New insights into the evolution of wheat avenin-like proteins in wild emmer wheat (*Triticum dicoccoides*)

Yujuan Zhang[a], Xin Hu[a,b], Shahidul Islam[a], Maoyun She[a], Yanchun Peng[a,b], Zitong Yu[a], Steve Wylie[a], Angela Juhasz[a], Mirza Dowla[a], Rongchang Yang[a], Jingjuan Zhang[a], Xiaolong Wang[a], Bernard Dell[a], Xueyan Chen[a,c], Eviatar Nevo[d,1], Dongfa Sun[b,1], and Wujun Ma[a,1]

[a]Australia–China Joint Centre for Wheat Improvement, Western Australian State Agriculture Biotechnology Centre, School of Veterinary and Life Sciences, Murdoch University, Perth, WA 6150, Australia; [b]College of Plant Science and Technology, Huazhong Agriculture University, Wuhan, China; [c]Crop Research Institute, Shandong Academy of Agricultural Sciences, Jinan, China; and [d]Institute of Evolution, University of Haifa, Mount Carmel, 3498838 Haifa, Israel

Contributed by Eviatar Nevo, October 22, 2018 (sent for review August 2, 2018; reviewed by Yong Q. Gu and Steven Xu)

Fifteen full-length wheat grain avenin-like protein coding genes (*TaALP*) were identified on chromosome arms 7AS, 4AL, and 7DS of bread wheat with each containing five genes. Besides the a- and b-type ALPs, a c type was identified in the current paper. Both a and b types have two subunits, named x and y types. The five genes on each of the three chromosome arms consisted of two x-type genes, two y-type genes, and one c-type gene. The a-type genes were typically of 520 bp in length, whereas the b types were of 850 bp in length, and the c type was of 470 bp in length. The *ALP* gene transcript levels were significantly up-regulated in *Blumeria graminis* f. sp. *tritici (Bgt)*-infected wheat grain caryopsis at early grain filling. Wild emmer wheat [(WEW), *Triticum dicoccoides*] populations were focused on in our paper to identify allelic variations of *ALP* genes and to study the influence of natural selection on certain alleles. Consequently, 25 alleles were identified for *TdALP-bx-7AS*, 13 alleles were identified for *TdALP-ax-7AS*, 7 alleles were identified for *TdALP-ay-7AS*, and 4 alleles were identified for *TdALP-ax-4AL*. Correlation studies on *TdALP* gene diversity and ecological stresses suggested that environmental factors contribute to the *ALP* polymorphism formation in WEW. Many allelic variants of *ALPs* in the endosperm of WEW are not present in bread wheat and therefore could be utilized in breeding bread wheat varieties for better quality and elite plant defense characteristics.

avenin-like proteins | ALP gene evolution | TdALP gene alleles | wild emmer wheat | natural selection

Prolamin superfamily proteins share a conserved pattern of cysteine residues, including the sulfur-rich prolamins of the *Triticeae*, the cereal α-amylase/trypsin inhibitors, 2S storage albumins, puroindolines, grain softness proteins, α-globulins, and a group of hydroxyproline-rich cell wall proteins, which might all have originated from a small number of ancestral genes. According to Shewry and Halford (1), the gliadins, members of the prolamin superfamily, include members with a large repetitive domain and a conserved set of cysteine residues (α- and γ-gliadins), members with a repetitive domain but no cysteine (ω-gliadins), and members with novel low molecular weight gliadins (LMWGs) also known as avenin-like proteins (ALPs) that contain a conserved cysteine pattern but with no repetitive domains (2).

LMWGs are proteo-lipid-like hydrophobic proteins, similar to albumin-like and globulin-like proteins (3, 4). Genes encoding LMWGs are located in bread wheat on chromosomes 7A, 4A, and 7D (3). This observation supports the 4A/7B chromosome interchange hypothesis because there is a similar chromosomal distribution of peroxidase genes (5–7). In 2001, Anderson and others (8) cloned five genes that shared complex relationships with the gliadins. One cloned gene, *11dc7*, corresponded to one group of LMWGs described by Salcedo et al. (4). Rocher et al. (9) reported two similar LMWG proteins, rye-15 and rye-18 that showed weak immune reactivity with antibodies in serum from celiac patients. Clark et al. described the identification of a functional class of genes relevant to wheat grain end use belonging to a novel glutenin/gliadin seed storage protein (10). Kan et al. (11) identified two highly expressed transcripts encoding a- and b-type ALPs in wheat but with typically much higher expression in the *Aegilops* species. Overexpression of type-b ALPs in transgenic wheat improved dough mixing properties (12). ALP-coding genes were mapped to the short arms of chromosomes 7A and 7D and to the long arm of chromosome 4A in bread wheat (13). Importantly, alleles on 7A have been found with differential effects on dough quality, and its allele-specific markers have been developed to track the allelic effects (13). Recently, wheat ALP proteins were discovered with a significant *Fusarium* head blight resistant function, which highlighted the divergent functions of this gliadin domain containing protein family (14).

Wild emmer wheat (WEW), *T. dicoccoides*, is the progenitor of cultivated tetraploid and hexaploid wheats. It evolved in the northern ecogeographical region of the upper Jordan River in the eastern Upper Galilee Mountains and Golan Heights. Here, we studied 21 WEW populations from across their natural range in Israel. These were screened for allelic variation of *ALP* genes with the aim of identifying alleles useful for bread wheat improvement and determining the regional ecological influences on

Significance

Wheat grain avenin-like proteins (ALPs) have functions for dough quality and antifungal activities. A genome-wide characterization of *ALP* encoding genes in bread wheat is conducted. Results showed that most *ALPs* are transcriptionally active in developing grains and are up-regulated upon *Bgt* infection. The allelic diversity of *ALPs* in 21 natural populations of wild emmer wheat (WEW) in Israel were studied. Many *ALP* allelic variations in WEW were associated with regional environmental adaption. Our findings demonstrate that the diversifying natural selection through climatic and edaphic factors was a major driving force for the allelic diversity of *ALP* genes. The results indicate that WEW harbors a high genetic diversity of *ALPs* utilizable for wheat improvement.

