Co-evolution of methods and thoughts in cereal domestication studies: a tale of barley (Hordeum vulgare)
Artem Pankin$^{1,2,3,4}$ and Maria von Korff$^{1,2,3,4}$

Five major cereals such as wheat, rice, maize, barley and sorghum were among the first Neolithic crops that facilitated the establishment of the early agricultural societies. Since then they have remained the staple source of calories for the majority of the human population. Ample archaeological and molecular evidence has provided important insights into the domestication history of cereals but the debates on the origin of cereal crops are still far from resolved. Here, we review the recent advances in applying genome sequencing technologies for deciphering the history of cereal domestication. As a model example, we demonstrate that the evolution of thoughts on barley domestication closely followed the development of views on the rise of agriculture in the Near East in general and greatly accelerated with the advent of the genomic technologies and resources available for barley research.

Addresses
$^1$ Institute of Plant Genetics, Heinrich-Heine-University, 40225 Düsseldorf, Germany
$^2$ Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany
$^3$ Cluster of Excellence on Plant Sciences “From Complex Traits towards Synthetic Modules”, 40225 Düsseldorf, Germany

Corresponding authors: Pankin, Artem (pankin@mpipz.mpg.de) and von Korff, Maria (korff@mpipz.mpg.de)

Postal address: Max Planck Institute for Plant Breeding Research, Carl-von-Linne weg 10, 50829 Cologne, Germany.

Introduction
Cereals have a long history of interaction with humans. At around 40–50 thousand years ago (kA), late Neanderthals consumed wild Triticum grains in a cooked form as suggested by the analysis of the dental calculus from multiple locations in Europe, the Near East, and Africa [1]. The first fossil remnants of plants exhibiting characteristics of the domesticated cereal crops have been discovered at the multiple archaeological sites dated to 8–10 kA — the historic period marking a transition from the hunter-gatherer lifestyle to the early sedentary agricultural societies. The most prominent examples are the non-brittle spikes of the proto-domesticated wheat (diploid, tetraploid and hexaploid Triticum species) and barley (Hordeum vulgare ssp. vulgare) discovered in the excavations in the Fertile Crescent area of the Near East; the non-shattering panicles of rice (Oryza sativa) from the Yangtze river valley, China; and the starchy enlarged seeds of maize (Zea mays ssp. mays) from the Central Balsas river valley, Mexico, which distinguish maize from its wild progenitor teosinte (Z. mays ssp. parviglumis) [2–6]. The archaeobotanical analyses of the fossil findings provided direct empirical evidence on the temporal scale and the location of the early events of cereal domestication history, which was packaged in the so-called domestication models [2,3,7]. However, to expand on the archaeological models, detailed understanding of genetic relationships between the cultivated and wild populations is crucial. In the pre-genomics era, the pioneering molecular studies suggested explicit demographic models of cereal domestication based on the population genetic analyses of nucleotide variation in isolated genes and genomic regions [8–12]. The molecular evolutionary models favored a monophyletic origin of crops and strong selective pressures under domestication. On the contrary, the archaeological models assumed a polycentric origin of domestication traits and a millennia-long process of fixation of these traits in the cultivated populations.

Advancements in the genome sequencing technologies enabled the elucidation of the genetic ancestry and evolutionary history of domesticated genotypes at the unprecedented resolution and scale [13,14*,15**]. It was anticipated that the genome-wide diversity estimates would provide a solid basis to the demographic history reconstruction. However, for many cereals, the interpretation of the revealed patterns remained controversial hinting at the complexity of their evolutionary history [15**,16*]. Population genomics has recently revealed intriguing patterns in the cultivated barley genomes that shed light on the trajectory of barley domestication [17**,18,19**]. We use barley as a model to demonstrate
how the development of genomic technologies gradually changed our views on the circumstances of domestication of the Neolithic Near Eastern crop.

**Genome resequencing in cereal domestication studies**

Genomes of cereals greatly vary in size. Asian rice and its sister African species (*Oryza glaberrima*) possess the smallest of the cereal genomes of ~0.4 gigabase pairs (Gbp) followed by the sorghum, maize and barley genomes of ~0.73 Gbp, 2.4 Gbp and 5.1 Gbp, respectively [20]. The hexaploid bread wheat (*Triticum aestivum*) resides on the other end of the spectrum with the enormous genome of ~17 Gbp [21]. For these crops and their wild ancestors, the genome sequences and the physical maps of varying degree of completion have become available [22–29]. The history and the current status of the cereal genomics have been discussed in greater details elsewhere [20].

