ORIGINAL ARTICLE

Geographic differentiation of domesticated einkorn wheat and possible Neolithic migration routes

A Brandolini[1], A Volante[1] and M Heun[2]

To analyse the spread of domesticated einkorn into Europe, 136 landraces, 9 wild einkorns and 3 *Triticum urartu* were fingerprinted by the diversity array technology sequence (DArT-seq) marker technology. The obtained 3455 single-nucleotide polymorphism (SNP) markers confirmed earlier results about the separation of wild and domesticated einkorn from *T. urartu* and about the pinpointing of the domesticated forms to the Karacadağ Mountains (Turkey). Further analyses identified two major domesticated landrace einkorn groups, one relating to the Prealpine region and the other to the Maghreb/Iberian region. The previously published four geographical provenance groups were mostly identified in our results. The earlier reported unique position of the Maghreb/Iberia einkorns cannot be confirmed, as the three landrace clusters we identified with STRUCTURE also occur in the remaining einkorn, although at different frequencies. The results are discussed with respect to the spreading of domesticated einkorn into Western Europe and two possible Neolithic migration routes are indicated.
Heredity (2016) **117**, 135–141; doi:10.1038/hdy.2016.32; published online 11 May 2016

INTRODUCTION

Domesticated einkorn wheat (*Triticum monococcum* L. ssp. *monococcum*) is one of the oldest crops in the world and part of the Neolithic package that started emerging ~ 10400 calibrated years before present (BP) in the Near East (Zohary and Hopf, 2000; Haldorsen *et al.*, 2011). Einkorn cannot compete in terms of yield with bread wheat, but its quality characteristics are unique (Abdel-Aal *et al.*, 1995; Hidalgo and Brandolini, 2014) and consumption of einkorn is very healthy for humans (Løje *et al.*, 2003; Brandolini and Hidalgo, 2011). In Europe, today probably < 1000 Ha are grown with einkorn, a trifling order of magnitude compared with the > 200 million Ha grown with bread wheat alone in 2013 (FAO, 2015). Accordingly, einkorn breeding is negligible in comparison with, for example, bread wheat, barley, rice and maize breeding. The few existing efforts are limited to mostly research-based activities that strive to analyse its unique traits and to make this healthy cereal accessible to consumers. As a positive by-product of this limited economic relevance, > 99.9% einkorn accessions stored in gene banks around the world are either wild einkorn (*T. boeoticum* subsp. *thaoudar*; henceforth *T. boeoticum*), feral einkorn (*T. boeoticum* subsp. *aegilopoides*, not tested here) or domesticated landrace einkorn (*T. monococcum* ssp. *monococcum*; hereafter *T. monococcum*). We follow the nomenclature of *thaoudar* and *aegilopoides* given by Schiemann (1948) that elucidated the confusion surrounding '*aegilopoides*'; see also Harlan and Zohary (1966) for the distinction between these two subspecies.

Probably because of einkorn's yield inferiority, the use of this wheat diminished after the Copper Age (Zaharieva and Monneveux, 2014) and this plant survived as a forgotten crop on marginal land throughout Europe, the Near East and Maghreb. At the end of the twentieth century, scattered einkorn populations were recorded in Turkey, the Balkans, France, Italy, Switzerland, Germany, Spain and Morocco (Perrino *et al.*, 1996). The fragmentation, isolation and marginalisation of einkorn support the assumption that the existing einkorn landraces reflect some of the genetic variation existing during the Copper Age, in addition to later natural adaptation to local growing conditions, random genetic drift and possible selection by farmers. The hypothesis that genetic information relating to the Neolithic spreading of crops has been maintained in landraces is also advocated in other publications. For example, Jones *et al.* (2011) assessed the evolutionary relationships between different European barley landraces by microsatellites, and presented a theory on the spread of barley cultivation, including the adaptation of the crop to new environments. Badaeva *et al.* (2015) analysed the C-banding pattern of several emmer wheat accessions from different Old World regions, and observed four major karyotype groups, each harbouring characteristic patterns, leading them to postulate the existence of different diffusion routes of this crop out of the Fertile Crescent.

Thus, the aim of this research was to analyse the genetic composition of domesticated landrace einkorn from Turkey, Europe and Maghreb in order to assess the existence of region-specific genetic patterns and to infer the possible Neolithic diffusion paths of this crop into Europe. To this end, 136 landraces of einkorn with rather well-known origin were selected. Nine accessions of wild einkorn and three accessions of *T. urartu* (the donor of the A genome to durum/emmer and bread wheats, not fertile with einkorn but morphologically almost identical to wild einkorn; Johnson and Dhaliwal, 1976; Heun *et al.*, 2008) were included into our analyses as reference points. The resulting 148 accessions were fingerprinted by DArT-seq DNA

[1]Consiglio per la Ricerca e la Sperimentazione in Agricoltura e l'analisi dell'economia Agraria, Unità per la Selezione dei Cereali e la Valorizzazione Delle Varietà Vegetali (CREA-SCV), Angelo Lodigiano, Italy and [2]Department of Ecology and Natural Resource Management (INA), Norwegian University of Life Sciences (NMBU), Ås, Norway
Correspondence: Dr A Brandolini, Consiglio per la Ricerca e la Sperimentazione in Agricoltura e l'analisi dell'economia Agraria, Unità per la Selezione dei Cereali e la Valorizzazione Delle Varietà Vegetali (CREA-SCV), Via R Forlani 3, 26866S Angelo Lodigiano (LO), Italy.
E-mail: andrea.brandolini@entecra.it

markers, a sequence-based improvement of the 'diversity array technology' (DArT) allowing whole-genome analyses (Wenzl et al., 2004).

