Vegetation History and Archaeobotany
https://doi.org/10.1007/s00334-022-00898-7

ORIGINAL ARTICLE

# 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

**Alice Iob**[1] · **Laura Botigué**[1]

Received: 2 March 2022 / Accepted: 15 August 2022


**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. *dicoccum* (emmer wheats) with a previously published 3,000 year 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 analysed. 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* (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 (Fuller et al. 2014; Allaby et al. 2017; 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) (Zaharieva et al. 2010; Özkan et al. 2011). 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 et al. 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

Communicated by J. Whitlam.

✉ Laura Botigué
laura.botigue@cragenomica.es

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

Published online: 25 November 2022

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 emmers from the northern and eastern Fertile Crescent suggested that southeastern Turkey and northern Syria were the centres of emmer domestication (Lev-Yadun et al. 2000; Özkan 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 (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) (11.6–10.7 ka 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) (10.2–8.3 ka 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 2,200 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 9,800 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 7,500–6,650 cal BP (Wendrich and Cappers 2005). Around 5,000 BP emmer reached Ethiopia (Helbæk 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. (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 9,000 BP (even though evidence of management of partially domestic emmer is known from Chogha Golan as far back as 9,800 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 (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 3,130–3,000 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-analyse 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).

## 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 (Zhou et al. 2020a). 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, $^{14}$C dated to 1130–1000 cal BC (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 (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 (Broad 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. (2020a) 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.

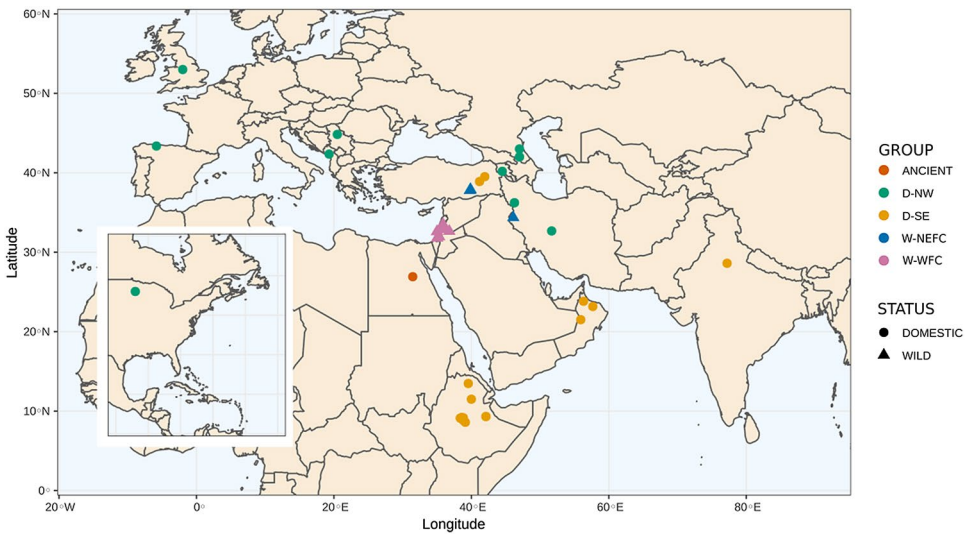


**Fig. 1** Map showing the distribution of sites of the samples analysed in this study, coloured according to the genetic clustering identified in the analyses. *WWFC* wild western Fertile Crescent, *WNEFC* wild northern and eastern Fertile Crescent, *DNW* domestic northwest route of dispersal, *DSE* domestic southeast route of dispersal. Samples from the WNEFC population come from only two sampling sites (7 samples, 2 sites). Sample W-LBN3 from Lebanon is not represented due to lack of geographical information

### 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. 2020a, https://bigd.big.ac.cn/search/?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-fertilising, 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 (randomised 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 neighbour 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 mis-incorporations 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 neighbour-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 analysed 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 analysing 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/vcf2treemix.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 neighbour joining, we kept transversions only and filtered with 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. (2020a) 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 coloured 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,