Author contributions: W.M. designed research; Y.Z., X.H., S.I., M.S., Y.P., Z.Y., M.D., R.Y., J.Z., X.W., and X.C. performed research; Y.Z., X.H., S.I., M.S., Y.P., Z.Y., A.J., R.Y., J.Z., and E.N. analyzed data; and Y.Z., S.W., B.D., D.S., and W.M. wrote the paper.

Reviewers: Y.Q.G., USDA/ARS Western Regional Research Center; and S.X., USDA-ARS Cereal Crops Research Unit.

The authors declare no conflict of interest.

Data deposition: The sequences of raw data have been deposited in the National Center for Biotechnology Information Sequence Read Archive, https://www.ncbi.nlm.nih.gov/, and data accession nos. MK061124–MK061172 are listed in *SI Appendix*, Table S13.

[1]To whom correspondence may be addressed. Email: w.ma@murdoch.edu.au, sundongfa1@mail.hzau.edu.cn, or nevo@evo.haifa.ac.il.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1812855115/-/DCSupplemental.

Published online December 10, 2018.

allele formation. The 21 Israeli populations used in this paper had previously been studied by Nevo and coworkers (15–22). They identified local and regional ecological differences, genetic differences, and allozymic polymorphisms. These early studies identified adaptive allozyme diversity induced by abiotic and biotic stresses, highlighting the influence of selection on the adaptive nature of allozymic variation, and thereby negating the neutral theory of evolution.

Results

Identification of *ALP* Homologous Genes from Wheat Genome Phylogeny. We start the results section with an analysis of *ALP* homologs from wheat genomes to place the *TdALP* analysis from WEW populations (*SI Appendix*, Fig. S1) in a broader context of the *Triticeae* and other plants. In allohexaploid bread wheat, 15 unique full-length *TaALP* genes cDNA were mapped to chromosome groups 4 and 7 (*SI Appendix*, Fig. S2). Besides *ALP* genes reported in published studies (13, 23, 24), all other genes were cloned and sequenced (*SI Appendix*, Table S1). Alignment of the translated amino acid sequences encoded by the 15 full-length *TaALP* genes showed that ALPs vary in length from 150 to 285 amino acids (*SI Appendix*, Fig. S3*A*). Their signal peptides were predicted and listed in *SI Appendix*, Table S2. According to the domain classification based on the pFam database, ALPs are characterized by possessing gliadin domains (PF13016) as well as α-amylase inhibitors and seed storage protein subfamily domains (PF00234) (25). Based on alignment analysis and comparison with the reported a- and b-type *TaALP* genes, a new type was found and named c type in this paper (Fig. 1 and *SI Appendix*, Fig. S3 *A–C*). The homogeneity of the *TaALP* genes within the same subgroup is based on their high sequence identities (>86.43%) (*SI Appendix*, Table S3). To investigate the evolutionary relationships among *ALP* genes from the *Triticeae*, a ML phylogeny was constructed based on the deduced amino acid sequence alignment of 46 genes, including ALP-related sequences.

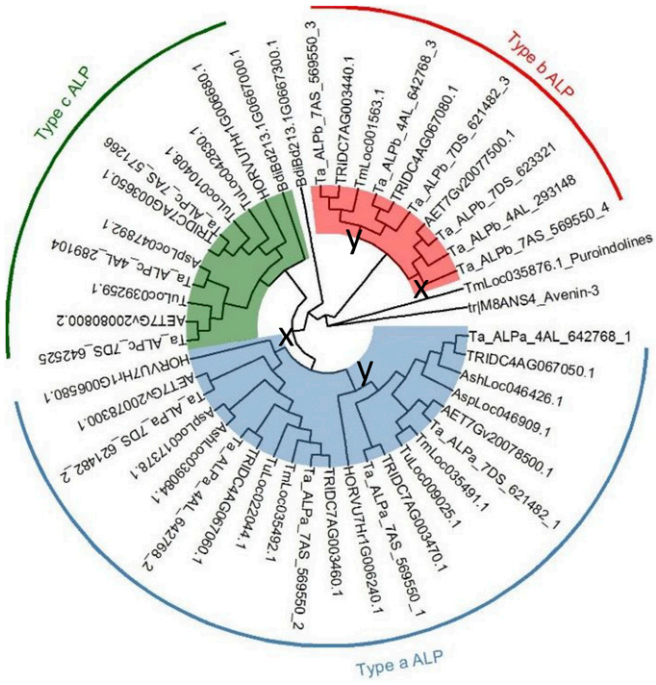


Fig. 1. Phylogenetic analyses of the ALPs. Maximum likelihood (ML) phylogeny of ALPs of the bread wheat (*Triticum aestivum*), *T. dicoccoides*, *Triticum urartu*, *Triricum monococcum*, *Aegilops speltoides*, *Aegilops. sharonesis*, *Aegilops tauschii*, *Brachypodium distachyon*, and *Hordeum vulgare* based on amino acid sequence alignment.

As shown (Fig. 1), besides the outgroup of puroindolines and avenin-3 in the monophyletic group, three major *ALP* gene clades, type a (blue), type b (red), and type c (green) for the 15 genes in bread wheat as well as one gene copy of *B. distachyon* were classified. Type-a *ALP* and type-b *ALP*s can be further divided into x and y subgroups. For the type-c clade, three *TaALP* genes on chromosomes 4A, 7A, and 7D from bread wheat and one orthologous barley gene *HvALP* on chromosome 7H, one from *T. dicoccoides* chromosome 7A, and genes from *T. monoccocum*, *T. urartu*, *A. speltoides*, *A. sharonesis*, *A. tauchii*, and *B. distachyon* were all closely related. Three subunits from *T. urartu* are clustered, indicating three type-c *ALP* genes in that species. Within the type-b clade y-type subgroup, one ortholog for *T. monococcum* and two homeologous genes on chromosomes 4A and 7A for *T. dicoccoides* were found. However, other orthologs of type-b ALPs were not identified in databases. For the x-type subgroup, one ortholog from *A. taucchii* and three homeologous *TaALP* genes from bread wheat genomes were identified. For the type-a clade y-type subgroups, 11 genes were identified—three from bread wheat, two from *T. dicoccoides* (4A and 7A), one ortholog from barley (7H), and one each from *T. monoccocum*, *T. urartu*, *A. speltoides*, *A. sharonesis*, and *A. tauchii*. The same patterns were found for the type-a ALP clade x-type subunits. All of the gene sequences used in this phylogenetic analysis (*SI Appendix*, Fig. S3*B*) contain a gliadin domain classified as PF13016. An unrooted ML phylogenetic analysis identified a wheat ALP clade (gray), avenin-3, a gliadin clade (purple), and a clade comprising millet, sorghum, and maize prolamins that have a gliadin domain (yellow). The *TaALP* clade was separated into three subgroups comprising type a, type b, and type c, whereas the rice prolamin, α-amylase inhibitor, grain softness protein, and puroindolines diverged from ALPs much earlier in the evolutionary history. This most recent common ancestor of the wheat and barley ALP clade (gray) and avenin-3 and gliadin clades (purple) can be traced to much earlier in the PF13016 domain evolutionary history (*SI Appendix*, Fig. S3*B*). Much earlier are the common ancestors of the members of the yellow clade for the millet, sorghum, and maize prolamins with a gliadin domain. As a result, *Triticeae* prolamins (ALPs, gliadins, and avenin-3) are more closely related to one another than to *Panicoideae* prolamins. A neighbor joining (NJ) analysis was performed on gliadin domains, AAI-LTSS domains, and LTP2 domains of monocots and lower plants (*SI Appendix*, Fig. S4).