MATERIALS AND METHODS

Plant material
A total of 136 landraces of einkorn were selected from the CREA-SCV gene bank following passport geographic criteria. Nine wild einkorn and three *Triticum urartu* accessions were also included as controls. The complete list of the 148 accessions and their origin is given in Supplementary Table S1.

DNA extraction and DArT-seq analysis
DNA extraction from 7-day-old seedlings (one plant per accession) was performed as described by Stein et al. (2001). DNA concentration was quantified with a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA, USA). The DNA was send to Australia and analysed with the DArT-seq technology (Kilian et al., 2012) by the company Diversity Arrays Technology (Canberra, Australia). The DArT-seq approach combines a DArT complexity reduction method targeting predominantly active genes and low copy sequences, and involves next-generation sequencing platforms. Single-nucleotide polymorphisms (SNPs) in defined DArT-seq markers were recorded and scored as '0' for the common allele and '1' for the alternative allele. When heterozygosity at a locus was recorded, a '2' was noted. Nine genotypes (that is, the very same DNA) were fingerprinted twice (unknown to the lab in Australia).

Genetic similarity analysis
From the filtered SNP DArT-seq data matrix, a pair-wise genetic similarity matrix based on Dice (1945) was computed and utilised to perform principal coordinates analysis with the software PAST3 (Hammer et al., 2001; Hammer and Harper, 2006). The similarity matrix was converted into a distance matrix (similarity minus 1, multiplied by − 1 gives distance) and used to generate neighbour-joining trees (Saitou and Nei, 1987) built with the MEGA version 4 software (Tamura et al., 2007). Bootstrap values ($n = 1000$) were computed with the PAST3 software and when superior to 70% (Hillis and Bull, 1993) these bootstrap values were added to the corresponding branches.

Structure v2.3.4 (Pritchard et al., 2000; Falush et al., 2003; Hubisz et al., 2009) was used to estimate the number of genetic clusters (K; from 1 to 10) best representing the data, and to assign each genotype to the different populations. As proposed by Nordborg et al. (2005) for selfing species, the data were analysed as haploids (more appropriate for a highly autogamous species of known phase that seldom reaches Hardy–Weinberg equilibrium), treating heterozygous loci as missing data. Structure simulations were carried out using an admixture model. The accessions were not preassigned to populations before the analyses, and were allocated to the K clusters with 10 iterations each (50 000 burn-in runs and 200 000 Markov chain Monte Carlo repeats), assuming admixture and independent alleles. The number of STRUCTURE clusters was determined as per Falush et al. (2003) and Evanno et al. (2005). The program CLUMPP v1.1.2 (Jakobsson and Rosenberg, 2007) allowed maximising the accuracy of the Q-values between the 10 independent runs, using the greedy function. The distribution of the individual Q-values was visualised with DISTRUCT v1.1 (Rosenberg, 2004). The final K was compared with significant bootstrapping (that is, > 70%; Hillis and Bull, 1993) in the neighbour-joining tree and seen in combination with principal coordinates analysis, because the results from STRUCTURE are best interpreted in conjunction with other different approaches (Anderson and Dunham, 2008; Thaulow et al., 2013; Thaulow et al., 2014).
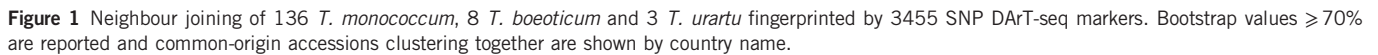
RESULTS

DArT-Seq-based genetic similarity
A total of 9241 SNP polymorphisms were identified by the DArT-seq genotyping approach across the 148 accessions tested. The reproducibility of the data was 99.8% (range: 97.7–100%). Heterozygous loci, scored as '2', represented 0.74% of total loci. Among the couples of the repeated genotypes, the average correlation coefficient was 0.881 (range: 0.841–0.922) and the heterozygous loci showed a mean frequency of 0.95% (range: 0.65–1.57), but the pair-wise comparison of each couple showed 66.8% mismatching when one sample was scored 2 (as the other was then scored either 1 or 0). Therefore, in most instances the score '2' was probably a consequence of fingerprinting errors, an interpretation also supported by the fact that einkorn is a strict inbreeder. The heterozygous results (that is, the '2' scores) were thus treated as missing data. Afterwards, the correlation coefficient between each couple of the 9 repeated genotypes was on average 0.9999 (range: 0.9996–1.0000). The markers with a proportion of missing information > 5% and the singletons (that is, polymorphisms that occur only in a single accession) in the domesticated einkorn data set were then removed from the matrix, and the remaining 3455 markers (still corresponding to > 500 000 data points) were used for further analyses. A comparison with a bread wheat consensus map built from SNP DaRTs data (A Kilian, unpublished data) showed 605 shared markers, covering the whole genome of the diploid wheats.