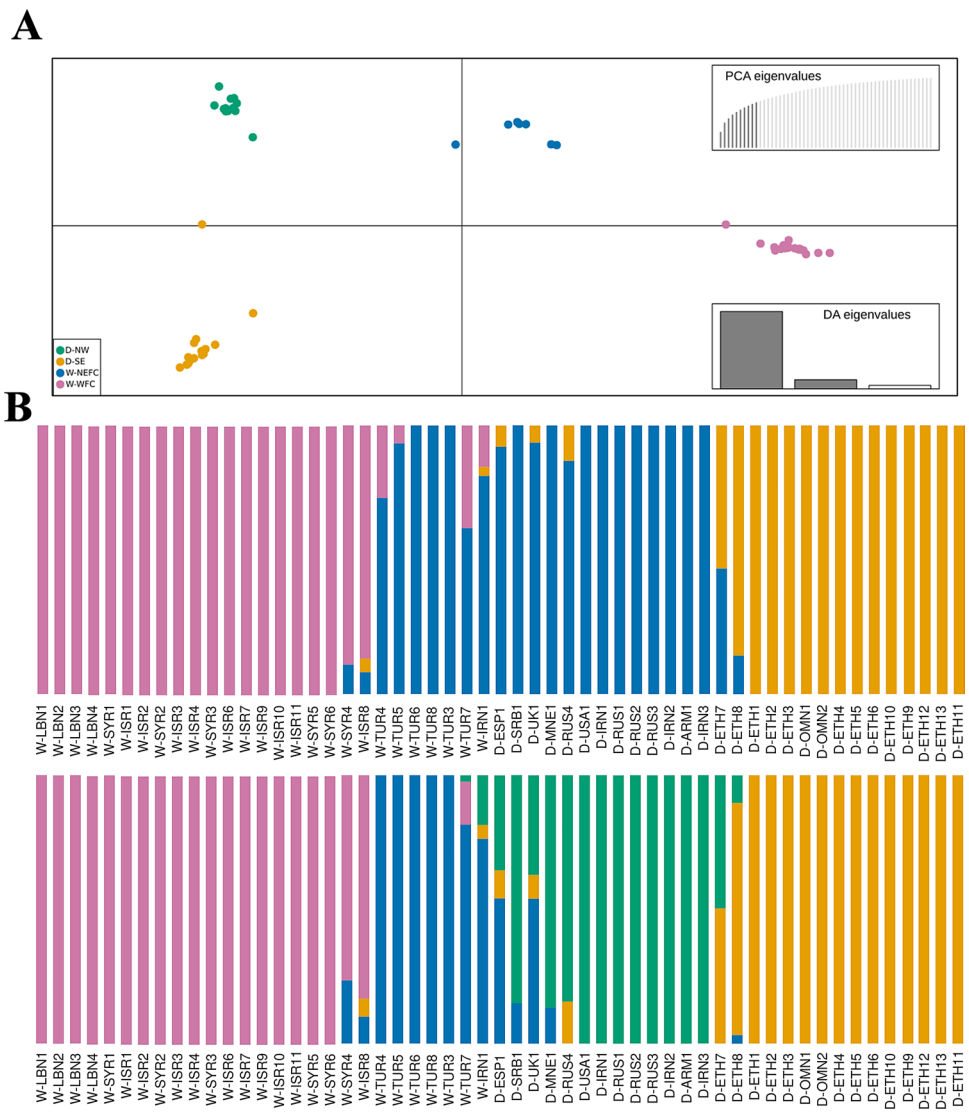


**Fig. 2** **a** Discriminant analysis of principal components (DAPC). Groups shown by their colours are arranged according to the PCA results and confirmed by this analysis. **b** ADMIXTURE analysis for the best values of K. Upper panel, K = 3; lower panel, K = 4. **c** Maximum likelihood tree, modern 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

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 sub-clades, 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 neighbour-joining tree (ESM Fig. S3).

