Wild emmer genome architecture and diversity elucidate wheat evolution and domestication

Raz Avni,1* Moran Nave,1 Omer Barad,2 Kobi Baruch,2 Sven O. Twvardziok,3 Heidrun Gundlach,3 Iago Hale,4 Martin Mascher,5.4 Manuel Spannagl,3 Krystalee Wiebe,5 Katherine W. Jordan,6 Guy Golan,7 Jasline Deek,1 Batsheva Ben-Zvi,1 Gil Ben-Zvi,5 Axel Himmelbach,5 Ron P. MacLachlan,7 Andrew G. Sharpe,7,10 Allan Fritz,8 Roi Ben-David,11 Hikmet Budak,12 Tzion Fahlina,13 Abraham Korol,13 Justin D. Faris14 Alvaro Hernandez,15 Mark A. Mikel,15 Avraham A. Levy,16 Brian Steffenson,17 Marco Macecafferi,18 Roberto Tuberosa,18 Luigi Cattivelli,19 Primetta Faccioli,19 Aldo Ceriotti,20 Khalil Kashkush,21 Mohammad Pourheirandish,22a Takao Komatsuda,22 Tamar Ilam,23 Hanan Sela,23 Amir Sharon,1,23 Nir Ohad,1 Daniel A. Chamovitz,1 Klaus F.X. Mayer,3,24 Nils Stein,5 Gil Ronen,2 Zvi Peleg,9 Curtis J. Pozniak,7 Eduard D. Akhunov,8 Assaf Distelfeld1,22†

Wheat (Triticum spp.) is one of the founder crops that likely drove the Neolithic transition to sedentary agrarian societies in the Fertile Crescent more than 10,000 years ago. Identifying genetic modifications underlying wheat’s domestication requires knowledge about the genome of its allo-tetraploid progenitor, wild emmer (T. turgidum ssp. dicoccoides). We report a 10.1-gigabase assembly of the 14 chromosomes of wild tetraploid wheat, as well as analyses of gene content, genome architecture, and genetic diversity. With this fully assembled polyploid wheat genome, we identified the causal mutations in Brittle Rachis 1 (TcBra1) genes controlling shattering, a key domestication trait. A study of genomic diversity among wild and domesticated accessions revealed genomic regions bearing the signature of selection under domestication. This reference assembly will serve as a resource for accelerating the genome-assisted improvement of modern wheat varieties.

---

Wheat (Triticum spp.) is one of the Neo- lithic founder crop plants, domesticated ~10,000 years ago in the Near-Eastern Fertile Crescent (1). The initial domestica- tion of the allo-tetraploid wild emmer wheat [WEW; T. turgidum ssp. dicoccoides (Körn.) Thell.; genome BBAA] and the subsequent evolution of hulled domesticated emmer wheat (DEW; T. turgidum ssp. dicoccum Schr.) led to the selec- tion of free-threshing durum wheat [T. turgidum ssp. durum (Desf. MacKey)] (2). Subsequently, hexaploid bread wheat (T. aestivum L., genome BBADD) arose from the hybridization of domest- icated emmer with the diploid Aegilops tauschii Coss. (genome DD) (3), indicating that WEW was the direct progenitor of all economically impor- tant domesticated wheats.

Crop domestication is generally marked by rapid modification of key traits (i.e., the “domestication syndrome”) followed by subsequent evolution (4) introducing further, incremental changes in morpho-physiological traits (5–7). The domestication of WEW involved traits related primarily to seed dormancy, spike morphology, and grain development. For example, whereas the spikes of WEW shatter at maturity (trait: brittle rachis (BR)), all domesticated wheat spikes remain intact (i.e., “non-BR spikes”), enabling easier harvest. Characterizing the genetic mechanisms underlying wheat domestication and improve- ment may provide insights into basic wheat biology and human cultural evolution (8) and offer oppor- tunities for further improvement (e.g., (9)). To date, few genes involved in wild emmer domest- ication have been discovered, mainly due to lack of a comprehensive reference genome. The wheat reference genome has been difficult to obtain owing to its polyploid nature, large size (estimated ~12 Gb for tetraploid wheat and ~17 Gb for hexa- ploid wheat), and high repeat content (~80%) (10). Although the gene space of wheat has been described, most intergenic regions remain frag- mented and unassigned to chromosomal loca- tions (e.g., (10)).