The neighbour-joining tree from the Dice distance matrix, along with the bootstrap values > 70 of the main branches, are presented in Figure 1. The three *T. urartu* accessions are clearly separated from all the other genotypes. Similarly, the *T. boeoticum* accessions split from the domesticated einkorns; however, the two Karacadağ accessions (ID752 and ID758) position themselves nearer to the *T. monococcum*. The tree shows also several landrace clusters of domesticated einkorn with a clear regional origin, often supported by high bootstrap values. Starting counterclockwise from the wild wheats, a small group with some former Yugoslavia accessions, and three Turkish einkorns (two belong to the free-threshing form *T. monococcum* subsp. *monococcum* convar. *sinskajae*, first described in 1926 by Zhukovski) are forming a supported branch. Next comes a cluster including some Caucasus accessions, a big group comprising almost all the Maghreb and Iberia accessions (bootstrap value: 100%), a group of accessions mostly from Albania and former Yugoslavia, a Greece–Turkey–Bulgaria group, a small mainly Italian branch and a big cluster largely made up of Austrian, Swiss, German and, to some extent, French einkorns. The accessions from Hungary and Romania, instead, are interspersed in the different groups. Some accessions had identical DNA patterns and three different situations can occur, as the indistinguishable accessions might come: (1) from the same country (for example, ID571, ID572, ID573, ID574 and ID575; ID195 and ID566); (2) from bordering countries (for example, ID139 and ID314; ID13 and ID432); or (3) from distant countries (for example, ID508 and ID517; ID118 and ID322).
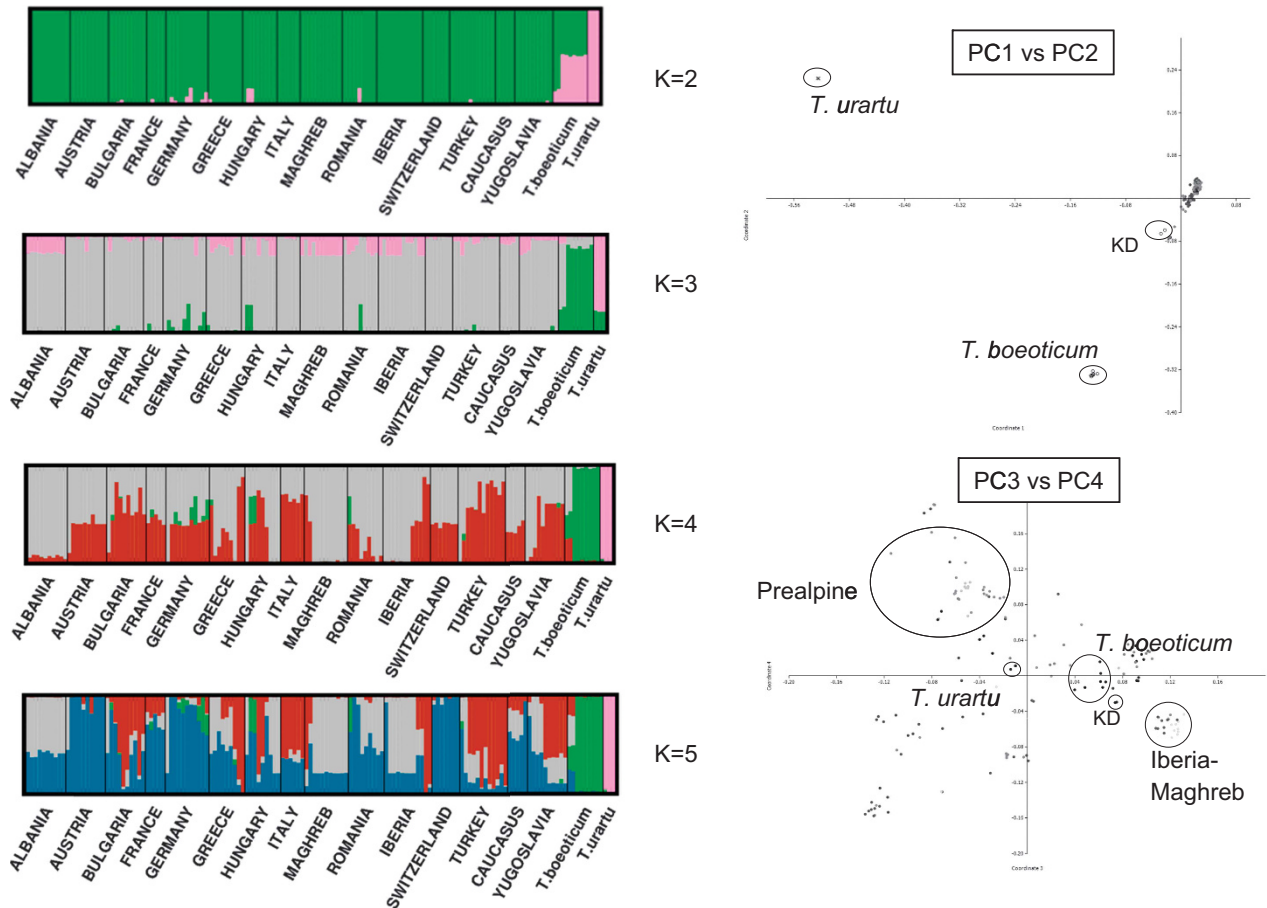
The principal coordinates analysis showed that the first four dimensions accounted for 69.4% of the total variation (Figure 2, right part). Dimension 1 (principal component 1) explained 38.4% of the variation and separated *T. urartu* accessions from all the other accessions, as well as *T. boeoticum* from *T. monococcum*; but as mentioned above, the two wild einkorn accessions coming from the Karacadağ were positioned close to the domesticated einkorns. The second dimension (principal component 2), which described 16.4% of the variation, again separated *T. urartu* from *T. boeoticum*, and both from domesticated einkorn; the Karacadağ wild einkorns now fell within the domesticated einkorn continuum. The third (principal component 3) and the fourth (principal component 4) dimensions explained 7.5% and 7.2% of the total variation, respectively, and separated the genotypes from Maghreb and Iberia as well as most of the accessions from areas close to the Alps (Germany, Austria, Switzerland and France, henceforth called Prealpine group, because einkorn grows well in the alpine foothills but not in the High Alps) from the remaining accessions. These results are in good agreement with the STRUCTURE analysis (shown in Figure 2, on the left side), as the Falush et al. (2003) and Evanno et al. (2005) approaches suggest

