	673294999
TRITD3Bv1G220990	Calcium sensing receptor, chloroplastic	3B	673298499	673302702
TRITD3Bv1G221000	Nuclear transport factor 2 family protein with RNA binding domain	3B	673326361	673329240
TRITD3Bv1G221030	IQ domain-containing protein	3B	673403512	673406550
TRITD3Bv1G221270	Hydroxyethylthiazole kinase G	3B	673904607	673905659
TRITD3Bv1G221280	cDNA clone:J013058P10, full insert sequence G	3B	673907153	673915841
TRITD3Bv1G275470	Serine/arginine repetitive matrix protein 2 isoform 1 G	3B	819381737	819389208
TRITD3Bv1G275540	F-box protein, putative (DUF295)	3B	819532784	819533689
TRITD3Bv1G275580	Aspartyl aminopeptidase, putative	3B	819614827	819618614
TRITD3Bv1G275600	Pentatricopeptide repeat-containing protein	3B	819619375	819622950
TRITD3Bv1G275690	Late embryogenesis abundant protein	3B	819767532	819773391
TRITD3Bv1G275720	Cytochrome P450, putative	3B	819862917	819869061
TRITD3Bv1G276000	rRNA N-glycosidase G	3B	820852357	820852755
TRITD3Bv1G276120	Late embryogenesis abundant protein	3B	820958682	820959744
TRITD3Bv1G276130	Late embryogenesis abundant protein	3B	820961396	820962602
TRITD3Bv1G276150	Pm3-like disease resistance protein	3B	820998939	821003159
TRITD3Bv1G276200	Powdery mildew resistance protein Pm3 G	3B	821012209	821016021
TRITD3Bv1G276220	BTB/POZ domain-containing protein FBL11	3B	821028731	821034998
TRITD3Bv1G276370	Pm3-like disease resistance protein	3B	821314812	821317046
TRITD3Bv1G276400	Disease resistance protein RPM1	3B	821392491	821643139
TRITD3Bv1G276600	Histone-lysine N-methyltransferase, H3 lysine-9 specific	3B	821849439	821851373
TRITD3Bv1G276610	F-box family protein	3B	821859804	821860961
TRITD3Bv1G276650	Villin	3B	821920996	821929803
TRITD4Av1G018670	2-oxoglutarate-dependent dioxygenase-related family protein	4A	41378541	41379640
TRITD4Av1G018720	Alpha-L-fucosidase 2	4A	41426673	41434011
TRITD4Av1G018730	ABC subfamily C transporter	4A	41442472	41448025
TRITD4Av1G018820	Mannose-1-phosphate guanyltransferase	4A	41547576	41552765
TRITD4Av1G018890	Protein transport protein GOT1	4A	41683076	41684917
TRITD4Av1G018900	S-acyltransferase	4A	41685855	41692421
TRITD4Av1G019020	Cysteine proteinase inhibitor	4A	42231763	42232119
TRITD4Av1G019160	Cysteine proteinase inhibitor	4A	42685929	42686279
TRITD4Av1G019170	Beta-xylosidase, putative	4A	42692395	42696529
TRITD4Av1G019420	Poly(A) RNA polymerase cid14 G	4A	43235140	43243857
TRITD4Av1G072050	Potassium transporter	4A	188183278	188194775
TRITD4Av1G072560	tRNA pseudouridine synthase G	4A	189970604	189978580
TRITD4Av1G078280	NAD(P)H-quinone oxidoreductase subunit I, chloroplastic	4A	209127935	209129258

TRITD4Av1G079370	Eukaryotic translation initiation factor 3 subunit A G	4A	214090906	214091404
TRITD4Av1G080010	Kinase family protein	4A	216932867	216939418
TRITD4Av1G085580	splicing factor-like protein	4A	239040720	239045018
TRITD4Av1G085760	PQ-loop repeat family protein / transmembrane family protein G	4A	239764612	239765349
TRITD4Av1G089930	Cytochrome c oxidase subunit 2 G	4A	254832711	254833139
TRITD4Av1G099100	Aberrant pollen transmission 1, putative, expressed	4A	287251650	287290120
TRITD4Av1G099300	ATP synthase subunit alpha G	4A	287648249	287648581
TRITD4Av1G112820	Dentin sialophosphoprotein-related, putative isoform 1 G	4A	339748475	339754559
TRITD4Av1G113600	WRKY transcription factor	4A	343541184	343557191
TRITD4Av1G115800	DCD (Development and Cell Death) domain protein G	4A	351478508	351478972
TRITD4Av1G120050	Formin-like protein G	4A	368879712	368880083
TRITD4Av1G120170	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein G	4A	369231476	369232911
TRITD4Av1G120180	Protein SUPPRESSOR OF GENE SILENCING 3	4A	369235062	369246805
TRITD4Av1G122130	ENHANCED DISEASE RESISTANCE protein (DUF1336)	4A	376558793	376595103
TRITD4Av1G123860	DNA topoisomerase G	4A	384136987	384137906
TRITD4Av1G124180	Beta-adaptin-like protein	4A	386375303	386381387
TRITD4Av1G124210	Glycerol-3-phosphate dehydrogenase G	4A	386465161	386465616
TRITD4Av1G125150	Zinc finger, PHD-finger	4A	391172616	391187327
TRITD4Av1G128080	Acyl-CoA-binding domain-containing protein 4	4A	403604916	403624254
TRITD4Av1G128580	Dihydroorotase G	4A	405932135	405932395
TRITD4Av1G136560	Galactoside 2-alpha-L-fucosyltransferase	4A	433236086	433237735
TRITD4Av1G136740	Leucine-rich repeat protein kinase family protein, putative	4A	433886117	433906883
TRITD4Av1G137020		4A	434783852	434784322
TRITD4Av1G137700	Alkaline alpha-galactosidase seed imbibition protein	4A	437345242	437345867
TRITD4Av1G138050	DWNN domain, a CCHC-type zinc finger TE?	4A	438130908	438131184
TRITD4Av1G140920	Laccase	4A	445010568	445013386
TRITD4Av1G141060	Protein prenyltransferase alpha subunit repeat-containing protein 1 G	4A	445275865	445279053
TRITD4Av1G141340	Proline transporter	4A	446189658	446203856
TRITD4Av1G141530	Alpha/beta-Hydrolases superfamily protein	4A	446545217	446547974
TRITD4Av1G153990	DUF1191 superfamily protein	4A	477470563	477471501
TRITD4Av1G154050	heat-inducible transcription repressor G	4A	477547711	477550920
TRITD4Av1G154400	Protein kinase	4A	478316635	478319681
TRITD4Av1G154480	GH3 family protein	4A	478582579	478584003
TRITD4Av1G154540		4A	478611138	478611671
TRITD4Av1G154860	Bis(5'-adenosyl)-triphosphatase	4A	479300757	479303300
TRITD4Av1G155000	Ethanolamine-phosphate cytidylyltransferase	4A	479530265	479538614
TRITD4Av1G155460	Calcium uniporter protein, mitochondrial	4A	481078823	481084082
TRITD4Av1G155730	Dirigent protein	4A	481977115	481978604
TRITD4Av1G155820	Electron transport complex subunit RsxC	4A	482400485	482403578
TRITD4Av1G156050	Type I inositol-1,4,5-trisphosphate 5-phosphatase	4A	482962547	482965328

TRITD4Av1G156130	Methylesterase	4A	483063700	483076979
TRITD4Av1G169680	Ribonuclease E inhibitor RraA/Dimethylmenaquinone methyltransferase G	4A	521186379	521186822
TRITD4Av1G169760	Methionyl-tRNA formyltransferase G	4A	521420515	521421485
TRITD4Av1G169800		4A	521444736	521445143
TRITD4Av1G170150	50S ribosomal protein L19, putative	4A	522166092	522168605
TRITD4Av1G170380	F-box family protein	4A	522428889	522431488
TRITD4Av1G170390	F-box family protein	4A	522440637	522441149
TRITD4Av1G170440	Cytochrome P450	4A	522499187	522508288
TRITD4Av1G170480	Phytoalexin-deficient 4-1 protein	4A	522516499	522520186
TRITD4Bv1G018150	Tubulin alpha chain	4B	45208574	45211075
TRITD4Bv1G018210	ERD (Early-responsive to dehydration stress) family protein	4B	45225275	45228845
TRITD4Bv1G018280	Major facilitator superfamily protein	4B	45534027	45535245
TRITD4Bv1G018290	Protease HtpX homolog G	4B	45541449	45542717
TRITD4Bv1G018410	AMSH-like ubiquitin thioesterase 1	4B	46039835	46045893
TRITD4Bv1G018420	GDP-mannose transporter	4B	46094352	46098151
TRITD4Bv1G018440	Epidermal patterning factor-like protein	4B	46099945	46100425
TRITD4Bv1G018500	Alpha/beta-hydrolase superfamily protein	4B	46207539	46212019
TRITD4Bv1G018870	Digeranylgeranylglyceryl phosphate synthase G	4B	47503402	47504921
TRITD4Bv1G018890	1-aminocyclopropane-1-carboxylate synthase 1	4B	47580213	47582033
TRITD4Bv1G019410	ethylene-dependent gravitropism-deficient and yellow-green-like 2 G	4B	49309986	49312043
TRITD4Bv1G019440	Ribosomal RNA small subunit methyltransferase E	4B	49413560	49416471
TRITD4Bv1G019450	Protein kinase family protein	4B	49416675	49419176
TRITD4Bv1G019550	Werner Syndrome-like exonuclease	4B	49528598	49529667
TRITD4Bv1G019610	1-deoxy-D-xylulose-5-phosphate synthase G	4B	49705678	49706052
TRITD4Bv1G019740	Growth-regulating factor	4B	50273283	50276701
TRITD4Bv1G019850	Thioredoxin-like protein AAED1, chloroplastic	4B	50627928	50629584
TRITD4Bv1G019930	Ribosomal RNA small subunit methyltransferase F	4B	50981004	50987925
TRITD4Bv1G019940	Lecithin-cholesterol acyltransferase-like 1	4B	50990193	50991726
TRITD4Bv1G019950	Serine carboxypeptidase family protein, expressed	4B	50992563	50996109
TRITD4Bv1G019960	Lecithin-cholesterol acyltransferase-like 1 G	4B	51005271	51005621
TRITD4Bv1G019990	F-box protein	4B	51247552	51249933
TRITD4Bv1G020030	Serine carboxypeptidase family protein, expressed	4B	51529942	51534365
TRITD4Bv1G020050	Core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein	4B	51567001	51568594
TRITD4Bv1G020060	Serine carboxypeptidase family protein, expressed	4B	51614914	51641108
TRITD4Bv1G020800	Protein MIZU-KUSSEI 1	4B	54262917	54263609
TRITD4Bv1G020840	Kinase, putative	4B	54290989	54294188
TRITD4Bv1G020940		4B	54342299	54346040
TRITD4Bv1G020960		4B	54512272	54514500
TRITD4Bv1G021030	GRF1-interacting factor-like protein	4B	54853020	54855818
TRITD4Bv1G021400	Dehydrin	4B	55707461	55708928

TRITD4Bv1G021460	cDNA clone:J013058P10, full insert sequence G	4B	55818928	55825009
TRITD4Bv1G021480	Protein phosphatase 2C family protein G	4B	55834231	55834567
TRITD4Bv1G021490	DNAJ heat shock N-terminal domain-containing protein G	4B	55836911	55837324
TRITD4Bv1G152530	F-box protein	4B	524066169	524075815
TRITD4Bv1G152550	RING/U-box superfamily protein	4B	524077212	524078938
TRITD4Bv1G152590	BTB/POZ domain-containing protein	4B	524203954	524206220
TRITD4Bv1G152640	carboxyl-terminal peptidase (DUF239)	4B	524328810	524330815
TRITD4Bv1G153010	DCD (Development and cell death) domain protein	4B	525565624	525568943
TRITD4Bv1G153070	NADH-cytochrome b5 reductase G	4B	525893362	525894866
TRITD4Bv1G153160	BTB/POZ domain-containing protein	4B	526144834	526148537
TRITD4Bv1G153190	Serine/threonine-protein phosphatase 7 long form-like protein	4B	526220370	526222540
TRITD4Bv1G153210	DNA repair protein UVH3 G	4B	526232969	526240661
TRITD4Bv1G153220	U3 small nucleolar RNA-associated protein 18-like protein	4B	526251964	526258347
TRITD4Bv1G153280	basic helix-loop-helix (bHLH) DNA-binding superfamily protein	4B	526335498	526336868
TRITD4Bv1G153320	Kinase-like	4B	526480611	526481429
TRITD4Bv1G153340	Kinase-like	4B	526492875	526494047
TRITD4Bv1G162650	DUF639 family protein	4B	555067354	555071198
TRITD4Bv1G162700	Exosome component 10	4B	555087552	555089555
TRITD4Bv1G162880	weak chloroplast movement under blue light protein (DUF827) G	4B	555752084	555809396
TRITD4Bv1G162900	B3 domain-containing protein	4B	555804746	555805476
TRITD4Bv1G162960	Photosynthetic NDH subcomplex B 2	4B	556037365	556038556
TRITD4Bv1G163000	Glutathione S-transferase	4B	556047381	556053320
TRITD4Bv1G163270	Subtilisin-like protease	4B	556963066	556966787
TRITD4Bv1G163300	Kinase family protein	4B	556977579	556979961
TRITD4Bv1G163370	Sporamin A G	4B	557353467	557360456
TRITD4Bv1G163410	Transcription repressor OFP12	4B	557458696	557459310
TRITD4Bv1G163570	High affinity cationic amino acid transporter 1 G	4B	557704249	557705019
TRITD4Bv1G163580	Late embryogenesis abundant D-like protein	4B	557706556	557707565
TRITD4Bv1G163730	Pentatricopeptide repeat-containing protein	4B	558021485	558023581
TRITD4Bv1G164470	Serine/threonine-protein kinase	4B	560274070	560282334
TRITD4Bv1G164550	P-loop containing nucleoside triphosphate hydrolases superfamily protein	4B	560453133	560457708
TRITD4Bv1G164850	Type I inositol-1,4,5-trisphosphate 5-phosphatase CVP2	4B	561871775	561874485
TRITD4Bv1G165010	Sulfate transporter	4B	562458695	562464000
TRITD4Bv1G165020	Kinase interacting (KIP1-like) family protein G	4B	562464721	562473764
TRITD4Bv1G165200	IQ domain-containing protein	4B	563187633	563189296
TRITD4Bv1G165540	MADS box transcription factor	4B	564165590	564198273
TRITD4Bv1G165610	Pectinesterase	4B	564416920	564417231
TRITD4Bv1G165670	1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase	4B	564572442	564575272
TRITD4Bv1G165680	Heat shock transcription factor	4B	564577723	564579889
TRITD4Bv1G173660	Ethylene-responsive transcription factor	4B	589855470	589856306

TRITD4Bv1G173700	Ethylene-responsive transcription factor	4B	589989694	590130409
TRITD4Bv1G173810	Ethylene-responsive transcription factor	4B	590205823	590206704
TRITD4Bv1G173970	Receptor-like protein kinase	4B	590618583	590624116
TRITD4Bv1G173980	Plasma membrane ATPase	4B	590624690	590629181
TRITD4Bv1G174000	Phytoene desaturase	4B	590725185	590729523
TRITD4Bv1G174830	Transmembrane protein 131	4B	593147147	593152070
TRITD4Bv1G175170	Dentin sialophosphoprotein-related, putative isoform 1	4B	594391936	594401211
TRITD4Bv1G175340	WD and tetratricopeptide repeat protein, putative	4B	594768288	594776658
TRITD4Bv1G175350	Zinc finger protein	4B	594780445	594781801
TRITD4Bv1G180650	Benzyl alcohol O-benzoyltransferase	4B	609130544	609132022
TRITD4Bv1G180690	Sulfotransferase	4B	609248904	609249999
TRITD4Bv1G180790	LOB domain-containing protein, putative	4B	609442225	609442788
TRITD4Bv1G180800	LOB domain-containing protein, putative	4B	609451308	609452174
TRITD4Bv1G180870	HR-like lesion-inducing protein-related protein	4B	609584823	609587615
TRITD4Bv1G180900	WAT1-related protein	4B	609623109	609666208
TRITD4Bv1G180910	Protein KINESIN LIGHT CHAIN-RELATED 3	4B	609672414	609675679
TRITD4Bv1G181190	Ethylene-responsive transcription factor	4B	610496047	610496415
TRITD4Bv1G181280	Phosphate transporter protein	4B	610606059	610607830
TRITD4Bv1G181310	Phosphate transporter protein	4B	610647818	610649395
TRITD4Bv1G181440	Zinc finger protein	4B	611015585	611024865
TRITD4Bv1G181480	WD40 repeat-like protein	4B	611161202	611166812
TRITD4Bv1G181960	Diacylglycerol kinase	4B	612677347	612680211
TRITD4Bv1G182080	DET1-and DDB1-associated protein 1	4B	612749732	612753816
TRITD4Bv1G182160	disease resistance protein (TIR-NBS-LRR class)	4B	613046204	613050761
TRITD4Bv1G182210	auxin response factor 1 G	4B	613148411	613150820
TRITD4Bv1G182270	C-terminal binding protein AN G	4B	613209467	613212554
TRITD4Bv1G182280	Chaperone protein dnaJ	4B	613218344	613220690
TRITD4Bv1G182310	S-adenosyl-L-methionine-dependent methyltransferase domain-containing protein	4B	613227051	613231844
TRITD4Bv1G182320	S-adenosyl-L-methionine-dependent methyltransferase domain-containing protein	4B	613234817	613237900
TRITD5Av1G018160	RNA helicase	5A	40081140	40082528
TRITD5Av1G018170	Eukaryotic translation initiation factor 3 subunit A	5A	40085042	40087781
TRITD5Av1G018420	Protein TRIGALACTOSYLDIACYLGLYCEROL 2, chloroplastic	5A	40660898	40663946
TRITD5Av1G018430	BTB/POZ domain-containing protein	5A	40668036	40671460
TRITD5Av1G018440	Histone-lysine N-methyltransferase	5A	40674333	40681194
TRITD5Av1G018510	Glutamate receptor	5A	40723921	40731666
TRITD5Av1G018520	Cationic amino acid transporter, putative	5A	40738945	40740866
TRITD5Av1G018530	Heavy metal transport/detoxification superfamily protein, putative	5A	40745485	40746593
TRITD5Av1G018560	Ankyrin repeat family protein	5A	40755308	40758238
TRITD5Av1G018890	Nuclear transcription factor Y subunit	5A	41696251	41700875
TRITD5Av1G018930	Homeobox leucine-zipper protein	5A	41807467	41814789

TRITD5Av1G018940	30S ribosomal protein S17	5A	41817899	41820012
TRITD5Av1G019110	Pectinesterase	5A	42129508	42133688
TRITD5Av1G019360	Ankyrin repeat family protein, putative	5A	42575306	42581417
TRITD5Av1G019430	Germin-like protein	5A	42682545	42683484
TRITD5Av1G019440	Zinc finger protein MAGPIE TE?	5A	42716525	42717103
TRITD5Av1G019450	Glutaredoxin family protein	5A	42774557	42774868
TRITD5Av1G019500	Pentatricopeptide repeat-containing protein	5A	42832989	42834854
TRITD5Av1G019540	Glutaredoxin family protein	5A	42872259	42872570
TRITD5Av1G019720	WPP domain-associated protein G	5A	43249900	43252932
TRITD5Av1G019730	Glutathione synthetase G	5A	43254026	43254909
TRITD5Av1G019750	Protoporphyrinogen oxidase	5A	43265254	43268592
TRITD5Av1G019830	RING/U-box superfamily protein	5A	43309827	43310315
TRITD5Av1G019900	Defensin protein	5A	43389525	43389926
TRITD5Av1G020020	nodulin MtN21 /EamA-like transporter family protein G	5A	43625004	43625234
TRITD5Av1G020290	ATP-dependent Clp protease ATP-binding subunit ClpB	5A	43885116	43886040
TRITD5Av1G020360	Cullin-associated NEDD8-dissociated protein 1	5A	44017365	44027510
TRITD5Av1G020460	cDNA clone:J013058P10, full insert sequence G	5A	44191911	44197803
TRITD5Av1G020500	BAG family molecular chaperone regulator 7 G	5A	44256955	44258645
TRITD5Av1G020640	DEAD-box ATP-dependent RNA helicase-like protein TE?	5A	44517425	44520969
TRITD5Av1G021050	BTB/POZ domain containing protein	5A	45168543	45169511
TRITD5Av1G021520	Valine--tRNA ligase	5A	46138975	46156855
TRITD5Av1G021560		5A	46162955	46165863
TRITD5Av1G021570	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	5A	46168609	46173149
TRITD5Av1G021600	Peptide transporter	5A	46208532	46214109
TRITD5Av1G021630	Stomatal closure-related actin-binding protein 1	5A	46469561	46475550
TRITD5Av1G021750	Receptor-like kinase	5A	46813871	46816203
TRITD5Av1G021770	Tuftelin-interacting protein 11	5A	46816590	46820362
TRITD5Av1G021790	Zinc finger protein LSD1	5A	46843995	46846717
TRITD5Av1G021830	NAC domain protein G	5A	46851976	46856331
TRITD5Av1G030000	Long chain acyl-CoA synthetase 1 G	5A	67557552	67558012
TRITD5Av1G030020	F-box family protein	5A	67644299	67645230
TRITD5Av1G030360	F-box protein family-like protein TE?	5A	68487775	68488485
TRITD5Av1G030580	Octanoyltransferase	5A	69093139	69094679
TRITD5Av1G030590	Receptor-like kinase	5A	69094828	69104162
TRITD5Av1G033370	Receptor protein kinase, putative	5A	76135924	76139625
TRITD5Av1G033530	Zinc finger CCCH domain-containing protein 62	5A	76481388	76485918
TRITD5Av1G033660	B3 domain-containing protein	5A	76788098	76792270
TRITD5Av1G033690	B3 domain-containing protein	5A	76802565	76805384
TRITD5Av1G033780	B3 domain-containing protein	5A	76978339	76981899
TRITD5Av1G033790	Hexosyltransferase	5A	76985684	76998342
TRITD5Av1G033840	Pentatricopeptide repeat-containing protein, putative	5A	77037726	77040656

TRITD5Av1G033850	Kinesin-like protein	5A	77046073	77059334
TRITD5Av1G034340	Disease resistance protein (NBS-LRR class) family	5A	78463706	78464575
TRITD5Av1G034350	Disease resistance protein	5A	78468984	78471904
TRITD5Av1G034460	F-box family protein	5A	78819616	78874035
TRITD5Av1G034780	Cytoplasmic tRNA 2-thiolation protein 2	5A	79240247	79242485
TRITD5Av1G034850	Calcium-transporting ATPase	5A	79417181	79421699
TRITD5Av1G034860	Protein misato-like protein 1	5A	79423712	79427327
TRITD5Av1G035130	FBD-associated F-box protein	5A	79983739	79985350
TRITD5Av1G035150	F-box domain containing protein, expressed	5A	79989726	79991373
TRITD5Av1G043450	Foldase protein prsA 1 G	5A	108456296	108464512
TRITD5Av1G043460	Purple acid phosphatase	5A	108480943	108483950
TRITD5Av1G043640	Importin subunit beta-1	5A	109195083	109198939
TRITD5Av1G043650	CASP-like protein	5A	109225762	109231299
TRITD5Av1G043790	50S ribosomal protein L2	5A	110040576	110042023
TRITD5Av1G043900	Receptor-like protein kinase	5A	110345771	110358296
TRITD5Av1G044080	ATP-dependent RNA helicase TE?	5A	110796564	110797753
TRITD5Av1G166120	Chymotrypsin inhibitor	5A	454003738	454003944
TRITD5Av1G166200	cDNA clone:J013058P10, full insert sequence G	5A	454143874	454153077
TRITD5Av1G166310	tRNA-splicing endonuclease subunit Sen54	5A	454206493	454210333
TRITD5Av1G166370	tRNA-splicing endonuclease subunit SEN54	5A	454224099	454225974
TRITD5Av1G166380	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	5A	454228655	454229893
TRITD5Av1G166410	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	5A	454280406	454281679
TRITD5Av1G166470	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	5A	454432850	454434114
TRITD5Av1G166510	Transcription factor GTE8	5A	454479675	454480703
TRITD5Av1G166760	Lipid transfer protein	5A	454862405	454863197
TRITD5Av1G166780	F-box domain containing protein TE?	5A	454876188	454877309
TRITD5Av1G166800	Ubiquitinyl hydrolase 1	5A	454945275	454949680
TRITD5Av1G167000	Auxin efflux carrier component	5A	455486558	455490389
TRITD5Av1G167010	F-box protein	5A	455535409	455536323
TRITD5Av1G167350	Receptor kinase-like protein	5A	456045496	456048676
TRITD5Av1G167410	Uridine kinase	5A	456319366	456324591
TRITD5Av1G167460	PP2A regulatory subunit TAP46	5A	456454840	456455907
TRITD5Av1G167480	Auxin efflux carrier component	5A	456516519	456639082
TRITD5Av1G167530	Purple acid phosphatase	5A	456685324	456690791
TRITD5Av1G254490	Casein kinase II subunit beta	5A	659139866	659142516
TRITD5Av1G254520	Invertase inhibitor	5A	659144318	659145928
TRITD5Av1G254540	Ribose-phosphate pyrophosphokinase	5A	659149211	659155599
TRITD5Av1G254600	C2 domain-containing family protein	5A	659310709	659312143
TRITD5Av1G254640	glycosyltransferase family exostosin protein	5A	659362911	659364128
TRITD5Av1G254650	UPF0176 protein YceA	5A	659365426	659367322
TRITD5Av1G254660		5A	659369649	659370293