In the genomics era, unraveling domestication history typically involves comparison of various facets of genomic variation in the large representative samples of the domesticated and wild progenitor species. Using phylogenetic and population genetic analyses, the ancestry of the domesticated genomes or specific domestication loci can be traced back to the predefined wild populations and geographic locations. This, combined with the archaeological findings, helps to hypothesize a specific scenario of when and where a crop originated and how it further evolved also known as a domestication model. The robustness of such inferences positively correlates with the representativeness of the investigated plant accessions, the accuracy of genotyping and allele frequency estimates and the density of interrogated polymorphisms. Owing to the rapidly sinking costs of high-throughput sequencing, whole-genome resequencing has become common in crop domestication studies since it provides reliable genotypic information at virtually every nucleotide position (except for the highly repetitive regions). In the domestication context, genome resequencing studies have been reported in maize, sorghum, and African and Asian rice; the latter being thus far the largest survey of ~8 million single nucleotide polymorphisms (SNP) in >1500 wild and domesticated Asian rice genotypes (Table 1) [13,14*,15**,30]. For the large barley and wheat genomes consisting of up to 84% of the mobile elements and other repeat structures, genome resequencing is not yet feasible. To circumvent this limitation, the barley exome sequencing assay, capturing variation only at the coding regions, was developed and used to resequence 91 wild and 137 domesticated barley genotypes [31,32*]. A larger collection of 433 wild and domesticated accessions has been resequenced using a custom enrichment assay interrogating a panel of ~500 000 SNPs [18]. The wheat exome capture assays are also available but have not yet been used to systematically investigate the phylogeographic patterns of wheat domestication [33,34].

**Barley domestication models**

Wild barley (*H. vulgare* ssp. *spontaneum*) is an immediate ancestor of the domesticated subspecies. Following domestication, cultivated barley spread from the Near Eastern region and became adapted to various ecogeographic conditions—it thrives even beyond the Arctic Circle. Despite adaptations to diverse cultivation environments and intensive breeding for various end-use qualities, the morphology of cultivated barley plants did not dramatically

### Table 1

<table>
<thead>
<tr>
<th>Species</th>
<th>Wild ancestor</th>
<th>Ploidy, 2n</th>
<th>Number of chromosomes, 1n</th>
<th>Type of assay</th>
<th>Size, Mb</th>
<th>Number of SNPs</th>
<th>Number of genotypes (domesticated/wild)</th>
<th>Domestication bottleneck</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barley</td>
<td><em>Hordeum</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>vulgare</em></td>
<td>2x</td>
<td>7</td>
<td>Genome</td>
<td>4980</td>
<td>&gt;15 M</td>
<td>5/1</td>
<td>n.d.</td>
<td>[22]</td>
</tr>
<tr>
<td></td>
<td><em>spontaneum</em></td>
<td></td>
<td></td>
<td>Exome</td>
<td>60 b</td>
<td>1.7 M</td>
<td>137/91</td>
<td>0.73</td>
<td>[32*]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Exome subset</td>
<td>13.8 c</td>
<td>0.54 M</td>
<td>89/344</td>
<td>0.52</td>
<td>[18]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Genome</td>
<td>373</td>
<td>8 M</td>
<td>1083/446</td>
<td>0.80</td>
<td>[15**]</td>
</tr>
<tr>
<td>Asian rice</td>
<td><em>Oryza</em></td>
<td>2x</td>
<td>12</td>
<td>Genome</td>
<td>316</td>
<td>2.3 M</td>
<td>93/3</td>
<td>n.d.</td>
<td>[14]</td>
</tr>
<tr>
<td></td>
<td><em>sativa</em></td>
<td></td>
<td></td>
<td>Exome</td>
<td>84 c</td>
<td>8.4 M</td>
<td>20/94</td>
<td>0.43</td>
<td>[25]</td>
</tr>
<tr>
<td>African rice</td>
<td><em>O. glaberrima</em></td>
<td>2x</td>
<td>12</td>
<td>Genome</td>
<td>626</td>
<td>4.9 M</td>
<td>35/7</td>
<td>0.62</td>
<td>[30]</td>
</tr>
<tr>
<td>Sorghum</td>
<td><em>Sorghum</em></td>
<td>2x</td>
<td>10</td>
<td>Genome</td>
<td>2300</td>
<td>21.1 M</td>
<td>58/14</td>
<td>0.83</td>
<td>[13]</td>
</tr>
<tr>
<td></td>
<td><em>bicolor</em></td>
<td></td>
<td></td>
<td>Exome</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>bicolor ssp. bicolor</em></td>
<td></td>
<td></td>
<td>Genome</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Zea mays ssp. parviglumis</em></td>
<td></td>
<td></td>
<td>Genome</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a Size of the physical map.