Transcriptional Analyses of *TaALP* Genes in Bread Wheat Under *Bgt* Infection. Gene expression dynamics of *TaALP* genes under biotic stress were studied to select gene loci for detailed evolutionary study. The relative expression of *TaALP* genes in the lemma and grain of Spitfire × Mace doubled haploid (DH) lines 130, 131, and 187 were studied at 2 and 10 d after pollination (DAP) under *Bgt* infection (*SI Appendix*, Fig. S5). Gene-specific primers were designed for the 15 *TaALP* genes (*SI Appendix*, Table S4A). The parent wheat cvs. Spitfire and Mace displayed allelic variations at three *TaALP* loci, whereas the four DH lines were selected with consistent allelic compositions for the *TaALP* genes (*SI Appendix*, Table S4B). The healthy wheat lines (DH line 241) were chosen as the control, and their *ALP* expressions at 2, 7, and 10 DAP were shown in *SI Appendix*, Table S4C. The *TaALP* genes were up-regulated in *Bgt* affected DH lines 130, 131, and 187 compared with the DH line 241 (*SI Appendix*, Fig. S5), and relative expression of diverse types of *TaALP* genes showed significant positive linear correlations under *Bgt* infection (*SI Appendix*, Table S5). Based on the transcriptional study above, we selected four *TaALP* genes (*bx*/*ay*/*ax-7AS* and *-ax-4AL*) for allele screening and evolution study across the 21 WEW populations.

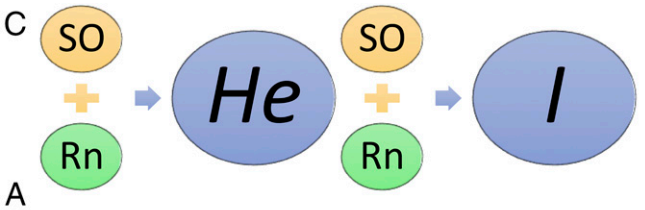
Gene Cloning and Sequencing Analyses of Four Selected *TdALP* Genes in WEW. Cloning and sequencing of the four selected *TdALP* genes in WEW (*TdALP-bx*/*ay*/*ax-7AS* and *-ax-4AL*) revealed a

EVOLUTION

surprisingly rich diversity. A total of 49 alleles were identified, including 25 *bx-7AS* genes, 13 *ax-7AS* genes, 7 *ay-7AS* genes, and 4 *ax-4AL* genes (*SI Appendix*, Fig. S6). For the *TdALP-bx-7AS* gene, among the 25 haplotypes (*SI Appendix*, Fig. S6*A*), 14 *bx-7AS* genes were assumed to be pseudogenes. Other alleles, *bx-7AS-a**, *-d*, *-g*, *-k*, *-m*, *-r*, *-s*, *-t*, *-x*, *-y*, and *-z*, were functional genes, and the amino acid translations indicated continuous reads from initiation to termination. Amino acid A/T replacement at position 12, a Q insertion at position 35, and I/S and M/W replacements at positions 58 and 60 occurred for genes *bx-7AS-g*, *-s*, *-t*, and *-r*. The Q/H replacement at position 205 also occurred in several alleles (*bx-7AS-m*, *-g*, *-s*, and *-t*). The Q insertion at position 35 occurred for alleles *bx-7AS-d* and *-k*. For *bx-7AS-k* encoded type-b ALP proteins, G/C replacement at the N-terminal region occurred. The coding sequences and deduced amino acid sequences of *TdALP-ax/ay-7AS* and *ax-4AL* were aligned, and the particular single-nucleotide polymorphisms (SNPs) and indels are shown in *SI Appendix*, Fig. S6 *B–D*. The amino acid alignments of the 11 functional alleles of *TdALP-bx-7AS* are shown in *SI Appendix*, Fig. S7*A*, whereas the type-a alleles are shown in *SI Appendix*, Fig. S7 *B–D*.

Population Genetics in Relation to Water and Edaphic Effects on *TdALP* Gene Diversity. The genetic diversity among different populations of WEW in Israel was assessed by comparing the *TdALP* gene alleles identified from each population and the corresponding *He* and *I* indices. As shown in *SI Appendix*, Table S6, 45 of 49 alleles were present in WEW populations in Israel. Overall, the four *TdALP* gene loci were polymorphic in most populations. The mean number of alleles per locus ranged from one to eight (*SI Appendix*, Table S6). The genetic variation of *TdALP* genes displayed a clear region-specific pattern, corresponding to a *He* index ranging from 0 to 0.64 and an *I* index from 0 to 1.26 (Fig. 2*A*).