TRITD5Av1G254740	Protein FLOWERING LOCUS T	5A	659472125	659473571
TRITD5Av1G254750	Alkyl transferase	5A	659484970	659486089
TRITD5Av1G254760	Trichome birefringence-like protein	5A	659487779	659489841
TRITD5Av1G254790	Benzyl alcohol O-benzoyltransferase	5A	659504338	659505927
TRITD5Av1G254880	ATP-dependent chaperone ClpB	5A	659710431	659716133
TRITD5Av1G254930	RING/U-box superfamily protein	5A	659763544	659766459
TRITD5Av1G255180	Apyrase	5A	660355616	660398268
TRITD5Av1G255250	RING/U-box superfamily protein	5A	660449239	660451283
TRITD5Av1G255310	RING/U-box superfamily protein	5A	660591296	660595655
TRITD5Bv1G010460	IQ domain-containing protein	5B	27686806	27691202
TRITD5Bv1G010540	Tropinone reductase-like protein	5B	27826992	27828730
TRITD5Bv1G010670	GTP binding protein	5B	28062157	28065700
TRITD5Bv1G010680	Xyloglucan endotransglucosylase/hydrolase	5B	28071830	28072790
TRITD5Bv1G011050	9-cis-epoxycarotenoid dioxygenase	5B	29602428	29604272
TRITD5Bv1G032160	F-box family protein	5B	89148283	89149851
TRITD5Bv1G032340	Cytoplasmic tRNA 2-thiolation protein 2	5B	89521328	89523549
TRITD5Bv1G032380	Cyclin family protein	5B	89591676	89593141
TRITD5Bv1G032580	Calcium-transporting ATPase	5B	89918064	89922526
TRITD5Bv1G032590	Protein misato-like protein 1	5B	89924528	89928162
TRITD5Bv1G032620	Two-component response regulator	5B	89971951	89974004
TRITD5Bv1G032770	Cation/H(+) antiporter	5B	90391158	90393701
TRITD5Bv1G032850	FBD-associated F-box protein	5B	90499901	90501513
TRITD5Bv1G032960	Methionine--tRNA ligase	5B	90813020	90818496
TRITD5Bv1G033190	Lipid transfer protein	5B	91303295	91303612
TRITD5Bv1G033220	Protease inhibitor/seed storage/lipid transfer protein family protein	5B	91414307	91414624
TRITD5Bv1G033230	Protease inhibitor/seed storage/lipid transfer protein family protein	5B	91417538	91417855
TRITD5Bv1G033270	Nucleobase-ascorbate transporter	5B	91494804	91502729
TRITD5Bv1G033300	tRNA-splicing ligase RtcB homolog 2 G	5B	91778231	91778842
TRITD5Bv1G033340	Zinc finger protein, putative	5B	91843910	91844629
TRITD5Bv1G033490	BTB/POZ domain-containing protein	5B	92239061	92242905
TRITD5Bv1G033550	Aspartic proteinase nepenthesin-1	5B	92539548	92540864
TRITD5Bv1G092850	Nucleolar RNA binding protein	5B	273637638	273642314
TRITD5Bv1G093120	Mediator complex, subunit Med7 G	5B	274169846	274170334
TRITD5Bv1G093130	MATE efflux family protein G	5B	274170845	274171449
TRITD5Bv1G093610	Regulatory-associated protein of tor 1	5B	275188220	275198933
TRITD5Bv1G093620	Pentatricopeptide repeat-containing protein, putative	5B	275200252	275201286
TRITD5Bv1G093730	Glycosyltransferase	5B	275615800	275618246
TRITD5Bv1G093820	Plant cadmium resistance protein	5B	275748084	275749712
TRITD5Bv1G134790	Cyclic nucleotide-gated channel	5B	402097272	402099567
TRITD5Bv1G134810	DUF1666 family protein	5B	402106782	402111650
TRITD5Bv1G134860	Pentatricopeptide repeat protein	5B	402266897	402269362

TRITD5Bv1G134930	RING/FYVE/PHD zinc finger superfamily protein	5B	402331149	402334008
TRITD5Bv1G134980	Pentatricopeptide repeat-containing protein	5B	402408267	402409766
TRITD5Bv1G135160	Vacuolar-processing enzyme	5B	402710128	402711736
TRITD5Bv1G135170	Calcium-binding family protein	5B	402749235	402749921
TRITD5Bv1G135190	O-fucosyltransferase family protein	5B	402752188	402755421
TRITD5Bv1G135610	Chloroplast lysine N-methyltransferase	5B	404104331	404106478
TRITD5Bv1G135620	Pentatricopeptide repeat-containing protein	5B	404107377	404109329
TRITD5Bv1G135630	Sulfate transporter 1	5B	404113013	404116384
TRITD5Bv1G135640	Beta purothionin	5B	404117264	404117946
TRITD5Bv1G135700	Copper-transporting ATPase	5B	404244445	404252862
TRITD5Bv1G135850	NBS-LRR class disease resistance protein	5B	404844760	404849603
TRITD5Bv1G232040	Actin depolymerizing factor	5B	652021111	652021774
TRITD5Bv1G232160	Sec14 cytosolic factor	5B	652180465	652185116
TRITD5Bv1G232370	Dof zinc finger protein	5B	652614066	652615678
TRITD5Bv1G232500	F-box family protein TE?	5B	652827121	652827499
TRITD5Bv1G232740	arabinogalactan protein 15 G	5B	653222737	653223141
TRITD5Bv1G232770	Condensin-2 complex subunit D3	5B	653244721	653250624
TRITD5Bv1G232830	Disease resistance protein RGA2 G	5B	653368036	653369175
TRITD5Bv1G232840	Pentatricopeptide repeat-containing protein	5B	653373031	653374526
TRITD5Bv1G232980	F-box protein-like	5B	653610384	653611334
TRITD5Bv1G233070	Autoinducer 2-binding protein LsrB G	5B	653830242	653834173
TRITD5Bv1G233120	disease resistance family protein / LRR family protein	5B	653862056	653863909
TRITD5Bv1G233130	receptor kinase 1	5B	653864788	653867715
TRITD5Bv1G233140	disease resistance family protein / LRR family protein	5B	653876345	653881714
TRITD5Bv1G233150	NBS-LRR resistance-like protein	5B	653903658	653906883
TRITD6Av1G026740	Cytochrome P450 family protein, expressed	6A	63155455	63158475
TRITD6Av1G026800	Ethylene-responsive transcription factor	6A	63327812	63329797
TRITD6Av1G026860	Ethylene-responsive transcription factor	6A	63520916	63522849
TRITD6Av1G026890	Proteasome subunit alpha type	6A	63650337	63653232
TRITD6Av1G026920	Lipid transfer protein	6A	63727732	63728409
TRITD6Av1G026980	Phosphopantetheine adenylyltransferase	6A	64002118	64004734
TRITD6Av1G027130	evolutionarily conserved C-terminal region 4 G	6A	64230181	64230927
TRITD6Av1G027170	Kinetochore protein nuf2, putative	6A	64361409	64369780
TRITD6Av1G027180	Heat Stress Transcription Factor family protein	6A	64370791	64371909
TRITD6Av1G027190	Calcium lipid binding protein, putative	6A	64382427	64392540
TRITD6Av1G027380	Pentatricopeptide repeat-containing protein	6A	64893281	64894417
TRITD6Av1G027400	Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial	6A	64946910	64950528
TRITD6Av1G027570	Pentatricopeptide repeat-containing protein	6A	65334320	65343477
TRITD6Av1G027820	Kinase-like protein	6A	65927051	65930428
TRITD6Av1G027860	Esterase/lipase/thioesterase-like protein	6A	66036613	66041197
TRITD6Av1G027880	Werner Syndrome-like exonuclease	6A	66044171	66051200

TRITD6Av1G027890	Werner Syndrome-like exonuclease	6A	66046287	66046637
TRITD6Av1G027920	Esterase/lipase/thioesterase-like protein	6A	66066220	66069187
TRITD6Av1G027930	Esterase/lipase/thioesterase-like protein G	6A	66083681	66089893
TRITD6Av1G028030	F-box family protein	6A	66194012	66196384
TRITD6Av1G028150	Topless-related protein 2	6A	66632834	66638521
TRITD6Av1G028160	BURP domain protein RD22	6A	66641066	66645122
TRITD6Av1G028330	RNA-binding family protein	6A	66975207	66978934
TRITD6Av1G028390	DegP protease-like	6A	67076910	67078885
TRITD6Av1G028710	Pentatricopeptide repeat-containing protein	6A	67611219	67612730
TRITD6Av1G028730	Signal recognition particle protein	6A	67615240	67618107
TRITD6Av1G028840	transmembrane protein, putative (DUF594)	6A	67827850	67829517
TRITD6Av1G028880	F-box domain containing protein TE?	6A	67838972	67840159
TRITD6Bv1G033140	FBD-associated F-box protein	6B	91670541	91672284
TRITD6Bv1G033160	Lil3 protein	6B	91763794	91765663
TRITD6Bv1G033180	Neutral/alkaline invertase	6B	91783025	91786751
TRITD6Bv1G033200	Poly(RC)-binding protein, putative	6B	91878187	91882888
TRITD6Bv1G033210	Adenylate kinase, putative	6B	91923074	91931773
TRITD6Bv1G033230	Cytochrome P450	6B	91948012	91949899
TRITD6Bv1G033240	Cytochrome P450	6B	91990578	91991700
TRITD6Bv1G033250	Pre-rRNA-processing protein ESF1 G	6B	91998128	92000384
TRITD6Bv1G033260	Pre-rRNA-processing protein ESF1 G	6B	92001821	92002402
TRITD6Bv1G033550		6B	92762102	92762410
TRITD6Bv1G033580	Bax inhibitor 1-like protein	6B	92821295	92825635
TRITD6Bv1G033630	AT hook motif DNA-binding family protein	6B	92892213	92896147
TRITD6Bv1G033670	Disease resistance protein RPM1	6B	93051607	93056261
TRITD6Bv1G033680	Pentatricopeptide repeat-containing protein	6B	93059180	93063126
TRITD6Bv1G033710	Aquaporin	6B	93074981	93075868
TRITD6Bv1G033830	Transducin family protein / WD-40 repeat family protein	6B	93364119	93379553
TRITD6Bv1G058460	Serine/threonine-protein phosphatase 7 long form-like protein	6B	164801334	164811805
TRITD6Bv1G058480	Leucine-rich repeat receptor-like protein kinase family protein, putative	6B	164853240	164855009
TRITD6Bv1G058500	U3 small nucleolar RNA-associated protein 18-like protein	6B	164857217	164858482
TRITD6Bv1G058550	Trypsin family protein with PDZ domain-containing protein	6B	164878691	164885201
TRITD6Bv1G058560	Leucine-rich repeat receptor-like protein kinase family protein, putative	6B	165006413	165009334
TRITD6Bv1G058570	Leucine-rich repeat receptor-like protein kinase family protein, putative	6B	165011532	165012149
TRITD6Bv1G058580	Leucine-rich repeat receptor-like protein kinase family	6B	165028941	165031733
TRITD6Bv1G058700	Leucine-rich repeat receptor-like protein kinase family protein	6B	165531597	165533447
TRITD6Bv1G058720	Peroxidase	6B	165533970	165540697
TRITD6Bv1G058750	Polyubiquitin	6B	165549158	165819561
TRITD7Av1G216820	F-box family protein	7A	578535970	578537070
TRITD7Av1G217110	1-aminocyclopropane-1-carboxylate synthase 9	7A	579263912	579267220

TRITD7Av1G217190	BTB/POZ/MATH-domain protein	7A	579688388	579689479
TRITD7Av1G217280	C2 calcium/lipid-binding and GRAM domain protein	7A	579992630	579995071
TRITD7Av1G249550	minichromosome maintenance (MCM2/3/5) family protein	7A	660040345	660042435
TRITD7Av1G249860	Embryo yellow protein	7A	660511896	660517211
TRITD7Av1G249890	Poly(A) RNA polymerase protein 2	7A	660575634	660586094
TRITD7Av1G250090	Pentatricopeptide repeat-containing protein	7A	660832948	660835029
TRITD7Av1G250380	BTB/POZ domain containing protein, expressed	7A	661308649	661309737
TRITD7Av1G250400	BTB/POZ domain containing protein	7A	661325665	661326753
TRITD7Av1G250410		7A	661330214	661330840
TRITD7Av1G250490	BTB/POZ domain containing protein, expressed	7A	661460270	661614679
TRITD7Av1G250500	disease resistance protein (TIR-NBS-LRR class)	7A	661611270	661612418
TRITD7Av1G250610	BTB/POZ domain containing protein, expressed	7A	661858690	661861297
TRITD7Av1G250620	BTB/POZ and MATH domain-containing protein 1	7A	661862430	661863539
TRITD7Av1G250650	BTB/POZ/MATH-domain protein G	7A	661902886	661904464
TRITD7Av1G250710	BTB/POZ domain containing protein, expressed	7A	662057794	662062545
TRITD7Av1G250730	BTB/POZ domain containing protein	7A	662098726	662100438
TRITD7Av1G250810	cyclin delta-3 G	7A	662316580	662317186
TRITD7Av1G250820	Inorganic pyrophosphatase 2	7A	662328133	662329337
TRITD7Av1G250920	DUF674 family protein	7A	662520095	662520896
TRITD7Av1G250930	Germin-like protein	7A	662535750	662536548
TRITD7Av1G251020	FBD-associated F-box protein	7A	662723788	662726103
TRITD7Av1G251090	Translation initiation factor IF-2 G	7A	662802795	662804462
TRITD7Av1G251110	Phosphoglycerate kinase	7A	662833745	662840879
TRITD7Av1G251120	carboxyl-terminal peptidase (DUF239)	7A	662872748	662875056
TRITD7Av1G251170	BTB/POZ domain containing protein, expressed	7A	662959624	662960697
TRITD7Av1G258720	SQUAMOSA promoter-binding protein-like protein	7A	678556723	678559162
TRITD7Av1G258840	Squamosa promoter binding protein-like protein	7A	679361439	679366960
TRITD7Av1G258870	Squamosa promoter binding protein-like protein	7A	679498486	679500999
TRITD7Av1G258970	Zinc finger CCCH domain-containing protein 4	7A	679838197	679838940
TRITD7Av1G258990	Homogentisate phytyltransferase	7A	679841147	679846173
TRITD7Av1G259130	4-coumarate--CoA ligase 2 G	7A	680101653	680103429
TRITD7Av1G259170	Mitochondrial carrier like protein	7A	680174252	680176674
TRITD7Av1G259240	AP2-like ethylene-responsive transcription factor	7A	680287955	680290609
TRITD7Av1G259290	DUF247 domain protein	7A	680393848	680394234
TRITD7Av1G259330	4-coumarate:CoA ligase	7A	680419022	680424646
TRITD7Av1G259370	Oxygen-dependent choline dehydrogenase	7A	680485867	680489880
TRITD7Av1G262910	Magnesium transporter MRS2-like protein G	7A	687170543	687175861
TRITD7Av1G262930	Lectin	7A	687190734	687191690
TRITD7Av1G262940	BTB/POZ domain containing protein	7A	687199174	687200352
TRITD7Av1G262960	Histone-lysine N-methyltransferase	7A	687219647	687226216
TRITD7Av1G263030	Protein kinase	7A	687379300	687381462

TRITD7Av1G263050	Protein kinase	7A	687410141	687411691
TRITD7Av1G263140	Protein kinase	7A	687536289	687538118
TRITD7Av1G263160	Protein kinase	7A	687561712	687562773
TRITD7Av1G263210	Protein kinase	7A	687596618	687598684
TRITD7Av1G263230	Carboxypeptidase	7A	687690521	687692325
TRITD7Av1G263240	Transcription elongation factor (TFIIS) family protein, putative	7A	687719135	687722131
TRITD7Av1G263250	DNA-3-methyladenine glycosylase, putative	7A	687728972	687732885
TRITD7Av1G263820	Wd40-repeat-like protein	7A	689302231	689311733
TRITD7Av1G263880	Citrate-binding protein	7A	689465424	689466219
TRITD7Bv1G030170	Galactoside 2-alpha-L-fucosyltransferase	7B	85023440	85025260
TRITD7Bv1G030300	Galactoside 2-alpha-L-fucosyltransferase	7B	85522459	85533440
TRITD7Bv1G030310	Galactoside 2-alpha-L-fucosyltransferase	7B	85550511	85552372
TRITD7Bv1G030450	Serpin family protein	7B	85847348	85849543
TRITD7Bv1G030490	Galactoside 2-alpha-L-fucosyltransferase	7B	85872129	85873912
TRITD7Bv1G030500	Aldose 1-epimerase	7B	85878602	85879533
TRITD7Bv1G030550		7B	85930084	85930725
TRITD7Bv1G030640	Transmembrane protein 53	7B	86373453	86377703
TRITD7Bv1G030760	Cyclin-dependent kinase inhibitor	7B	86522205	86523383
TRITD7Bv1G030890	NAD(P)H dehydrogenase C1	7B	86860038	86867273
TRITD7Bv1G030900	RING/U-box superfamily protein	7B	86869571	86874223
TRITD7Bv1G030920	Scarecrow transcription factor family protein	7B	86891739	86895811
TRITD7Bv1G056080	Glycosyltransferase	7B	156462428	156463825
TRITD7Bv1G056240	Thioredoxin-like family protein	7B	156792804	156793293
TRITD7Bv1G056250	CDK5RAP3-like protein	7B	156793465	156796722
TRITD7Bv1G056500	Seed maturation protein LEA 4	7B	157854628	157855010
TRITD7Bv1G172730	O-glucosyltransferase rumi like	7B	546030307	546036749
TRITD7Bv1G172890	carboxyl-terminal peptidase (DUF239)	7B	546792050	546794150
TRITD7Bv1G172950	Galactokinase	7B	546907947	546916432
TRITD7Bv1G173020	Type 1 membrane protein-like G	7B	547065347	547068145
TRITD7Bv1G173030	Disease resistance protein RPM1	7B	547070248	547073599
TRITD7Bv1G173200	Auxin response factor	7B	547633254	547641282
TRITD7Bv1G173260	Protein MID1-COMPLEMENTING ACTIVITY 1 G	7B	547800009	547800761
TRITD7Bv1G173290	3-phosphoinositide-dependent protein kinase-1-like G	7B	547962424	547965500
TRITD7Bv1G220030	Eukaryotic aspartyl protease family protein	7B	685001684	685003075
TRITD7Bv1G220130	Cytochrome P450	7B	685128478	685133332
TRITD7Bv1G220150	Argonaute	7B	685165723	685168076
TRITD7Bv1G220200	Phox (PX) domain-containing protein	7B	685304450	685309670
TRITD7Bv1G220220	Ubiquitin system component Cue protein, putative	7B	685317256	685319809
TRITD7Bv1G220270	Cytochrome P450 family protein, expressed	7B	685326836	685328356
TRITD7Bv1G220340	Cytochrome P450 family protein, expressed	7B	685512921	685514557
TRITD7Bv1G220380	Cytochrome P450 family protein, expressed	7B	685585857	685587386

TRITD7Bv1G220390	Ubiquitin system component Cue protein G	7B	685662538	685665001
TRITD7Bv1G220420	Cytochrome P450, putative	7B	685699937	685701532
TRITD7Bv1G220450	Cytochrome P450 family protein, expressed	7B	685843395	685846203
TRITD7Bv1G220460	Cytochrome P450 family protein	7B	685876223	685879318
TRITD7Bv1G220640	Cytochrome P450 family protein, expressed	7B	686158028	686159704
TRITD7Bv1G220650	Cytochrome P450 family protein, expressed	7B	686174320	686175957
TRITD7Bv1G220690	Elongator complex protein 4	7B	686298047	686300486
TRITD7Bv1G220700	Protein kinase, putative	7B	686301118	686303699
TRITD7Bv1G220840	Endo-1,31,4-beta-D-glucanase	7B	686788223	686790933
TRITD7Bv1G220870		7B	686887266	686887691
TRITD7Bv1G220910	Methyltransferase family protein	7B	686951348	686953465

S. Table 4: Introgression from WSL to DSE; genomic windows showing closer relationship between WSL and DSE than between WNL and DSE, and high genetic distance between DSE and DNW.

Chrom	start	end	sites	Dxy WSL-DSE	Dxy WNL-DSE	Dxy DSE-DNW
1B	307000001	309000000	459	0.0939	0.1192	0.5749
2A	197000001	199000000	2147	0.1387	0.1648	0.4846
2A	198000001	200000000	2489	0.1405	0.1601	0.4255
2B	669000001	671000000	364	0.3218	0.3595	0.3327
3B	148000001	150000000	1981	0.2197	0.3396	0.3536
3B	223000001	225000000	131	0.4258	0.4654	0.3337
4B	95000001	97000000	1635	0.2605	0.2984	0.4005
4B	96000001	98000000	2207	0.2436	0.2745	0.444
4B	97000001	99000000	1753	0.1967	0.2733	0.4285
4B	107000001	109000000	564	0.1813	0.2011	0.338
4B	189000001	191000000	816	0.1658	0.3369	0.3576
4B	190000001	192000000	653	0.1498	0.413	0.442
4B	191000001	193000000	661	0.1351	0.4125	0.4388
4B	192000001	194000000	811	0.1389	0.4286	0.4569
4B	193000001	195000000	835	0.1419	0.4155	0.4466
4B	194000001	196000000	753	0.138	0.4264	0.456
4B	195000001	197000000	631	0.1309	0.4451	0.4735
4B	196000001	198000000	633	0.1309	0.407	0.4349
4B	197000001	199000000	756	0.1422	0.3999	0.4237

4B	198000001	200000000	763	0.1373	0.4183	0.4455
4B	199000001	201000000	847	0.1282	0.4219	0.4554
4B	200000001	202000000	990	0.1417	0.407	0.4379
4B	201000001	203000000	942	0.1509	0.4023	0.4304
4B	202000001	204000000	725	0.1356	0.4137	0.4399
4B	203000001	205000000	734	0.1264	0.4309	0.4588
4B	204000001	206000000	742	0.1323	0.4422	0.4651
4B	205000001	207000000	670	0.1379	0.4228	0.4384
4B	206000001	208000000	718	0.1545	0.3851	0.4083
4B	207000001	209000000	790	0.1433	0.3584	0.3869
4B	208000001	210000000	935	0.1402	0.3625	0.3872
4B	209000001	211000000	1029	0.1347	0.4043	0.4303
4B	210000001	212000000	910	0.1282	0.4267	0.4554
4B	211000001	213000000	921	0.1356	0.3986	0.4275
4B	212000001	214000000	959	0.1392	0.3984	0.4231
4B	213000001	215000000	974	0.1536	0.39	0.4161
4B	214000001	216000000	1079	0.15	0.3901	0.4162
4B	215000001	217000000	831	0.1482	0.3947	0.4172
4B	216000001	218000000	1026	0.146	0.3913	0.4204
4B	217000001	219000000	1147	0.1353	0.4161	0.4446
4B	218000001	220000000	681	0.1314	0.4101	0.4375
4B	219000001	221000000	733	0.1312	0.4074	0.4359
4B	220000001	222000000	1005	0.1324	0.4168	0.447
4B	221000001	223000000	869	0.1454	0.4064	0.4331
4B	222000001	224000000	664	0.1433	0.4139	0.4399
4B	223000001	225000000	563	0.1346	0.4168	0.4424
4B	224000001	226000000	676	0.1447	0.4134	0.4445
4B	225000001	227000000	969	0.1454	0.4235	0.4586
4B	226000001	228000000	1120	0.1375	0.4156	0.4465
4B	227000001	229000000	982	0.1349	0.4003	0.4313
4B	228000001	230000000	987	0.1451	0.4083	0.4386

4B	229000001	231000000	1165	0.1461	0.4018	0.4266
4B	230000001	232000000	1124	0.1458	0.4027	0.4265
4B	231000001	233000000	906	0.1468	0.4049	0.4291
4B	232000001	234000000	880	0.1356	0.402	0.4289
4B	233000001	235000000	983	0.1313	0.4125	0.4369
4B	234000001	236000000	804	0.1356	0.4197	0.4425
4B	235000001	237000000	675	0.1478	0.3988	0.4309
4B	236000001	238000000	666	0.147	0.4099	0.4424
4B	237000001	239000000	434	0.1481	0.4232	0.4532
4B	238000001	240000000	495	0.1403	0.4005	0.4267
4B	239000001	241000000	906	0.1343	0.3932	0.4195
4B	240000001	242000000	1102	0.1393	0.3824	0.4075
4B	241000001	243000000	1258	0.1385	0.3926	0.4244
4B	242000001	244000000	1086	0.1358	0.4155	0.4487
4B	243000001	245000000	874	0.1373	0.4172	0.4493
4B	244000001	246000000	879	0.1384	0.4154	0.4493
4B	245000001	247000000	965	0.1364	0.4186	0.4422
4B	246000001	248000000	1045	0.1334	0.4136	0.4325
4B	247000001	249000000	915	0.1243	0.3887	0.4116
4B	248000001	250000000	836	0.1268	0.3802	0.4072
4B	249000001	251000000	838	0.1331	0.4157	0.4347
4B	250000001	252000000	676	0.1462	0.373	0.3856
4B	251000001	253000000	620	0.1562	0.3457	0.3658
4B	252000001	254000000	804	0.1405	0.4124	0.4367
4B	253000001	255000000	1015	0.1404	0.4293	0.4581
4B	254000001	256000000	1130	0.1441	0.4033	0.4367
4B	255000001	257000000	895	0.1406	0.4055	0.4359
4B	256000001	258000000	707	0.1464	0.4054	0.4219
4B	257000001	259000000	732	0.145	0.4006	0.42
4B	258000001	260000000	530	0.1324	0.421	0.4473
4B	259000001	261000000	256	0.1448	0.381	0.4089

4B	260000001	262000000	410	0.1462	0.4011	0.4254
4B	261000001	263000000	724	0.1349	0.433	0.4548
4B	262000001	264000000	797	0.1352	0.4537	0.4778
4B	263000001	265000000	857	0.1362	0.4261	0.4545
4B	264000001	266000000	936	0.1437	0.3959	0.4216
4B	265000001	267000000	1040	0.1511	0.3913	0.4166
4B	266000001	268000000	934	0.1376	0.3943	0.4242
4B	267000001	269000000	761	0.1331	0.377	0.4032
4B	268000001	270000000	745	0.1474	0.3715	0.3971
4B	269000001	271000000	763	0.1376	0.4018	0.4334
4B	270000001	272000000	851	0.1363	0.4057	0.4355
4B	271000001	273000000	792	0.1497	0.4016	0.4281
4B	272000001	274000000	557	0.1529	0.3877	0.4128
4B	273000001	275000000	595	0.1486	0.3822	0.4058
4B	274000001	276000000	943	0.1388	0.4135	0.4366
4B	275000001	277000000	776	0.1372	0.4357	0.4643
4B	276000001	278000000	257	0.1454	0.4508	0.4938
4B	279000001	281000000	225	0.119	0.4521	0.4905
4B	280000001	282000000	462	0.1244	0.4184	0.4506
4B	281000001	283000000	559	0.135	0.4149	0.444
4B	282000001	284000000	872	0.137	0.4196	0.4495
4B	283000001	285000000	901	0.1341	0.3908	0.4161
4B	284000001	286000000	963	0.1394	0.3572	0.3791
4B	285000001	287000000	929	0.1386	0.3787	0.4035
4B	286000001	288000000	764	0.1362	0.4113	0.4399
4B	287000001	289000000	658	0.1342	0.3959	0.4262
4B	288000001	290000000	654	0.1344	0.3683	0.4024
4B	289000001	291000000	1102	0.1409	0.3844	0.4192
4B	290000001	292000000	1170	0.1386	0.4009	0.4327
4B	291000001	293000000	1180	0.1391	0.427	0.4547
4B	292000001	294000000	1100	0.1457	0.4191	0.4471