b >10 read coverage in 95% of accessions.

c >6 read coverage in at least 2 accessions.

d M, millions.

e Fraction of nucleotide diversity retained by the domesticated species compared to their wild ancestors.
Box 1 Barley domestication syndrome.

A domestication syndrome is a suite of morphological and physiological traits that differentiate cultivated from wild forms [73]. In barley, spike brittleness remains the only known trait that reliably distinguishes wild and cultivated plants. The brittle wild-type spikes facilitate dissemination of the seeds, whereas the tough rachises of cultivated barley, which prevent spontaneous disarticulation of the mature spikes into spikelets, are an adaptive innovation ensuring efficient harvesting [63]. The other morphological characters such as plant height, shape and size of seeds, width of spikes and flag leaves, number and angle of side shoots, and synchrony of flowering and germination have been mentioned as the candidate traits of cereal domestication syndrome [45,74]. However, variation of these traits in wild and cultivated barley awaits a statistical assessment. Such traits as reduced vernalization and photoperiod sensitivity, fertility of lateral spikes and hulless caryopses are specific only to certain groups of cultivated barley and therefore apparently represent the improvement but not domestication traits [73,75].

diverge from the wild form. Spike brittleness, as a diagnostic trait of the barley domestication syndrome, has been used by archaeologists to date the early onset of barley domestication [35]. (Box 1) It is controlled by two adjacent genetic loci Btr1 and 2 (also known as Btr1 and 2) and the non-brittle spikes are produced by the recessive alleles of either of these genes [36].

Archaeobotanical studies suggested that harvesting of barley plants with wild-type brittle spikes occurred already during the Upper Paleolithic Period in the Rift Valley of modern Israel at the site radiocarbon dated to ~23 kA [37,38]. The first non-brittle barley spikes were discovered in the Fertile Crescent in excavation sites dated to ~10 kA [35]. The Fertile Crescent as a primary habitat comprises most of the modern wild barley diversity [39]. Although the modern range of its documented occurrence spans the whole region between North Africa, Western Anatolia and the Tibetan Plateau. At least nine wild barley populations have been identified in the Fertile Crescent and two additional populations across the Zagros mountains [18,32*,40,41] (Figure 1).

It has been largely accepted that the major events of barley domestication unfolded in the Fertile Crescent and the adjacent areas. However, other regions such as the Horn of Africa, Morocco and the Tibetan plateau were discussed in the literature as alternative centers of domestication [42–44]. One of the first studies that used DNA marker technology to systematically investigate the origin of barley using large samples of wild (317) and domesticated (57) accessions was based on the amplified fragment length polymorphism (AFLP) markers, which are anonymous genomic fragments scored as a binary-coded presence/absence pattern [45]. Phylogenetic clustering of 400 AFLP alleles discovered that barley cultivars and landraces from various geographic locations formed a single monophyletic group. That seemed to refute the earlier hypotheses of multiple barley domestication centers. Wild barley populations collected from the area of modern Israel and Jordan appeared genetically closer to the cultivated gene pool and were therefore deemed as the likely primary ancestors. The distributions of several AFLP alleles in cultivated barley closely resembled those found in wild barley from Western Iran, which was considered as the evidence of a secondary gene flow between wild and cultivated germplasm. These findings aligned well with the then-prevailing idea that the domestication of the Near Eastern cereals has been a rapid process confined to the defined core regions within the Fertile Crescent [46]. Thus, the model of a monophyletic barley origin has long persisted in the scientific community.