To investigate the association of genetic diversity of *TdALP* genes in WEW with environmental variables, first, one-tailed Spearman correlation analysis was performed to analyze the association of genetic diversity of *TdALP* (*He* and *I*) with various environmental variables (Fig. 2*B* and *SI Appendix*, Table S7). The results showed that eight variables, including altitude, mean annual temperature, mean August temperature, mean January temperature, mean annual rainfall (Rn), mean annual evaporation, interannual variability of rainfall and soil type (So), and mean interannual variability of rainfall, were significantly correlated with the *TdALP* genotype (Fig. 2*B*). Second, the eight variables listed above plus two further geographical parameters were tested by backward multiple linear regression (MR) analysis for which 15 of the 21 WEW populations were included (*SI Appendix*, Table S8). When the three-variable model was used (*SI Appendix*, Table S9), a significant regression equation was found: $F(3, 11) = 8.99$, $P < 0.01$ with an R^2 of 0.710, $t(15) = 5.62$, and $P < 0.01$. Accordingly, latitude (50%), Rn (99%), and So (56%) were identified as the variables that could explain the highest proportion of the *TdALP* genetic diversity (*He*) among different populations. Noteworthy is the standardized coefficient for the Rn variable calculated as 99%, which suggests that Rn dominates the other two variables in their relevance to *TdALP* diversity. As such, a two-variable model was applied to the dataset (*SI Appendix*, Table S10). A significant regression equation was found: $F(2, 12) = 10.69$, $P < 0.01$ with an R^2 of 0.640, $t(15) = 5.90$, and $P < 0.001$. The results showed that Rn and So could explain 65% and 35% of *TdALP* diversity (*He*), respectively, which suggest that the two-variable model fits the dataset much better than the three-variable model. To further validate that Rn and So are variables mostly associated with the *TdALP* genetic diversity, other backward MR analyses were performed. Similarly, a significant regression equation was found: $F(2, 12) = 11.59$, $P < 0.01$ with an R^2 of 0.659, $t(15) = 4.98$, and $P < 0.001$ (*SI Appendix*, Table S11). The results showed that Rn and So contribute 57% and 47%, respectively, to the genetic diversity (*I*) (Fig. 2*C*). These results are comparable to the *He* index calculation (Fig. 2*C*).



A

	na*	ne*	*He*	*I*
Qazrin	2.00	1.87	0.46	0.66
Yehudiyya	5.00	2.57	0.60	1.12
Gamla	3.00	2.78	0.64	1.06
Rosh-Pinna	1.00	1.00	0.00	0.00
Ammiad-85	1.33	1.33	0.17	0.23
Tabigha	4.67	3.10	0.67	1.26
Mt. Gilboa	1.67	1.67	0.33	0.46
Mt. Gerizim	1.67	1.60	0.31	0.44
Gitit	2.00	1.73	0.42	0.61
Kokhav-Hashahar	2.67	2.67	0.61	0.96
Taiyiba	2.67	2.44	0.58	0.92
Sanhedriyya	2.00	2.00	0.50	0.69
Bet-Meir	2.00	2.00	0.50	0.69
J'aba	2.33	2.33	0.56	0.83
Amirim	2.00	2.00	0.50	0.69
Nesher	1.00	1.00	0.00	0.00
Beit-Oren	2.00	1.80	0.44	0.64
Daliyya	1.00	1.00	0.00	0.00
Bat-Shelomo	1.00	1.00	0.00	0.00
Kabara	2.00	2.00	0.50	0.69
Mt. Hermon	1.00	1.00	0.00	0.00

B

	He	*I*
He	1	
I	0.989**	1
Ln	0.079	0.093
Lt	-0.285	-0.251
AI	-0.467*	-.490*
Tm	0.446*	0.435
Ta	0.514*	0.519*
Tj	0.435	0.449*
Td	0.113	0.093
Tdd	0.392	0.392
Rn	-0.564*	-.525*
Rd	-0.368	-0.356
Hul4	-0.288	-0.316
Huan	-0.139	-0.154
Dw	0.077	0.109
Sh	-0.208	-0.253
Th	-0.427	-0.427
Trd	0.264	0.3
Ev	0.497*	0.501*
Sz	0.142	0.174
Ma	0.245	0.273
So	0.559*	0.594**
Rv	0.495*	0.478*
Rr	0.256	0.265

Fig. 2. Population genetics of WEW based on *TdALP* diversity. (*A*) Genetic indices of 21 WEW populations; (*B*) Spearman rank correlations of genetic indices at each WEW and climatic variables; (*C*) So and Rn are two environmental parameters which can best predict the *TdALP* diversity in WEW populations. Note: ** Correlation is significant at the 0.01 level (one tailed). * Correlation is significant at the 0.05 level (one tailed). He, Nei's—1973 gene diversity; I, Shannon's Information index (46); na, Observed number of alleles; ne, Effective number of alleles (45).

***TdALP-bx-7AS* Gene Clustering Analysis and Correlation with Environmental Factors.** A NJ analysis was performed based on 25 *TdALP-bx-7AS* gene sequence alignments (Fig. 3*A*). Natural selection pressure on the *bx-7AS* gene in WEW was examined by measuring the ratio of nonsynonymous to synonymous substitutions (dN:dS = ω) (26). The silent alleles (−) and the functional alleles (+) are listed in *SI Appendix*, Table S6. The branch ω value of *Haplo 4*, *9*, *12*, and *14* is 0.45, for *Haplo 15* and *25* the branch ω value is 0.0001, for *Haplo 1*, *7*, and *10* the branch ω value is 0.0001, for *Haplo 2*, *8*, *17*, and *23* the branch ω value is 2.37, for *Haplo 18*, *21*, and *22* the branch ω value is 2.23, for *Haplo 3*, *5*, *6*, *10*, *11*, *13*, *16*, and *19* the branch ω value is 1.93, and for *Haplo 24* the branch ω value is 1, displaying a neutral selection. The results indicated that all of the functional alleles (ω < 1 branches) are under purifying selection (except for *Haplo 11*, clustered with other silent alleles), whereas all of the