4B	293000001	295000000	907	0.1469	0.3709	0.3984
4B	294000001	296000000	926	0.1384	0.3993	0.42
4B	295000001	297000000	695	0.14	0.4153	0.4239
4B	296000001	298000000	515	0.145	0.3799	0.3877
4B	297000001	299000000	741	0.1373	0.4208	0.4412
4B	298000001	300000000	869	0.1339	0.4325	0.4566
4B	299000001	301000000	992	0.1302	0.4432	0.4662
4B	300000001	302000000	1176	0.1382	0.416	0.4385
4B	301000001	303000000	1059	0.1391	0.3904	0.4192
4B	302000001	304000000	892	0.1308	0.4226	0.4518
4B	303000001	305000000	882	0.1379	0.4022	0.427
4B	304000001	306000000	859	0.1389	0.3781	0.4024
4B	305000001	307000000	911	0.1526	0.4025	0.4194
4B	306000001	308000000	880	0.1573	0.4055	0.4223
4B	307000001	309000000	867	0.1489	0.4042	0.4287
4B	308000001	310000000	1052	0.1441	0.4073	0.4344
4B	309000001	311000000	1143	0.1442	0.4089	0.4347
4B	310000001	312000000	1010	0.1435	0.3938	0.4192
4B	311000001	313000000	621	0.1357	0.3867	0.4148
4B	312000001	314000000	475	0.1598	0.4055	0.4291
4B	313000001	315000000	582	0.1653	0.3856	0.3995
4B	314000001	316000000	539	0.1475	0.4114	0.4244
4B	315000001	317000000	840	0.1389	0.4385	0.4666
4B	316000001	318000000	1023	0.1377	0.4198	0.4453
4B	317000001	319000000	925	0.1312	0.3981	0.4177
4B	318000001	320000000	1107	0.1378	0.4074	0.4365
4B	319000001	321000000	1091	0.1336	0.4334	0.4603
4B	320000001	322000000	890	0.1407	0.4473	0.4718
4B	321000001	323000000	639	0.1488	0.4262	0.464
4B	322000001	324000000	560	0.1316	0.4121	0.4497
4B	323000001	325000000	769	0.147	0.4061	0.4376

4B	324000001	326000000	849	0.1445	0.4081	0.4347
4B	325000001	327000000	898	0.1255	0.4531	0.4838
4B	326000001	328000000	855	0.1236	0.4476	0.478
4B	327000001	329000000	457	0.1316	0.3951	0.4202
4B	328000001	330000000	658	0.1389	0.4113	0.4447
4B	329000001	331000000	910	0.1464	0.4018	0.4314
4B	330000001	332000000	861	0.1484	0.3738	0.4087
4B	331000001	333000000	912	0.1372	0.3926	0.4332
4B	332000001	334000000	866	0.1264	0.4116	0.4474
4B	333000001	335000000	788	0.1245	0.3923	0.4238
4B	334000001	336000000	637	0.1363	0.3715	0.4027
4B	335000001	337000000	685	0.1442	0.3855	0.4156
4B	336000001	338000000	703	0.1343	0.4219	0.4498
4B	337000001	339000000	617	0.1293	0.4347	0.4611
4B	338000001	340000000	779	0.1339	0.4137	0.4405
4B	339000001	341000000	598	0.1308	0.4125	0.4274
4B	340000001	342000000	240	0.1349	0.4122	0.3801
4B	341000001	343000000	358	0.14	0.4327	0.4358
4B	342000001	344000000	515	0.1396	0.4002	0.4263
4B	343000001	345000000	901	0.1298	0.4291	0.459
4B	344000001	346000000	1136	0.1289	0.4398	0.4635
4B	345000001	347000000	1057	0.1412	0.4041	0.4234
4B	346000001	348000000	932	0.1392	0.3991	0.4186
4B	347000001	349000000	895	0.1316	0.4023	0.425
4B	348000001	350000000	1083	0.1257	0.44	0.4658
4B	349000001	351000000	1089	0.1266	0.4517	0.4779
4B	350000001	352000000	1165	0.1281	0.4219	0.4497
4B	351000001	353000000	1195	0.1306	0.4272	0.461
4B	352000001	354000000	1226	0.1369	0.4375	0.4753
4B	353000001	355000000	1247	0.1367	0.4385	0.4723
4B	354000001	356000000	883	0.1344	0.4363	0.4501

4B	355000001	357000000	723	0.1175	0.4601	0.4331
4B	356000001	358000000	873	0.1149	0.4604	0.4594
4B	357000001	359000000	1003	0.1296	0.4211	0.4445
4B	358000001	360000000	925	0.1417	0.4209	0.4437
4B	359000001	361000000	792	0.1382	0.4546	0.4793
4B	360000001	362000000	813	0.1306	0.4561	0.4804
4B	361000001	363000000	678	0.1353	0.4575	0.4888
4B	362000001	364000000	625	0.1407	0.4293	0.4667
4B	363000001	365000000	659	0.1341	0.4151	0.4498
4B	364000001	366000000	847	0.1208	0.4357	0.4703
4B	365000001	367000000	892	0.1172	0.4258	0.4579
4B	366000001	368000000	726	0.1284	0.3935	0.4261
4B	367000001	369000000	849	0.1348	0.3768	0.4064
4B	368000001	370000000	966	0.1464	0.3657	0.3921
4B	369000001	371000000	878	0.1575	0.3827	0.41
4B	370000001	372000000	900	0.1422	0.4325	0.4591
4B	371000001	373000000	972	0.1339	0.4354	0.4616
4B	372000001	374000000	971	0.1346	0.4254	0.4485
4B	373000001	375000000	863	0.1401	0.4244	0.4499
4B	374000001	376000000	779	0.1403	0.4046	0.427
4B	375000001	377000000	942	0.1425	0.4038	0.4267
4B	376000001	378000000	810	0.1586	0.4047	0.4339
6A	177000001	179000000	1327	0.4189	0.4435	0.3705
6A	179000001	181000000	1228	0.4398	0.4521	0.5362
6A	180000001	182000000	932	0.3799	0.4137	0.5449
6A	181000001	183000000	952	0.427	0.4529	0.5365
6A	182000001	184000000	857	0.4023	0.4263	0.5186
6A	183000001	185000000	779	0.3404	0.3619	0.473
6A	184000001	186000000	780	0.3793	0.395	0.4804
6A	185000001	187000000	745	0.4074	0.4294	0.4798
6A	186000001	188000000	824	0.3801	0.3946	0.4444

6A	187000001	189000000	745	0.3716	0.3839	0.4389
6A	188000001	190000000	696	0.3692	0.3914	0.4593
6A	189000001	191000000	921	0.3856	0.4081	0.4746
6A	190000001	192000000	949	0.3947	0.4147	0.4663
6A	191000001	193000000	686	0.3917	0.4035	0.4488
6A	192000001	194000000	634	0.3729	0.3893	0.4351
6A	193000001	195000000	800	0.3426	0.3742	0.4713
6A	194000001	196000000	937	0.3109	0.349	0.5581
6A	195000001	197000000	710	0.3384	0.3749	0.6062
6A	196000001	198000000	693	0.4454	0.4661	0.5766
6A	197000001	199000000	880	0.4191	0.4426	0.5458
6A	198000001	200000000	766	0.3513	0.3743	0.4789
6A	199000001	201000000	666	0.3157	0.3357	0.4457
6A	200000001	202000000	664	0.3318	0.3527	0.4683
6A	201000001	203000000	678	0.343	0.3583	0.4443
6A	202000001	204000000	643	0.3308	0.3426	0.4129
6A	203000001	205000000	705	0.3466	0.3471	0.4186
6A	204000001	206000000	742	0.3648	0.3669	0.4395
6A	205000001	207000000	621	0.3441	0.3616	0.43
6A	206000001	208000000	534	0.3059	0.3268	0.4008
6A	207000001	209000000	568	0.3239	0.3415	0.4338
6A	208000001	210000000	642	0.3261	0.3346	0.4585
6A	256000001	258000000	621	0.2658	0.3506	0.3563
6A	257000001	259000000	723	0.2965	0.3546	0.3685
6A	382000001	384000000	480	0.2974	0.3181	0.3314
6A	383000001	385000000	555	0.2911	0.3179	0.3361
6A	418000001	420000000	613	0.1618	0.3269	0.3371
6A	419000001	421000000	633	0.178	0.3659	0.3767
6A	420000001	422000000	500	0.1669	0.3254	0.3362
6A	425000001	427000000	472	0.1578	0.3462	0.353
6A	511000001	513000000	1356	0.3338	0.3655	0.3375

6B	190000001	192000000	2616	0.1603	0.3128	0.3837
6B	208000001	210000000	2577	0.1839	0.3175	0.3505
6B	429000001	431000000	363	0.2106	0.3631	0.3546
7A	523000001	525000000	1691	0.2457	0.2827	0.3376
7B	161000001	163000000	1952	0.1958	0.2	0.3843
7B	162000001	164000000	2127	0.1825	0.2204	0.4372
7B	196000001	198000000	3048	0.1894	0.1896	0.3321
7B	203000001	205000000	1344	0.167	0.1953	0.4382
7B	204000001	206000000	1663	0.1558	0.1957	0.4701
7B	205000001	207000000	1361	0.159	0.1924	0.4376
7B	232000001	234000000	1781	0.2653	0.2787	0.3333

S. Table 5: Genes affected by moderate/high impact variants within the windows in S.Table 4. Genes affected by high impact variants are reported in bold.

Gene	*Gene description*	*Chrom*	*Gene start (bp)*	*Gene end (bp)*
TRITD1Bv1G103000	28S ribosomal protein S29, mitochondrial G	1B	307085502	307093442
TRITD1Bv1G103010	Transmembrane protein, putative G	1B	307094254	307095540
TRITD1Bv1G103100	Coatomer subunit gamma	1B	307261540	307268236
TRITD1Bv1G103330	Flotillin-1	1B	307910850	307912431
TRITD1Bv1G103360	Glucose-induced degradation protein 8-like protein	1B	308049623	308056812
TRITD2Av1G083230	Receptor-like protein kinase	2A	197203217	197205235
TRITD2Av1G083320	GRAS transcription factor	2A	197623930	197628314
TRITD2Av1G083500	Formin-like protein	2A	198050715	198053674
TRITD2Av1G083890	transmembrane protein, putative (DUF1218)	2A	199219834	199220590
TRITD2Av1G083940	50S ribosomal protein L28	2A	199458865	199463563
TRITD2Av1G084010	Beta-amylase	2A	199489879	199492516
TRITD2Bv1G222460	Kinase	2B	669332906	669336626
TRITD2Bv1G222510	Glutamate receptor	2B	669557138	669560065
TRITD2Bv1G222530	Cytochrome P450 family protein	2B	669831858	669833541
TRITD2Bv1G222540	Cytochrome P450 family protein	2B	669831858	669833541
TRITD2Bv1G222590	NAC domain-containing protein	2B	670080794	670083844
TRITD3Bv1G054280	DUF674 family protein	3B	148001286	148002592
TRITD3Bv1G054290	NADPH:adrenodoxin oxidoreductase, mitochondrial	3B	148004229	148009655
TRITD3Bv1G054300	DUF674 family protein	3B	148016029	148016887
TRITD3Bv1G054340	Pentatricopeptide repeat-containing protein	3B	148127191	148133107
TRITD3Bv1G054350	DUF674 family protein	3B	148134379	148139778
TRITD3Bv1G054360	Alpha/beta-Hydrolases superfamily protein	3B	148140554	148141549
TRITD3Bv1G054450	Alpha/beta-Hydrolases superfamily protein	3B	148140554	148141549
TRITD3Bv1G054590	F-box family protein	3B	148735515	148805147
TRITD3Bv1G054680	F-box family protein	3B	148735515	148805147
TRITD3Bv1G054820	nine-cis-epoxycarotenoid dioxygenase 4 G	3B	149227033	149227656
TRITD3Bv1G055000	Receptor-like kinase, putative	3B	149718227	149720956
TRITD3Bv1G055020	GDSL esterase/lipase	3B	149725864	149727418

TRITD3Bv1G055060	centrosomal protein of 135 kDa-like protein G	3B	149859092	149866624
TRITD4Bv1G033620	Serpin	4B	95121965	95123410
TRITD4Bv1G033630	B3 domain-containing protein	4B	95127922	95130059
TRITD4Bv1G033800	Transducin/WD-like repeat-protein	4B	95364062	95371261
TRITD4Bv1G033920	Calcium-binding EF hand family protein G	4B	95546340	95546792
TRITD4Bv1G034450	ABC transporter G family member	4B	97207282	97210677
TRITD4Bv1G034570	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein	4B	97557682	97558380
TRITD4Bv1G034940	Receptor-type tyrosine-protein phosphatase	4B	98483290	98487901
TRITD4Bv1G035000	ATP-dependent RNA helicase RhlB	4B	98821192	98831536
TRITD4Bv1G037930	D-amino acid dehydrogenase 3	4B	107110437	107112022
TRITD4Bv1G038090	Acetyltransferase, GNAT family protein, expressed	4B	107666560	107667108
TRITD4Bv1G038230	F-box family protein	4B	108231254	108233508
TRITD4Bv1G038360	F-box family protein	4B	108231254	108233508
TRITD4Bv1G038480	Mitochondrial transcription termination factor-like	4B	108367509	108368681
TRITD4Bv1G038650	FAD-binding Berberine family protein	4B	108667462	108669096
TRITD4Bv1G063180	Dihydroxy-acid dehydratase	4B	189291036	189291929
TRITD4Bv1G063190	Xylanase inhibitor protein 1	4B	189431716	189439277
TRITD4Bv1G064320	MYB transcription factor	4B	193159847	193161112
TRITD4Bv1G064680	Pentatricopeptide repeat-containing protein	4B	194057755	194060552
TRITD4Bv1G065410	Peptide chain release factor 1	4B	194057755	194060552
TRITD4Bv1G065650	50S ribosomal protein L2	4B	197602273	197602739
TRITD4Bv1G065660		4B		
TRITD4Bv1G065750	MADS-box transcription factor family protein	4B	197994133	197997006
TRITD4Bv1G066540	Lipase	4B	201514852	201516112
TRITD4Bv1G067860	Receptor protein kinase, putative	4B	208103964	208107930
TRITD4Bv1G068330	Serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit B alpha isoform	4B	209938266	209938616
TRITD4Bv1G069260	MIP18 family protein G	4B	214415514	214416104
TRITD4Bv1G069390	Heat shock protein 70, conserved site-containing protein G	4B	215011185	215011846
TRITD4Bv1G069590	Protein DEK G	4B	215849182	215857028
TRITD4Bv1G069730	Ubiquitin carboxyl-terminal hydrolase	4B	216357191	216365085
TRITD4Bv1G069820	Glycerophosphodiester phosphodiesterase 1 G	4B	216758079	216758949
TRITD4Bv1G070390	NADH dehydrogenase subunit 4 G	4B	219182219	219182413
TRITD4Bv1G070620	ATPase subunit 8	4B	219272789	219273261
TRITD4Bv1G070650	30S ribosomal protein S12	4B	219300625	219301002
TRITD4Bv1G070770	Non-lysosomal glucosylceramidase	4B	219706181	219706919
TRITD4Bv1G071120	Splicing factor u2af large subunit, putative	4B	221278864	221289559
TRITD4Bv1G071720	P-loop containing nucleoside triphosphate hydrolases superfamily protein	4B	224356317	224361258
TRITD4Bv1G073110	Hexosyltransferase	4B	231234065	231238151
TRITD4Bv1G073250	Pectin lyase-like superfamily protein	4B	233245424	233248769
TRITD4Bv1G074790	CLK4-associating serine/arginine rich	4B	241461835	241468545
TRITD4Bv1G075330	Aberrant pollen transmission 1, putative, expressed	4B	243756663	243779713
TRITD4Bv1G075690	Beta-adaptin-like protein	4B	244645846	244651890
TRITD4Bv1G076550	Protein SUPPRESSOR OF GENE SILENCING 3	4B	246591705	246606089
TRITD4Bv1G077210	Agenet domain containing protein	4B	249869173	249875531
TRITD4Bv1G078800	RNA-binding protein, putative	4B	257339883	257341410
TRITD4Bv1G080720	Protein phosphatase-2c, putative	4B	263695984	263697754
TRITD4Bv1G080890	Retinoblastoma-related protein G	4B	264168335	264177849
TRITD4Bv1G081530	S-acyltransferase G	4B	267507491	267516566
TRITD4Bv1G081570	Kinase family protein	4B	267728951	267730766
TRITD4Bv1G082040	tRNA-2-methylthio-N(6)-dimethylallyladenosine synthase	4B	269274326	269277844

TRITD4Bv1G082360	Protein transport protein SEC23	4B	270738750	270756907
TRITD4Bv1G083350	3-ketoacyl-CoA synthase	4B	274851563	274854601
TRITD4Bv1G084630	Chloroplastic group IIA intron splicing facilitator CRS1, chloroplastic	4B	280784225	280794965
TRITD4Bv1G086490	aldehyde dehydrogenase 10A8 G	4B	288556355	288557427
TRITD4Bv1G087000	inactive purple acid phosphatase-like protein G	4B	290021633	290024736
TRITD4Bv1G087360	DNA-directed RNA polymerase I subunit rpa49	4B	291769917	291772731
TRITD4Bv1G088420	Alpha/beta-Hydrolases superfamily protein	4B	296785394	296801421
TRITD4Bv1G088690	Zinc finger, CCHC-type	4B	298136532	298138166
TRITD4Bv1G089430	Hydroxysteroid dehydrogenase, putative	4B	300794910	300803147
TRITD4Bv1G089770	1-phosphatidylinositol-3-phosphate 5-kinase	4B	302231888	302270453
TRITD4Bv1G090440	S-acyltransferase	4B	307010942	307036771
TRITD4Bv1G090800	Villin	4B	308709568	308722499
TRITD4Bv1G091090	1-phosphatidylinositol-3-phosphate 5-kinase FAB1A G	4B	309474638	309474886
TRITD4Bv1G091890	1-phosphatidylinositol-3-phosphate 5-kinase FAB1A G	4B	309474638	309474886
TRITD4Bv1G095040	Nuclear cap-binding protein subunit 1	4B	328688611	328707327
TRITD4Bv1G095820	Protein ENHANCED DISEASE RESISTANCE 2-like G	4B	332896442	332925408
TRITD4Bv1G096530	30S ribosomal protein S4 G	4B	335916750	335916995
TRITD4Bv1G096600	WAPL (Wings apart-like protein regulation of heterochromatin) protein G	4B	336124027	336126319
TRITD4Bv1G097210	Ubiquitin carboxyl-terminal hydrolase family protein, expressed	4B	338739811	338788353
TRITD4Bv1G098130	ATPase family AAA domain-containing protein 1	4B	343141535	343152001
TRITD4Bv1G098670	Aspartic acid-rich protein	4B	344805788	344806585
TRITD4Bv1G099410	Transcription factor jumonji (JmjC) domain-containing	4B	347895132	347903271
TRITD4Bv1G101340	Transcription initiation factor TFIID subunit 4b	4B	357329134	357355156
TRITD4Bv1G101720	Paladin	4B	358279289	358305775
TRITD4Bv1G102740	TUDOR-SN protein 1 isoform 2 G	4B	362417693	362427278
TRITD4Bv1G103630	Fanconi anemia group I protein	4B	366101018	366106964
TRITD4Bv1G104200	Sucrose synthase	4B	368676388	368712076
TRITD4Bv1G104730	tRNA pseudouridine synthase	4B	371063288	371086998
TRITD4Bv1G104990	Uroporphyrinogen decarboxylase	4B	372092733	372096313
TRITD4Bv1G105380	VP1/ABI3-like 3 G	4B	373238329	373241179
TRITD4Bv1G105390	Poly (A) RNA polymerase cid14	4B	373241828	373245576
TRITD4Bv1G105670		4B		
TRITD4Bv1G105740	Insulin-degrading enzyme	4B	373968755	373981334
TRITD4Bv1G105850	Pentatricopeptide repeat-containing protein	4B	374912240	374913948
TRITD4Bv1G106440	extra-large G-like protein, putative (DUF3133)	4B	377687830	377689896
TRITD4Bv1G106550	Photosystem II Psb27 protein	4B	377927963	377928469
TRITD6Av1G072260	Protein FLX-like 2 G	6A	178114781	178118250
TRITD6Av1G072900	Receptor-like protein kinase	6A	179702981	179708289
TRITD6Av1G073760	Kinase family protein	6A	181559538	181562419
TRITD6Av1G073970	GDSL esterase/lipase	6A	182500880	182503114
TRITD6Av1G073990		6A	182504303	182505214
TRITD6Av1G074010	GDSL esterase/lipase	6A	182531659	182532868
TRITD6Av1G074160	Subtilisin-like protease	6A	183004196	183019687
TRITD6Av1G074210	U11/U12 small nuclear ribonucleoprotein 25 kDa protein	6A	183088617	183090605
TRITD6Av1G075310	U11/U12 small nuclear ribonucleoprotein 25 kDa protein	6A	183088617	183090605
TRITD6Av1G075320	transmembrane protein G	6A	185530197	185540976
TRITD6Av1G075400	Kynurenine formamidase	6A	185677558	185679778
TRITD6Av1G076030	Nuclear pore complex protein NUP205	6A	187296822	187330561
TRITD6Av1G076780	Glycosyltransferase G	6A	189407527	189413386
TRITD6Av1G077430		6A	191581634	191581982

TRITD6Av1G079280	transmembrane protein, putative (Protein of unknown function, DUF642)	6A	195951059	195952361
TRITD6Av1G079810	Pleiotropic drug resistance ABC transporter	6A	198015998	198022394
TRITD6Av1G080190	Copper ion binding protein G	6A	198816617	198854437
TRITD6Av1G080200	Mitochondrial substrate carrier family protein G	6A	198859002	198876136
TRITD6Av1G080380	Coatomer beta' subunit	6A	199307061	199317318
TRITD6Av1G080520	Ubiquitin-conjugating enzyme E2 B G	6A	199516432	199519223
TRITD6Av1G081580	Myosin heavy chain-related protein G	6A	202466118	202471338
TRITD6Av1G081590	Respiratory burst oxidase	6A	202473638	202479216
TRITD6Av1G082070	Ethylene-responsive transcription factor	6A	204146758	204147506
TRITD6Av1G096660	Poly(A) polymerase	6A	256129464	256140041
TRITD6Av1G097160	50S ribosomal protein L22	6A	258092363	258095168
TRITD6Av1G131740	Cytosolic Fe-S cluster assembly factor NBP35	6A	382054443	382057276
TRITD6Av1G131750	Kelch repeat-containing F-box family protein	6A	382060926	382061987
TRITD6Av1G132690	TGN-related, localized SYP41-interacting protein, putative G	6A	384421098	384428011
TRITD6Av1G143110	Hexosyltransferase	6A	418050709	418058214
TRITD6Av1G143270	MYB	6A	418453514	418457241
TRITD6Av1G143280	MYB	6A	418597870	418598622
TRITD6Av1G143400	60S ribosomal protein L14, putative	6A	418982514	418985721
TRITD6Av1G143960	Receptor-like protein kinase	6A	420740885	420745230
TRITD6Av1G146010	Myb transcription factor	6A	425633292	425634329
TRITD6Av1G146340	response regulator 1	6A	426658150	426660366
TRITD6Av1G178250	DUF506 family protein	6A	511247000	511248051
TRITD6Av1G178290	Eukaryotic aspartyl protease family protein	6A	511417846	511419820
TRITD6Av1G178330	Eukaryotic aspartyl protease family protein, putative	6A	511417846	511419820
TRITD6Av1G178540	ATP synthase mitochondrial F1 complex assembly factor 1	6A	511955592	511958678
TRITD6Av1G178850	Ubiquitin-conjugating enzyme E2, putative	6A	512570447	512571191
TRITD6Av1G178860	Ubiquitin-conjugating enzyme E2	6A	512570447	512571191
TRITD6Av1G178880	WD repeat-containing protein-like protein	6A	512595865	513084796
TRITD6Bv1G066050	Fatty acid hydroxylase superfamily	6B	190892874	190898092
TRITD6Bv1G066280	Protein kinase-like	6B	191964258	191965391
TRITD6Bv1G070750	Phosphoglycerate kinase	6B	208295102	208298267
TRITD7Av1G193440	ABC transporter-like family-protein	7A	523723145	523729000
TRITD7Av1G193730	Homeobox protein bel1-like protein	7A	524564810	524571209
TRITD7Bv1G057320	ATPase subunit 4	7B	161081505	161082083
TRITD7Bv1G057370	Galactan beta-1,4-galactosyltransferase GALS1	7B	161198037	161201019
TRITD7Bv1G057430	Jasmonate O-methyltransferase	7B	161331563	161332917
TRITD7Bv1G057560	LOB domain protein	7B	162322557	162323953
TRITD7Bv1G058090	F-box domain containing protein, expressed	7B	163602551	163607973
TRITD7Bv1G067860	3-oxoacyl-reductase	7B	196215829	196219954
TRITD7Bv1G067870	Potassium transporter	7B	196220811	196224671
TRITD7Bv1G067920	Mitochondrial transcription termination factor family protein, putative	7B	196287944	196289830
TRITD7Bv1G068280	Beta-glucosidase, putative	7B	197437680	197439783
TRITD7Bv1G070020	E3 ubiquitin-protein ligase	7B	203606746	203607432
TRITD7Bv1G070810	Receptor-like protein kinase	7B	205906138	205923870
TRITD7Bv1G070820	Receptor-like protein kinase	7B	205906138	205923870
TRITD7Bv1G070840	Protein kinase family protein	7B	205974824	205981140
TRITD7Bv1G070860	Methionyl-tRNA formyltransferase	7B	205989852	205991662
TRITD7Bv1G070980	Ubiquitin-associated domain-containing family protein	7B	206251444	206254998
TRITD7Bv1G071080	Chaperone protein htpG family protein	7B	206524048	206527805
TRITD7Bv1G071100	Pentatricopeptide repeat-containing protein	7B	206534518	206536311
TRITD7Bv1G071140		7B	206632826	206634395

TRITD7Bv1G079980	Receptor protein kinase, putative	7B	232379389	232386194
TRITD7Bv1G079990	Stress up-regulated Nod 19 protein	7B	232387150	232390904
TRITD7Bv1G080040	Retinoblastoma-related protein	7B	232553490	232559000
TRITD7Bv1G080200	Transcription factor IIIB subunit	7B	232789939	232792509
TRITD7Bv1G080290	Pentatricopeptide repeat-containing protein	7B	232837985	232841560
TRITD7Bv1G080310	Pentatricopeptide repeat-containing protein	7B	232837985	232841560
TRITD7Bv1G080450	Major facilitator superfamily protein	7B	233831889	233833172

S. Table 6: intersection of distance and selection statistics for windows showing introgression from WSL to DSE. Windows highlighted in bold are in the top 5% values of XP-EHH (MAX SCORE XPEHH, % GT XPEHH = 5), as reported by Selcan norm, indicating selection in DSE compared to DNW.