A competing idea of the diffused non-centric origin of crops — already expressed by the early botanists critically revising the concept of the Vavilovian ‘centers of origin’ — was later formalized in the protracted model of domestication [47–50]. This model was substantiated by a large corpus of the archaeobotanical evidence and a statistical framework [7,51–53]. Regarding the Near Eastern crops, it postulates that the fixation of domestication traits in cultivated populations was a slow process and their origin was not restricted to a certain geographic center. The concepts and the statistical methods of the protracted model were heavily criticized [54–58].

The molecular evidence indicating that cultivated barley could descend from several distinct lineages came from the population structure analysis that split 25 wild barley genotypes into two populations — ‘eastern’ and ‘western’ — separated by the Zagros Mountain range based on the genotype data of 684 SNPs from 18 genes [59]. It revealed that the ‘western’ and ‘eastern’ wild barley populations proportionately contributed to the ancestry of cultivated barley genotypes. The ancestry assignments bisected the cultivation areas of the Near Eastern and Asian barley landraces into the western and eastern clusters relative to the Zagros Mountains [59,60*]. This model fit well into the long-established separation of cultivated barley into the ‘occidental’ (btr1btr1Btr2btr2) and ‘oriental’ (Btr1Btr1btr2btr2) types based on the Btr allelic status and also found support in later studies [61,62].

Cloning of the genes underlying the Btr1 and Btr2 loci identified nucleotide deletions in their coding regions, which truncated open reading frames and rendered the non-brittle phenotype [63**]. The mutant btr alleles were independent and private to cultivated barley. A geographic distribution of their frequencies followed the occidental-oriental model. However, the btr2 mutation did not seem to originate in the proposed eastern cluster of barley domestication. Geographic locations of the putative ancestral wild barley haplotypes pointed to the origin of both mutations in the western horn of the Fertile Crescent. Apparently, the btr2 was spread from its putative place of origin in the Northern Levant to the east of
Western Asia with the migration of the early farmers and subsequently enriched in the eastern part of barley cultivation range. Admixture patterns in the ancient human DNA suggested the gene flow between Neolithic Levantine and Zagros highland farmers hinting at the existence of such migration route [64**].

In contrast with the analyses based on the sparsely distributed DNA markers, the genome-wide analyses using genotypes of thousands of markers densely covering all chromosomes revealed intricate patterns of genetic ancestry. The first detailed genome-wide ancestry scan in barley was based on a set of 6152 SNPs screened in 803 diverse landraces using Illumina genotyping platforms. It revealed that the modern cultivated barley genomes consist of a mosaic of fragments that descended from wild barley populations from the both sides of the Zagros Mountains [17**]. (Figure 1) Another genome-wide analysis based on ~500 000 SNPs, which originated from targeted resequencing of 433 wild and cultivated barley accessions, arrived at a similar conclusion [18]. The contribution of wild populations to the different landraces was not uniform; the landrace populations were enriched for genetic material originating from the proximal wild populations [17**]. Nevertheless, the mosaic patterns of different cultivated barley populations and individual genotypes were strikingly similar; the wild population from the Southern Levant contributed > that 50% of the genetic material in all landraces [17**,18]. This raises a hypothesis that a single admixed lineage might have been at the root of all examined domesticated genotypes thus contesting the model of the independent Eastern and
Western domestication lineages. Intriguingly, another study discovered that the 6000-year-old cultivated barley excavated at Yoram Cave in Israel was not genetically distinct from the modern-day landraces based on the genome and exome sequencing data, which tentatively suggests a similar ancestral composition [18**]. This corroborates the hypothesis that the admixed progenitor lineage existed already at the early stages of barley cultivation. Alternatively, similar mosaic admixture patterns could occur independently in the western and eastern domestication clusters but this solution is less parsimonious and unlikely.