Figure 1 Neighbour joining of 136 *T. monococcum*, 8 *T. boeoticum* and 3 *T. urartu* fingerprinted by 3455 SNP DArT-seq markers. Bootstrap values ⩾70% are reported and common-origin accessions clustering together are shown by country name.

the presence of $K = 5$ groups. A step-wise consideration of the STRUCTURE results in Figure 2 shows that at $K = 2$, *T. urartu* and, partially, *T. boeoticum* are already separated from all the other accessions; at $K = 3$, the two wild species split; at $K = 4$, the main separation between the southern and the continental einkorns is emerging; and at $K = 5$, the Prealpine and the Spain-Maghreb groups are clearly defined. An important observation is the position of the Karacadağ accessions, two true *T. boeoticum* that however marker-wise are much more related to the domesticated einkorns.

In Figure 3, country-wise frequencies are represented in a map, after the assemblage of a database combining each accession origin and Q matrix information identified by STRUCTURE for the $K = 5$ model; pie sizes are proportional to the number of accessions included. It becomes evident that *T. urartu* (pink, cluster 1) is unique, whereas wild einkorn contains a huge green cluster (cluster 2), and also the three clusters (grey, blue and red) that are predominant among domesticated einkorns and are distributed in a country-specific way.

DISCUSSION

Genetic similarity
DArT markers measure genome-wide genetic similarity and became mapped and merged with microsatellites in einkorn (Jing *et al.*, 2009). The markers applied here, SNP DArT-seq, are a sequence-based

improvement of the former hybridisation-based approach and allow even higher resolution because of the combination with next-generation sequencing. Our stringent criteria reduced the original number of DArT-seq markers to ~40%. The resulting 3455 markers provided on average 3 markers per centimorgan (cM), as the 7 einkorn chromosomes cover roughly 1000 cM (Tänzler *et al.*, 2002; Singh *et al.*, 2007). This resolution should allow a detailed genetic similarity measurement of the selected 148 genotypes. In fact, the observed genetic difference between *T. urartu* and the einkorn accessions is huge, as expected (Heun *et al.*, 2008). Wild einkorn is also very clearly separated from domesticated einkorn, and the Karacadağ accessions act as bridge, as shown initially by Heun *et al.* (1997). Therefore, the analysed DArT-seq markers are confirming the earlier amplified fragment length polymorphism-based analyses in full. We conclude here that genetic similarity is well described by the applied SNP DArT-seq markers, and focus now on the differentiation between the 136 domesticated einkorn landraces.

Country of origin
The country of origin of the einkorn landraces was obtained from the gene banks who provided most material and from L. Peña-Chocarro, who collected einkorns in Spain and Morocco. We initially assumed as a basis of our analyses that the country of origin information was correct, but we are aware that some factors (for example, not well

Figure 2 STRUCTURE analyses with *K* ranging from 2 to 5 (on the left) and principal coordinates analysis (principal component 1 (PC1) vs PC2 and PC3 vs PC4, on the right) of 136 *T. monococcum*, 8 *T. boeoticum* and 3 *T. urartu* based on 3455 SNP DArT-seq markers.

documented seed exchange) violate this assumption. In fact, accession misidentification is common even in well-managed gene banks (Goncharov, 2011). Accessions with identical molecular fingerprints allowed us distinguishing three different situations. In the first (identical accessions belonging to the same geographical area) and second (identical accessions from neighbouring countries), the respective accessions probably belong to the same widespread landrace, beyond modern country borders. In the third situation (limited to five einkorn pairs with identical fingerprint, but very different country of origin), the most probable explanation is poorly documented seed exchange and/or sample confusion in the years after collection. A good example is represented by accessions ID118 from Morocco and ID322 from Bulgaria that are identical and cluster in a group rich of Bulgarian genotypes; both are classified as var. *macedonicum*, hinting that the Balkans are their most probable home (Szabó and Hammer, 1996). In conclusion, our study will not focus on individual accessions and their assumed county of origin, but instead we will work with country averages, and even then, some insecurity remains. Therefore, country averages are only used to deduct the possible Neolithic spread of domesticated einkorn into Europe. Furthermore, we primarily concentrate on the main distinction into a Prealpine group and a Maghreb/Iberia group.