chr	*start*	*end*	*dxy WSL-DSE*	*dxy WNL-DSE*	*dxy DSE-DNW*	*TajimaD*	*MAX SCORE XPEHH*	*MIN SCORE XPEHH*	*% GT XPEHH*	*% LT XPEHH*
2A	198000001	200000000	0,1405	0,1601	0,4255	-1,8185	1,2537	-1,5676	100	100
3B	148000001	150000000	0,2197	0,3396	0,3536	0,6086	2,2795	-1,2602	100	100
4B	96000001	98000000	0,2436	0,2745	0,444	-1,6319	2,6894	-1,7330	100	100
4B	190000001	192000000	0,1498	0,413	0,442	-0,7912	1,8513	-1,2624	100	100
4B	192000001	194000000	0,1389	0,4286	0,4569	-0,8221	1,7811	-1,1862	100	100
4B	194000001	196000000	0,138	0,4264	0,456	-1,0821	1,8780	-0,8938	100	100
4B	196000001	198000000	0,1309	0,407	0,4349	-1,2442	1,9992	-1,3678	100	100
4B	198000001	200000000	0,1373	0,4183	0,4455	-0,9454	2,1298	-1,0799	100	100
4B	**200000001**	**202000000**	**0,1417**	**0,407**	**0,4379**	**-1,0052**	**2,3957**	**-1,1791**	**5**	**100**
4B	**202000001**	**204000000**	**0,1356**	**0,4137**	**0,4399**	**-1,0432**	**2,2601**	**-1,0903**	**5**	**100**
4B	204000001	206000000	0,1323	0,4422	0,4651	-1,1035	2,0743	0,0982	100	100
4B	**206000001**	**208000000**	**0,1545**	**0,3851**	**0,4083**	**-0,7753**	**2,6653**	**-1,3396**	**5**	**100**
4B	208000001	210000000	0,1402	0,3625	0,3872	-0,9024	1,4521	-1,1052	100	100
4B	210000001	212000000	0,1282	0,4267	0,4554	-0,9585	1,6136	-1,3818	100	100
4B	**212000001**	**214000000**	**0,1392**	**0,3984**	**0,4231**	**-0,9126**	**2,8884**	**-1,2087**	**5**	**100**
4B	214000001	216000000	0,15	0,3901	0,4162	-0,7423	2,5141	-1,6468	100	100
4B	**216000001**	**218000000**	**0,146**	**0,3913**	**0,4204**	**-0,6824**	**2,6491**	**-1,2841**	**5**	**100**
4B	218000001	220000000	0,1314	0,4101	0,4375	-1,0275	1,0181	-0,9300	100	100
4B	**220000001**	**222000000**	**0,1324**	**0,4168**	**0,447**	**-1,1103**	**2,4497**	**-1,0667**	**5**	**100**
4B	222000001	224000000	0,1433	0,4139	0,4399	-0,6910	1,0534	-1,3395	100	100
4B	224000001	226000000	0,1447	0,4134	0,4445	-0,7136	1,3126	-0,3626	100	100
4B	**226000001**	**228000000**	**0,1375**	**0,4156**	**0,4465**	**-0,8533**	**2,3717**	**-1,1475**	**5**	**100**
4B	228000001	230000000	0,1451	0,4083	0,4386	-0,8520	2,0963	-0,4437	100	100
4B	230000001	232000000	0,1458	0,4027	0,4265	-0,8684	1,9014	-1,4390	100	100
4B	**232000001**	**234000000**	**0,1356**	**0,402**	**0,4289**	**-1,1993**	**2,3939**	**-0,9021**	**5**	**100**
4B	234000001	236000000	0,1356	0,4197	0,4425	-0,9737	1,7155	-1,7051	100	100
4B	238000001	240000000	0,1403	0,4005	0,4267	-0,8561	0,5414	-0,0823	100	100
4B	**240000001**	**242000000**	**0,1393**	**0,3824**	**0,4075**	**-0,9288**	**2,3449**	**-0,3448**	**5**	**100**
4B	**242000001**	**244000000**	**0,1358**	**0,4155**	**0,4487**	**-1,0830**	**2,2626**	**-1,3565**	**5**	**100**
4B	244000001	246000000	0,1384	0,4154	0,4493	-0,8743	1,2078	-2,0921	100	100

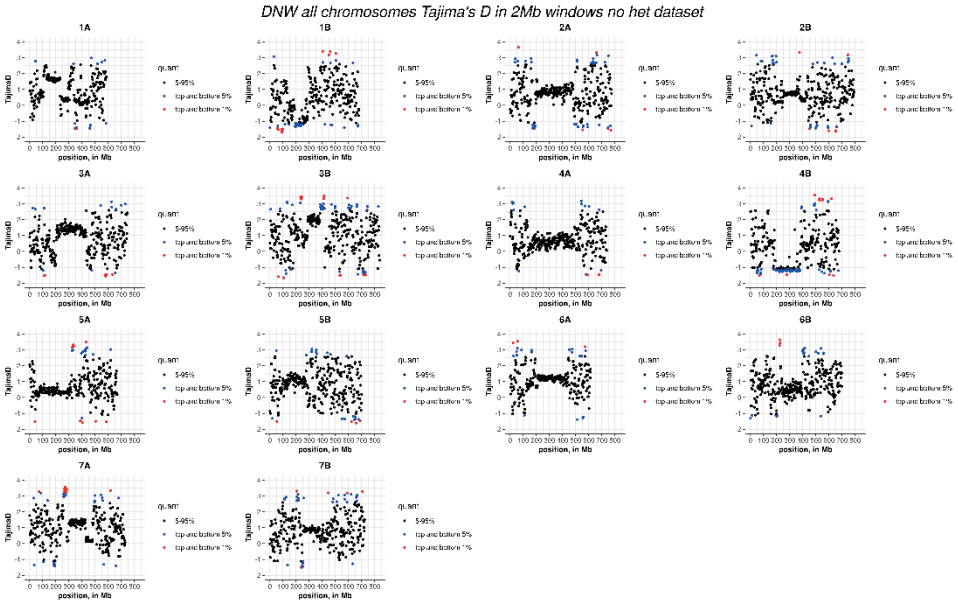
4B	246000001	248000000	0,1334	0,4136	0,4325	-1,2366	2,0153	-0,4321	100	100
4B	**248000001**	**250000000**	**0,1268**	**0,3802**	**0,4072**	**-1,0855**	**2,8032**	**-0,1669**	**5**	**100**
4B	250000001	252000000	0,1462	0,373	0,3856	-0,5904	2,0869	-1,0567	100	100
4B	252000001	254000000	0,1405	0,4124	0,4367	-1,1052	1,5615	-0,9383	100	100
4B	254000001	256000000	0,1441	0,4033	0,4367	-0,9897	1,7175	-0,9574	100	100
4B	256000001	258000000	0,1464	0,4054	0,4219	-1,1410	1,2377	-1,1172	100	100
4B	258000001	260000000	0,1324	0,421	0,4473	-0,9665	1,5799	-0,4522	100	100
4B	262000001	264000000	0,1352	0,4537	0,4778	-0,9532	1,8298	-1,1035	100	100
4B	264000001	266000000	0,1437	0,3959	0,4216	-0,9965	1,4966	-1,4496	100	100
4B	266000001	268000000	0,1376	0,3943	0,4242	-0,9040	1,8218	-0,3913	100	100
4B	268000001	270000000	0,1474	0,3715	0,3971	-0,7007	0,8239	-0,4136	100	100
4B	**270000001**	**272000000**	**0,1363**	**0,4057**	**0,4355**	**-0,8758**	**2,6638**	**-0,6141**	**5**	**100**
4B	**272000001**	**274000000**	**0,1529**	**0,3877**	**0,4128**	**-0,9527**	**2,2360**	**-1,5645**	**5**	**100**
4B	274000001	276000000	0,1388	0,4135	0,4366	-0,9599	1,4595	-0,7994	100	100
4B	280000001	282000000	0,1244	0,4184	0,4506	-0,8355	1,7711	-0,4047	100	100
4B	282000001	284000000	0,137	0,4196	0,4495	-1,0126	2,4667	-0,3998	5	100
4B	284000001	286000000	0,1394	0,3572	0,3791	-0,6072	1,9144	-1,0328	100	100
4B	286000001	288000000	0,1362	0,4113	0,4399	-0,8305	1,5664	-1,2981	100	100
4B	288000001	290000000	0,1344	0,3683	0,4024	-1,0257	1,5759	-1,4276	100	100
4B	**290000001**	**292000000**	**0,1386**	**0,4009**	**0,4327**	**-1,0257**	**2,1883**	**-1,1034**	**5**	**100**
4B	292000001	294000000	0,1457	0,4191	0,4471	-0,9256	1,9998	-1,5554	100	100
4B	294000001	296000000	0,1384	0,3993	0,42	-0,6927	1,6163	-1,5794	100	100
4B	**296000001**	**298000000**	**0,145**	**0,3799**	**0,3877**	**-0,5165**	**2,0755**	**-0,7070**	**5**	**100**
4B	298000001	300000000	0,1339	0,4325	0,4566	-0,8401	1,4255	-1,3941	100	100
4B	300000001	302000000	0,1382	0,416	0,4385	-0,8368	1,9847	-1,2904	100	100
4B	302000001	304000000	0,1308	0,4226	0,4518	-1,0310	2,0197	-0,6971	100	100
4B	304000001	306000000	0,1389	0,3781	0,4024	-0,9488	2,0385	-0,7341	100	100
4B	306000001	308000000	0,1573	0,4055	0,4223	-1,0240	1,9891	-1,4965	100	100
4B	**308000001**	**310000000**	**0,1441**	**0,4073**	**0,4344**	**-0,6732**	**2,4844**	**-1,7375**	**5**	**100**
4B	310000001	312000000	0,1435	0,3938	0,4192	-1,1278	2,1583	-0,8905	100	100
4B	312000001	314000000	0,1598	0,4055	0,4291	-0,3706	1,5767	-3,1294	100	5
4B	**314000001**	**316000000**	**0,1475**	**0,4114**	**0,4244**	**-1,1586**	**2,2164**	**-2,2866**	**5**	**100**
4B	316000001	318000000	0,1377	0,4198	0,4453	-0,9012	0,9371	-1,8740	100	100
4B	318000001	320000000	0,1378	0,4074	0,4365	-0,9718	2,3861	-0,8098	100	100
4B	320000001	322000000	0,1407	0,4473	0,4718	-1,1682	1,2965	-0,8164	100	100
4B	322000001	324000000	0,1316	0,4121	0,4497	-0,6464	1,9391	-0,1208	100	100
4B	324000001	326000000	0,1445	0,4081	0,4347	-0,8920	2,0764	-0,8789	100	100
4B	326000001	328000000	0,1236	0,4476	0,478	-0,9319	1,2964	-0,5983	100	100
4B	328000001	330000000	0,1389	0,4113	0,4447	-0,9319	1,0126	-1,1697	100	100
4B	330000001	332000000	0,1484	0,3738	0,4087	-0,9379	1,9742	-0,8715	100	100
4B	332000001	334000000	0,1264	0,4116	0,4474	-1,0850	1,8278	-1,1514	100	100
4B	334000001	336000000	0,1363	0,3715	0,4027	-0,9089	1,7886	-0,9514	100	100
4B	**336000001**	**338000000**	**0,1343**	**0,4219**	**0,4498**	**-1,3316**	**2,6727**	**-0,8325**	**5**	**100**

4B	338000001	340000000	0,1339	0,4137	0,4405	-1,2107	2,2179	-0,4020	100	100
4B	340000001	342000000	0,1349	0,4122	0,3801	-0,8273	1,6370	-0,3259	100	100
4B	342000001	344000000	0,1396	0,4002	0,4263	-1,1831	1,8774	-0,4933	100	100
4B	**344000001**	**346000000**	**0,1289**	**0,4398**	**0,4635**	**-0,9442**	**2,3167**	**-0,6291**	**5**	**100**
4B	346000001	348000000	0,1392	0,3991	0,4186	-1,1870	2,2776	-0,8237	100	100
4B	348000001	350000000	0,1257	0,44	0,4658	-1,0084	1,3492	-0,9640	100	100
4B	350000001	352000000	0,1281	0,4219	0,4497	-0,9837	1,5945	-1,0249	100	100
4B	**352000001**	**354000000**	**0,1369**	**0,4375**	**0,4753**	**-1,0749**	**2,2637**	**-0,5168**	**5**	**100**
4B	354000001	356000000	0,1344	0,4363	0,4501	-0,6612	2,2413	-1,3204	100	100
4B	**356000001**	**358000000**	**0,1149**	**0,4604**	**0,4594**	**-0,7124**	**2,6320**	**-0,6380**	**5**	**100**
4B	358000001	360000000	0,1417	0,4209	0,4437	-1,1747	1,4882	-1,6250	100	100
4B	360000001	362000000	0,1306	0,4561	0,4804	-1,0085	2,0641	-0,3586	100	100
4B	362000001	364000000	0,1407	0,4293	0,4667	-0,9637	1,8162	-0,7633	100	100
4B	364000001	366000000	0,1208	0,4357	0,4703	-0,8157	1,6143	-1,0701	100	100
4B	366000001	368000000	0,1284	0,3935	0,4261	-1,0953	1,8412	-0,0104	100	100
4B	368000001	370000000	0,1464	0,3657	0,3921	-0,7270	1,7621	-2,2834	100	100
4B	370000001	372000000	0,1422	0,4325	0,4591	-1,1003	1,8319	-0,9343	100	100
4B	372000001	374000000	0,1346	0,4254	0,4485	-1,0789	1,5907	-1,1853	100	100
4B	374000001	376000000	0,1403	0,4046	0,427	-0,3717	1,5395	-1,9899	100	100
4B	376000001	378000000	0,1586	0,4047	0,4339	-1,6140	1,0003	-1,1172	100	100
6A	180000001	182000000	0,3799	0,4137	0,5449	-1,6140	1,4024	-1,5956	100	100
6A	182000001	184000000	0,4023	0,4263	0,5186	-1,5426	1,4502	-2,1918	100	100
6A	184000001	186000000	0,3793	0,395	0,4804	-1,5333	1,9104	-1,5556	100	100
6A	186000001	188000000	0,3801	0,3946	0,4444	-1,3292	1,2331	-1,9844	100	100
6A	188000001	190000000	0,3692	0,3914	0,4593	-1,4498	1,2469	-1,4161	100	100
6A	190000001	192000000	0,3947	0,4147	0,4663	-1,4088	1,4031	-2,3042	100	100
6A	192000001	194000000	0,3729	0,3893	0,4351	-1,4199	1,3797	-1,5786	100	100
6A	194000001	196000000	0,3109	0,349	0,5581	-1,5912	1,3917	-1,5483	100	100
6A	196000001	198000000	0,4454	0,4661	0,5766	-1,6504	1,1479	-1,4905	100	100
6A	198000001	200000000	0,3513	0,3743	0,4789	-1,5425	1,8110	-1,3542	100	100
6A	200000001	202000000	0,3318	0,3527	0,4683	-1,4218	1,5667	-1,5674	100	100
6A	202000001	204000000	0,3308	0,3426	0,4129	-1,3892	1,8496	-1,4351	100	100
6A	204000001	206000000	0,3648	0,3669	0,4395	-1,4417	0,9337	-2,0909	100	100
6A	206000001	208000000	0,3059	0,3268	0,4008	-1,5681	1,6434	-1,3485	100	100
6A	208000001	210000000	0,3261	0,3346	0,4585	-1,4275	2,0963	-1,9441	100	100
6A	256000001	258000000	0,2658	0,3506	0,3563	-1,0697	0,7469	-1,7126	100	100
6A	382000001	384000000	0,2974	0,3181	0,3314	-1,0758	0,6060	-2,1689	100	100
6A	418000001	420000000	0,1618	0,3269	0,3371	-0,4101	1,3345	-1,5742	100	100
6A	420000001	422000000	0,1669	0,3254	0,3362	-0,6074	1,1356	-2,2405	100	100
6B	190000001	192000000	0,1603	0,3128	0,3837	-1,6751	1,6120	-1,3239	100	100
6B	208000001	210000000	0,1839	0,3175	0,3505	0,7179	1,1896	-1,6091	100	100
7B	162000001	164000000	0,1825	0,2204	0,4372	-0,4004	0,9378	-2,4448	100	5
7B	196000001	198000000	0,1894	0,1896	0,3321	-1,7294	1,7069	-1,6059	100	100

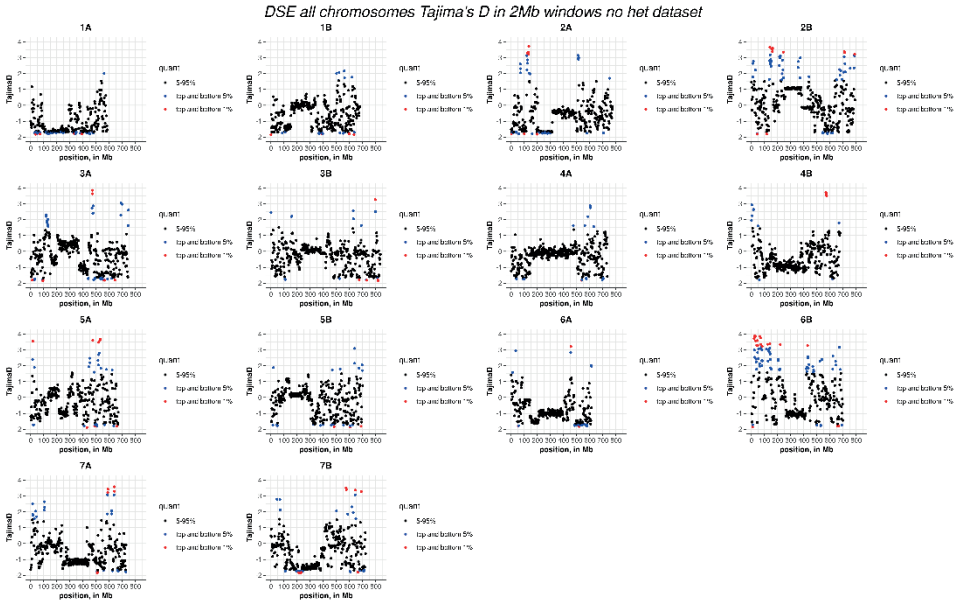
7B	204000001	206000000	0,1558	0,1957	0,4701	-1,7257	1,8599	-1,8881	100	100
7B	**232000001**	**234000000**	**0,2653**	**0,2787**	**0,3333**	**-1,8103**	**2,4279**	**-0,3046**	**5**	**100**

S. Figure 1: Tajima's D values along the chromosomes, calculated in 2Mb windows. Top 5% and top 1% values are colored in blue and red, respectively. 1A: DNW, 1B: DSE.

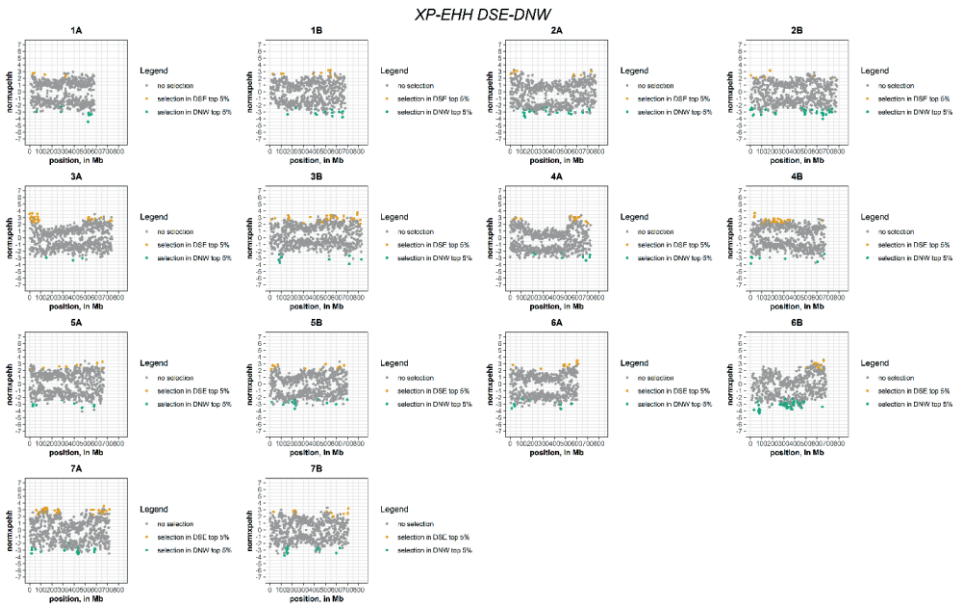
1A



1B



S. Figure 2: XP-EHH between DSE and DNW. Top positive values (orange) indicate selection in DSE, top negative values (green) indicate selection in DNW.



7.3. Ancient wheat genomes illuminate domestication, dispersal, and diversity

Alice Iob[1], Michael F. Scott[2], Laura Botigué[1]

1 Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Campus UAB, Bellaterra, Barcelona, Spain.

2 School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK

In "The wheat genome" Springer, *in press*

https://link.springer.com/book/9783031382925

1. Shining a light on the past: the promise of ancient DNA

Ancient DNA (aDNA) has fostered a revolution in evolutionary genomics, as it allows direct observation of historical molecular diversity (Der Sarkissian et al. 2014). Previously, hypotheses were based solely on the observation of modern genetic diversity, which is the end effect of thousands of years of evolution, with the main caveat that the same pattern of genetic variation is often consistent with different historical scenarios (Lawson et al., 2018). The analysis of aDNA allows the genomic characterization of populations at different points in time, adding a fundamentally new dimension to evolutionary studies (Gutaker and Burbano 2017; Orlando et al. 2021).

The very first aDNA analysis was conducted on a mitochondrial sequence of a museum-preserved quagga (Higuchi et al. 1984). Since then, the field of archaeogenomics has rapidly flourished (Morozova et al. 2016), allowing for a better understanding of human, animal and plant evolutionary history. Recent advances in this field include sedimentary, epigenetic, pathogens and microbiome aDNA analysis (Key et al. 2020; Parducci et al. 2017; Pedersen et al. 2014; Spyrou et al. 2019; Warinner et al. 2014).

aDNA has already had a remarkable impact on our understanding of human history, shedding light on important patterns of migration (Lacan et al. 2011), admixture (Yang et al. 2020), adaptation (Marciniak and Perry 2017), population dispersal, expansion and decline (Nielsen et al. 2017). Notably, aDNA gave fundamental contribution to our knowledge about the genetic relationships between modern humans and their extinct relatives Neanderthals (Weyrich, Dobney, and Cooper 2015) and Denisovans (Krause et al. 2010; Reich et al. 2010), the latter of which have only been identified through aDNA analysis. Similar insights have been gained in other animals, such as dogs (Botigue et al. 2017; Leathlobhair et al. 2018), cattle (Daly et al. 2018; Verdugo et al. 2019), pigs (Frantz et al. 2019), and horses (Gaunitz et al. 2018). These studies have led to a reassessment of previous evidence and an overturning of the existing narrative (Librado, P., Khan, N., Fages 2021).

Now, aDNA promises a similar revolution in our understanding of how crops have been domesticated and spread around the globe, and the ways that these processes have shaped genetic diversity. By revealing how crops have adapted to new environments and what genetic diversity has been lost, aDNA can also set a basis for future breeding strategies (di Donato et al. 2018; Pont, Wagner, Kremer, Orlando, et al. 2019). Crop archaeogenomics is still in its infancy but aDNA from several

important crops has been analysed, including maize (Ramos-Madrigal et al. 2016), barley (Mascher et al. 2016; Palmer et al. 2009), cotton (Palmer et al. 2012), bean (Trucchi et al. 2021), sunflower (Wales et al. 2019), sorghum (Smith et al. 2019), watermelon (Renner et al. 2019), and emmer wheat (Scott et al. 2019).

In this chapter, we first give a very brief overview of the history of wheat cultivation and the key genetic changes involved. aDNA promises unique insights in this area. We review the wheat aDNA studies carried out so far and their contribution to understanding phenomena that have shaped wheat genomes. To conclude, we discuss the key open questions in this field and discuss the limitations posed by wheat's large polyploid genome and idiosyncratic preservation. Our goal is to give an overview of the important answered and unanswered questions in the history of wheat cultivation and the promise of aDNA for resolving them.

2. A brief history of wheat cultivation

Human societies have relied on wheat for thousands of years. Thus, the history of wheat domestication, geographic expansion, and cultivation has cross-disciplinary significance (Figure 1). Understanding how wheat genetic diversity has been shaped also has contemporary relevance due to its continued nutritional and economic importance. Archaeogenomic studies aim to give new information about at least three key aspects of this process: domestication, dispersal, and gene flow between different wheat species. To contextualise contributions from archaeogenomics, we briefly overview these basic tenets of wheat cultivation history.

Figure 1 *Wheat has been culturally important for millennia and DNA extracted from ancient specimens can reveal how humans have shaped crop genetic diversity. Left: Facsimile of a vignette on the tomb of Sennedjem and Iineferti showing grain harvest in the abundant fields of the next life (painted by Charles K Wilkinson in 1922CE, original ca. 1295–1213BCE, public domain image from the Metropolitan Museum of Art). Right: Archaeological specimens of dessicated emmer wheat chaff from Egypt. Photo from Dorian Q. Fuller, University College London, Institute of Archaeology.*

2.1. Domestication

Wild tetraploid emmer wheat was one of the first species to be domesticated (Haas, Schreiber, and Mascher 2018), during the so-called Neolithic Transition, in parallel with humans' shift from hunting and gathering to agriculture and animal husbandry (Diamond 2002). The quintessential trait for cereal domestication is the loss of rachis brittleness: in wild cereals, the spikelets disarticulate spontaneously from the rachis upon maturity, ensuring seed dispersal and germination. In domestic cereals the rachis is non-brittle; spikelets remain attached, allowing easier harvesting but requiring subsequent sowing in the following season in order to germinate. Because plants with a non-brittle rachis depend on human action for dispersal, this phenotype has been used to define domestication in cereals (Abbo et al. 2014; Snir et al. 2015). Loss-of-function mutations in the *TtBtr1-A* and *TtBtr1-B* genes on chromosomes 3A and 3B are the main determinants of such phenotype (Avni et al. 2017; Nave et al. 2019). Therefore, alleles at these two loci essentially distinguish wild from domesticated emmer wheat. Other traits that are favourable in the human-mediated environment and most likely deleterious in a wild environment (Kantar et al. 2017; Purugganan and Fuller 2009), give a more broad definition of the "domestication syndrome" (Larson et al. 2014), like the loss of seed dormancy and larger seed sizes (Haas, Schreiber, and Mascher 2018; Zohary 2013).

Wild emmer wheat has a very restricted distribution, growing only in the Fertile Crescent region of SW Asia (N. I. Vavilov et al. 1992). The exact location of the emergence of domestic emmer has been a long-standing controversy. In the 2000's early genetic studies started addressing this issue, with the so-called "cradle of agriculture" theory (Lev-Yadun et al., 2000). Further genetic studies had pointed to the Northern Fertile Crescent and specifically to the Karaca Dağ Mountain region as the centre of domestication of emmer wheat (M. C. Luo et al. 2007; Ozkan et al. 2002, 2005), mostly based on the higher similarities between the genomes of the modern domestic landraces and the wild emmer from the Northern Levant, compared to that of the Southern Levant (Avni et al. 2017).

However, this monophyletic origin has been challenged, with increasing evidence that different wild populations have contributed to domestic wheats. Several authors argue that domestic emmer wheat arose from an admixed wild population and that mutations for domestication traits appeared in different chromosomes at different times and possibly in different places (Civáň et al., 2013; Jorgensen et al., 2017; Oliveira et al., 2020). This is in line with the observation that the domestic phenotype, which requires at least two independent recessive mutations, took millennia to be established (Avni et al., 2017; Fuller et al., 2014). As testified by the archaeological record, wild emmer wheat was first exploited in the Southern Levant, where

increasing, even though small, proportions of phenotypically domestic emmer wheat are found at different archaeological sites as early as during Early Pre-Pottery Neolithic B (8700 – 8200 BCE) (Arranz-Otaegui et al. 2018). However, domesticated emmer is found in very high proportions in the Northern Levant starting from the Middle/Late Pre-Pottery Neolithic B (8200–6300 BCE) (Arranz-Otaegui et al. 2016). This indicates that wild emmer was managed (a phenomenon often regarded as "pre-domestication cultivation") (Fuller et al., 2010) long before the domestic forms emerged, and that probably wild populations from across the Fertile Crescent contributed to the domestic pool. (Feldman and Kislev 2007).