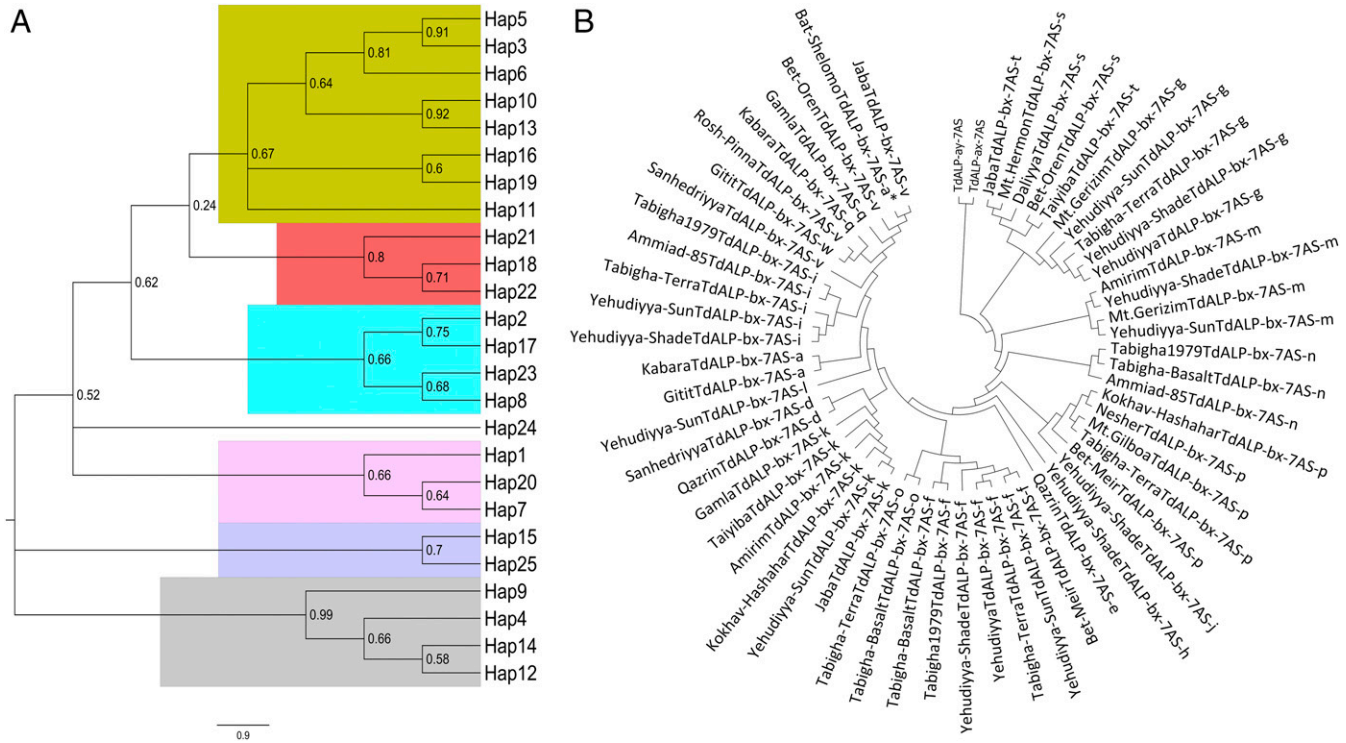


Fig. 3. Phylogenetic analysis of the *TdALP-bx-7AS* gene in WEW populations. (*A*) NJ phylogenetic analysis and natural selection tests of 25 haplotypes of the *TaALP-bx-7AS* gene in WEW; (*B*) UPGMA phylogenetic analysis based on *TaALP-bx-7AS* gene variation in 21 WEW populations.

silent alleles ($\omega > 1$ branches) are under positive selection. The *t* test results (one tailed with equal variance) (*SI Appendix*, Table S12) showed that the environmental factors had a P value < 0.05, indicating significant correlations of the *TdALP-bx-7AS* functional allele/silent allele (±) with environmental factors. The microenvironments selecting functional alleles (+) were significantly different from the microenvironments favoring silent alleles (−). The P values of the mean number of rainy days, the mean number of dew nights in summer, and the mean annual evaporation were >0.05, indicating no significant correlations.

Unweighted Pair Group Method with Arithmetic Mean Phylogenetic Analysis of *TdALP-bx/ay/ax-7AS* in WEW Populations. Genetic variation of *TdALP-bx-7AS* among different WEW populations was analyzed. An unweighted pair group method with arithmetic mean (UPGMA) tree was developed based on *TdALP-bx/ax/ay-7AS* sequence alignments (Fig. 3*B*). In total, eight alleles (*bx-7AS-f*, *-g*, *-h*, *-i*, *-j*, *-k*, *-l*, and *-m*) were identified in the Yehudiyya population. These alleles were distributed among different branches, suggesting a significant degree of genetic variation. Some alleles are related to adaption to shade (*-j* and *-h*) under tree canopies and sun (*-k* and *-l*) between trees (17, 20–22). Some alleles collected from different WEW populations clustered together. For example, alleles *bx-7AS-g* (Yehudiyya, Tabigha terra rossa, and Mt. Gerizm), *-s* (Bet Oren, Daliyya, and Mt. Hermon), and *-t* (J'aba and Taiyiba) were grouped together. The *bx-7AS-m* allele was found in Yehudiyya (both sun and shade), Mt. Gerizm, and Amirim, whereas *-n* was present in Tabigha basalt and Ammiad-85. The *bx-7AS-p*s from Nesher, Tabigha terra rossa, Mt. Gilboa, Kokhav Hashahar, and Bet Meir (mostly xeric populations) were grouped together with *-j* from Yehudiyya shade, suggesting they were related populations from sites of similar ecogeographic backgrounds with respect to rainfall and terra rossa soil. The remaining 12 *bx-7AS* genes were from populations of diverse ecogeographic backgrounds, and these were clustered together in a separate group. Notably, *bx-7AS-f* dominates those populations from Yehudiyya (allele frequency of 28.22%, *SI Appendix*, Table S13) and ranks the highest in Tabigha (with both terra rossa and basalt soils). Specifically, taking So into consideration, alleles *bx-7AS-g*, *-p*, and *-i* were found in Tabigha terra rossa not in the abutting basalt soil. In contrast, *-n* was found in Tabigha basalt soil, demonstrating adaption to this So. The results of the UPGMA analysis for *TdALP-ax-7AS* and *TdALP-ay-7AS* genes are shown in *SI Appendix*, Figs. S8 and S9, respectively.