Two main routes out of Turkey

The STRUCTURE analyses indicated *K* = 5 as the most probable distinction of the whole data set. The above-mentioned separation of

T. urartu (containing almost purely cluster 1), wild einkorn (contains a huge cluster 2, but also at low frequency the three remaining clusters) and domesticated einkorn suggests that the landrace einkorn accessions are described mostly by the remaining three clusters (called landrace clusters) with varying frequencies. The frequency of the landrace-related colours (that is, clusters) along the trail from Turkey to Bulgaria/Greece via former Yugoslavia to Hungary shown in Figure 3 is well fitting the assumed spread of agriculture along an inland path (Davison *et al.*, 2006; Bocquet-Appel *et al.*, 2009). The presence of the green cluster in several country groups (Turkey, Bulgaria, former Yugoslavia, Hungary, Austria, Germany and France) along this path further bolsters our conclusions.

The further diffusion of early agriculture from Hungary into central Europe is connected to the Linearbandkeramik culture, as indicated by human ancient DNA profiles from Mesolithic, Neolithic Starcevo and Linearbandkeramik sites (Szécsényi-Nagy *et al.*, 2015). Kreuz *et al.* (2005), working with archaeobotanical remains, showed that the Linearbandkeramik was focussed on einkorn and emmer, and that einkorn was the dominant cereal. Kreuz *et al.* (2005) discuss reasons for this fact by considering agronomic traits, like yield and lodging tolerance. We can be more specific here: the analysed einkorn landraces from Switzerland, Austria and Germany, and in part France, which were mostly collected in the Prealpine region, must have undergone some random drift and/or selection, as two of the Structure colours are largely lost. The second main path of the spread of
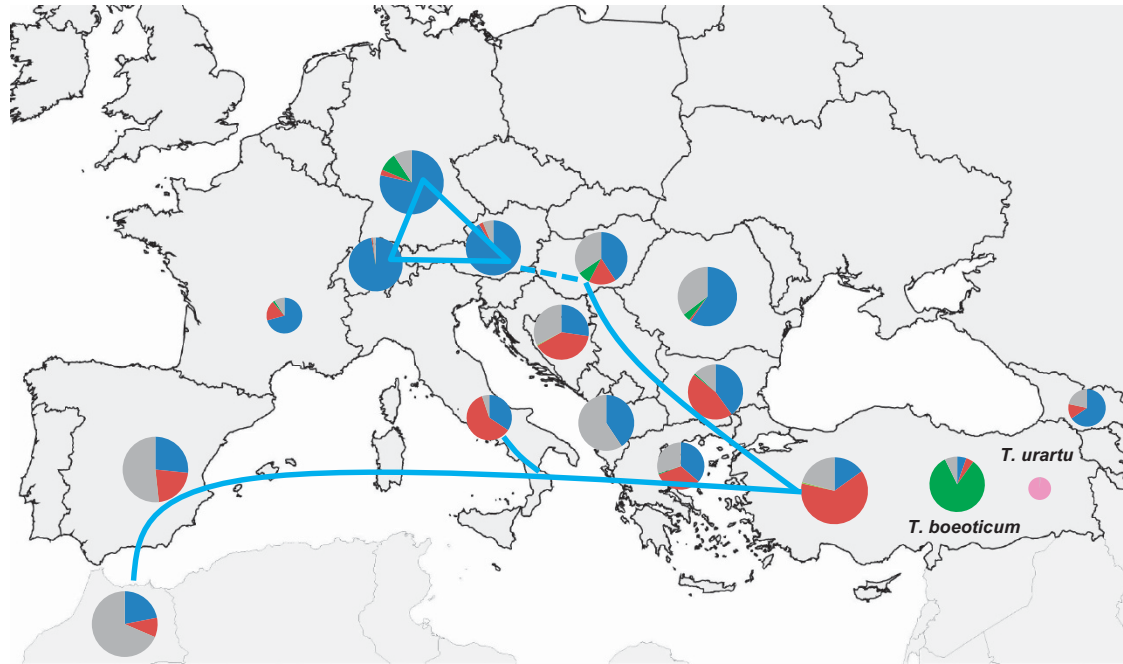
Figure 3 Geographical distribution of population structure in *T. monococcum*, *T. boeoticum* and *T. urartu* accessions based on 3455 SNP DArT-seq markers as revealed by STRUCTURE in a model assuming five clusters ($K = 5$). The pie charts summarise the results of the accessions from each region, with the proportional membership of the alleles to each one of the five clusters. The size of each pie is relative to the number of accessions per country. Two possible Neolithic migration routes are outlined.