The role of introgression from wild to domestic wheat has been demonstrated by several studies e.g. (Cheng et al. 2019; Pont, Leroy, Seidel, and Tondelli 2019; Przewieslik-Allen et al. 2021), even though the context in which these introgression events took place remains unknown.

Overall, archaeology and genetics point to a slow and geographically widespread domestication process in which both the Northern Levant and the Southern Levant played an important role.

2.2. Evolution

Domestic emmer wheat (*Triticum turgidum subsp. dicoccon*) gave rise to today's most economically important wheats: tetraploid durum wheat (*Triticum turgidum subsp. durum)* and hexaploid bread wheat (*Triticum turgidum subsp. aestivum*). These descendants differ from their ancestor in one character of great agricultural importance: the free-threshing phenotype. Emmer is a hulled, non-free-threshing wheat, and the extraction of seeds from husks requires substantial mechanical processing. On the other hand, durum and bread wheat are naked and free-threshing: as the spikelets disarticulate from the rachis they fall apart, releasing the seeds without further processing. While durum wheat is tetraploid (BBAA), bread wheat is hexaploid (BBAADD) and evolved from the hybridization of tetraploid wheat with the diploid wild goatgrass (*Aegilops tauschii*), donor of the D subgenome (Haas, Schreiber, and Mascher 2018; Pont, Leroy, Seidel, and Tondelli 2019). The tetraploid that contributed the B and A subgenomes to bread wheat has been a matter of debate (Sharma et al. 2019), but considering the need for multiple mutations to determine the free-threshing phenotype, the most supported (and most parsimonious) models indicate that hybridization with *Ae. tauschii* occurred with a free-threshing tetraploid (Y. Zhou et al. 2020).

The emergence of modern wheat is therefore the result of three processes: I) domestication of wild emmer wheat, associated with the loss of rachis brittleness; II) crop evolution (often also referred to as crop improvement under cultivation), which

includes the emergence of the free-threshing phenotype and adaptation to new ecological niches; III) allopolyploidization between a free-threshing tetraploid with *A. tauschii*, giving rise to bread wheat. We summarise these changes in Figure 2.

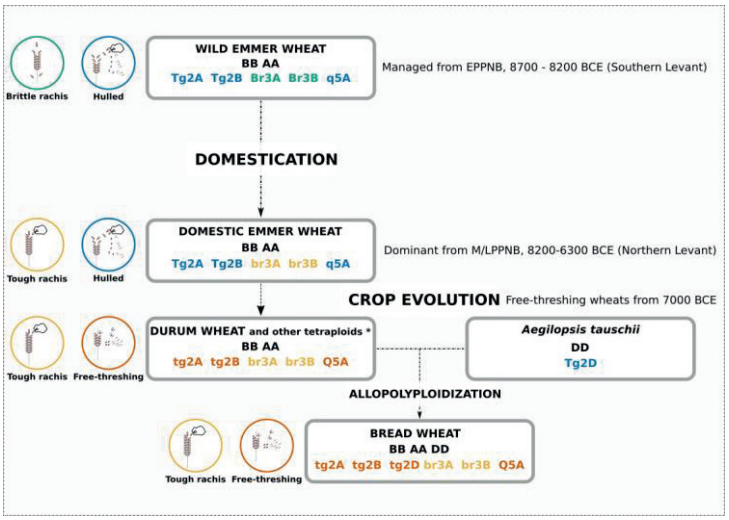


Figure 2 *Schematic representation of the domestication and evolution of the most economically important wheats today, showing important phenotypes and the mutations that determine them. Basic information about the appearance of the different wheats in the archaeological record is given on the right. The small white hand represents the investment of human labour in processing the harvest. *For simplicity, we use the common name "durum wheat" for all free-threshing tetraploids, but other common names are used for free-threshing tetraploids and it is not known which was involved in this allopolyploid event. This scheme is an adaptation of the model proposed by Sharma et al., (2019)*

Perhaps surprisingly, hulled wheats continued to be used for thousands of years after the appearance of free-threshing durum wheat and bread wheat. The slow and regionally specific shifts in wheat usage probably reflect cultural practices and preferences (Nesbitt and Samuel 1996). Also, increasing archaeological evidence shows that early farmers relied on a wide range of other domestic wheats for their subsistence, including einkorn, spelt, and *T. timopheevi* alongside emmer and free-threshing wheats (Özbaşaran et al. 2018). This is in accordance with the evidence for intra and interspecific introgression that has been detected in modern wheat (Cheng et al. 2019; Y. Zhou et al. 2020).

3. Archaeogenomics of wheat

Wheat archaeogenomics is a powerful tool to investigate how wild wheat evolved into domestic forms and how these domestic wheat varieties adapted to different ecological niches and cultural preferences through history.

However, the limitations and the characteristics of ancient genomes have to some extent impacted the approach taken in this research field. Before high quality reference genomes were available, most studies avoided whole genome analysis and used a target and amplification strategy. This mitigates the challenges of a large genome but gives much less rich genomic information. Furthermore, the primers used for amplification mask the characteristic patterns of degradation that are useful for ruling out contamination by confirming the antiquity of the DNA. Unlike these amplification methods, whole genome libraries can also be re-analysed to get more data without further destructive sampling of rare material. For these reasons, amplification approaches are no longer recommended for ancient samples (Gutaker and Burbano 2017; Prüfer and Meyer 2015).

We first overview wheat aDNA studies that use amplification and then describe the first two whole-genome analyses. Even though wheat archaeogenomics is in a germinal stage, the results have shifted our understanding of wheat genetics in important ways.

3.1. Target gene amplification

The most common use of target gene amplification has been to interrogate key genes or to identify wheat remains at the species level.

The x and y copies of the *Glu1* loci were often the focus of early studies. These genes, present in all wheat subgenomes, are located in the long arms of chromosome 1 and encode for the High Molecular Weight Glutenin Subunits (HMW-GSs), storage proteins present in the starchy endosperm cells of wheat. Allelic varieties in these genes impact the properties of dough for bread making. Because of its effect over bread quality, the evolution of the HMW genes can provide insights into the nature of human selective pressures during wheat evolution (Allaby, Banerjee, and Brown 1999). In this manuscript, authors surveyed these loci in a collection of modern and ancient wheats, constructed a phylogenetic tree and obtained time estimates by using a substitution rate to calibrate the observed variation. By comparing the genetic variability for x and y copies in each genome they were able to determine that the genetic variability in these loci for the cultivated species predates domestication, pointing to either incomplete lineage sorting, multiple domestication events or introgression after domestication. Another study used a similar approach with the

same loci to inquire about the origins of spelt (Blatter, Jacomet, and Schlumbaum 2002). They surveyed a collection of modern and ancient bread wheat and spelt specimens and determined that the high genetic variability of spelt compared to that of bread wheat in the A and B genomes are compatible with the origin of spelt being a hybridization event between bread wheat and hulled tetraploid emmer.

HMW genes have also been used to identify wheat remains at the subspecies level and inform about its dispersal. Without associated chaff, it is difficult to distinguish between free-threshing wheats (e.g. bread wheat or durum wheat). Bilgic et al., (2016) targeted the HMW promoter region in 8,400-year-old specimens from a notorious Neolithic site in central Turkey, Çatalhöyük, to determine whether the genetic variability characteristic of the D genome could be recovered, as a proof of that wheat being hexaploid. The finding of HMW subunits from the A, B and D genomes is quite remarkable, since it evidences the presence of hexaploid wheat at a very early point in time and highlights the importance of this settlement in the expansion of hexaploid wheat cultivation. Another study used the Internal Transcribed Spacer regions (ITS1 and ITS2) and the Inter-Genic Spacer region (IGS) from the nuclear ribosomal DNA for species level identification (C. Li et al. 2011). They also found early evidence for hexaploid wheat in NW China around 1760-1540 BCE.

These results highlight the high diversity of wheats consumed by humans during early agricultural expansion. Free-threshing naked wheats first appear in the archaeological record between 7,000-5,500 BCE (Feldman and Kislev 2007). Early naked wheats co-existed with domestic and wild emmer populations (Bilgic et al. 2016), giving opportunities for genetic exchange. Along with the protracted period of emmer domestication, this probably explains the higher genetic diversity on A and B subgenomes of modern bread wheat compared to the D subgenome (Cheng et al. 2019). This demonstrates how the details of agricultural history directly impact modern wheat diversity and breeding. Moreover, other wild *Triticum* species gave rise to domestic forms during the Neolithic. These include the diploid einkorn wheat, *Triticum monococcum* subsp. *monococcum,* that emerged from wild einkorn, *Triticum monococcum* subsp. *Aegilopoides* (Nesbitt and Samuel 1996), spelt (*Triticum spelta*), an hulled hexaploid, and tetraploid *T. timopheevii* (domesticated from *T. timopheevii araraticum*) (Wagenaar 1966), only recently classified thanks to aDNA analysis.

This last wheat exemplifies the value of aDNA to gain insights on certain domestication processes. Briefly, due to the technical difficulties in the identification of *timopheevii,* for a long time its existence was questioned, and it was often unclassified, or ascribed to other wheat species, such as "New Glume Wheat". Recently, archaeological remains described as "New Glume Wheat" have been designated as domestic *Triticum timopheevii* based on aDNA evidence (Czajkowska et

al. 2020). The authors used the *Ppd1* locus to identify G genome alleles in "New Glume Wheat" remains. This study has sparked the interest of the archaeobotanical community. Decades have passed since the first classification of an archaeological specimen to "New Glume Wheat". It was not until numerous remains of this type of wheat were found in several Neolithic and Bronze Age archaeological sites in northern Greece and compared with other locations (Jones, Valamoti, and Charles 2000) that archaeologists were able to describe the distinctive features of this wheat (Ulaş and Fiorentino 2021). Nevertheless, identification based on grain morphology is still problematic.

The identification of New Glume Wheat as domestic *T. timopheevii* thanks to ancient DNA analysis has had important ramifications on our understanding of the complexity of the domestication process in SW Asia and the confirmation that multiple species evolved into domestic forms, moving away from the "founder crops" theory. *T. timopheevii* was actually cultivated for a very long period of time in certain regions. New efforts are now being undertaken to revisit archaeobotanical assemblages and reassess the relative abundance of plant species, with the expectation that many grains classified as emmer wheat will now be classified as *T. timopheevii*.

The HMW loci were also used, together with the ribulose 1,5 biphosphate carboxylase (rbcL) and the chloroplast microsatellite WCT12 in the chloroplast genome to study the viability of DNA extraction on ancient plant specimens (Fernández et al. 2013). In this study 126 grains of naked wheat in different preservation conditions (charred, partially charred and waterlogged) were analysed (Figure 3 shows different preservation conditions of ancient wheat samples). Results showed that DNA extraction from totally charred remains is virtually impossible, while DNA amplification of modern contaminants is pervasive. Unfortunately, almost all of the most ancient archaeological wheat specimens are charred, which is a severe limitation for future aDNA studies.

Figure 3 *Examples of different preservation conditions of archaeobotanical wheat. Left: charred emmer wheat seeds from the Vinča culture in Serbia (middle/late Neolithic; c. 5400-4600/4500 BC), published in Filipovic (2014). Right: Waterlogged chaff remains of Triticum cf. durum/turgidum from the end of the 5th millennium BC at the site of Les Bagnoles. Photo by Raül Soteras, AgriChange Project, reproduced with permission.*

As mentioned above, one important limitation of amplification-based studies is the confidence with which one can rule out contamination. Commonly used indicators such as the fragment length distribution or deamination patterns are difficult to assess in target specific PCR amplification studies. In addition, Allaby et al. reported PCR jumping, probably related with the shortness of some fragments. Their results showed patterns of linked diversity that did not exist in the modern pool and had to manually rearrange the observed diversity so it would match known modern haplotypes with the subsequent potential biases.

Different strategies have been used to increase confidence in the antiquity of the data. Allaby et al. replicated the results *in situ* with the same specimen and produced blanks with each extraction run. Czajkowska et al., performed the extractions in laboratory facilities where no wheat had been processed before, hoping to preclude contamination. Bilgic et al., processed all samples in two different facilities, so that replication of the results acts as a proof of authenticity. In spite of this, even if contamination can be ruled out, it is not possible to distinguish deamination patterns from true polymorphisms. Therefore, phylogenetic analyses and interpretation of the accumulation of variation through time should be taken with caution unless transitions (C/T or G/A SNPs) are excluded.

3.2. Whole genome analyses

As with modern wheat samples, the genomic scale of archaeological wheat genetics has been expanded since the publication of reference genomes (Table 1). Nevertheless, only two studies have so far reported whole genome sequence from archaeological wheat specimens. One has been the analysis of several bread wheat remains from China to infer dispersal into the region (Wu et al. 2019). The earliest bread wheat remains found in China date to approximately 4,500 years ago in the north-western part of the country, but the most interesting aspect of its dispersal is that upon its arrival, wheat had to adapt to a wide variety of climatic conditions. Ancient wheat from two archaeological sites within the Xinjiang winter-spring wheat zone was analysed. Even though coverage was extremely low (0.25-0.01x), the authors were able to call more than 7,000 SNP sites, compare them with modern data from neighbouring regions and provide new evidence on wheat dispersal in China, a still controversial topic. Their results were consistent with one of the routes that had been previously suggested: an early dispersal into the Qinjianh Tibetan plateau, based on the highest genetic similarities between the ancient samples and the modern ones from that region. Conversely, another ancient route that advocated for an introduction towards the eastern region was not supported. However, more data is

needed to determine whether different gene pools were introduced to China and to confirm that modern landraces correspond with ancient ones from the same area.

Table 1 *Genomic information available for wheats and relatives mentioned in the text. This is not a comprehensive list of wheat species/subspecies*

Species Name	Genome(s)	Genome Size	Common Name	Key Phenotypes	Reference Genome(s)
Aegilops tauschii	D	4Gb	Tausch's goatgrass		(Luo et al., 2017)
Triticum urartu	A	4.5Gb	Wild Red Einkorn	Brittle rachis, hulled	(Ling et al. 2018)
Triticum monococcum	A^m	5.7Gb	Wild Einkorn	Brittle rachis, Hulled	NA
			Einkorn	Non-brittle rachis, Hulled	NA
Triticum turgidum	BA	12Gb	Wild Emmer	Brittle rachis, Hulled	(Avni et al. 2017) (Zhu et al. 2019)
			Emmer	Non-brittle rachis, Hulled	NA
			Durum	Non-brittle rachis, free-threshing	(Maccaferri et al. 2019)
Triticum timopheevii	GA	5.7Gb	Wild Timopheev's wheat	Brittle rachis, Hulled	NA
			Timopheev's wheat	Non-brittle rachis, Hulled	NA
Triticum aestivum	BAD	17Gb	Spelt	Non-brittle rachis, Hulled	(Walkowiak et al. 2020)
			Bread / Common	Non-brittle rachis, free-threshing	(Appels et al. 2018) (Alonge et al. 2020) (Walkowiak et al. 2020)

Another whole genome analysis of archaeobotanical specimens looked at two desiccated samples of 3,000-year-old emmer wheat chaff (Figure 4) from Egypt (Scott et al. 2019) to investigate early wheat dispersal and introgression from wild populations. The ancient samples were used to genotype exonic SNPs that segregate in modern accessions, at which coverage was 0.48x after quality control, yielding ~100,000 high confidence genotypes. The authors used a haplotype-based approach to overcome as much as possible the limitations of aDNA analysis of polyploid species. Nearby sites, that are not broken apart by recombination, form co-inherited blocks called haplotypes. A "haplotype reference panel" combines information from multiple modern genomes to characterise the haplotypic variation at each genomic location (McCarthy et al. 2016). In the analysis of ancient data, when a sufficient number of genotypes can be identified within a region, it is possible to assign a known haplotype (or no known haplotype, as may be the case when ancient diversity has gone lost in existing populations) to the ancient sample. At this point, non-sequenced genotypes within the region can be deduced based on haplotype assignment, a method called imputation. Haplotypes are relatively long in wheat (Walkowiak et al. 2020) because selfing tends not to break apart haplotypes as much as outcrossing. As a consequence, low coverage data is more likely to yield enough sites to assign an individual to a haplotype. This method allowed Scott et al., (2019) to identify genomic tracts tens of megabases long containing hundreds of genotypes that matched a modern sample in the haplotype reference panel. These included regions where important domestication QTLs had been identified, such that the domestication allele can be imputed and the phenotype inferred. In contrast, other genomic regions did not match anything in the haplotype reference panel.

Figure 4 *Dessicated emmer wheat chaff from Hememiah North Spur (Egypt) 14C dated 1300-1000 BC, analyzed by Scott et al., 2019. Photo by Chris J. Stevens, reproduced with permission.*

The data essentially confirmed that genetic changes associated with domestication were completed by 3 000 years ago, prior to emmer wheat dispersal to Egypt. Nevertheless, the ancient Egyptian sample carried more 'unique' haplotypes than any other domesticated sample in the dataset, indicating regions where genetic diversity has been lost. It's not yet possible to state whether this lost variation is associated with adaptation to local environmental conditions or confers other useful traits. Nevertheless, these results highlight geographic and genomic regions that may harbour genetic diversity that has been used in the past and therefore might be useful in the present and future.

Moreover, while the highly repetitive nature of the wheat genome increases the chances of misalignment issues and subsequent inflated heterozygosity, Scott et al. (2019) found that the estimated heterozygosity of the ancient sample fell within the range of the modern samples. This suggests that reliable genotypes can be obtained from ancient wheat, providing appropriate quality filters are used to restrict attention to sites that do not suffer from alignment problems.

Important results from this study concern early emmer wheat dispersal. Ancient routes of dispersal generally define modern population structure and overall genetic similarity but, with the changing usage of different wheat species and the adoption of modern elite varieties, we have little grasp of historical population dispersal and replacement. Contemporary emmer wheat subpopulations (landraces) reflect the dispersal outside of SW-Asia to the West (Mediterranean), to the Balkans (Eastern Europe), to Transcaucasia (Caucasus) and towards India and the Arabian peninsula (Indian Ocean) (Avni et al. 2017). The authors found that the ancient sample from Egypt resembles modern cultivars from the Indian Ocean subgroup, indicating a connection between early emmer dispersal to the East (across the Iranian Plateau and into the Indus valley) and to the South-West (Nile Valley). This is particularly interesting in light of the fact that Ethiopia currently represents a region of genetic isolation and differentiation for tetraploid wheat. This ancient Egyptian sample also has signatures of gene flow with wild populations in the Southern Levant, which could have occurred during dispersal towards Egypt or during Egyptian conquests in the Rammesside era. We expect further aDNA studies to connect historical events with changes to wheat genetics. Answering these questions will not only bring a deeper understanding of wheat evolution, but also human history, which has been intimately linked to wheat cultivation for millennia.

Overall, the field of wheat archaeogenomics has yet to reach its full potential. However, the field is primed for new advances with the availability of reference genomes and a wealth of resequenced modern landraces for comparison. While the prospects for studying DNA from charred remains are poor, many desiccated or

waterlogged samples have great potential for further study. Archaeological research on waterlogged sites is increasing, which promises new material to complement the specimens currently in museums and collections.

4. Analysing degraded DNA from ancient polyploid wheat

Degradation and contamination are key complications for the reliable analysis of ancient DNA. To mitigate these problems, specific methods have been developed for sample preparation and downstream analysis (reviewed in Orlando et al., 2021). Even with appropriate methodology, DNA from ancient and historical samples cannot be used for all the applications that modern sequence data allows. We briefly overview these general principles of ancient DNA analysis, before discussing the specific issues posed by wheat, as all these factors should be considered during study design and analysis. We expect future methodological improvements to address these challenges, raising the possibility of resolving further important questions in the history of wheat domestication and evolution.

4.1. aDNA damage

A prominent difference between ancient and modern DNA is that ancient DNA is much more fragmented prior to extraction (Figure 5a). Most DNA fragmentation occurs rapidly after death (Kistler et al., 2017), as the DNA 'backbone' breaks down through a process called 'hydrolytic depurination', which is biochemically predicted to occur more rapidly with exposure to water and high temperatures (Lindahl 1993). Thus, local preservation and environmental conditions are key in determining DNA yield and quality in different samples. Nevertheless, fruitful DNA sequencing has been conducted from plant tissue that is thousands of years old and from tropical and warm environments (Fornaciari et al. 2018; Mascher et al. 2016; Ramos-Madrigal et al. 2016; Renner et al. 2019). Overall, excellent DNA preservation has been reported from plant remains in desiccated and waterlogged conditions (Kistler et al., 2020).

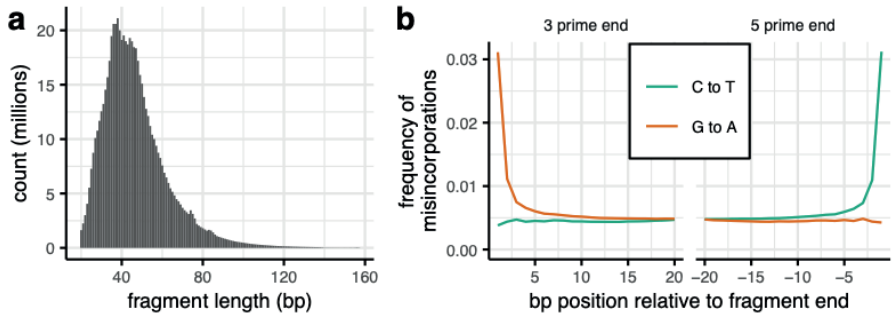


Figure 5 *Characteristic patterns of DNA degradation in sequence from a 3,000-year-old emmer wheat sample* (Scott et al. 2019). *(a) shows the raw distribution of fragments sizes and (b) shows misincorporations relative to the reference genome after alignment. In this case, the sequenced library was partially UDG treated such that the misincorporations caused by post-mortem damage are confined to a few base pairs at the fragment ends, which are removed for further analysis.*

Besides fragmentation, the DNA sequence itself undergoes modifications. Notably, a proportion of cytosine residues lose an amine group, becoming uracil residues, which code as thymine during sequencing (Briggs et al. 2007). This hydrolytic deamination occurs more commonly on the single stranded overhangs of the fragmented DNA molecules. As a result, when aligned to a reference genome, sequenced ancient DNA has a higher proportion of C-to-T misincorporations at the 5` end of each fragment. Double-stranded DNA libraries will also show a higher proportion of the complementary misincorporation, G-to-A, at the 3` end of each fragment after alignment.

These characteristic patterns of degradation found in ancient samples can be useful to the analysis, as they are proof of the sample antiquity. Therefore, the most common approach, is to carry out a protocol developed for partial UDG treatment (Rohland et al. 2015). With this method, uracil-DNA-glycosylase (UDG), is used to remove uracils (Briggs et al. 2010) in the inner region of the fragments, but not at their ends. In this way, some amount of damage is maintained, but it is confined to the fragment ends (Figure 5b). Similarly, the distribution of fragment lengths is used to confirm that the sequenced DNA is ancient, where large fragments may indicate contamination. Finally, paired-end sequencing of short fragments will often result in the same base pair being sequenced twice, which can be used to improve confidence in the sequence (Jonsson et al. 2014).

Standard bioinformatic protocols have been established for processing fragmented and damaged DNA. In general, standard approaches have been established for

mapping short-read data to reference genomes and automated tools/pipelines are available for ancient genotypes calling for downstream analyses (Peltzer et al. 2016; Schubert et al. 2014). Common methods involve trimming off all the base pairs at the end of fragments that are potentially affected by damage (Jonsson et al. 2014) and verifying that analyses are unaffected when transitions (SNPs where the two alleles are either C/T or G/A and that can include post-mortem damage) are excluded (Korneliussen, T. S. Albrechtsen and Nielsen 2014). We further note that 'reference bias' (preferential alignment of reads carrying the same allele as the reference) is stronger in ancient data due to the shorter fragment size, so correction methods should be used (Günther and Nettelblad 2019).

For all these reasons, whole genome sequencing has become the standard in ancient DNA studies, while PCR-based approaches are no longer considered unless for very specific goals such as genome identification, since they do not allow to verify the presence of these important patterns of post-mortem damage and to exclude contamination.

Contamination is a significant concern in ancient DNA studies. Because the amount of DNA preserved in ancient samples tends to be low, relatively small amounts of contamination from contemporary material can overwhelm the target DNA in the library (Renaud et al. 2019). Extraction and manipulation of ancient DNA therefore requires specialised facilities with protocols that minimize contamination by modern DNA (Fulton 2012). Standard practice is to create a control sequencing library without using the sample tissue (an 'extraction blank'). The data from controls is analysed alongside the main sample to quantify the contamination and spurious signals likely to have been introduced during DNA extraction. Contamination can also come from microbial decomposers that invade tissues after death. A simple estimate for overall contamination is the percentage of reads that can be aligned to the reference genome of the targeted species, although other methods are available (Peyrégne and Prüfer 2020).

So far, the percentage of endogenous DNA (the DNA of interest) reported in whole genome studies of ancient plants has been high, compared to animal studies. For example, reported endogenous fractions have been 33%-66% in emmer wheat (Scott et al. 2019), 5%-90% in bread wheat (Wu et al. 2019), 7%-54% (mean 44%) in common bean (Trucchi et al. 2021) and 70% in maize (Ramos-Madrigal et al. 2016).

Degradation and contamination limit the applications of ancient DNA, relative to modern DNA. Firstly, the fraction of endogenous DNA in well-preserved ancient DNA libraries is far below that of modern DNA (which usually is >99%). Because endogenous fragments are short, the sequencer will often read through the DNA

fragment and continue onto the adapter sequences used for library preparation. Sequenced adapter fragments must thus be discarded. Furthermore, if the sequencing has been performed for paired-ends, the forward and reverse reads will overlap (and are then collapsed into a consensus sequence). Given the low endogenous content and the short fragments, more sequence data is needed to reach reasonable coverage. Nevertheless, when small amounts of DNA are present in the sample, it may not be possible to keep sequencing to increase the coverage, since the library gradually yields diminishing returns as more duplicate reads are sequenced (Link et al. 2017). For all these reasons, coverage tends to be significantly lower in aDNA studies, when compared to the expectations for modern data.

Overall, due to low coverage and short fragments in ancient DNA, a typical approach is to identify variable sites (e.g., SNPs) using modern samples only, then use ancient DNA alignments to genotype the ancient samples. Fortunately, this approach often yields sufficient high-quality genotypes to perform analyses of interest, such as estimating genome-wide relatedness, introgression, and population genetic parameters.

4.2. Large polyploid wheat genomes

The large genome of wheat (17 gigabases for bread wheat) implies that whole-genome sequencing of each wheat sample requires more resources compared to other organisms with smaller genomes. This cost is exacerbated in ancient DNA studies by the lower fraction of endogenous DNA, which requires further sequencing effort to obtain the same genomic coverage. In wheat, pre-designed probes are available for exons and promoters (Gardiner et al. 2019; Jordan et al. 2015), which reduce sequencing costs by enriching for sequences that are captured by the probes used. In ancient DNA, capture can enrich endogenous DNA (Hofreiter et al. 2015) but increase clonality and introduce biases towards the sequence on the probes (Ávila-Arcos et al. 2011). Exome-wide capture has not been reported for an ancient wheat. However, targeted capture might be useful since short aDNA fragments can give little information about repetitive regions.