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 northeastern 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

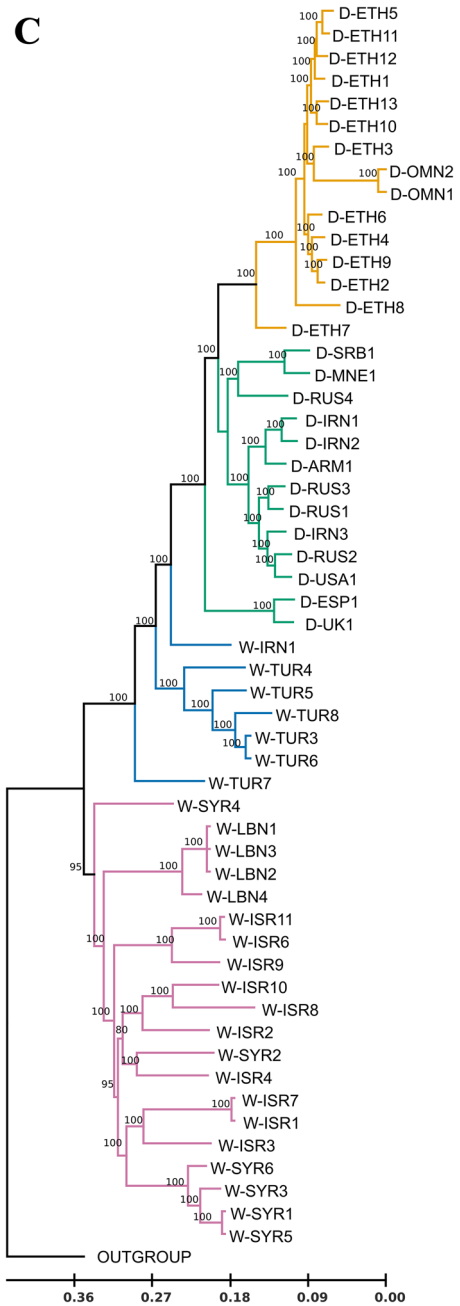


**Fig. 2** (continued)

the allelic variability in the genes responsible for a non-shattering rachis (TtBTR1A and TtBTR1B) (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 themelves (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.

## 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 emmers 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.

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

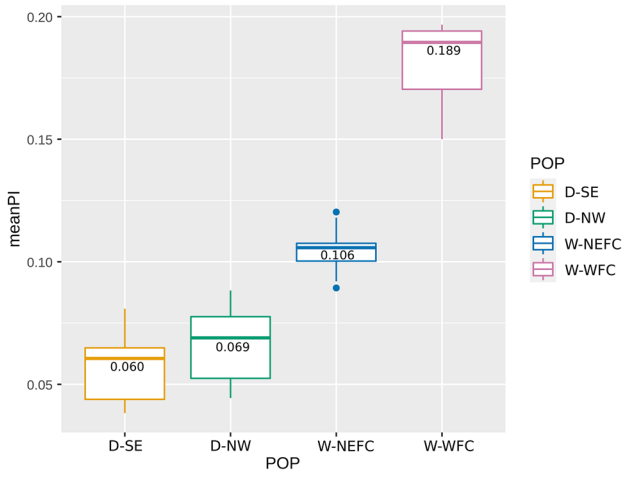


**Fig. 3** Nucleotide diversity (pi) in different populations, showing that the WWFC population is the most diverse, while WNEFC is less so. Domestic populations, DNW and DSE, have similar levels of diversity

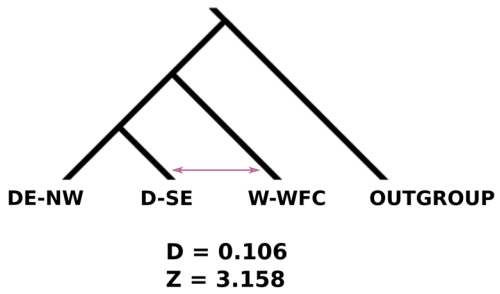


**Fig. 4** Schematic representation of Patterson's D (ABBA-BABA) statistical test results; the arrow shows the direction of gene flow between WWFC and DSE

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.