Wild emmer accession “Zavitan” was chosen for this genome assembly to leverage the genetic data already collected for this line (7, 11). The WEW reference genome, constructed by whole-genome shotgun (WGS) sequencing of various insert-size libraries (12), produced contigs with an N50 of 57,376 base pairs (bp) and scaffolds with an N50 of 6,955,168 bp (table S1) (13). The scaffolds were validated with genetic data and combined with three-dimensional (3D) chromosome conformation capture sequencing (HiC) data (14), enabling construction of chromosome-scale assemblies (pseudomolecules) (Fig. 1, A and B). The resulting 10.5-Gb genome assembly is composed of 14 pseudomolecule sequences representing the 14 chromosomes of WEW (10.1 Gb) and one group of unassigned scaffolds (0.4 Gb). The gaps between scaffolds, estimated to represent ~1.5 Gb of the genome (13), are likely the result of techni- cally difficult-to-sequence or difficult-to-assemble regions (Fig. 1, C and D). A strength of this assem- bly strategy is its ability, despite these gaps, to construct pseudomolecules with correct scaffold orienta- tions, thereby enabling whole-genome analyses in polyploid wheat with high physical resolution.

In addition to gene models from other grasses, RNA sequencing reads generated from 20 different combinations of WEW tissues and developmental stages were used to annotate protein-coding genes in the WEW assembly (13). We identified 65,012 high-confidence (HC) gene models (fig. S1 and tables S2 and S3), and validation with the BUSCO (15) gene set (table S4) indicated that the assembly captures 98.4% of the WEW gene complement.

The density of HC genes was up to 14-fold higher in the distal compared to the pericentromeric regions of chromosome arms, confirming the reported gene density gradient along the centromere- telomere axis (10) (Fig. 2, A and B). Of the 62,813 genes assigned to chromosomes, 30.4% were ex- pressed under all 20 conditions (i.e., tissue types + time points), 48% were expressed in at least one condition but not under all conditions, and 21.6% were expressed at a low level or not at all. Both the mean expression level and expression breadth (i.e., the number of conditions under which gene expression was detected) per gene were higher for genes in the proximal regions of chromosome arms (Fig. 2, C and D, and fig. S2). The mean lower expression levels in the distal regions may be due to the higher proportion of condition-specific genes expressed at a low level in these chromosomal regions, whereas in the proximal regions, a greater proportion of highly
Avni et al., Science 357, 93–97 (2017) 7 July 2017

expressed housekeeping genes can be found. The total numbers of genes expressed per subgenome were roughly equivalent (Fig. S3A), and only a slight but consistent higher expression ratio of genes from the A subgenome was observed (a 5% increase, on average; fig. S3B). This supports previous findings of a lack of genome-wide transcriptional dominance of individual subgenomes in bread wheat (10).

Homology analysis between the two WEW subgenomes revealed distinct homeologous pairs for most genes (45,386, 72.3%), another 9123 (14.5%) consisting of potentially duplicated genes within one of the subgenomes, and an unclassified remainder (e.g., singleton genes) (Fig. 2E, table S5, and database S1). Expression of homeologous gene pairs revealed a strong positive correlation across all samples (fig. S4). Both homologs were expressed at least once in 18,306 pairs (36,612 genes), but expression of 2188 pairs was not detected in any sample. A portion of the homeologous gene pairs was exclusively expressed in only one of the two subgenomes, with expression of only the A subgenome gene for 1084 pairs and of only the B subgenome gene for 1115 pairs. These genes were functionally enriched for protein phosphorylation and general metabolic processes (table S6), suggesting that subgenome-regulated gene expression may contribute to adaptation of wheat varieties.

Most (82.2%) of the WEW assembly was annotated as transposable elements (TEs), and overall TE content appears to be similar for the two subgenomes (Fig. 2F, fig. S5, and table S7). The majority (7.1 Gb) of TEs derive from long-terminal repeat–retrotransposons (LTR-RTs) that can be divided into three groups: Ty1/copia-like, Ty3/gypsy-like, and unclassified (table S7). Mutations within full-length (FL) LTR-RTs can serve as a clock to track their activity in the genome over evolutionary time (16). In WEW, the age distribution of all FL-LTR-RT insertions suggests the occurrence of a general amplification wave around 1.5 million years ago (Ma), with the overall age distribution patterns of Ty3/gypsy-like and unclassified LTR-RT insertions appearing similar in both the A and B subgenomes (fig. S6 and database S2). The amplification wave of Ty1/copia-like elements, however, occurred more recently, around 0.5 Ma, which corresponds to the estimated time of hybridization of the A and B subgenomes (10). The B subgenome also exhibits an older Ty1/copia-like maximum dating roughly 1.2 Ma, indicating a subgenome-specific amplification wave prior to hybridization (Fig. 2G).