agriculture and its founding crops followed the Mediterranean coast and finally reached Maghreb and Iberia. Cortés Sánchez *et al.* (2012) date the first existence of Neolithic settlements in southern Iberia and the Maghreb to 7500 calibrated years BP. Einkorn was domesticated by 10 400 calibrated years BP in the Karacadağ (Haldorsen *et al.*, 2011) and therefore this migration path might have taken ~3000 years. However, as the start of the spreading out of Turkey might be connected to a short cooling event at 8200 calibrated years BP (Weninger *et al.*, 2006, recently criticised by Flohr *et al.*, 2015) the actual time would be less than one millennium. Oliveira *et al.* (2011) genotyped 50 einkorns with 16 nuclear- and 5 chloroplast-derived microsatellites and argued that the history of the Western Mediterranean einkorns (from the Iberian Peninsula and Morocco) was different from the rest of Europe. We confirm that the Maghreb/Iberia einkorns contain a different frequency of alleles (summarised by the STRUCTURE colours) that are however already present in the Turkish accessions tested. The long distance of the Western Mediterranean einkorns from their wild and domesticated relatives in Turkey, their fast arrival in the Iberian Peninsula (Zapata *et al.*, 2004), the peculiar growing conditions (very southern, yet relatively cold mountain climate), the special use of the plant, for example, as straw for thatched roof (Peña-Chocarro and Zapata, 1998; Peña-Chocarro *et al.*, 2009) and the continued utilisation of this crop for food/feed into present times can explain the observed shift in allelic frequencies. However, we do not agree calling the Western Mediterranean a separate einkorn gene pool, as Oliveira *et al.* (2011) do. The Western Mediterranean chloroplast haplotype 24 (their main argument) is separated from type 22 (occurring in the Fertile Crescent) by a single mutation (according to Supplementary Figure S1 of Oliveira *et al.*, 2011) and this mutation could have easily occurred and accumulated during the past several thousand years in Western Mediterranean. We must also

remember that during the early Neolithic the Iberian Peninsula was not characterised by just one farming practice (Antolín *et al.*, 2015).

Geographical provenance groups

Szabó and Hammer (1996) reviewed the taxonomy of hulled wheats and suggested four 'geographical provenance groups' for einkorn: *Helotinum* covering Asia Minor, Transcaucasia and the Crimea, *Alemanum* with Germany and Switzerland, *Ibericum* with Spain, Southern France and Maghreb and *Transsylvanicum* with Eastern Carpathians and possibly the Balkan Peninsula. Our data agree with *Alemanum* in general terms, but we might add that the Swiss einkorn accessions seem to be more typical than the German accession as a whole, as the latter might include material from Central and Northern Germany. We consider this group as Prealpine, because these einkorn accessions seem to be adapted to the foothills of the Alps, experiencing rain, frost and a prolonged vegetative period. Some French accessions might experience such conditions too. The group *Ibericum* is also supported by our data in general terms. The group *Transsylvanicum* is covered by our accessions from Rumania that have a distinctive *K* colour composition (see Figure 3). The growing conditions are special and the uniqueness could be the result of the immigrations of central European farmers and their crops into this rather isolated area, but a definitive answer cannot be given here. Szabó and Hammer (1996) considered that the Balkan Peninsula might be part of the *Transsylvanicum* geographical provenance group, but our data do not suggest such an inclusion, as the einkorns from Bulgaria, former Yugoslavia and so on are containing the red *K* cluster (see Figure 3) that is almost absent in Romania. Albanian einkorns also lack the red cluster, and might be more similar to the Rumanian einkorn, but no sensible explanation

is available. The fourth group, *Helotinum*, cannot be addressed here, yet the geographical provenance groups presented by Szabó and Hammer (1996) can be mostly identified in our results.

In conclusion, our research confirms the identification of the Karacadağ Mountains in Turkey as the most likely area of transition from wild to domesticated einkorn. Domesticated einkorn landraces highlight the presence of regional groups; in particular, two major clusters occur, one defining the Prealpine region and the other the Maghreb/Iberian region. However, it is important to notice that all the three landrace-related *K* colours are already present at low frequency in the wild einkorn and at high frequencies in the domesticated einkorns from Turkey. We observe no discontinuity from East (Turkey) to West (Maghreb/Iberian) with the >3000 genome-wide SNP DArT-seq markers and cannot agree with Oliveira *et al.* (2011), who base their results on 21 microsatellites. Our data support the hypothesis of two different Neolithic migration routes linked to the spread of domesticated einkorn into Western Europe, one continental and the other maritime. We warn that country of origin information for single accessions can be misleading and suggest that even country averages should be used with care.

DATA ARCHIVING

The SNP DArT-seq data set used for the analyses of this paper is deposited as Excel file at the Dryad repository as doi:10.5061/dryad.39d23.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

Abdel-Aal E-SM, Hucl P, Sosulski FW (1995). Compositional and nutritional characteristics of spring einkorn and spelt wheats. *Cereal Chem* **72**: 621–624.