Genetic Distance Analyses Among Different WEW Populations. Pairwise genetic distances (p distance), based on the normalized gene sequence identity of *TdALP-bx/ay/ax-7AS* in WEW populations, were calculated using MEGA 7.0 software (*SI Appendix*, Table S14). Overall, the target populations demonstrated a close distance with each other, ranging from 0.013 to 0.291 (*SI Appendix*, Table S14). The genetic distance between WEW populations was found for J'aba, >0.195 with Bet Meir, Gamla, Gitit, Mt. Gilboa, Qazrin, Sanhedriyya, Tabigha 1979, Taiyiba, and Kokhav Hashahar. The maximum pairwise genetic distance (p distance = 0.291) among different populations was identified between populations J'aba and Kokhav Hashahar, indicating significant variations in closer but marginal steppic populations. However, one interesting phenomenon was observed, some physically distant populations displayed relatively lower p distances than some physically close populations. For example, populations Gamla and Yehudiyya sun, which were separated by only 9.5 km (*SI Appendix*, Fig. S1), have a p distance of 0.212. In contrast, the p distance between Mt. Gerizim and Gamla (156 km, *SI Appendix*, Fig. S1) is 0.155, and, most significantly, Yehudiyya sun and J'aba (130 km, *SI Appendix*, Fig. S1) have the lowest p distance (0.013) (*SI Appendix*, Table S14). Noteworthy, the p distance between Yehudiyya sun and Yehudiyya shade is 0.133, whereas that of Tabigha terra rossa and basalt is 0.163

(*SI Appendix*, Table S14), suggesting that the So plays a more significant role than temperature.

Discussion

Origin, Mechanism, and Phylogeny of *ALP* Gene Evolution. The ALPs' antifungal functions revealed by Zhang et al. (14) indicate a vital importance to identify genetic diversity of *ALPs* for potential exploration in wheat breeding. Similar to most grain storage protein genes, the current paper revealed that abundant *ALP* alleles have accumulated during evolution. Previous studies have shown that unequal crossover or gene slippage of insertions or deletions of blocks often happened during duplication events for *HMW-GS* genes (27, 28), which might also help to explain the emergence, expansion, and the allelic variations of the *ALP* genes. Domain replication observed for *ALP* genes (b type) serves a similar function to gene duplication, establishing gene variability driven by evolutionary forces. As for many protease inhibitor gene families, instead of complete gene duplication, including promoter and terminator sequences and possible reintegration at a distinct locus, there is duplication of the inhibitory domain sequence with the domains remaining fused (29–31). Fifteen full-length *TaALP* genes were clustered into three major subgroups (types a, b, and c) in our phylogenetic analyses (Fig. 1 and *SI Appendix*, Fig. S3*B*). In addition, our results revealed the existence of intra- and interchromosomal *ALP* genes in WEW and bread wheat. There are three copies of type-c genes in *T. urartu* (Fig. 1), whereas only one copy in each chromosome of bread wheat and WEW, indicating that wheat-specific *ALP* gene duplication/elimination events most likely occurred in a diploid wheat ancestor, leading to the loss of these genes in WEW as well as in bread wheat. Similar wheat-specific gene duplication events and/or chromosomal translocations are also likely to be responsible for the origin of the multi-*ALP* genes.

The Importance of Natural Population in Highlighting Genetic Adaptations. Experimental populations evolving under natural selection represent an important resource for studying the genetic basis of adaptation. Our analysis of *TdALP* gene variation and evolution in WEW was based on ecogenetic analyses of Israeli and Golan Heights populations as was demonstrated previously in a study of allozyme evolution in these populations (15–22). Our results demonstrated that polymorphisms in *ALP* genes in WEW correlated with the ecogeographic distribution of the genotypes. Observations were consistent with previous results on *HMW-GS*, *LMW-GS*, gliadins, and α-/β-amylase inhibitors (32–37). Some geographically close populations were very different in their *TdALP* structures at the considered loci (*SI Appendix*, Table S6). An example can be found at Tabigha, now designated the evolution slope where two divergent Sos—the calcareous terra rossa soil and the volcanic basaltic soil (15, 38)—influenced the *TdALP* composition of WEW populations occupying these different soils. Alleles *bx-7AS-g*, *ay-7AS-b*, and *ax-7AS-c* occurred only in plants from terra rossa soil, whereas *bx-7AS-n*, *ay-7AS-c*, and *ax-7AS-b* occurred only in basaltic soil. The absence of a significant relationship between geographic separation and genetic distance attests to a sharp local ecological differentiation rather than a gradual change in allele frequencies across the range of WEW in Israel. Genetic diversity did not follow the simple isolation by the distance model of Wright et al. (39). Quite often, a greater genetic difference occurred between physically close populations than between distant populations. This was clearly demonstrated by the proximal populations located at Tabigha (two Sos) (39) and Yehudiyya (sun vs. shade) (17, 20–22). Thus, the genetic structure of WEW populations in Israel is mosaic. This patchy genetic distribution appears to reflect the underlying ecological, climatic, edaphic, and biotic heterogeneities on both micro- and macroscales (16, 18, 40). Microenvironmental variation coupled with a limited migration of *T. dicoccoides* may explain the dramatic genetic divergence of the two populations at the Tabigha site (15, 38). Specific SNP positions detected in *TdALP* genes were found to be highly effective in distinguishing genotypes and populations of WEW originating from diverse ecogeographic sites. These results suggest that genetic variations at these SNP positions in the *TdALP* were, at least, partly ecologically determined.

Natural Selection of *TdALP-bx-7AS* Genes in WEW. Significant diversities at the *TdALP-bx-7AS* gene locus were detected both between and within WEW populations. The *bx-7AS* genes were naturally selected across populations supported by a different ratio of dN:dS (ω) (26). Environmental factors significantly correlate with the functional and silent alleles for the *bx-7AS* locus (*SI Appendix*, Table S12). A sharp genetic divergence over short geographic distances compared with a small genetic divergence between long geographic distances also suggested that the SNPs were subjected to natural selection, and ecological factors played an important evolutionary role in gene polymorphism formation (15–22). Natural selection of orthologous genes can be assessed by comparing the ratio of ω in protein coding sequences (41). Ecological stresses have often been proposed as inducing active and rapid evolutionary changes. Compared with positive natural selection, purifying selection acts against mutations that have deleterious effects on protein structure. The ω value of *Haplo 24* equals 1, displaying neutral selection (Fig. 3*A*). The results indicated that all of the functional alleles are smaller than 1, suggesting that natural selection may have eliminated most of the deleterious effects caused by purifying selection (Fig. 3*A*). On the other hand, all of the silent alleles and one functional allele *Haplo 11* are greater than 1, suggesting that positive selection gives rise to dominant alleles in several WEW populations (Fig. 3*A*). Altitude plays a significant role in population *TdALP* divergence as evidenced by the populations of Mt. Hermon, Rosh Pinna, Gamla, Bat Shelomo, and Tabigha, located at altitudes of 1300, 700, 200, 75, and 0 m, respectively (*SI Appendix*, Table S7). The results showed that the populations located below 700 m (Tagbiha, 0 m) tend to have a higher level of genetic diversity with the *He* and *I* being 0.67 and 1.26, respectively (Fig. 2*B*). In contrast, populations collected above 900 m, such as Mt. Hermon, were not polymorphic at all (Fig. 2*B*). Along with altitude, several other environmental factors differed among these populations, such as abiotic climatic conditions, water availability, So, and biotic factors including parasites, pathogens, and competitors (*SI Appendix*, Table S7) (16, 42).