In order to better characterize the nature of the DSE group, we included publicly available data on emmers 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 neighbour 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.

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

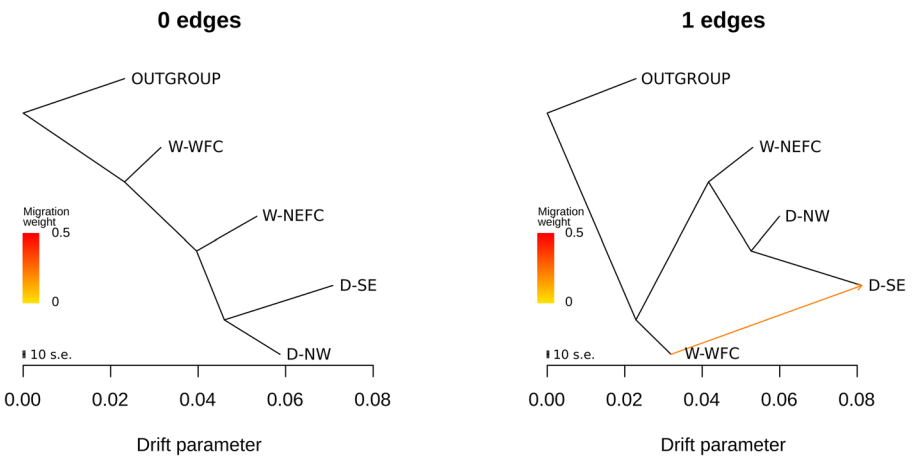


**Fig. 5** TreeMix maximum likelihood analysis of the modern whole genome dataset allowing for 0 (left panel) and 1 (right panel) migration events, showing gene flow between WWFCSL and DSE (arrow). The colour of the arrow is proportional to the amount of migration