Ploidy and the high identity between subgenomes, estimated to be as high as 97 – 98% , supposes another challenge for ancient DNA studies. Even with modern samples, wheat resequencing studies can only reliably observe genomic regions that can be unambiguously aligned using the read lengths available. The shorter fragment length of ancient DNA places a practical limit on the portion of the genome that can be directly observed by mapping to reference genomes.

Heterozygosity is commonly used as an indicator of misalignment problems. Because wheat is predominantly selfing (Golenberg 1988), most sites should be homozygous in most individuals. However, various structural variants can cause reads from different genomic regions in the sample to be aligned to the same position in the reference genome (Figure 6) with high mapping-quality scores, thus passing quality filters. As a consequence, sample heterozygosity will be inflated after calling genotypes. A common solution is to remove variants that are heterozygous in multiple samples e.g. (Gardiner et al. 2019; He et al. 2019). Recent data indicate that undetected gene duplicates are common within wheat subgenomes on reference assemblies (Alonge et al. 2020). In general, polyploid wheat resequencing data will suffer from additional misalignments due to homeologous sequences on different subgenomes, but reliable gentypes can be obtained from both modern and ancient wheat providing appropriate quality filters are used to restrict attention to sites that do not suffer from alignment problems. Nevertheless, we emphasise that care should be taken when measuring heterozygosity in polyploid wheats, especially from ancient genomes. This is unfortunate because heterozygosity is a common indicator of outcrossing and genetic variation in the population, changes to which are key questions in the history of cultivation practices (Smith et al. 2019; Trucchi et al. 2021).

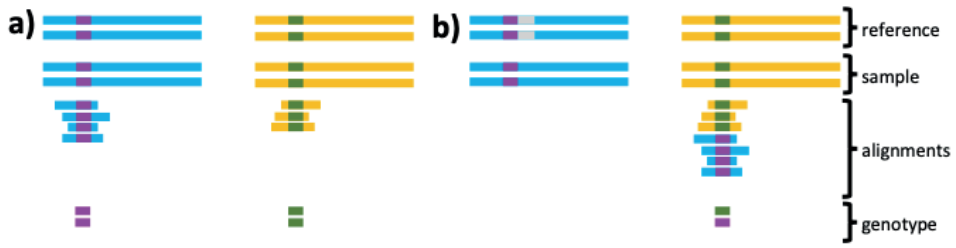


Figure 6 *False heterozygosity introduced by mis-mappings to the reference. Here we consider two genomic regions (blue and yellow), which are homeologues or duplicated regions that are relatively similar to one another. A site in each region is genotyped (coloured purple and green). In (a), the sample is similar to the reference so that reads can be aligned to the correct region and the genotype calls are all homozygous, as expected for most sites in a largely selfing species. In (b) there is a difference between the reference genome and sequenced genome (indicated in grey). The sample reads from the blue genomic region in (b) are best aligned to the yellow region of the reference. This results in a heterozygous genotype call, while all the true genotypes are homozygous. Thus, inaccurate reference genome assemblies, deletions, insertions, or duplications can all result in spurious heterozygous genotypes.*

5. The future of the past: open questions and prospects for wheat aDNA

Crop archaeogenomics has already proved to be a powerful tool to investigate phenomena such as domestication, crop dispersal, and subsequent adaptation (Logan Kistler et al. 2020; Orlando et al. 2021). Studies on bean (Trucchi et al. 2021), sunflower (Wales et al. 2019) and sorghum (Smith et al. 2019) showed that the 'domestication bottleneck' (i.e. the initial loss of genetic diversity associated with domestication), may not be as intense as previously assumed. Ancient DNA analysis has been used to trace the origin of some important winemaking grape cultivars (Ramos-Madrigal et al. 2019), and brought insights on the genetic basis of potato adaptation to the European climate (Gutaker, Weiß, et al. 2019). In maize, adaptation to climatic constraints (selected from ancient standing variation within the domestic forms) has been identified as the main driver of modern differentiation between populations (Da Fonseca et al. 2015; Swarts et al. 2017).

5.1. Open questions in domestication

In recent years, some paradigms of domestication have been challenged by new scientific discoveries, and wheat represents a good example of such changing perspectives.

Because now we know that domestic forms took thousands of years to dominate archaeological assemblages and that different wild populations seem to contribute to modern diversity, it is likely that wheat domestication was not as severe, abrupt, or geographically restricted as expected under the assumption of a "domestication bottleneck" (see section 2). The presence of peculiar haplotypes in an ancient emmer wheat sample from Egypt showed that possibly genetic diversity has been lost after emmer wheat domestication and dispersal to Egypt (Scott et al., 2019), in line with what has been found for other species e.g. (Trucchi et al. 2021). In the case of wheat, more ancient samples are needed to determine the association (or lack of thereof) between domestication and losses of genetic diversity.

Second, it is unclear whether there is a monophyletic "centre of domestication" for emmer wheat in the Northern Levant. The contribution of the Southern Levant gene pool to domestic emmer has been detected in several studies, but its origin remains unsolved. Whether emmer was domesticated from a proto-domestic admixed population, or if early domestic populations benefited from extensive gene flow from the wild is still to be revealed. It has been proposed that the high genetic similarity of modern domestic to Turkish wild emmer could be explained by a feralization of the very first proto-domestic population (Civáň et al., 2013; Oliveira et al., 2020). The

analysis of wild and domestic samples from this region dating back to PPN and Neolithic could help determine the origin of the domestic pool, and its relationships with ancient and extant wild populations.

The recent genetic identification of domesticated *T. timopheevii* has triggered a re-evaluation of its importance and abundance in the archaeological record. This effort will be greatly aided by a genetic survey of the modern wild specimens, together with ancient seeds. In general, it will be interesting to use ancient and modern genetic data to compare the origins in space and time of parallel domestication events in wheat (emmer wheat, einkorn wheat, and *T. timopheevii)*.

Prospects for the analysis of DNA from fully charred remains are poor, which limits the direct genetic analysis to unveil some of the earliest and most crucial events in wheat domestication. Nevertheless, we expect that improvements in the modelling of genomic evolution and the increasing availability of waterlogged remains will allow to test alternative scenarios on top of addressing questions concerning adaptation and spread of wheat.

5.2. Open questions in dispersal and adaptation

The dispersal of wheat was accompanied by adaptation to different environments, leading to the evolutionary success of this species. An interesting example is adaptation to altitude along certain dispersal routes. Wild emmer wheat from the Northern Levant, the closest to all domestic landraces, is always found at high altitude. Its dispersal towards Egypt entailed cultivation at sea level, but emmer wheat grown on the Ethiopian plateau is cultivated at high altitudes again. There are two possible routes of dispersal leading to Ethiopia, one through Africa and another through the Iranian plateau and the Arabian Peninsula. The first one would entail a second adaptation event to high altitudes. The other would have always been cultivated at high altitudes, but there would require a longer dispersal route. How did emmer wheat arrive to Ethiopia? The analysis of desiccated specimens from the Arabian Peninsula, Sudan and ideally Iran could help to answer this question, as well as potentially unveiling genetic mechanisms for adaptation to high altitude.

5.3. Open questions in hybridisation and speciation

Archaeological data increasingly suggests that different wheat species were used in a complex geographical mosaic that shifted through time. Given that several wheat species, i.e. emmer, einkorn, naked wheats and *Timopheevi* (and wild relatives) coexisted in the same area for millennia, we can ask how much genetic exchange was ongoing in Neolithic settlements. While the vast majority of wheat cultivated today is

bread wheat, other free-threshing hexaploids such as the Indian dwarf wheat or the Yunan wheat could have arisen from different hybridization events, since the phylogeny of the A and B genomes differs from that of the D genome (Y. Zhou et al. 2020). Furthermore, forms such as *Triticum compactum* (Club Wheat) have been described (e.g. Kaplan et al., 1992), even though it is unclear whether these morphotypes are the product of different hybridizations events or the consequence of differential selective pressures. A comparison of the D subgenome in ancient hexaploids with modern *Aegilops* specimens could tackle this question and narrow down the geographic origin where these hybridizations occurred.

Even more intriguingly, we can speculate whether introgressed genetic variation between different wheats was important for crop evolution and adaptation to different environments such as adaptation to northern latitudes or to heat stress. Einkorn wheat and spelt were important crops in central and northern Europe. On the other hand, hexaploid free-threshing wheats such as Indian dwarf wheat and *Triticum compactum* are more commonly found in warm environments. Studying changes in allele frequencies with the spread of these crops into new environments would identify candidate adaptive regions, whose phenotypic effects and usefulness could be analysed through crossing and genetic mapping. Learning from the phylogenetic relationship between ancient wheat specimens would greatly increase the power to detect the genomic regions conferring adaptation to those traits.

Furthermore, besides the impact that archaegenomics has on our understanding of the past, it has also the potential to set the basis for future food security (Pont, Wagner, Kremer, Orlando, et al. 2019), conservation and breeding strategies, in the current context of climate change (di Donato et al. 2018). During the dispersal of domestic plants, crops adapted to a multitude of environments, and aDNA can reveal genetic diversity present in historical landraces but lost from the modern domestic pool (e.g., Scott et al., 2019). Detecting signals of positive selection in such lost diversity may therefore be particularly valuable, especially when it is the source of adaptations to extreme environments. After its identification, such diversity can be prioritised for preservation or introduced to modern cultivars via breeding if still present in seed banks, landraces, or wild relatives (di Donato et al. 2018). Plant aDNA studies can lead to the identification of lost crops and their wild relatives, revealing their genetic makeup. Such knowledge could set the ground for *de novo* domestications and ultimately aid in the diversification of our food system, which currently relies on a rather small number of domestic species (Estrada et al. 2018). Finally, aDNA can be informative of past plant-pathogens interactions and their co-evolution e.g. (Yoshida et al. 2013), providing valuable insights for crop management (di Donato et al. 2018; Estrada et al. 2018; Przelomska, Armstrong, and Kistler 2020).

In conclusion, archaeogenomics allows interrogation of a plethora of questions about wheat evolutionary history, such as population continuity and demographic changes through time, identification of climatic or cultural conditions that correspond to germplasm shifts and relationships with other wheats. We expect these questions to be addressed in future aDNA studies. Overall, answering these questions will not only bring a deeper understanding of wheat evolution, but will also aid answering questions about human cultural evolution and trade.

Bibliography

Abbo, S., Pinhasi van-Oss, R., Gopher, A., Saranga, Y., Ofner, I., & Peleg, Z. (2014). Plant domestication versus crop evolution: A conceptual framework for cereals and grain legumes. Trends in Plant Science, 19(6), 351–360. https://doi.org/10.1016/j.tplants.2013.12.002

Allaby, R. G., Banerjee, M., & Brown, T. A. (1999). Evolution of the high molecular weight glutenin loci of the A, B, D, and G genomes of wheat. Genome, 42(2), 296–307. https://doi.org/10.1139/g98-114

Alonge, M., Shumate, A., Puiu, D., Zimin, A., & Salzberg, S. L. (2020). Chromosome-Scale Assembly of the Bread Wheat Genome Reveals Thousands of Additional Gene Copies. Genetics, genetics.303501.2020. https://doi.org/10.1534/genetics.120.303501

Arranz-Otaegui, A., Colledge, S., Zapata, L., Teira-Mayolini, L. C., & Ibáñez, J. J. (2016). Regional diversity on the timing for the initial appearance of cereal cultivation and domestication in southwest Asia. Proceedings of the National Academy of Sciences of the United States of America, 113(49), 14001–14006. https://doi.org/10.1073/pnas.1612797113

Arranz-Otaegui, A., González Carretero, L., Roe, J., & Richter, T. (2018). "Founder crops" v. wild plants: Assessing the plant-based diet of the last hunter-gatherers in southwest Asia. Quaternary Science Reviews, 186, 263–283. https://doi.org/10.1016/j.quascirev.2018.02.011

Ávila-Arcos, M. C., Cappellini, E., Romero-Navarro, J. A., Wales, N., Moreno-Mayar, J. V., Rasmussen, M., Fordyce, S. L., Montiel, R., Vielle-Calzada, J. P., Willerslev, E., & Gilbert, M. T. P. (2011). Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. Scientific Reports, 1. https://doi.org/10.1038/srep00074

Avni, R., Nave, M., Barad, O., Baruch, K., Twardziok, S. O., Gundlach, H., Hale, I., Mascher, M., Spannagl, M., Wiebe, K., Jordan, K. W., Golan, G., Deek, J., Ben-zvi, B., Ben-zvi, G., Himmelbach, A., Maclachlan, R. P., Sharpe, A. G., Komatsuda, T., … Distelfeld, A. (2017). Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. Science, 97(July), 93–97. https://doi.org/10.1126/science.aan0032

Bilgic, H., Hakki, E. E., Pandey, A., Khan, M. K., & Akkaya, M. S. (2016). Ancient DNA from 8400 Year-Old Çatalhöyük Wheat: Implications for the Origin of Neolithic Agriculture. Plos One, 11(3), e0151974. https://doi.org/10.1371/journal.pone.0151974

Blatter, R. H. E., Jacomet, S., & Schlumbaum, A. (2002). Spelt-specific alleles in HMW glutenin genes from modern and historical European spelt (Triticum spelta L.). Theoretical and Applied Genetics, 104(2–3), 329–337. https://doi.org/10.1007/s001220100680

Botigue, L. R., Song, S., Scheu, A., Gopalan, S., Pendleton, A. L., Zeeb-lanz, A., Arbogast, R., Oetjens, M., Taravella, A. M., Serege, T., Burger, J., Kidd, J. M., Veeramah, K. R., Bobo, D., Daly, K., & Unterla, M. (2017). Ancient European dog genomes reveal continuity since the Early Neolithic. Nature Communications, May. https://doi.org/10.1038/ncomms16082

Briggs, A. W., Stenzel, U., Johnson, P. L. F., & Pääbo, S. (2007). Patterns of damage in genomic DNA sequences from a Neandertal. Proceedings of the National Academy of Sciences, 104, 14616–14621. https://doi.org/doi: 10.1073/pnas.0704665104.

Briggs, A. W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., & Pääbo, S. (2010). Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. Nucleic Acids Research, 38. https://doi.org/doi: 10.1093/nar/gkp1163

Cheng, H., Liu, J., Wen, J., Nie, X., Xu, L., Chen, N., Li, Z., Wang, Q., Zheng, Z., Li, M., Cui, L., Liu, Z., Bian, J., Wang, Z., Xu, S., Yang, Q., Appels, R., Han, D., Song, W., … Jiang, Y. (2019). Frequent intra- and inter-species introgression shapes the landscape of genetic variation in bread wheat. Genome Biology, 20(1), 1–16. https://doi.org/10.1186/s13059-019-1744-x

Civáň, P., Ivaničová, Z., & Brown, T. A. (2013). Reticulated origin of domesticated emmer wheat supports a dynamic model for the emergence of agriculture in the fertile crescent. PLoS ONE, 8(11), 1–11. https://doi.org/10.1371/journal.pone.0081955

Czajkowska, B. I., Bogaard, A., Charles, M., Jones, G., Kohler-Schneider, M., Mueller-Bieniek, A., & Brown, T. A. (2020). Ancient DNA typing indicates that the "new" glume wheat of early Eurasian agriculture is a cultivated member of the Triticum timopheevii group. Journal of Archaeological Science, 123(October), 105258. https://doi.org/10.1016/j.jas.2020.105258

Da Fonseca, R. R., Smith, B. D., Wales, N., Cappellini, E., Skoglund, P., Fumagalli, M., Samaniego, J. A., Carøe, C., Ávila-Arcos, M. C., Hufnagel, D. E., Korneliussen, T. S., Vieira, F. G., Jakobsson, M., Arriaza, B., Willerslev, E., Nielsen, R., Hufford, M. B., Albrechtsen, A., Ross-Ibarra, J., & Gilbert, M. T. P. (2015). The origin and evolution of maize in the Southwestern United States. Nature Plants, 1(January), 1–5. https://doi.org/10.1038/nplants.2014.3

Daly, K. G., Delser, P. M., Mullin, V. E., Scheu, A., Mattiangeli, V., Teasdale, M. D., Hare, A. J., Burger, J., Verdugo, M. P., Collins, M. J., Kehati, R., & Erek, C. M. (2018). Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. Science , 85–88, 361, 24–27. https://doi.org/https://www.science.org/doi/10.1126/science.aas9411

Der Sarkissian, C., Allentoft, M. E., Avila-Arcos, M. C., Barnett, R., Campos, P. F., Cappellini, E., Ermini, L., Fernandez, R., da Fonseca, R., Ginolhac, A., Hansen, A. J., Jonsson, H., Korneliussen, T., Margaryan, A., Martin, M. D., Moreno-Mayar, J. V., Raghavan, M., Rasmussen, M., Velasco, M. S., … Orlando, L. (2014). Ancient genomics. Philosophical Transactions of the Royal Society B: Biological Sciences, 370(1660), 20130387–20130387. https://doi.org/10.1098/rstb.2013.0387

di Donato, A., Filippone, E., Ercolano, M. R., & Frusciante, L. (2018). Genome sequencing of ancient plant remains: Findings, uses and potential applications for the study and improvement of modern crops. Frontiers in Plant Science, 9(April). https://doi.org/10.3389/fpls.2018.00441

Diamond, J. (2002). Evolution, consequences and future of plant and animal domestication. Nature, 418(6898), 700–707. https://doi.org/10.1038/nature01019

Estrada, O., Breen, J., Richards, S. M., & Cooper, A. (2018). Ancient plant DNA in the genomic era. Nature Plants, 4(7), 394–396. https://doi.org/10.1038/s41477-018-0187-9

Feldman, M., & Kislev, E. M. (2007). Domestication of emmer wheat and evolution of free-threshing tetraploid wheat. Isr J Plant Sci, 55, 207–221.

Fernández, E., Thaw, S., Brown, T. A., Arroyo-Pardo, E., Buxò, R., Serret, M. D., & Araus, J. L. (2013). DNA analysis in charred grains of naked wheat from several archaeological sites in Spain. Journal of Archaeological Science, 659–670. https://doi.org/doi: 10.1016/j.jas.2012.07.014

Filipovic, D. (2014). Southwest Asian founder- and other crops at Neolithic sites in Serbia. Bulgarian E-Journal of Archaeology, 4(2), 229–247.

Fornaciari, R., Fornaciari, S., Francia, E., Mercuri, A. M., & Arru, L. (2018). Panicum spikelets from the Early Holocene Takarkori rockshelter (SW Libya): Archaeo-molecular and -botanical investigations. Plant Biosystems, 152(1), 1–13. https://doi.org/10.1080/11263504.2016.1244117

Frantz, L. A. F., Haile, J., Lin, A. T., Scheu, A., Benecke, N., Alexander, M., Linderholm, A., Mullin, V. E., Daly, K. G., Battista, V. M., Price, M., Gron, K. J., Alexandri, P., Arbogast, R., Adrian, B., Barnett, R., Bartosiewicz, L., Baryshnikov, G., Bonsall, C., … Stoddart, S. (2019). Ancient pigs reveal a near-complete genomic turnover following their introduction to Europe. Pnas, 116. https://doi.org/10.1073/pnas.2008793117

Fuller, D. Q., Allaby, R. G., & Stevens, C. (2010). Domestication as innovation: The entanglement of techniques, technology and chance in the domestication of cereal crops. World Archaeology, 42(1), 13–28. https://doi.org/10.1080/00438240903429680

Fuller, D. Q., Denham, T., Arroyo-Kalin, M., Lucas, L., Stevens, C. J., Qin, L., Allaby, R. G., & Purugganan, M. D. (2014). Convergent evolution and parallelism in plant domestication revealed by an expanding archaeological record. Proceedings of the National Academy of Sciences of the United States of America, 111(17), 6147–6152. https://doi.org/10.1073/pnas.1308937110

Fulton, T. L. (2012). Setting Up an Ancient DNA Laboratory. In Shapiro, B. and Hofreiter, M. (eds) Ancient DNA: Methods and Protocols. (pp. 1–11). https://doi.org/doi: 10.1007/978-1-61779-516-9_1.

Gardiner, L. J., Brabbs, T., Akhunov, A., Jordan, K., Budak, H., Richmond, T., Singh, S., Catchpole, L., Akhunov, E., & Hall, A. (2019). Integrating genomic resources to present full gene and putative promoter capture probe sets for bread wheat. GigaScience, 8(4), 1–13. https://doi.org/10.1093/gigascience/giz018

Gaunitz, C., Fages, A., Hanghøj, K., Albrechtsen, A., Khan, N., Schubert, M., Seguin-orlando, A., Owens, I. J., Felkel, S., Bignon-lau, O., Damgaard, P. D. B., Mittnik, A., Wallner, B., Merz, V., Merz, I., Zaibert, V., & Willerslev, E. (2018). Ancient genomes revisit the ancestry of domestic and Przewalski ' s horses. April, 1–5.

Golenberg, E. M. (1988). Outcrossing rates and their relationship to phenology in Triticum dicoccoides . Theoretical and Applied Genetics, 75, 937–944. https://doi.org/https://doi.org/10.1007/BF00258057

Günther, T., & Nettelblad, C. (2019). The presence and impact of reference bias on population genomic studies of prehistoric human populations. PLoS Genetics, 15(7), 1–20. https://doi.org/10.1371/journal.pgen.1008302

Gutaker, R. M., & Burbano, H. A. (2017). Reinforcing plant evolutionary genomics using ancient DNA. Current Opinion in Plant Biology, 36, 38–45. https://doi.org/10.1016/j.pbi.2017.01.002

Gutaker, R. M., Weiß, C. L., Ellis, D., Anglin, N. L., Knapp, S., Fernández-alonso, J. L., Prat, S., & Burbano, H. A. (2019). The origins and adaptation of European potatoes reconstructed from historical genomes. Nature Ecology & Evolution, 3(July). https://doi.org/10.1038/s41559-019-0921-3

Haas, M., Schreiber, M., & Mascher, M. (2018). Domestication and crop evolution of wheat and barley: Genes, genomics, and future directions. Journal of Integrative Plant Biology, XXXX(Xxxx). https://doi.org/10.1111/jipb.12737

He, F., Pasam, R., Shi, F., Kant, S., Keeble-Gagnere, G., Kay, P., Forrest, K., Fritz, A., Hucl, P., Wiebe, K., Knox, R., Cuthbert, R., Pozniak, C., Akhunova, A., Morrell, P. L., Davies, J. P., Webb, S. R., Spangenberg, G., Hayes, B., … Akhunov, E. (2019). Exome sequencing highlights the role of wild-relative introgression in shaping the adaptive landscape of the wheat genome. Nature Genetics, 51(5), 896–904. https://doi.org/10.1038/s41588-019-0382-2

Higuchi, R., Bowman, B., Freiberger, M., Ryder, O. A., & Wilson, A. C. (1984). DNA sequences from the quagga, an extinct member of the horse family. Nature, 312, 282–284.

Hofreiter, M., Paijmans, J. L. A., Goodchild, H., Speller, C. F., Barlow, A., Fortes, G. G., Thomas, J. A., Ludwig, A., & Collins, M. J. (2015). The future of ancient DNA: Technical advances and conceptual shifts. BioEssays, 37(3), 284–293. https://doi.org/10.1002/bies.201400160

Jones, G., Valamoti, S., & Charles, M. (2000). Early crop diversity: A "new" glume wheat from northern Greece. Vegetation History and Archaeobotany, 9(3), 133–146. https://doi.org/10.1007/BF01299798

Jonsson, H., Korneliussen, T., Margaryan, A., Martin, M. D., Moreno-Mayar, J. V. Raghavan, M. Rasmussen, M., Velasco, M. S., & Orlando, L. (2014). Ancient genomics. Philosophical Transactions of the Royal Society B: Biological Sciences, 370. https://doi.org/https://doi.org/10.1098/rstb.2013.0387

Jordan, K. W., Wang, S., Lun, Y., Gardiner, L. J., MacLachlan, R., Hucl, P., Wiebe, K., Wong, D., Forrest, K. L., Sharpe, A. G., Sidebottom, C. H. D., Hall, N., Toomajian, C., Close, T., Dubcovsky, J., Akhunova, A., Talbert, L., Bansal, U. K., Bariana, H. S., … Akhunov, E. (2015). A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. Genome Biology, 16(1), 1–18. https://doi.org/10.1186/s13059-015-0606-4

Jorgensen, C., Luo, M. C., Ramasamy, R., Dawson, M., Gill, B. S., Korol, A. B., Distelfeld, A., & Dvorak, J. (2017). A high-density genetic map of wild emmer wheat from the karaca dağ region provides new evidence on the structure and evolution of wheat chromosomes. Frontiers in Plant Science, 8(October), 1–13. https://doi.org/10.3389/fpls.2017.01798

Kantar, M. B., Nashoba, A. R., Anderson, J. E., Blackman, B. K., & Rieseberg, L. H. (2017). The Genetics and Genomics of Plant Domestication. BioScience, XX(X), 1–12. https://doi.org/10.1093/biosci/bix114

Kaplan, L., Smith, M. B., & Sneddon, L. A. (1992). Cereal Grain Phytoliths of Southwest Asia and Europe. In G. Rapp & S. C. Mulholland (Eds.), Phytolith Systematics in Archaeological and Museum Science (pp. 149–174). Springer US. https://doi.org/10.1007/978-1-4899-1155-1_8

Key, F. M., Posth, C., Esquivel-gomez, L. R., Hübler, R., Spyrou, M. A., Neumann, G. U., Furtwängler, A., Sabin, S., Burri, M., Wissgott, A., Lankapalli, A. K., Vågene, Å. J., Meyer, M., Nagel, S., Tukhbatova, R., Khokhlov, A., Chizhevsky, A., Hansen, S., Belinsky, A. B., … Krause, J. (2020). Emergence of human-adapted Salmonella enterica is linked to the Neolithization process. Nature Ecology & Evolution, 4(March). https://doi.org/10.1038/s41559-020-1106-9

Kistler, L., Ware, R., Smith, O., Collins, M., & Allaby, R. G. (2017). A new model for ancient DNA decay based on paleogenomic meta-analysis. Nucleic Acids Research, 45, 6310–6320. https://doi.org/doi: 10.1093/nar/gkx361.

Kistler, Logan, Bieker, V. C., Martin, M. D., Pedersen, M. W., Ramos Madrigal, J., & Wales, N. (2020). Ancient Plant Genomics in Archaeology, Herbaria, and the Environment. Annual Review of Plant Biology, 71, 605–629. https://doi.org/10.1146/annurev-arplant-081519-035837

Korneliussen, T. S. Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics, 15, 1–13. https://doi.org/doi: 10.1186/s12859-014-0356-4.

Krause, J., Fu, Q., Good, J. M., Viola, B., Shunkov, M. V, & Derevianko, A. P. (2010). The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. 464(April), 894–897. https://doi.org/10.1038/nature08976

Lacan, M., Keyser, C., Ricaut, F., Brucato, N., Duranthon, F., & Guilaine, J. (2011). Ancient DNA reveals male diffusion through the Neolithic Mediterranean route. https://doi.org/10.1073/pnas.1100723108/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1100723108

Larson, G., Piperno, D. R., Allaby, R. G., Purugganan, M. D., Andersson, L., Arroyo-Kalin, M., Barton, L., Climer Vigueira, C., Denham, T., Dobney, K., Doust, A. N., Gepts, P., Gilbert, M. T. P., Gremillion, K. J., Lucas, L., Lukens, L., Marshall, F. B., Olsen, K. M., Pires, J. C., … Fuller, D. Q. (2014). Current perspectives and the future of domestication studies. Proceedings of the National Academy of Sciences, 111(17), 6139–6146. https://doi.org/10.1073/pnas.1323964111

Lawson, D. J., van Dorp, L., & Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. Nature Communications, 9(1), 1–11. https://doi.org/10.1038/s41467-018-05257-7

Leathlobhair, M. N., Perri, A. R., Irving-pease, E. K., & Witt, K. E. (2018). The evolutionary history of dogs in the Americas. 85(July), 81–85.