Genetic Distance and Evolution of *TdALP* in WEW. The relationship between *TdALP* genetic distance and geographical distance indicated that the estimates of genetic distance (p distance) were geographically independent. Sharp genetic divergences (long p distance) over very short geographic distances against small genetic divergence (short p distance) between long physical distances were observed. For example, the genetic distance between populations of Tabigha (terra rossa and basalt) and Gamla located only about 9.5 km apart from p distance = 0.212, was 16 times higher than that between Yehudiyyasun and J'aba (130 km, *SI Appendix*, Fig. S1). Environmental stress can greatly influence plant susceptibility to herbivores and pathogens, and drought stress can promote outbreaks of fungal diseases and plant-eating insects (43). Different herbivore-related and pathogen-related selection pressures at these ecological locations may influence polymorphism of insect-resistant and pathogen-resistant loci in WEW (36). It can be concluded that the variation in *ALP* genetic diversity between populations is due to selective forces. The genetic structure of WEW populations in Israel is a mosaic (16, 17, 19, 33, 34, 44, 45). Thus, higher levels of polymorphisms and genetic variations of *TdALP* within and between populations can be explained as adaptive complexes generated by natural selection and coevolution with biotic or abiotic pressure.

Conclusions and Prospects. Our molecular characterization of the *TdALP* gene family in WEW allows several conclusions to be made about the origin of *ALP* genes. Future challenges of crop improvement can be overcome by effectively utilizing the immense resources of genetic diversity unraveled by the evolution and allele analysis in natural populations of the wheat progenitors. The drivers of *ALP* allelic variations in WEW populations appear to be intimately linked to the environment in which the populations originated. These results suggest: (*i*) during the evolutionary history of WEW, diversifying natural selection through climatic (e.g., annual rain fall and temperature) and edaphic factors (So) was a major agent of genetic structure and differentiation at *TdALP* loci; and (*ii*) WEW populations harbor large amounts of genetic diversity exploitable for wheat improvement. Furthermore, at the transcriptional level, we found that most members of this multifunction large gene family are transcriptionally active at multiple stages of bread wheat development as well as under conditions of pathogen infection (powdery mildew). The allelic diversity associated with the germplasm-originating environmental conditions may provide a solution to fight the negative impact of the global warming complexities on modern wheat production. These genetic resources provide potential values for improving wheat cultivars under uncertain environmental conditions in the future.

Materials and Methods

The geographical sources of 21 WEW (*T. dicoccoides*) populations analyzed are shown (*SI Appendix*, Fig. S1). WEW seeds were obtained from the GenBank of the Institute of Evolution, University of Haifa, Mount Carmel, Haifa, Israel. Mace × Spitfire DH population wheat lines were used for *ALP* gene transcriptional analysis. Details of experimental procedures, such as *ALP* genes' cloning, phylogenetic analysis, quantitative reverse transcription polymerase chain reaction, population genetic analysis, and statistical analysis, are described in *SI Appendix, Materials and Methods*.

ACKNOWLEDGMENTS. We thank Clare Johnson from the Australia Grain Research & Development Corporation for her support and invaluable suggestions to this research. Murdoch University Strategic Scholarship funded Yujuan's PhD study. This paper was jointly funded by the Australian Grains Research & Development Corporation Project UMU00043 and Murdoch University internal research funds.