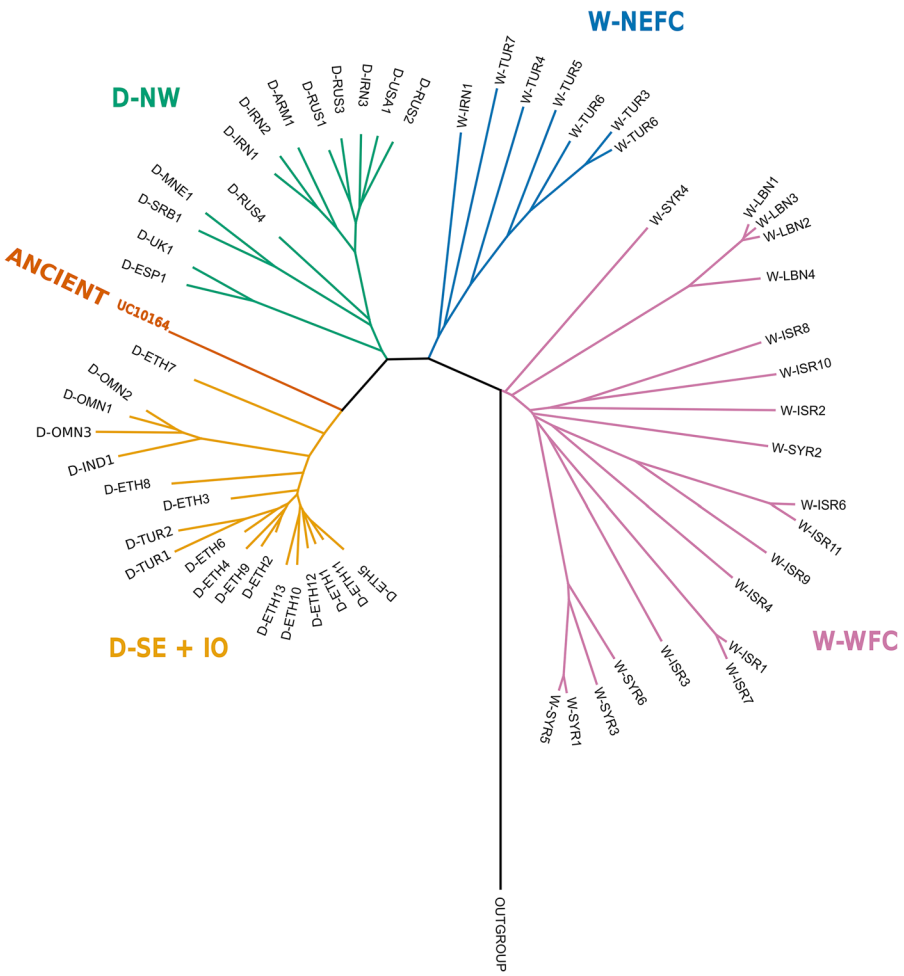


**Fig. 6** Neighbour joining phylogenetic tree of the extended dataset, including the Indian Ocean samples and the ancient Egyptian sample. Bootstrap 100 replicates, all nodes have bootstrap values are above 90

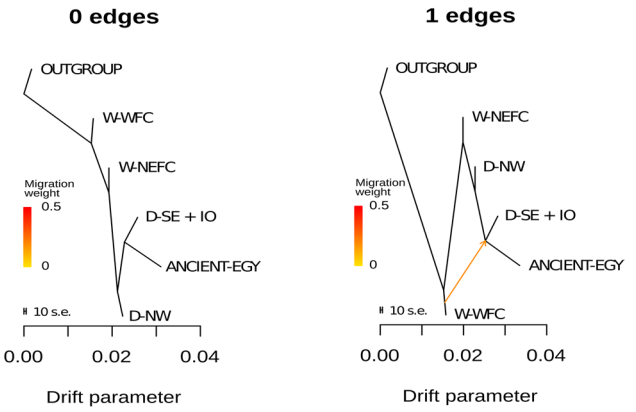


**Fig. 7** TreeMix maximum likelihood analysis of the entire dataset including the Indian Ocean and the ancient sample, allowing for 0 (left) and 1 (right) migration events, showing gene flow between WWFC and the node to modern DSE + IO and the ancient sample (arrow). The colour of the arrow is proportional to the amount of migration

(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).

## 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 9,000 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 (Zhou et al. 2020b), 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 4,700–4,500 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 raises 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 (Özkan 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 3,000 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://doi.org/10.1007/s00334-022-00898-7.

**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.



## References

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19:1,655–1,664. https://doi.org/10.1101/gr.094052.109.vidual

Allaby RG, Stevens C, Lucas L, Maeda O, Fuller DQ (2017) Geographic mosaics and changing rates of cereal domestication. Philos Trans R Soc B 372:1–10. https://doi.org/10.1098/rstb.2016.0429

Arranz-Otaegui A, Colledge S, Zapata L, Teira-Mayolini LC, Ibáñez JJ (2016) Regional diversity on the timing for the initial appearance of cereal cultivation and domestication in southwest Asia. Proc Natl Acad Sci USA 113:14,001–14,006. https://doi.org/10.1073/pnas.1612797113

Avni R, Nave M, Barad O et al (2017) Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. Science 357(6346):93–97. https://doi.org/10.1126/science.aan0032

Broad Institute (2019) Picard Toolkit. https://broadinstitute.github.io/picard/

Civáň P, Ivaničová Z, Brown TA (2013) Reticulated origin of domesticated emmer wheat supports a dynamic model for the emergence of agriculture in the fertile crescent. PLoS ONE 8:e81955. https://doi.org/10.1371/journal.pone.0081955