Various demographic scenarios — both related and unrelated to domestication process — may have resulted in the mosaic patterns observed in the cultivated barley genomes (Figure 1). One possibility is that the hypothetical wild progenitor population could have had a highly admixed ancestry that was passed down to the cultivated lineage. The admixed wild barley genotypes frequently occur at the contact zones between the modern wild populations [18,32**,41]. The second hypothesis is that the wild progenitor lineage was not admixed and the recurrent gene flow from wild into the (proto-)domesticated populations happened during the transition to cultivation gradually creating the heterogeneous admixture patterns. The third and perhaps the likeliest scenario is a combination of the ancestral population structure and the gene flow. Statistical testing of these and other more complex demographic models will help determine the best fit to the existing empirical data and thus suggest the likeliest demographic history of barley domestication [65].

In other crops, the mosaic ancestry of the domesticated genomes has not been reported. The mosaic ancestry model poses intriguing questions for future studies: to which extent the mosaic ancestry is relevant for domestication versus the neutral gene flow or further adaptations; and whether the other Fertile Crescent crops carry similar patterns in their genomes, which would indicate parallel trajectories of their domestication.

Concluding remarks and future perspectives

The state-of-the-art genomic resources enabled surveys of genetic diversity in many cereal crops and its wild relatives at the ultimate resolution probing variation at millions of nucleotide positions. Using barley as an example, we demonstrated that with the arrival of more detailed and dense genomic datasets the increasingly complex and reticulate models of domestication history emerged. A similar trend is observed in some other cereal crops [15**,16].

Genome studies provided compelling evidence that hybridization and introgression played a primary role in the cereal domestication and further adaptation. A highly admixed ancestral population(s) seemed to be a progenitor of modern barley domesticates. In rice, japonica subspecies contributed to the domestication of indica through introgression [15**]. The maize landraces from Mexico comprised introgressions from conspecific wild relative Z. mays ssp. mexicana, which putatively led to the adaptation of maize to the highland environments [60,67]. In emmer wheat (Triticum turgidum), phylogenetic evidence suggested reticulate domestication history [68].

In the domestication models that implicate long-distance migration and introgression events, distinguishing single and multiple origins of crops becomes a non-trivial task since the genealogies of the domestication loci and the rest of the genome may not always match. Arguably, it may remain a semantic preference until a series of events that gave rise to the cultivated forms is understood in much greater detail. To this end, it is critical that, in future studies, we engage in discovering and characterizing additional domestication genes, which contributed to the emergence of the cultivated forms. In cereals, genome scans identified multiple genomic regions putatively affected by selection under domestication, which comprised hundreds of candidate genes [13,15**,18]. Further exploring these regions to identify novel domestication genes and specific mutations targeted by selection followed by their phenotypic characterization will greatly promote our understanding of domestication syndrome traits and their genetic basis [63**,69]. Reconstructing genealogies of the selected mutations will illuminate the history of the origin of individual domestication traits thus providing a more targeted approach to modeling the domestication process. Furthermore, sequencing of ancient cereal DNA presents another promising avenue for untangling the complex domestication histories [19**,70–72]. The paleogenomics data combined with the archaeobotanical findings will inform the domestication models with a more precise empirical evidence on the location and timing of events that led to the assembly of the domestication syndrome.

Acknowledgements

We apologize to colleagues whose work could not be included in this review because of the space constraints. This work was supported by the Max Planck Society and by Deutsche Forschungsgemeinschaft (DFG) grants SPP1530 (‘Flowering time control: from natural variation to crop improvement’) and the Excellence Cluster EXC1028. A.P. received a fellowship from the Max Planck Society. We thank the anonymous referee for the constructive criticism of the earlier version of the manuscript.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as: * of special interest ** of outstanding interest


16. The model suggesting introgression of domestication genes from japonica to indica, which facilitated indica domestication, based on the genomics sequencing of a large panel of wild and cultivated rice accessions.


20. The first evidence that landrace barley genotypes consist of distinct blocks that descended from wild barley populations from various parts of the Fertile Crescent. The reticulate history of landrace barley genotypes is unraveled using a combination of the illumina genotyping assays.


23. Genome sequencing of the ancient barley DNA isolated from the landrace accession from the Yarmouch Cave, Israel suggests that the ancient and modern-day landraces are not genetically distinct.


36. Exome-based comparison of the nucleotide diversity in wild and domesticated barley.