To illustrate the potential of this assembly in the genetic dissection of agronomically important traits, we targeted the domestication trait of nonshattering spike using a population derived from a cross between Zavitan and domesticated durum wheat (cultivar “Svevo”) (11). This analysis revealed genomic regions regulating the BR phenotype, including two major loci on WEW chromosomes 3A and 3B (15.5 and 32.5 Mb, respectively) containing homologies to the Btr1 and Btr2 genes controlling BR in barley (17). Among a complex of gene duplication clusters (three or four copies for each gene in the A and B subgenomes), we identified the orthologous wheat genes (chromosome-3A: TtBtr1-A and TtBtr2-A; chromosome-3B: TtBtr1-B and TtBtr2-B) (Fig. 3, fig. S7, and table S8) (19).

The domesticated (i.e., Svevo) TtBtr1-A and TtBtr2-B alleles carry mutations predicted to disrupt the structures of the encoded proteins and are likely loss-of-function alleles (Fig. 3A). Notably, no polymorphisms were detected between the coding regions of the Zavitan and Svevo TtBtr2-A or TtBtr2-B alleles, leading us to postulate that the combination of the mutations in the two TtBtr1 genes could be complementary to achieve the non-BR domesticated phenotype. We thus developed a pair of near-isogenic lines (NILs), each carrying one functional allele (TtBtr1-A or TtBtr1-B) in the background of Svevo. Both NILs exhibited an intermediate BR phenotype, in which the upper part of the spike was brittle and the lower part was nonbrittle. Scanning electron microscopy (SEM) confirmed a smooth abscission site typical of WEW spikelets in the scar tissues of...
spikelets from the upper rachises of both NILs (Fig. 3B), whereas the lower nonbrittle spikelets had rough abscission sites similar to those of Svevo.

Thus, these two homozygous recessive mutations in orthologs of the Btr1 gene (but not in TtBtr2) appear to be minimally required for transforming the BR morphology of WEW. Diversity analysis of 113 WEW, 85 DEW, and 9 durum accesses showed that all domesticated (DEW and durum) accesses carry the loss-of-function alleles for both TtBtr1 genes (TtBtr1-A and TtBtr1-B) (Table S8). Relative to the fixation of the non-BR phenotype in barley, requiring only one mutation in either the Btr1 or Btr2 genes (17), we speculate that selection for the non-BR phenotype in wheat may have been more gradual, as it requires at least two homozygous recessive mutations. This is supported by the archaeological records, which suggest that spike indehiscence took several millennia to become established (18).

To detect additional regions of the WEW genome under domestication selection, we examined DNA variation by exome capture sequencing (19) in a set of wild and domesticated emmer wheat accesses (fig. S8 and Table S10). The phylogenetic tree of these genotypes (Fig. 4A) shows a clear separation between the domesticated and wild wheat accesses. WEW accesses from Southern (Israel, Syria, and Lebanon) and Northern (Turkey) Levant clustered separately, whereas the DEW accesses were clustered to regions of the Indian Ocean, Mediterranean, Eastern Europe, and Caucasus. Turkish accesses were present in each of main DEW groups, which is consistent with the diffusion of domesticated accesses from a single site of origin (20). Notably, the closest domesticated accession to WEW was from Turkey, the putative location of wheat domestication (20).

Overall, only minor loss of genetic diversity is seen among DEW genotypes (mean nucleotide diversity in domesticated emmer, \( \pi_N = 1.1 \times 10^{-3} \)) as compared to their wild ancestors (mean nucleotide diversity in wild emmer, \( \pi_W = 1.3 \times 10^{-3} \)), consistent with observations in maize (21) and barley (22) (database S3). Similar to maize (23), the transition from wild to the domesticated form of tetraploid wheat was also accompanied by a shift in allele frequencies toward more common alleles, likely due to recent population bottlenecks (fig. S9). We investigated the patterns of genetic diversity (\( \pi_N/\pi_W \) diversity ratio), site frequency spectra (Tajima’s D), and genetic differentiation (\( F_{ST} \)) across the genome (Fig. S10).