1. Shewry PR, Halford NG (2002) Cereal seed storage proteins: Structures, properties and role in grain utilization. *J Exp Bot* 53:947–958.
2. Kasarda DD, Adalsteins E, Lew EJL, Lazo GR, Altenbach SB (2013) Farinin: Characterization of a novel wheat endosperm protein belonging to the prolamin superfamily. *J Agric Food Chem* 61:2407–2417.
3. Ewart JAD (1975) Isolation of a Cappelle-Desprez gliadin. *J Sci Food Agric* 26: 1021–1025.
4. Salcedo G, Prada J, Aragoncillo C (1979) Low MW gliadin-like proteins from wheat endosperm. *Phytochemistry* 18:725–727.
5. Kobrehel K, Feillet P (1975) Identification of genomes and chromosomes involved in peroxidase synthesis of wheat seeds. *Can J Bot* 53:2336–2344.
6. Anderson JA, Ogihara Y, Sorrells ME, Tanksley SD (1992) Development of a chromosomal arm map for wheat based on RFLP markers. *Theor Appl Genet* 83: 1035–1043.
7. Devos KM, Dubcovsky J, Dvořák J, Chinoy CN, Gale MD (1995) Structural evolution of wheat chromosomes 4A, 5A, and 7B and its impact on recombination. *Theor Appl Genet* 91:282–288.
8. Anderson O, Hsia C, Adalsteins A, Lew EL, Kasarda D (2001) Identification of several new classes of low-molecular-weight wheat gliadin-related proteins and genes. *Theor Appl Genet* 103:307–315.
9. Rocher A, Calero M, Soriano F, Méndez E (1996) Identification of major rye secalins as coeliac immunoreactive proteins. *Biochim Biophys Acta* 1295:13–22.
10. Clarke BC, Hobbs M, Skylas D, Appels R (2000) Genes active in developing wheat endosperm. *Funct Integr Genomics* 1:44–55.
11. Kan Y, et al. (2006) Transcriptome analysis reveals differentially expressed storage protein transcripts in seeds of *Aegilops* and wheat. *J Cereal Sci* 44:75–85.
12. Ma F, et al. (2013) Overexpression of avenin-like b proteins in bread wheat (*Triticum aestivum* L.) improves dough mixing properties by their incorporation into glutenin polymers. *PLoS One* 8:e66758.
13. Chen XY, et al. (2016) Genetic characterization of cysteine-rich type-b avenin-like protein coding genes in common wheat. *Sci Rep* 6:30692.
14. Zhang Y, et al. (2018) Wheat avenin-like protein and its significant Fusarium head blight resistant functions. bioRxiv:10.1101/406694. Preprint, posted September 13, 2018.
15. Nevo E, Brown A, Zohary D, Storch N, Beiles A (1981) Microgeographic edaphic differentiation in allozyme polymorphisms of wild barley (*Hordeum spontaneum, Poaceae*). *Plant Syst Evol* 138:287–292.
16. Nevo E, Beiles A (1989) Genetic diversity of wild emmer wheat in Israel and Turkey: Structure, evolution, and application in breeding. *Theor Appl Genet* 77:421–455.
17. Li Y, et al. (2000) Microsatellite diversity correlated with ecological-edaphic and genetic factors in three microsites of wild emmer wheat in North Israel. *Mol Biol Evol* 17: 851–862.
18. Golenberg EM, Nevo E (1987) Multilocus differentiation and population structure in a selfer, wild emmer wheat, *Triticum dicoccoides*. *Heredity* 58:451–456.
19. Nevo E, Beiles A, Krugman T (1988) Natural selection of allozyme polymorphisms: A microgeographical differentiation by edaphic, topographical, and temporal factors in wild emmer wheat (*Triticum dicoccoides*). *Theor Appl Genet* 76:737–752.
20. Nevo E, Beiles A, Krugman T (1988) Natural selection of allozyme polymorphisms: A microgeographic climatic differentiation in wild emmer wheat (*Triticum dicoccoides*). *Theor Appl Genet* 75:529–538.
21. Nevo E (1988) Genetic diversity in nature. *Evolutionary Biology* (Springer, Boston), pp 217–246.
22. Li YC, et al. (2000) Natural selection causing microsatellite divergence in wild emmer wheat at the ecologically variable microsite at Ammiad, Israel. *Theor Appl Genet* 100: 985–999.
23. Clarke BC, Phongkham T, Gianibelli MC, Beasley H, Bekes F (2003) The characterisation and mapping of a family of LMW-gliadin genes: Effects on dough properties and bread volume. *Theor Appl Genet* 106:629–635.
24. Subburaj S, et al. (2016) Molecular characterization and evolutionary origins of farinin genes in *Brachypodium distachyon* L. *J Appl Genet* 57:287–303.
25. Finn RD, et al. (2016) The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res* 44:D279–D285.
26. Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591.
27. Hassani M, Gianibelli M, Shariflou M, Sharp P (2005) Molecular structure of a novel y-type HMW glutenin subunit gene present in *Triticum tauschii*. *Euphytica* 141:191–198.
28. Liu Y, Xiong ZY, He YG, Shewry PR, He GY (2007) Genetic diversity of HMW glutenin subunit in Chinese common wheat (*Triticum aestivum* L.) landraces from Hubei province. *Genet Resour Crop Evol* 54:865–874.
29. Christeller JT (2005) Evolutionary mechanisms acting on proteinase inhibitor variability. *FEBS J* 272:5710–5722.
30. Laskowski M, Jr, Qasim MA (2000) What can the structures of enzyme-inhibitor complexes tell us about the structures of enzyme substrate complexes? *Biochim Biophys Acta* 1477:324–337.
31. Gojobori T, Ikeo K (1994) Molecular evolution of serine protease and its inhibitor with special reference to domain evolution. *Philos Trans R Soc Lond B Biol Sci* 344:411–415.
32. Nevo E, Nishikawa K, Furuta Y, Gonokami Y, Beiles A (1993) Genetic polymorphisms of α- and β-amylase isozymes in wild emmer wheat, Triticum dicoccoides, in Israel. *Theor Appl Genet* 85:1029–1042.
33. Nevo E, Pagnotta MA, Beiles A, Porceddu E (1995) Wheat storage proteins: Glutenin DNA diversity in wild emmer wheat, *Triticum dicoccoids*, in Israel and Turkey. 3. Environmental correlates and allozymic associations. *Theor Appl Genet* 91:415–420.
34. Nevo E, Payne PI (1987) Wheat storage proteins: Diversity of HMW glutenin subunits in wild emmer from Israel: 1. Geographical patterns and ecological predictability. *Theor Appl Genet* 74:827–836.
35. Pagnotta MA, Nevo E, Beiles A, Porceddu E (1995) Wheat storage proteins: Glutenin diversity in wild emmer, *Triticum dicoccoides*, in Israel and Turkey. 2. DNA diversity detected by PCR. *Theor Appl Genet* 91:409–414.
36. Wang JR, et al. (2008) Molecular evolution of dimeric alpha-amylase inhibitor genes in wild emmer wheat and its ecological association. *BMC Evol Biol* 8:91.
37. Wang JR, et al. (2010) The impact of single nucleotide polymorphism in monomeric alpha-amylase inhibitor genes from wild emmer wheat, primarily from Israel and Golan. *BMC Evol Biol* 10:170.
38. Wang X, et al. (2018) Genomic adaptation to drought in wild barley is driven by edaphic natural selection at the Tabigha Evolution Slope. *Proc Natl Acad Sci USA* 115: 5223–5228.
39. Wright S (1943) Isolation by distance. *Genetics* 28:114–138.
40. Fahima T, et al. (1999) RAPD polymorphism of wild emmer wheat populations, *Triticum dicoccoides*, in Israel. *Theor Appl Genet* 98:434–447.
41. Hurst LD (2009) Fundamental concepts in genetics: Genetics and the understanding of selection. *Nat Rev Genet* 10:83–93.
42. Nevo E, Golenberg E, Beiles A, Brown AH, Zohary D (1982) Genetic diversity and environmental associations of wild wheat, *Triticum dicoccoides*, in Israel. *Theor Appl Genet* 62:241–254.
43. Mattson WJ, Haack RA (1987) The role of drought in outbreaks of plant-eating insects. *Bioscience* 37:110–118.
44. Nevo E, Carver BF, Beiles A (1991) Photosynthetic performance in wild emmer wheat, *Triticum dicoccoides*: Ecological and genetic predictability. *Theor Appl Genet* 81:445–460.
45. Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. *Genetics* 49:725–738.
46. Lewontin RC (1972) *Testing the Theory of Natural Selection* (Nature Publishing Group, London).