Anderson EC, Dunham KK (2008). The influence of family groups on inferences made with the program Structure. *Mol Ecol Resour* **8**: 1219–1229.

Antolín F, Jacomet S, Buxó R (2015). The hard knock life. Archaeobotanical data on farming practices during the Neolithic (5400–2300 cal BC) in the NE of the Iberian Peninsula. *J Archaeol Sci* **61**: 90–104.

Badaeva ED, Keilwagen J, Knüpffer H, Waßermann L, Dedkova OS, Mitrofanova OP *et al.* (2015). Chromosomal passports provide new insights into diffusion of emmer wheat. *PLoS One* **10**: e0128556.

Bocquet-Appel JP, Naji S, Vander Linden M, Kozlowski JK (2009). Detection of diffusion and contact zones of early farming in Europe from the space-time distribution of 14C dates. *J Archaeol Sci* **36**: 807–820.

Brandolini A, Hidalgo A (2011). Einkorn (*Triticum monococcum*) flour and bread. In: Preedy VR, Watson RR, Patel VB (eds), *Flour and Breads and Their Fortification in Health and Disease Prevention*. Academic Press: London, pp 79–88.

Cortés Sánchez M, Jiménez Espejo FJ, Simón Vallejo MD, Gibaja Bao JF, Carvalho AF, Martínez-Ruiz F *et al.* (2012). The Mesolithic–Neolithic transition in southern Iberia. *Quaternary Res* **77**: 221–234.

Davison K, Dolukhanov P, Sarson GR, Shukurov A (2006). The role of waterways in the spread of the Neolithic. *J Archaeol Sci* **33**: 641–652.

Dice LR (1945). Measures of the amount of ecologic association between species. *Ecology* **26**: 297–302.

Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14**: 2611–2620.

Falush D, Stephens M, Pritchard JK (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**: 1567–1587.

FAO (2015). http://faostat.fao.org/site/567/DesktopDefault.aspx?PageID=567#ancor (last accessed 15 September 2015).

Flohr P, Fleitmann D, Matthews R, Matthews W, Black S (2015). Evidence of resilience to past climate change in Southwest Asia: Early farming communities and the 9.2 and 8.2 ka events. *Quaternary Sci Rev* **136**: 23–39.

Goncharov NP (2011). Genus Triticum L. taxonomy: the present and the future. *Plant Syst Evol* **295**: 1–11.

Haldorsen S, Akan H, Çelik B, Heun M (2011). The climate of the Younger Dryas as a boundary for einkorn domestication. *Veg Hist Archaeobot* **20**: 305–318.

Hammer Ø, Harper DAT (2006). *Paleontological Data Analysis*. Blackwell: Oxford.

Hammer Ø, Harper DAT, Ryan PD (2001). PAST: paleontological statistics software package for education and data analysis. *Palaeontol Electron* **4**: 9.

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

Heun M, Schäfer-Pregl R, Klawan D, Castagna R, Accerbi M, Borghi B *et al.* (1997). Site of einkorn wheat domestication identified by DNA fingerprinting. *Science* **278**: 1312–1314.

Heun M, Haldorsen S, Vollan K (2008). Reassessing domestication events in the Near East: einkorn and *Triticum urartu*. *Genome* **51**: 444–451.

Hidalgo A, Brandolini A (2014). Nutritional properties of einkorn wheat. *J Sci Food Agr* **94**: 601–612.

Hillis DM, Bull JJ (1993). An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst Biol* **42**: 182–192.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009). Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour* **9**: 1322–1332.

Jakobsson M, Rosenberg NA (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**: 1801–1806.

Jing HC, Bayon C, Kanyuka K, Berry S, Wenzl P, Huttner E *et al.* (2009). DArT markers: diversity analyses, genomes comparison, mapping and integration with SSR markers in *Triticum monococcum*. *BMC Genomics* **10**: 458–465.

Johnson BL, Dhaliwal HS (1976). Reproductive isolation of *Triticum boeoticum* and *Triticum urartu* and the origin of the tetraploid wheat. *Am J Bot* **63**: 1088–1094.

Jones H, Civáň P, Cockram J, Leigh FJ, Smith LMJ, Jones MK *et al.* (2011). Evolutionary history of barley cultivation in Europe revealed by genetic analysis of extant landraces. *BMC Evol Biol* **11**: 320.

Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H *et al.* (2012). Diversity arrays technology: a generic genome pro filing technology on open platforms. In: Pompanon F, Bonin A (eds), *Data Production and Analysis in Population Genomics: Methods and Protocols, Methods in Molecular Biology* vol. 888, Springer: New York, NY, USA, pp 67–89.

Kreuz A, Marinova E, Schäfer E, Wiethold J (2005). A comparison of early Neolithic crop and weed assemblages from the Linearbandkeramik and the Bulgarian Neolithic cultures: differences and similarities. *Veg Hist Archaeobot* **14**: 237–258.

Løje H, Møller B, Laustsen AM, Hansen Å (2003). Chemical composition, functional properties and sensory profiling of einkorn (*Triticum monococcum* L.). *J Cereal Sci* **37**: 231–240.

Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H *et al.* (2005). The pattern of polymorphism in Arabidopsis thaliana. *PLoS Biol* **3**: e196.

Oliveira HR, Jones H, Leigh F, Lister DL, Jones MK, Peña-Chocarro L (2011). Phylogeography of einkorn landraces in the Mediterranean basin and Central Europe: population structure and cultivation history. *Archaeol Anthropol Sci* **3**: 327–341.

Peña-Chocarro L, Zapata L (1998). Hulled wheats in Spain: history of minor cereals. In: Jaradat AA (ed) *Triticeae III. Proceedings of the Third International Triticeae Symposium, Aleppo, Syria, 4–8 May 1997*. Science Publishers, Inc.: Enfield, pp 45–52.

Peña-Chocarro L, Zapata L, Gonzáles Urquijo JE, Ibáñez JJ (2009). Einkorn (*Triticum monococcum* L) cultivation in mountain communities of the western Rif (Morocco): an ethnoarchaeological project. In: Fairnbairn AS, Weiss E (eds), *From Foragers to Farmers: Papers in Honour of Gordon C. Hillman*. Oxbow Books: Oxford, pp 103–111.

Perrino P, Laghetti G, D'Antuono LF, Al Ajlouni M, Kanbertay M, Szabó AT *et al.* (1996). Ecogeographical distribution of hulled wheat species. In: Padulosi S, Hammer K, Heller J (eds), *Hulled Wheats. Promoting the Conservation and Use of Underutilized and Neglected Crops. 4. Proceedings of the First International Workshop on Hulled Wheats, 21–22 July 1995, Castelvecchio Pascoli, Tuscany, Italy*. International Plant Genetic Resources Institute: Rome, pp 102–120.

Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

Rosenberg NA (2004). DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes* **4**: 137–138.

Saitou N, Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–425.

Schiemann E (1948). Weizen, Roggen, Gerste: Systematik Geschichte und Verwendung, Verlag Gustav Fischer, Jena, Germany.

Singh K, Ghai M, Garg M, Chhuneja P, Kaur P, Schnurbusch T *et al.* (2007). An integrated molecular linkage map of diploid wheat based on a *Triticum boeoticum* x *T. monococcum* RIL population. *Theor Appl Genet* **115**: 301–312.

Stein N, Herren G, Keller B (2001). A new DNA extraction method for high-throughput marker analysis in a large-genome species such as Triticum aestivum. *Plant Breed* **120**: 354–356.

Szabó AT, Hammer K (1996). Notes on the taxonomy of Farro: *Triticum monococcum*, *T. dicoccum*, and *T. spelta*. In: Padulosi S, Hammer K, Heller J (eds), *Hulled Wheats. Promoting the Conservation and Use of Underutilized and Neglected Crops. 4. Proceedings of the First International Workshop on Hulled Wheats, 21–22 July 1995, Castelvecchio Pascoli, Tuscany, Italy*. International Plant Genetic Resources Institute: Rome, pp 2–40.

Szécsényi-Nagy A, Brandt G, Keerl V, Jakucs J, Haak W, Möller-Rieker S *et al.* (2015). Tracing the genetic origin of Europe's first farmers reveals insights into their social organization. *P Roy Soc B Biol Sci* **282**: 20150339.

Tänzler B, Esposti RF, Vaccino P, Brandolini A, Effgen S, Heun M *et al.* (2002). A molecular linkage map of Einkorn wheat: mapping of storage-protein and soft-glume genes and bread making QTLs. *Genet Res* **80**: 131–143.

Tamura K, Dudley J, Nei M, Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596–1599.

Thaulow J, Borgstrøm R, Heun M (2013). Brown trout population structure highly affected by multiple stocking and river diversion in a high mountain national park. *Conserv Genet* **14**: 145–158.

Thaulow J, Borgstrøm R, Heun M (2014). Genetic analyses of European minnow, *Phoxinus phoxinus*, in a river system outside its native range indicate multiple invasions from different sources. *Fisheries Manag Ecol* **21**: 75–81.

Weninger B, Alram-Stern E, Bauer E, Clare L, Danzeglocke U, Jöris O *et al.* (2006). Climate forcing due to the 8200 cal yr BP event observed at Early Neolithic sites in the eastern Mediterranean. *Quaternary Res* **66**: 401–420.

Wenzl P, Carling J, Kudrna D, Jaccoud D, Huttner E, Kleinhofs A *et al.* (2004). Diversity Arrays Technology (DArT) for whole-genome profiling of barley. *Proc Natl Acad Sci USA* **101**: 9915–9920.

Zaharieva M, Monneveux P (2014). Cultivated einkorn wheat (*Triticum monococcum* L. subsp. *monococcum*): the long life of a founder crop of agriculture. *Gen Res Crop Evol* **61**: 677–706.

Zapata L, Peña-Chocarro L, Pérez-Jordá G, Stika HP (2004). Early neolithic agriculture in the Iberian Peninsula. *J World Prehist* **18**: 283–325.

Zohary D, Hopf M (2000). *Domestication of Plants in the Old World* 3rd edn. Oxford University Press.

Supplementary Information accompanies this paper on Heredity website (http://www.nature.com/hdy)