Coward F, Shennan S, Colledge S, Conolly J, Collard M (2008) The spread of Neolithic plant economies from the Near East to northwest Europe: a phylogenetic analysis. J Archaeol Sci 35:42–56. https://doi.org/10.1016/j.jas.2007.02.022

Cuny J, Mouton M (2009) La transition vers la période sassanide dans la péninsule d'Oman: chronologie et modes de peuplement. In: Schiettecatte J, Robin CJ (eds) L'Arabie à la veille de l'Islam: Bilan Clinique. De Boccard, Paris, pp 91–133

Danecek P, Auton A, Abecasis G et al (2011) The variant call format and VCFtools. Bioinformatics 27:2,156–2,158

Danecek P, Bonfield JK, Liddle J, etal, (2021) Twelve years of SAMtools and BCFtools. GigaScience 10:1–4. https://doi.org/10.1093/gigascience/giab008

Fuller DQ, Denham T, Arroyo-Kalin M et al (2014) Convergent evolution and parallelism in plant domestication revealed by an expanding archaeological record. Proc Natl Acad Sci USA 111:6,147–6,152. https://doi.org/10.1073/pnas.1308937110

Green RE, Krause J, Briggs AW et al (2010) A draft sequence of the Neandertal genome. Science 328(5979):710–722. https://doi.org/10.1126/science.1188021

Helbæk H (1970) The plant husbandry of Hacilar. In: Mellaart J (ed) Excavations at Hacilar. University Press, Edinburgh, pp 189–244

Jorgensen C, Luo M-C, Ramasamy R 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. Front Plant Sci 8:1–13. https://doi.org/10.3389/fpls.2017.01798

Kabukcu C, Asouti E, Pöllath N, Peters J, Karul N (2021) Pathways to plant domestication in Southeast Anatolia based on new data from aceramic Neolithic Gusir Höyük. Sci Rep 11:1–15. https://doi.org/10.1038/s41598-021-81757-9

Kislev M, Nadel D, Carmi I (1992) Epipalaeolithic (19,000 BP) cereal and fruit diet at Ohalu II, Sea of Galilee, Israel. Rev Palaeobot Palynol 73:161–166

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1,547–1,549

Lev-Yadun S, Gopher A, Abbo S (2000) The cradle of agriculture. Science 288(5471):1,602–1,603

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows—Wheeler transform. Bioinformatics 25:1,754–1,760. https://doi.org/10.1093/bioinformatics/btp324

Luo M-C, Yang Z-L, You FM, Kawahara T, Waines JG, Dvorak J (2007) The structure of wild and domesticated emmer wheat populations, gene flow between them, and the site of emmer domestication. Theor Appl Genet 114:947–959. https://doi.org/10.1007/s00122-006-0474-0

Maccaferri M, Harris NS, Twardziok SO et al (2019) Durum wheat genome highlights past domestication signatures and future improvement targets. Nat Genet 51:885–895. https://doi.org/10.1038/s41588-019-0381-3

Malinsky M, Matschiner M, Svardal H (2021) Dsuite—Fast *D*-statistics and related admixture evidence from VCF files. Mol Ecol Resour 21:584–595. https://doi.org/10.1111/1755-0998.13265

Mani BR (2004) Further evidence on Kashmir Neolithic in the light of recent excavations at Kanishkapura. J Interdiscip Stud Hist Archaeol 1:137–142

Nave M, Avni R, Çakır E et al (2019) Wheat domestication in light of haplotype analyses of the *Brittle rachis* 1 genes (BTR1-A and BTR1-B). Plant Sci 285:193–199. https://doi.org/10.1016/j.plantsci.2019.05.012

Oliveira HR, Jacocks L, Czajkowska BI, Kennedy SL, Brown TA (2020) Multiregional origins of the domesticated tetraploid wheats. PLoS ONE 15:e0227148. https://doi.org/10.1371/journal.pone.0227148