Lev-Yadun, Simcha; Gopher, Avi; Abbo, S. (2000). The Cradle of Agriculture. Science, 288(June).

Li, C., Lister, D. L., Li, H., Xu, Y., Cui, Y., Bower, M. A., Jones, M. K., & Zhou, H. (2011). Ancient DNA analysis of desiccated wheat grains excavated from a Bronze Age cemetery in Xinjiang. Journal of Archaeological Science, 38(1), 115–119. https://doi.org/10.1016/j.jas.2010.08.016

Librado, P., Khan, N., Fages, A. et al. (2021). The origins and spread of domestic horses from the Western Eurasian steppes. Nature, 598. https://doi.org/10.1038/s41586-021-04018-9

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. Nature, 362, 709–715. https://doi.org/doi: 10.1038/362709a0

Link, V., Kousathanas, A., Veeramah, K., Sell, C., Scheu, A., & Wegmann, D. (2017). ATLAS: Analysis Tools for Low-depth and Ancient Samples. BioRxiv, 33(16), 1–7.

Luo, M. C., Yang, Z. L., You, F. M., Kawahara, T., Waines, J. G., & Dvorak, J. (2007). The structure of wild and domesticated emmer wheat populations, gene flow between them, and the site of emmer domestication. Theoretical and Applied Genetics, 114(6), 947–959. https://doi.org/10.1007/s00122-006-0474-0

Marciniak, S., & Perry, G. H. (2017). Harnessing ancient genomes to study the history of human adaptation. Nature Publishing Group, 18(11), 659–674. https://doi.org/10.1038/nrg.2017.65

Mascher, M., Schuenemann, V. J., Davidovich, U., Marom, N., Himmelbach, A., Hübner, S., Korol, A., David, M., Reiter, E., Riehl, S., Schreiber, M., Vohr, S. H., Green, R. E., Dawson, I. K., Russell, J., Kilian, B., Muehlbauer, G. J., Waugh, R., Fahima, T., … Stein, N. (2016). Genomic analysis of 6,000-year-old cultivated grain illuminates the domestication history of barley. Nature Genetics, 48(9), 1089–1093. https://doi.org/10.1038/ng.3611

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., & Ripatti, S. (2016). A reference panel of 64,976 haplotypes for genotype imputation. Nature Genetics, 48, 1279–1283.

Morozova, I., Flegontov, P., Mikheyev, A. S., Bruskin, S., Asgharian, H., Ponomarenko, P., Klyuchnikov, V., Kumar, G. P. A., Prokhortchouk, E., Gankin, Y., Rogaev, E., Nikolsky, Y., Baranova, A., Elhaik, E., & Tatarinova, T. V. (2016). Toward high-resolution population genomics using archaeological samples. DNA Research, 23(4), 295–310. https://doi.org/10.1093/dnares/dsw029

Nave, M., Avni, R., Çakır, E., Portnoy, V., Sela, H., Pourkheirandish, M., Ozkan, H., Hale, I., Komatsuda, T., Dvorak, J., & Distelfeld, A. (2019). Wheat domestication in light of haplotype analyses of the Brittle rachis 1 genes (BTR1-A and BTR1-B). Plant Science, 285(May), 193–199. https://doi.org/10.1016/j.plantsci.2019.05.012

Nesbitt, M., & Samuel, D. (1996). From staple crop to extinction? The archaeology and history of the hulled wheat. In S. Padulosi, K. Hammer, & J. Heller (Eds.), Hulled wheats, promoting the conservation and used of underutilized and neglected crops. (pp. 40–99).

Nielsen, R., Akey, J. M., Jakobsson, M., Pritchard, J. K., Tishkoff, S., & Willerslev, E. (2017). Tracing the peopling of the world through genomics. Nature, 541. https://doi.org/10.1038/nature21347

Oliveira, H. R., Jacocks, L., Czajkowska, B. I., Kennedy, S. L., & Brown, T. A. (2020). Multiregional origins of the domesticated tetraploid wheats. PLoS ONE, 15(1), 1–20. https://doi.org/10.1371/journal.pone.0227148

Orlando, L., Allaby, R., Pontus, S., Der Sarkissian, C., Stockhammer, P. W., Ávila-Arcos, M. C., Fu, Q., Krause, J., Willerslev, E., Stone, A., & Christina, W. (2021). Ancient DNA analysis. Nature Reviews Genetics, 1–26. https://doi.org/10.1038/s43586-020-00011-0

Özbaşaran, M., Duru, G., Stiner, M. C., & Esin, U. (2018). The early settlement at Aşıklı Höyük: essays in honor of Ufuk Esin. Ege Yayınları.

Ozkan, H., Brandolini, A., Pozzi, C., Effgen, S., Wunder, J., & Salamini, F. (2005). A reconsideration of the domestication geography of tetraploid wheats. Theoretical and Applied Genetics, 110, 1052– 1060.

Ozkan, H., Brandolini, A., Schafer-Pregl, R., & Salamini, F. (2002). AFLP analysis of a collection of tetraploid wheat indicated the origin of emmer and hard wheat domestication in south- eastern Turkey. Molecular Biology and Evolution, 19, 1797–1801.

Palmer, S. A., Clapham, A. J., Rose, P., Freitas, F. O., Owen, B. D., Beresford-Jones, D., Moore, J. D., Kitchen, J. L., & Allaby, R. G. (2012). Archaeogenomic evidence of punctuated genome evolution in gossypium. Molecular Biology and Evolution, 29(8), 2031–2038. https://doi.org/10.1093/molbev/mss070

Palmer, S. A., Moore, J. D., Clapham, A. J., Rose, P., & Allaby, R. G. (2009). Archaeogenetic evidence of ancient nubian barley evolution from six to two-row indicates local adaptation. PLoS ONE, 4(7), 2–8. https://doi.org/10.1371/journal.pone.0006301

Parducci, L., Bennett, K. D., Ficetola, G. F., Alsos, I. G., Suyama, Y., Wood, J. R., & Pedersen, M. W. (2017). Ancient plant DNA in lake sediments. New Phytologist, 214(3), 924–942. https://doi.org/10.1111/nph.14470

Pedersen, J. S., Valen, E., Velazquez, A. M. V., Parker, B. J., Rasmussen, M., Lindgreen, S., Lilje, B., Tobin, D. J., Kelly, T. K., Vang, S., Andersson, R., Jones, P. A., Hoover, C. A., Tikhonov, A., Prokhortchouk, E., Rubin, E. M., Sandelin, A., Gilbert, M. T. P., Krogh, A., … Orlando, L. (2014). Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. 454–466. https://doi.org/10.1101/gr.163592.113.Freely

Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., & Nieselt, K. (2016). EAGER: efficient ancient genome reconstruction. Genome Biology, 17, 60. https://doi.org/10.1186/s13059-016-0918-z

Peyrégne, S., & Prüfer, K. (2020). Present-Day DNA Contamination in Ancient DNA Datasets. BioEssays, 42(1–11). https://doi.org/doi: 10.1002/bies.202000081

Pont, C., Leroy, T., Seidel, M., Tondelli, A., Duchemin, W., Armisen, D., Lang, D., Bustos-Korts, D., Goué, N., Balfourier, F., Molnár-Láng, M., Lage, J., Kilian, B., Özkan, H., Waite, D., Dyer, S., Letellier, T., Alaux, M., Russell, J., … Ozkan, H. (2019). Tracing the ancestry of modern bread wheats. Nature Genetics, 51(5), 905–911. https://doi.org/10.1038/s41588-019-0393-z

Pont, C., Wagner, S., Kremer, A., Orlando, L., Plomion, C., & Salse, J. (2019). Paleogenomics: reconstruction of plant evolutionary trajectories from modern and ancient DNA. Genome Biology, 20(1), 29. https://doi.org/10.1186/s13059-019-1627-1

Prüfer, K., & Meyer, M. (2015). Comment on "Late Pleistocene human skeleton and mtDNA link Paleoamericans and modern Native Americans." Science, 347(6224), 835a. https://doi.org/10.1126/science.1260617

Przelomska, N. A. S., Armstrong, C. G., & Kistler, L. (2020). Ancient Plant DNA as a Window Into the Cultural Heritage and Biodiversity of Our Food System. Frontiers in Ecology and Evolution, 8(March), 1–8. https://doi.org/10.3389/fevo.2020.00074

Przewieslik-Allen, A. M., Wilkinson, P. A., Burridge, A. J., Winfield, M. O., Dai, X., Beaumont, M., King, J., Yang, C. yun, Griffiths, S., Wingen, L. U., Horsnell, R., Bentley, A. R., Shewry, P., Barker, G. L. A., & Edwards, K. J. (2021). The role of gene flow and chromosomal instability in shaping the bread wheat genome. Nature Plants, 7(2), 172–183. https://doi.org/10.1038/s41477-020-00845-2

Purugganan, M. D., & Fuller, D. Q. (2009). The nature of selection during plant domestication. Nature, 457(7231), 843–848. https://doi.org/10.1038/nature07895

Ramos-Madrigal, J., Runge, A. K. W., Bouby, L., Lacombe, T., Samaniego Castruita, J. A., Adam-Blondon, A. F., Figueiral, I., Hallavant, C., Martínez-Zapater, J. M., Schaal, C., Töpfer, R., Petersen, B., Sicheritz-Pontén, T., This, P., Bacilieri, R., Gilbert, M. T. P., & Wales, N. (2019). Palaeogenomic insights into the origins of French grapevine diversity. Nature Plants, 5(6), 595–603. https://doi.org/10.1038/s41477-019-0437-5

Ramos-Madrigal, J., Smith, B. D., Moreno-Mayar, J. V., Gopalakrishnan, S., Ross-Ibarra, J., Gilbert, M. T. P., & Wales, N. (2016). Genome Sequence of a 5,310-Year-Old Maize Cob Provides Insights into the Early Stages of Maize Domestication. Current Biology, 26(23), 3195–3201. https://doi.org/10.1016/j.cub.2016.09.036

Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., Viola, B., Briggs, A. W., Stenzel, U., Johnson, P. L. F., Maricic, T., Good, J. M., Marques-bonet, T., Alkan, C., Fu, Q., Mallick, S., Li, H., Meyer, M., Eichler, E. E., & Stoneking, M. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. https://doi.org/10.1038/nature09710

Renaud, G., Schubert, M., Sawyer, S., & Orlando, L. (2019). Authentication and Assessment of Contamination in Ancient DNA. In Shapiro, B. et al. Ancient DNA: Methods and Protocols (pp. 163–194). Springer US. https://doi.org/doi: 10.2307/j.ctt183pd9z.20

Renner, S. S., Pérez-Escbar, O. A., Silber, M. V., Nesbitt, M., Preick, M., Hofreiter, M., & Chomicki, G. (2019). A 3500-year-old leaf from a Pharaonic tomb reveals that New Kingdom Egyptians were cultivating domesticated watermelon. BioRxiv. https://doi.org/doi: 10.1101/642785

Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., & Reich, D. (2015). Partial uracil – DNA – glycosylase treatment for screening of ancient DNA. Philosophical Transactions of the Royal Society B: Biological Sciences, 370(1660). https://doi.org/10.1098/rstb.2013.0624

Schubert, M., Ermini, L., Sarkissian, C. Der, Jónsson, H., Ginolhac, A., Schaefer, R., Martin, M. D., Fernández, R., Kircher, M., McCue, M., Willerslev, E., & Orlando, L. (2014). Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. Nature Protocols, 9(5), 1056–1082. https://doi.org/10.1038/nprot.2014.063

Scott, M. F., Botigué, L. R., Brace, S., Stevens, C. J., Mullin, V. E., Stevenson, A., Thomas, M. G., Fuller, D. Q., & Mott, R. (2019). A 3,000-year-old Egyptian emmer wheat genome reveals dispersal and domestication history. Nature Plants, 5(11), 1120–1128. https://doi.org/10.1038/s41477-019-0534-5

Sharma, J. S., Running, K. L. D., Xu, S. S., Zhang, Q., Peters Haugrud, A. R., Sharma, S., McClean, P. E., & Faris, J. D. (2019). Genetic analysis of threshability and other spike traits in the evolution of cultivated emmer to fully domesticated durum wheat. Molecular Genetics and Genomics, 294(3), 757–771. https://doi.org/10.1007/s00438-019-01544-0

Smith, O., Nicholson, W. V, Kistler, L., Mace, E., Clapham, A., Rose, P., Stevens, C., Ware, R., Samavedam, S., Barker, G., Jordan, D., Fuller, D. Q., & Allaby, R. G. (2019). A domestication history of dynamic adaptation and genomic deterioration in Sorghum. Nature Plants, 5(April). https://doi.org/10.1038/s41477-019-0397-9

Snir, A., Nadel, D., Groman-Yaroslavski, I., Melamed, Y., Sternberg, M., Bar-Yosef, O., & Weiss, E. (2015). The origin of cultivation and proto-weeds, long before neolithic farming. PLoS ONE, 10(7), 1–12. https://doi.org/10.1371/journal.pone.0131422

Spyrou, M. A., Bos, K. I., Herbig, A., & Krause, J. (2019). Ancient pathogen genomics as an emerging tool for infectious disease research. Nature Reviews Genetics, 20(JUNE). https://doi.org/10.1038/s41576-019-0119-1

Swarts, K., Gutaker, R. M., Benz, B., Blake, M., Bukowski, R., Holland, J., Kruse-peeples, M., Lepak, N., Prim, L., Romay, M. C., Ross-ibarra, J., & Sanchez-gonzalez, J. D. J. (2017). Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America. 515(August), 512–515. https://doi.org/10.1126/science.aam9425 Estimating

Trucchi, E., Benazzo, A., Lari, M., Iob, A., Vai, S., Nanni, L., Bellucci, E., Bitocchi, E., Raffini, F., Xu, C., Jackson, S. A., Lema, V., Babot, P., Oliszewski, N., Gil, A., Neme, G., Michieli, C. T., De Lorenzi, M., Calcagnile, L., … Bertorelle, G. (2021). Ancient genomes reveal early Andean farmers selected common beans while preserving diversity. Nature Plants, 7(2), 123–128. https://doi.org/10.1038/s41477-021-00848-7

Ulaş, B., & Fiorentino, G. (2021). Recent attestations of " new " glume wheat in Turkey : a reassessment of its role in the reconstruction of Neolithic agriculture. Vegetation History and Archaeobotany, 30(5), 685–701. https://doi.org/10.1007/s00334-020-00807-w

Vavilov, N. I., Vavylov, M. I., Vavílov, N. Í., & Dorofeev, V. F. (1992). Origin and geography of cultivated plants. Cambridge University of Press.

Verdugo, M. P., Mullin, V. E., Scheu, A., Mattiangeli, V., Daly, K. G., Delser, P. M., Hare, A. J., Burger, J., Collins, M. J., Kehati, R., Hesse, P., & Fulton, D. (2019). Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. 176(July), 173–176. https://doi.org/http://doi.org/10.5281/zenodo.3206663

Wagenaar, E. B. (1966). Studies on the Genome Constitution of Triticum timopheevi Zhuk . II. The T . timopheevi Complex and Its Origin. Society for the Study of Evolution, 20(2), 150–164. https://www.jstor.org/stable/2406569

Wales, N., Akman, M., Watson, R. H. B., Sánchez Barreiro, F., Smith, B. D., Gremillion, K. J., Gilbert, M. T. P., & Blackman, B. K. (2018). Ancient DNA reveals the timing and persistence of organellar

genetic bottlenecks over 3000 years of sunflower domestication and improvement. Evolutionary Applications, 12(December 2017), 1–16. https://doi.org/10.1111/eva.12594

Wales, N., Akman, M., Watson, R. H. B., Sánchez Barreiro, F., Smith, B. D., Gremillion, K. J., Gilbert, M. T. P., & Blackman, B. K. (2019). Ancient DNA reveals the timing and persistence of organellar genetic bottlenecks over 3,000 years of sunflower domestication and improvement. Evolutionary Applications, 12(1), 38–53. https://doi.org/10.1111/eva.12594

Walkowiak, S., Gao, L., Monat, C., Haberer, G., Kassa, M. T., Brinton, J., Ramirez-Gonzalez, R. H., Kolodziej, M. C., Delorean, E., Thambugala, D., Klymiuk, V., Byrns, B., Gundlach, H., Bandi, V., Siri, J. N., Nilsen, K., Aquino, C., Himmelbach, A., Copetti, D., … Pozniak, C. J. (2020). Multiple wheat genomes reveal global variation in modern breeding. Nature, 588(7837), 277–283. https://doi.org/10.1038/s41586-020-2961-x

Warinner, C., Rodrigues, J. F. M., Vyas, R., Trachsel, C., Shved, N., Grossmann, J., Radini, A., Hancock, Y., Tito, R. Y., Fiddyment, S., Speller, C., Hendy, J., Charlton, S., Luder, H. U., Salazar-garcía, D. C., Eppler, E., Seiler, R., Hansen, L. H., Alfredo, J., … Willerslev, E. (2014). Pathogens and host immunity in the ancient human oral cavity. Nature Publishing Group, 46(4). https://doi.org/10.1038/ng.2906

Weyrich, L. S., Dobney, K., & Cooper, A. (2015). Ancient DNA analysis of dental calculus. Journal of Human Evolution, 79, 119–124. https://doi.org/10.1016/j.jhevol.2014.06.018

Wu, X., Ding, B., Zhang, B., Feng, J., Wang, Y., Ning, C., Wu, H., Zhang, F., Zhang, Q., Li, N., Zhang, Z., Sun, X., Zhang, Q., Li, W., Liu, B., Cui, Y., & Gong, L. (2019). Phylogenetic and population structural inference from genomic ancestry maintained in present-day common wheat Chinese landraces. Plant Journal, 99(2), 201–215. https://doi.org/10.1111/tpj.14421

Yang, M. A., Fan, X., Sun, B., Chen, C., Lang, J., Ko, Y., Tsang, C., Chiu, H., Wang, T., Bao, Q., Wu, X., Hajdinjak, M., Ko, A. M., Ding, M., Cao, P., Yang, R., Liu, F., Nickel, B., Dai, Q., … Fu, Q. (2020). Ancient DNA indicates human population shifts and admixture in northern and southern China. Science, 288(July), 282–288. https://doi.org/10.1126/science.aba0909

Yoshida, K., Schuenemann, V. J., Cano, L. M., Pais, M., Mishra, B., & Sharma, R., et al. (2013). (2013). The rise and fall of the Phytophthora infestans lineage that triggered the Irish potato famine. Elife, 2, 1–25. https://doi.org/doi: 10.7554/eLife.00731.001

Zhou, Y., Zhao, X., Li, Y., Xu, J., Bi, A., Kang, L., Xu, D., Chen, H., Wang, Y. ge Y. Y. ge Y. Y., Wang, Y. ge Y. Y. ge Y. Y., Liu, S., Jiao, C., Lu, H., Wang, J., Yin, C., Jiao, Y., & Lu, F. (2020). Triticum population sequencing provides insights into wheat adaptation. Nature Genetics, 52(12), 1412–1422. https://doi.org/10.1038/s41588-020-00722-w

Zohary, D. (2013). Domestication of Crop Plants. Encyclopedia of Biodiversity: Second Edition, March 2018, 657–664. https://doi.org/10.1016/B978-0-12-384719-5.00199-4

7.4 Crop archaeogenomics: A powerful resource in need of a well-defined regulation framework (DOI: 10.1002/ppp3.10233)

Received: 22 February 2021 | Revised: 1 September 2021 | Accepted: 6 September 2021

DOI: 10.1002/ppp3.10233

OPINION

Plants People Planet PPP Open Access

Crop archaeogenomics: A powerful resource in need of a well-defined regulation framework

Alice Iob | Laura Botigué

Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Barcelona, Spain

Correspondence
Laura Botigué, Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, CRAG building, campus UAB, 08193 Cerdanyola, Barcelona, Spain.
Email: laura.botigue@cragenomica.es

Societal Impact Statement

Crop archaeogenomics has rapidly flourished in recent years, leading to a new way of understanding the past and bringing answers to important questions about human history in relation to plant management and food production. Furthermore, the knowledge derived from the analysis of ancient crops can contribute to the development of a more sustainable future. However, the extant legal framework presents a number of challenges when applied to this research field, particularly in the current scenario of disparities in scientific outcomes between countries. We expose the uncertainties of the legal framework and the factors that maintain or exacerbate these inequalities, as well as possible solutions.

Summary

Crop archaeogenomics is a flourishing field that has greatly benefited from next-generation sequencing technologies. Ancient and historical plant remains are currently considered genetic resources and as such are subject to legal frameworks like those implemented by the Nagoya Protocol. In addition to the challenges in complying with genetic resource regulations that crop archaeogenomics share with other basic plant research disciplines, there are additional difficulties specific to this interdisciplinary field that includes science and humanities, namely, the need to comply with two different legislations before accessing the samples (one for genetic resources and one for cultural heritage), along with a high risk of not obtaining DNA. As a result, most studies to date have been done on samples for which the laws regulating genetic resources did not apply, sometimes avoiding the need of reaching Access and Benefit Sharing agreements with the country that originally provided the samples. This phenomenon is likely to worsen in the future, as the archaeological record is a limited resource and competition between laboratories will only widen the gap between developed and developing economies. Because crop archaeogenomics is a new and promising scientific field, it is desirable to begin a dialogue with other basic biological research fields to facilitate the implementation of these agreements so that basic sciences can easily utilize these biological samples while ensuring the rights of all parties involved.

KEYWORDS

access and benefit sharing, crop archaeogenomics, genetic resources, Nagoya protocol, policy

1 | INTRODUCTION

The field of archaeogenomics has come a long way since the publication of a mitochondrial sequence from a museum-preserved quagga in 1984 (Higuchi et al., 1984) and has recently flourished thanks to the development of genomic sequencing technologies.

Although most archaeogenomic studies have so far been performed in animals (including hominins), plant archaeogenomics has gained increasing attention in the last decade, because few samples are a powerful resource to investigate important events such as plant domestication. Processes like selection, local adaptation and early dispersal of domestic forms have been investigated in major crops, such as barley (Mascher et al., 2016; Palmer et al., 2009), bread wheat (Li et al., 2011), cotton (Palmer et al., 2012), maize (da Fonseca et al., 2015), emmer wheat (Scott et al., 2019) and bean (Trucchi et al., 2021). Plant ancient DNA analyses study both past human–environment interactions and the evolutionary forces that shaped modern crops. Furthermore, knowledge obtained from crop archaeogenomics holds great potential to aid in the development of new conservation strategies, breeding programmes and agricultural practices in response to climate change and human pressure on the environment (Di Donato et al., 2018; Estrada et al., 2018; Hofman et al., 2015; Pont et al., 2019; Przelomska et al., 2020).

With the advent of ancient DNA analysis, legal and ethical issues have arisen. These include the ethical implications of undertaking the genetic analysis of ancient humans, their cultural bonds with present-day communities, competition between research groups, hoarding of material from the Global South, and limits in experiment replication due to the intrinsic value of the sample coupled with their finite availability, both for human (Bardill et al., 2018; Elliott, 2009; Paradise & Andrews, 2007; Wagner et al., 2020) and archaeofaunal remains (Pálsdóttir et al., 2019).

In the case of crop archaeogenomics, many factors (including stochasticity) play a very important role in determining the preservation and the cultural value of the remains (see Box 1 for more details). Usually, the most promising samples consist of seeds or leaves (small amounts of material), and DNA extraction is most often a destructive experimental procedure that effectively impoverishes the archaeobotanical record. Moreover, it is not possible to know what fraction of the archaeobotanical record contains DNA, but so far it seems to be modest, especially for charred remains (Nistelberger et al., 2016). The relative abundance of ancient material has shielded this field from the competition and hoarding that have occurred in human archaeogenomics (Makarewicz & Nimrod Marom, 2017; Morris, 2017). However, the growing interest in the discoveries from crop archaeogenomics, coupled with the constant technical research for improvements (e.g., Brown et al., 2014; Gamba et al., 2016; Lendvay et al., 2018), could rapidly change that.

We discuss here aspects of this research field under the Convention on Biological Diversity (CBD) and Nagoya Protocol (NP) international framework. The field of ancient crop genomics shares many aspects with other basic research disciplines, such as the non-monetary nature of the results, but it has also some peculiarities, which include (1) the obligation to abide by both heritage and genetic regulation and the inadequacies of CBD definitions when it comes to archaeobotanical remains; (2) limitations of the current legal framework for Access and Benefit Sharing (ABS) agreements, and (3) possible solutions to improve the current situation.

BOX 1 Assessing the value of archaeobotanical samples. Examples from (Zohary et al., 2012)

1. *Uneven geographic distribution of the archaeological record*: The abundance of archaeological sites is uneven across the globe. Although the archaeological record provides an overall reliable overview of crop domestication in Europe, Southwest Asia and the Mediterranean basin, other regions have not been as explored, such as South Asia or in Africa, south of the Sahara.
2. *Type of sample*: Seeds are, together with microscopic pollen, the more resilient parts of the plants and thus represent the most abundant form of crop remains from early farming villages in the old world. On the contrary, horticultural vegetables that furthermore may undergo different food processing techniques are very rare because they have fewer chances to survive as archaeological remains.
3. *Conservation status*: The conservation status will also affect the value of the sample. Samples are usually charred, less commonly desiccated (when located in conditions of extreme dryness such as desertic areas), or waterlogged (e.g., in lakes or wells in anaerobic conditions). They can also be mineralized, in phytolites and coprolites. Desiccated remains have so far proven to be more likely to contain ancient DNA than charred remains do.
4. *Uneven temporal distribution of the archaeological record*: Within a given archaeological site the abundance of samples can vary between historical periods or sections excavated. This happens for instance at the site of Troy; although it is rich in plant remains in sections dated to the Middle Bronze Age, only a few specimens belonging to the Early Bronze age and Neolithic have been recovered.
5. *Additional value*: some specimens have an added cultural value. The material used to make artefacts (e.g., necklaces made of seeds and baskets made of leaves) is not only informative about the plants themselves but also about the culture that fabricated those artefacts, trade networks, etc. Another example would be for plant remains found in excavations or burials of historical characters notorious to society. Results from these studies would have higher chances to be disseminated by the media because of their capacity to engage society, rather than the strictly scientific findings from those studies.

This opinion piece focuses on archaeobotanical macroremains of crops because of their important role in human development and history. However, we note that plant archaeogenomics includes the analysis of many kinds of remains, among others plant DNA in ancient sediments, coprolites and wood remains. For a more detailed read on plant archaeogenomics, a review covering this topic has been recently published elsewhere (Kistler et al., 2020).