Consistent with the expected effects of selection on genetic diversity patterns, the regions harboring TtBtr1-A and TtBtr1-B genes were outliers in at least one of the three diversity scans (database S4). The TtBtr1-A region also showed a low Tajima’s D value (\( -0.82; <5th \text{ percentile} \)) within the population of DEW (Fig. 4B), a result often associated with a domestication bottleneck (24).

Depending on the diversity scan, between 32 and 154 genomic regions, spanning 0.6% (68 Mb) to 3.1% (372 Mb) of the WEW genome, emerge as regions potentially affected by selection. Regions of the wheat genome detected in all three selection scans were significantly enriched (>95th percentile)
for nonsynonymous single-nucleotide polymorphisms (SNPs) (Fig. 4C), indicating that domestication preferentially enriched variants with possible functional effects in coding regions. A Gene Ontology (GO) enrichment analysis of the genes in the regions identified in the diversity scans (table S11) included genes involved in response to auxin stimulus (GO: 0009733), consistent with the evidence of strong selective sweeps on auxin-responsive genes in both maize and rice (25). Thus, as with the non-BR phenotype in barley and wheat, convergent evolution has likely played a role in crop domestication.

In the face of global challenges, agricultural research and plant breeding will be essential to increase crop yields. The availability of enhanced genomic resources like the WEW assembly, capturing both genic and intergenic regions, will underpin and accelerate global efforts in gene discovery, functional characterization, and breeding.

REFERENCES AND NOTES
13. See supplementary materials.

ACKNOWLEDGMENTS
Supported by Israel Science Foundation (grants 999/12, 1824/15), the German-Israeli Foundation for Scientific Research and Development (GIF) grant I-1221-315.13. Integrated DNA Technologies, Inc. (IDT), the United States-Israel Binational Science Foundation (BSF grants 2013396 and 2014049), USDA NIFA (National Institute of Food and Agriculture) (grant 2016-67013-24473), German Ministry of Education and Research (grants 031A536, 0315954, 0314000, 031A536), Genome Canada and Genome Prairie, Saskatchewen Ministry of Agriculture and Western Grains Research Foundation (CTAG2 project), U.S. Agency for International Development Middle East Research and Cooperation (grant M34-037), Italian Ministry of Education and Research Flagship InterOmics (PR05), and CREA-interomics projects. Additional support was funded from the Manna Center Program for Food Safety and Security in Tel Aviv University, Montana Plant Sciences endowment, the Lieberman-Okinow endowment at the University of Minnesota, and the Charles and Tillie Lubin Flagship InterOmics (PB05), and CREA-interomics projects.

SUPPLEMENTARY MATERIALS
www.sciencemag.org/content/357/6346/93/suppl/DC1
Materials and Methods
Figs. S1 to S14 Tables S1 to S7
References (26–59)
Databases S1 to S7
21 February 2017, accepted 8 June 2017
10.1126/science.aan0032
Wild emmer genome architecture and diversity elucidate wheat evolution and domestication


Science 357 (6346), 93-97.
DOI: 10.1126/science.aan0032

Genomics and domestication of wheat
Modern wheat, which underlies the diet of many across the globe, has a long history of selection and crosses among different species. Avni et al. used the Hi-C method of genome confirmation capture to assemble and annotate the wild allotetraploid wheat (Triticum turgidum). They then identified the putative causal mutations in genes controlling shattering (a key domestication trait among cereal crops). They also performed an exome capture-based analysis of domestication among wild and domesticated genotypes of emmer wheat. The findings present a compelling overview of the emmer wheat genome and its usefulness in an agricultural context for understanding traits in modern bread wheat.

Science, this issue p. 93

ARTICLE TOOLS
http://science.sciencemag.org/content/357/6346/93

SUPPLEMENTARY MATERIALS
http://science.sciencemag.org/content/suppl/2017/07/06/357.6346.93.DC1

REFERENCES
This article cites 55 articles, 27 of which you can access for free
http://science.sciencemag.org/content/357/6346/93#BIBL

PERMISSIONS
http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service