Özkan H, Brandolini A, Schäfer-Pregl R, Salamini F (2002) AFLP analysis of a collection of tetraploid wheats indicates the origin of emmer and hard wheat domestication in southeast Turkey. Mol Biol Evol 19:1,797–1,801

Özkan H, Willcox G, Graner A, Salamini F, Kilian B (2011) Geographic distribution and domestication of wild emmer wheat (*Triticum dicoccoides*). Genet Resour Crop Evol 58:11–53. https://doi.org/10.1007/s10722-010-9581-5

Pankin A, Altmüller J, Becker C, von Korff M (2018) Targeted resequencing reveals genomic signatures of barley domestication. New Phytol 218:1,247–1,259. https://doi.org/10.1111/nph.15077

Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet 8:e1002967. https://doi.org/10.1371/journal.pgen.1002967

Purcell S, Neale B, Todd-Brown K et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575

R Core Team (2021) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, https://www.r-project.org/

Riehl S, Zeidi M, Conard NJ (2013) Emergence of agriculture in the foothills of the Zagros mountains of Iran. Science 341(6141):65–67. https://doi.org/10.1126/science.1236743

Salunkhe A, Tamhankar S, Tetali S, Zaharieva M, Bonnett D, Trethowan R, Misra SC (2013) Molecular genetic diversity analysis in emmer wheat (*Triticum dicoccon* Schrank ) from India. Genet Resour Crop Evol 60:165–174. https://doi.org/10.1007/s10722-012-9823-9

Scott MF, Botigué LR, Brace S et al (2019) A 3,000-year-old Egyptian emmer wheat genome reveals dispersal and domestication history. Nat Plants 5:1,120–1,128. https://doi.org/10.1038/s41477-019-0534-5

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:e0131422. https://doi.org/10.1371/journal.pone.0131422

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1,312–1,313

Stevens CJ, Murphy C, Roberts R, Lucas L, Silva F, Fuller DQ (2016) Between China and South Asia: a middle Asian corridor of crop dispersal and agricultural innovation in the Bronze Age. Holocene 26:1,541–1,555. https://doi.org/10.1177/0959683616650268

Van der Auwera GA, O'Connor BD (2020) Genomics in the cloud: using Docker, GATK, and WDL in Terra, 1st edn. O'Reilly Media Inc, Sebastopol

Van Zeist W, de Roller GJ (2003) The Çayönü archaeobotanical record. In: van Zeist W (ed) Reports on Archaeobotanical Studies in the Old World. The Groningen Institute of Archaeology. University of Groningen, Groningen, pp 143–166

Vavilov N (1926) Studies on the origin of cultivated plants. Institut Botanique Appliquée et d'Amélloration des Plantes, Leningrad

Weiss E, Kislev ME, Simchoni O, Nadel D (2004) Small-grained wild grasses as staple food at the 23 000-year-old site of Ohalo II, Israel. Econ Bot (suppl) 58:S125–S134

Wendrich WZ, Cappers RTJ (2005) Egypt's earliest granaries: evidence from the Fayum. Egyptian Archaeol 27:12–15

Zaharieva M, Ayana NG, Al Hakimi A, Misra SC, Monneveux P (2010) Cultivated emmer wheat (*Triticum dicoccon* Schrank), an old crop with promising future: a review. Genet Resour Crop Evol 57:937–962. https://doi.org/10.1007/s10722-010-9572-6

Zhou X, Yu J, Spengler RN et al (2020a) 5,200-year-old cereal grains from the eastern Altai Mountains redate the trans-Eurasian crop exchange. Nat Plants 6:78–87. https://doi.org/10.1038/s41477-019-0581-y

Zhou Y, Zhao X, Li Y et al (2020b) *Triticum* population sequencing provides insights into wheat adaptation. Nat Genet 52:1,412–1,422. https://doi.org/10.1038/s41588-020-00722-w