2 | THE LEGAL FRAMEWORK

Geneticists can access archaeobotanical remains either by establishing a collaboration with an archaeobotanist or, when the samples have already been deposited, with a curator of a museum or an herbarium. In the first case, the decision to collaborate will depend on the archaeological project, its director or the archaeobotanist, depending on the institutions involved. In the case of museum or herbarium collections, the decision process varies, but usually involves, an internal or external evaluation of the request as per general disciplinary standards: scientific need, nature of the project and research potential against the need to preserve the collections.

In cases where the collaboration is established between multiple countries, the movement of the samples must be granted following two national laws: that covering archaeological material and, if they exist, that of genetic resources. For the former, export permits have to be requested and accepted by Antiquity Authorities or Heritage Institutions, even though exceptions exist (e.g., Egypt restricts the export of all materials). Agreements usually last for a known time, during which researchers can study the archaeological material. If the analyses entail the destruction of the samples, it is usually harder to obtain the authorization.

Regarding the genetic nature of the material, two international agreements have also been established. The United Nations (UN) CBD entered into force in 1993 with the goals of conserving biological diversity, promoting the sustainable use of its components, and the fair and equitable ABS arising from their utilization. So far, it has been signed by 196 Parties, with the notable absence of the United States, among other countries (Cooper & Noonan-Mooney, 2013). The NP emerged to provide a legal framework to implement the third objective of the CBD, entered into force in 2014 and has 128 Parties. Briefly, the NP sets out core obligations for its contracting parties through mutually agreed terms concerning access to genetic resources, benefit sharing and compliance. It also covers traditional knowledge associated with genetic resources, taking into account the role of indigenous people and local communities (UN CBD, 2011). Each country implements NP with its own national ABS policies and laws, as the CBD recognizes each country's sovereignty over its own genetic resources.

The ultimate intent of these international agreements is to avoid undue appropriation of natural resources and *unfair distribution of benefits arising from their exploitation*. Thus, these agreements were born out of the necessity for international legal and ethical standards that act upon the increasing globalization and environmental degradation, while promoting the wellness of humanity.

3 | INADEQUACIES OF THE CBD DEFINITIONS AND THE NP FRAMEWORK FOR ARCHAEOBOTANICAL REMAINS

The CBD defines *genetic material* as "any material of plant, animal, microbial or other origin containing functional units of heredity", and defines *genetic resources* as "genetic material of actual or potential value" (Cooper & Noonan-Mooney, 2013). In addition, under the NP, authorization to use the genetic resources must be obtained *before* accessing them. In the study of crop archaeogenomics, this is particularly daunting because it is not possible to know whether the sample of interest contains DNA without processing it first. Indeed, it is not at all clear if ancient plant remains should be considered genetic resources. Foremost, ancient biological samples include a long list of remains as well as archaeological artefacts, most of which do not contain endogenous DNA at all. Second, if endogenous DNA is still present inside the sample, it is highly fragmented and degraded (e.g., Jónsson et al., 2013; Pääbo et al., 2004), containing numerous nucleotide misincorporations which prevent its functionality (only two notable exceptions have been described so far; Sallon et al., 2008; Shen-Miller et al., 1995).

Under the present legal framework, *the intention* to undertake a genetic analysis on an ancient plant specimen turns this sample into a genetic resource. Archaeological plant material is regulated only by Cultural and Heritage laws up until the time a researcher wants to do a genetic analysis on this material. At this point, the sample acquires a *potential value* that turns it into genetic resources according to the CBD terms. Nevertheless, most likely the sample cannot be considered to be *genetic material* because it does not contain DNA, especially if defined as *functional units of heredity*. This high risk of not obtaining DNA from archaeobotanical remains and the additional bureaucratic burden to comply with genetic, cultural and heritage regulations has led to the avoidance of NP-regulated archaeobotanical remains. The main downside of this legal framework is the requirement to enter into agreements *before* accessing the sample and thus be allowed to test the feasibility of the research project (i.e., the presence of analysable DNA in the sample). As a result, at times, the country of origin of the sample, provider country henceforth, is not considered at all in the study. This prevents the involvement of researchers from this country and promotes the *unfair distribution of benefits arising from the use of resources* that the very NP was established to avoid.

4 | THE NEED TO PROMOTE AN ABS AGREEMENT

Although crop archaeogenomics is a basic science field and no monetary profit can be envisioned out of it, it is still necessary to ensure

that research and scientific collaborations follow fundamental ethical principles. All parties involved in a given study should share intellectual and technological benefits, consistent with the intent of the NP to promote developing economies' technical capacities and science building. The field of crop archaeogenomics is usually led by researchers from the Global North even though many plant remains originate from the Global South. There are various reasons embedded in the research area that explain this: First, the generation of the data is a very expensive process. It requires the construction of an exclusively dedicated laboratory facility and continuous expense in laboratory material to minimize the risk of contamination from external sources and cross-contamination between samples (e.g., Krause, 2010; Pääbo et al., 2004). It also requires investment for sequencing and bioinformatic analyses. All these factors, along with the intrinsic uncertainty of obtaining (good amounts of) target DNA, leave developing economies out of the race of leading archaeogenomic studies, even those that have a rich archaeological record.

Second, the data to analyse consist of digital sequence information (DSI) to be processed along with other DSI available in public databases. Currently though, many institutions in developing economies cannot afford the subscription fees for many scientific journals where the information to locate the DSI in repositories is found, hindering their access to new knowledge and the genomic data (Djikeng, 2012; Helmy et al., 2016). Besides, the lack of basic infrastructures, such as fast and stable internet connection to download the DSI, as well as access to high-performance computers further hinders the data analysis process. International initiatives such as eIFL (*eIFL - Open Access Programme*), INASP (*Inasp - Homepage*) and Research4Life (*Research4life Organization*) have incentivized access to science and scientific advancement in the so-called Global South, even if extant inequalities remain difficult to overcome (Powell et al., 2020).

The ultimate consequence of these factors is a movement of ancient biological samples from the Global South to the Global North. Samples are then processed and analysed by researchers in countries that do have the infrastructures to do so (Morris, 2017; Prendergast & Sawchuk, 2018) (Figure 1). Without an ABS agreement, Global South countries that provided the samples risk having a mere testimonial role in the study, if they are represented at all. Considering the studies mentioned in Figure 1 in at least 70% of the cases (8 of 12) where samples were originally excavated in a country of the Global South, no author is affiliated with an institution of a developing economy. When present, authors from institutions in developing economies did not play a major role in the research, as reflected in the affiliation of the first and last authors.

5 | SOLUTIONS WITHIN THE NP

Recently, a new concept has been coined to refer to the colonialist practices that still occur in science, the helicopter research (Minasny et al., 2020). More and more voices are being raised about the need to maintain certain ethical rules when performing research involving

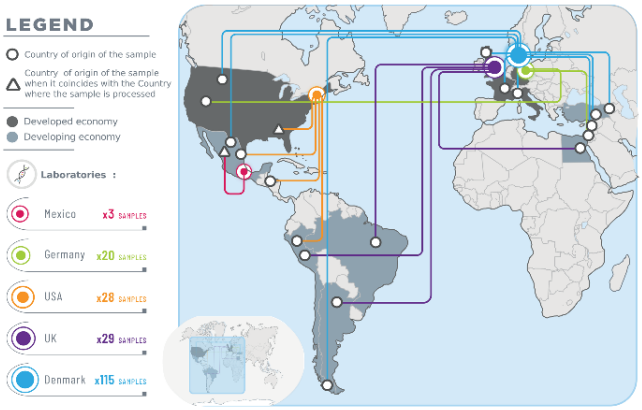


FIGURE 1 Movement of crop macroremains from country of origin to ancient DNA facilities up to 2019. The movement of samples from country of origin to the country where they were processed. The dimension of the dots depends on the number of the resulting scientific publications. This map has a per country resolution, the location of dots/triangles is not related to exact archaeological sites/labs coordinates inside the country. The number of samples analysed is reported on the left panel, next to the country where the analysis took place. The genomic studies considered in this figure are those included in (Kistler et al., 2020), hence reflecting studies published up to December 2019, not a complete survey of ancient DNA studies on plants. Studies about wood have been excluded, not being about crops

countries in the Global South. The NP, through the implementation of the ABS agreement, is the perfect venue to ensure that these non-monetary benefits are included when planning a research project. However, in the specific case of archaeogenomics, there are a large number of samples that were collected before the NP entered into force and therefore escape this regulation. In addition, the need to draft an agreement before accessing samples combined with the high risk of not obtaining DNA from them has a deterrent effect that results in avoiding the selection of NP-regulated specimens.

As other authors pointed out, one solution would be to have specific regulations for ancient (and historical) plant material within the NP. However, this is hard to accomplish, at least in the near future (Sherman & Henry, 2020). A more interdisciplinary strategy is more likely to succeed. There is an ongoing debate around the burcaucratic burden that basic plant research faces to comply with ABS (e.g., Kursar, 2011; Schindel, 2010; Watanabe, 2015). Certain solutions have been proposed, such as mechanisms that would allow differentiating between monetary and non-monetary benefits (Rourke, 2018). We also advocate a solution that goes beyond the specific case of non-modern plant specimens and unites all basic biological research.

We propose two actions that would have a positive impact in the present circumstances. First, add a clause in the NP for all basic science disciplines that do not entail monetary benefits, allowing exploratory analyses on the samples for a short time frame to determine the feasibility of the proposed project. This is similar to the regulations applied in social sciences, where samples are allowed to be studied for a restricted period. Second, a standardized ABS agreement, designed to avoid helicopter research detailing the participation and recognition of all the members involved in the project. The standardization will facilitate the process of sample acquisition for researchers while ensuring intellectual contribution, recognition and transfer of technical and scientific expertise to researchers in the provider country of the samples. An example of access standardization is provided by the multilateral ABS system and standard material transfer agreements of the International Treaty on Plant Genetic Resources for Food and Agriculture (2001) (Rourke, 2018).

6 | CONCLUSIONS

An appropriate legal framework for crop archaeogenomic studies is of particular importance because the current situation promotes the use of plant remains that are not subject to the NP or a national law regulating genetic resources. This is even more true because the distribution of archaeological remains is unbalanced between the Global North and the Global South, the former counting on extensive collections of foreign specimens that predate international agreements. In some instances, this implies the total exclusion of the provider country in the research study, and therefore the impossibility of sharing the benefits of the ongoing research.

Care must be taken, however, to avoid mere testimonial participation by the provider country, and agreements made between the parties should ensure the active participation of researchers from that country in the study. Indeed, we think that standardization of an ABS agreement for non-monetary benefits focused on avoiding helicopter research practices would have a very positive impact on that regard. This standard ABS form could even exist beyond the framework of the NP and be used more broadly by researchers from the Global North and South before undertaking international collaborations.

Given that crop archaeogenomics is a novel field, which can and likely will grow in the future, we think this is the right time to open a dialogue about possible options to aid easy access to resources for scientists while preserving the archaeological record and granting benefit sharing and inclusion of research institutions in developing economies. Standards in accessing the samples coupled with active sharing of the benefits deriving from the research could help to take away disparities between scientific communities in different countries and support the development of this field worldwide. Even if CBD and NP were drafted to recognize the sovereignty of each country over its genetic resources, we strongly believe that in the end all of us should gain benefit from the reconstruction of past phenomena such as plant domestication, which can be seen as a form of collective heritage of humankind.

ACKNOWLEDGEMENTS

We would like to thank Dr. Amaia Arranz-Otaegui and Dr. Alex Weide for helpful suggestions regarding archaeobotanical samples, the team of the Buzón Protocolo de Nagoya for their insights on the legal aspects of the NP and CBD international agreements, and Elizabeth Rovere for graphical suggestions.

AUTHOR CONTRIBUTIONS

L.B. planned and designed the research. A.I. and L.B. revised the scientific literature and conducted the research of the current international agreements. A.I. and L.B. wrote the manuscript.

FUNDING INFORMATION

A.I. is a FPI fellow (PRE2018-083529). L. B is a Ramón y Cajal Fellow. (RYC2018-024770-I) both fellowships funded by the Ministerio de Ciencia e Innovación—Agencia Estatal de Investigación/Fondo Social Europeo. We acknowledge financial support from the Spanish Ministry of Science and Innovation-State Research Agency (AEI), through the "Severo Ochoa Programme for Centres of Excellence in R&D" SEV-2015-0533 and CEX2019-000902-S. This work was also supported by the CERCA programme by the Generalitat de Catalunya.

DATA AVAILABILITY STATEMENT

Data sharing not applicable - no new data generated, or the article describes entirely theoretical research

ORCID

Alice Iob https://orcid.org/0000-0002-7505-3612
Laura Botigué https://orcid.org/0000-0001-7114-5168

REFERENCES

Bardill, B. J., Bader, A. C., Garrison, N. A., Bolnick, D. A., Raff, J. A., & Walker, A. (2018). Advancing the ethics of palaeogenomics. *Science*, *360*(6387), 384–385. https://doi.org/10.1126/science.aaq1131

Brown, T. A., Cappellini, E., Kistler, L., Lister, D. L., Oliveira, H. R., Wales, N., & Schlumbaum, A. (2014). Recent advances in ancient DNA research and their implications for archaeobotany. *Vegetation History and Archaeobotany*, *24*(1), 207–214. https://doi.org/10.1007/s00334-014-0489-4

Cooper, H. D., & Noonan-Mooney, K. (2013). Convention on biological diversity. In *Encyclopedia of biodiversity* (Second ed.) (pp. 306–319). https://doi.org/10.1016/B978-0-12-384719-5.00418-4

da Fonseca, R. R., Smith, B. D., Wales, N., Cappellini, E., Skoglund, P., Fumagalli, M., Samaniego, J. A., Carøe, C., Ávila-Arcos, M. C., Hufnagel, D. E., Korneliussen, T. S., Vieira, F. G., Jakobsson, M., Arriaza, B., Willerslev, E., Nielsen, R., Hufford, M. B., Albrechtsen, A., Ross-Ibarra, J., & Gilbert, M. T. P. (2015). The origin and evolution of maize in the southwestern United States. *Nat Plants*, *1*(January), 1–5. https://doi.org/10.1038/nplants.2014.3

di Donato, A., Filippone, E., Ercolano, M. R., & Frusciante, L. (2018). Genome sequencing of ancient plant remains: Findings, uses and potential applications for the study and improvement of modern crops. *Frontiers in Plant Science*, *9*(April). https://doi.org/10.3389/fpls.2018.00441

Djikeng, S. (2012). *Genomics applications for the developing world*. https://doi.org/10.1007/978-1-4614-2182-5

Elliott, L. M. (2009). Property rights of ancient DNA: The impact of cultural importance on the ownership of genetic information. *International Journal of Cultural Property*, *16*(2), 103–129. https://doi.org/10.1017/S0940739109090183

Estrada, O., Breen, J., Richards, S. M., & Cooper, A. (2018). Ancient plant DNA in the genomic era. *Nat Plants*, *4*(7), 394–396. https://doi.org/10.1038/s41477-018-0187-9

Gamba, C., Hanghøj, K., Gaunitz, C., Alfarhan, A. H., Alquraishi, S. A., Al-Rasheid, K. A. S., Bradley, D. G., & Orlando, L. (2016). Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. *Molecular Ecology Resources*, *16*(2), 459–469. https://doi.org/10.1111/1755-0998.12470

Helmy, M., Awad, M., & Mosa, K. A. (2016). Limited resources of genome sequencing in developing countries: Challenges and solutions. *Appl Transl Genom*, *9*, 15–19. https://doi.org/10.1016/j.atg.2016.03.003

Higuchi, R., Bowman, B., Freiberger, M., Ryder, O. A., & Wilson, A. C. (1984). DNA sequences from the quagga, an extinct member of the horse family. *Nature*, *312*, 282–284. https://doi.org/10.1038/312282a0

Hofman, C. A., Rick, T. C., Fleischer, R. C., & Maldonado, J. E. (2015). Conservation archaeogenomics: Ancient DNA and biodiversity in the Anthropocene. *Trends in Ecology & Evolution*, *30*(9), 540–549. https://doi.org/10.1016/j.tree.2015.06.008

Inasp - homepage. (n.d.). https://www.inasp.info/

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). MapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, *29*(13), 1682–1684. https://doi.org/10.1093/bioinformatics/btt193

Kistler, L., Bieker, V. C., Martin, M. D., Pedersen, M. W., Ramos Madrigal, J., & Wales, N. (2020). Ancient plant genomics in archaeology, herbaria, and the environment. *Annual Review of Plant Biology*, *71*, 605–629. https://doi.org/10.1146/annurev-arplant-081519-035837

Krause, J. (2010). From genes to genomes: What is new in ancient DNA? In *Mitteilungen Der Gesellschaft Für Urgeschichte* (Vol. 19) (pp. 11–33).

Kursar, T. A. (2011). What are the implications of the Nagoya protocol for research on biodiversity? *Bioscience*, *61*(4), 256–257. https://doi.org/10.1525/bio.2011.61.4.2

Lendvay, B., Hartmann, M., Brodbeck, S., Nievergelt, D., Reinig, F., Zoller, S., Parducci, L., Gugerli, F., Büntgen, U., & Sperisen, C. (2018). Improved recovery of ancient DNA from subfossil wood—Application to the world's oldest late glacial pine forest. *The New Phytologist*, *217*(4), 1737–1748. https://doi.org/10.1111/nph.14935

Li, C., Lister, D. L., Li, H., Xu, Y., Cui, Y., Bower, M. A., Jones, M. K., & Zhou, H. (2011). Ancient DNA analysis of desiccated wheat grains excavated from a bronze age cemetery in Xinjiang. *Journal of Archaeological Science*, *38*(1), 115–119. https://doi.org/10.1016/j.jas.2010.08.016

Makarewicz, C., & Nimrod Marom, G. B.-O. (2017). Ensure equal access to ancient DNA. *Nature*, *548*, 158. https://doi.org/10.1038/548158a

Mascher, M., Schuenemann, V. J., Davidovich, U., Marom, N., Himmelbach, A., Hübner, S., Korol, A., David, M., Reiter, E., Riehl, S., Schreiber, M., Vohr, S. H., Green, R. E., Dawson, I. K., Russell, J., Kilian, B., Muehlbauer, G. J., Waugh, R., Fahima, T., … Stein, N. (2016). Genomic analysis of 6,000-year-old cultivated grain illuminates the domestication history of barley. *Nature Genetics*, *48*(9), 1089–1093. https://doi.org/10.1038/ng.3611

Minasny, B., Fiantis, D., Mulyanto, B., Sulaeman, Y., & Widyatmanti, W. (2020). Global soil science research collaboration in the 21st century: Time to end helicopter research. *Geoderma*, *373*(May), 114299. https://doi.org/10.1016/j.geoderma.2020.114299

Morris, A. G. (2017). Ancient DNA comes of age, but still has some teenage problems. *South African Journal of Science*, *113*(9–10), 9–10. https://doi.org/10.17159/sajs.2017/a0232

Nistelberger, H. M., Smith, O., Wales, N., Star, B., & Boessenkool, S. (2016). The efficacy of high-throughput sequencing and target enrichment on charred archaeobotanical remains. *Scientific Reports*, *6*(0316), 1–11. https://doi.org/10.1038/srep37347

Pääbo, S., Poinar, H., Serre, D., Jaenicke-Després, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L., & Hofreiter, M. (2004). Genetic analyses from ancient DNA. *Annual Review of Genetics*, *38*(1), 645–679. https://doi.org/10.1146/annurev.genet.37.110801.143214

Palmer, S. A., Clapham, A. J., Rose, P., Freitas, F. O., Owen, B. D., Beresford-Jones, D., Moore, J. D., Kitchen, J. L., & Allaby, R. G. (2012). Archaeogenomic evidence of punctuated genome evolution in gossypium. *Molecular Biology and Evolution*, *29*(8), 2031–2038. https://doi.org/10.1093/molbev/mss070

Palmer, S. A., Moore, J. D., Clapham, A. J., Rose, P., & Allaby, R. G. (2009). Archaeogenetic evidence of ancient nubian barley evolution from six to two-row indicates local adaptation. *PLoS ONE*, *4*(7), 2–8. https://doi.org/10.1371/journal.pone.0006301

Pálsdóttir, A. H., Bläuer, A., Rannamäe, E., Boessenkool, S., & Hallsson, J. H. (2019). Not a limitless resource: Ethics and guidelines for destructive sampling of archaeofaunal remains. *Royal Society Open Science*, *6*(10), 7–9. https://doi.org/10.1098/rsos.191059

Paradise, J., & Andrews, L. (2007). Tales from the crypt: Scientific, ethical, and legal considerations for biohistorical analysis of deceased historical figures. *Temple Journal of Science, Technology & Environmental Law*, *26*(2), 223–299.

Pont, C., Wagner, S., Kremer, A., Orlando, L., Plomion, C., & Salse, J. (2019). Paleogenomics: Reconstruction of plant evolutionary trajectories from modern and ancient DNA. *Genome Biology*, *20*(1), 29. https://doi.org/10.1186/s13059-019-1627-1

Powell, A., Johnson, R., & Herbert, R. (2020). *Achieving an equitable transition to open access for researchers in lower and middle-income countries*. ICSR Perspectives. https://doi.org/10.2139/ssrn.3624778

Prendergast, M. E., & Sawchuk, E. (2018). Boots on the ground in Africa's ancient DNA 'revolution': Archaeological perspectives on ethics and best practices. *Antiquity*, *92*(363), 803–815. https://doi.org/10.15184/aqy.2018.70

Przelomska, N. A. S., Armstrong, C. G., & Kistler, L. (2020). Ancient plant DNA as a window into the cultural heritage and biodiversity of our food system. *Frontiers in Ecology and Evolution*, *8*(March), 1–8. https://doi.org/10.3389/fevo.2020.00074

research4life organization. (n.d.). https://www.research4life.org/

Rourke, M. F. (2018). Access and benefit-sharing in practice: Non-commercial research scientists face legal obstacles to accessing genetic resources. *The Journal of Science Policy & Governance*, *13*(1), 1–20.

Sallon, S., Solowey, E., Cohen, Y., Korchinsky, R., Egli, M., Woodhatch, I., Simchoni, O., & Kislev, M. (2008). Germination, genetics, and growth of an ancient date seed. *Science*, *320*, 1464. https://doi.org/10.1126/science.1153600

Schindel, D. (2010). Biology without borders. *Nature*, *467*, 779–781. https://doi.org/10.1038/467779a

Scott, M. F., Botigué, L. R., Brace, S., Stevens, C. J., Mullin, V. E., Stevenson, A., Thomas, M. G., Fuller, D. Q., & Mott, R. (2019). A 3,000-year-old Egyptian emmer wheat genome reveals dispersal and domestication history. *Nature Plants*, *5*(11), 31–34. https://doi.org/10.1038/s41477-019-0534-5

Shen-Miller, J., Mudgett, M. B., Schopf, J. W., Clarke, S., & Berger, R. (1995). Exceptional seed longevity and robust growth: Ancient sacred lotus from China. *American Journal of Botany*, *82*(11), 1367–1380. https://doi.org/10.1002/j.1537-2197.1995.tb12673.x

Sherman, B., & Henry, R. J. (2020). The Nagoya protocol and historical collections of plants. *Nature Plants*, *6*, 430–432. https://doi.org/10.1038/s41477-020-0657-8

Trucchi, E., Benazzo, A., Lari, M., Iob, A., Vai, S., Nanni, L., Bellucci, E., Bilocchi, E., Raffini, F., Xu, C., Jackson, S. A., Lema, V., Babot, P., Oliszewski, N., Gil, A., Neme, G., Michieli, C. T., De Lorenzi, M., Calcagnile, L., … Bertorelle, G. (2021). Ancient genomes reveal early Andean farmers selected common beans while preserving diversity. *Nature Plants*, *7*(2), 123–128. https://doi.org/10.1038/s41477-021-00848-7

UN Convention on Biological Diversity. (2011). Access to genetic resources and the fair and equitable sharing of benefits arising convention on Nagoya protocol on access to genetic resources and the fair and equitable sharing of benefits arising from their utilization to the convention on biological diversity. *Annual Review of Entomology*, *12*(3), 1–320. https://doi.org/10.1146/annurev.ento.48.091801.112645

Wagner, J. K., Colwell, C., Claw, K. G., Stone, A. C., Bolnick, D. A., Hawks, J., Brothers, K. B., & Garrison, N. A. (2020). Fostering responsible research on ancient DNA. *American Journal of Human Genetics*, *107*(2), 183–195. https://doi.org/10.1016/j.ajhg.2020.06.017

Watanabe, M. E. (2015). The Nagoya protocol on access and benefit sharing. *Bioscience*, *65*(6), 543–550. https://doi.org/10.1093/biosci/biv056

Zohary, D., Hopf, M., & Weiss, E. (2012). *Domestication of plants in the old world*. Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199549061.001.0001

How to cite this article: Iob, A., & Botigué, L. (2022). Crop archaeogenomics: A powerful resource in need of a well-defined regulation framework. *Plants, People, Planet*, *4*(1), 44–50. https://doi.org/10.1002/ppp3.10233

ACKNOWLEDGEMENTS

I would like to express my gratitude to those who contributed to my professional and personal growth in last few years.

First, I would like to thank my supervisor Laura Botigué for giving me the opportunity to conduct my research on such a fascinating topic. Despite the many challenges we faced during these past years, she never lost trust in me, even when I doubted myself.

To all the colleagues and friends at CRAG, you have been invaluable for me. Without you all, these years would have not been the same.
To Oriol, a small portion of this thesis is indebted to all the help you provided me during your experience at CRAG. You did not only help me solving issues and improving my coding skills, you also brought laughter to those challenging moments.
To Manuel, for your amazing enthusiasm towards science and life.
To Ioanna, for sharing *some* frustration, and a lot of joyful moments.
To the Animal Genomics group past and present members, who shared so much with me. To the ones who welcomed me from the first day, Lino, Maria, Yron, Lurdes, and to the ones that supported me during most of my PhD. Cristina, Magí, Ferran, Jesus, Antonia and Taina, I'll never forget "chisme" coffee breaks in the office, beers at Vila and all the good times we spent together.
To the people who shared shorter yet meaningful moments on this journey. Uxue, Kostas, Max, Marc, Laia, Elisa, Elena and Giacomo. It was a pleasure spending time with you and creating lasting memories together.

Voglio anche ringraziare i miei amici di Barcellona: Costanza, Marta, Umberto, Rachele, Marianna e Nicola. Le volte in cui mi avete ascoltata, supportata e sopportata non si possono contare. Ce l'abbiamo fatta!

Naturalmente un ringraziamento va anche ai miei amici di tutta la vita, Daisy, Elizabeth, Cristina, Dario,Cristiano e Alex che, nonstante la distanza, mi hanno accompagnata in tutte le tappe più importanti.

A Simone, per essermi stato accanto con amore e pazienza nella parte forse più difficile di questo percorso. Grazie di cuore.

A tutta la mia grande famiglia, e soprattutto a mia madre, mio padre e mia sorella Anna, che hanno sostenuto le mie scelte e mi hanno aiutata con coraggio ed entusiasmo a fare passi importati, guardandomi sempre le spalle. Senza di voi non avrei fatto nemmeno metà della strada che mi ha portata fin qua.

Par finì, a me none Meri, la me seconde mame. A je simpri stade un esempli di furce, coragjo e tenerece dute intune persone. Mi plasares tant che mi podes viodi ancje uè, co ai finalmentri finude "le scuele".