N.I. Vavilov's Theory of Centres of Diversity in the Light of Current Understanding of Wheat Diversity, Domestication and Evolution

J. DVORAK[1], M.-Ch. LUO[1] and E.D. AKHUNOV[1, 2]

[1]Department of Plant Sciences, University of California, CA 95616 Davis, USA;
[2]Department of Plant Pathology, Kansas State University, KS 66506 Manhattan, USA;
e-mail: jdvorak@ucdavis.edu

Abstract: N.I. Vavilov hypothesized that the geographical centres of diversity of crops indicate their geographical centres of origin. Vavilov's conclusions about the geographical origins of einkorn, durum and common wheat agree well with current population and molecular genetic studies when macro-geography is used but agree poorly when they are examined at higher resolution. We examined the causes of such disagreements for tetraploid emmer wheat and hexaploid common and club wheat. Molecular studies suggest that emmer was domesticated in the Diyarbakir region in south-eastern Turkey. Nucleotide diversity of wild emmer in the Diyarbakir region estimated earlier was compared with nucleotide diversity of wild and domesticated emmer across their distribution estimated here. Although domesticated emmer is only half as diverse as wild emmer, it is more diverse than the ancestral wild emmer population in the Diyarbakir region. Its centre of diversity is in the Mediterranean and does not coincide with the geographical centre of emmer origin. A similar disagreement exists in hexaploid wheat. Its centre of molecular diversity is in Turkey, which is west of the putative site of its origin in Transcaucasia and north-western Iran. It is shown that the primary cause of the disagreements between geographical centres of crop diversity and geographical centres of crop origin is gene flow from an ancestor subsequently to crop origin, which modifies the geographical pattern of crop diversity. When such gene flow takes place and when crop is domesticated in a peripheral population of the ancestor, the centre of crop diversity and the centre of crop origin are unlikely to coincide.

Keywords: domestication; emmer; gene flow; nucleotide diversity; *Triticum aestivum*; *Triticum dicoccoides*

Under the leadership of N.I. Vavilov, the Institute of Plant Industry proposed an ambitious crop breeding program and equally ambitious plant exploration program to support it. The exploration trips generated a broad view of the geographical distribution of phenotypic diversity of individual crops and their wild progenitors and led Vavilov to formulate his famous theory of geographical centres of crop diversity. He postulated that crop diversity usually has a geographical centre at which it is the highest and hypothesized that the centre with the highest diversity of a crop is the geographical region of its origin. He recognized that morphological diversity of crops overlapped in specific areas and proposed that those were the geographical centres of agriculture origins (see English translation of Vavilov's writings in Vavilov 1992).

Although the concept of crop diversity centres has had enormous impact on the evolution of thinking about crop variation and the origins of

agriculture, most of those ideas have since evolved. It has been repeatedly pointed out that the centre of crop diversity may or may not coincide with the centre of crop origin (Schiemann 1939; Gokgol 1941; Zohary 1970; Harlan 1971, 1975; Zohary & Hopf 1994). Vavilov's pioneering work and work of other plant explorers who went in his footsteps and collected, catalogued, and preserved local varieties and landraces generated an invaluable resource for studies of crop domestication with the current theoretical and experimental tools and facilitated the examination of the relationships between crop diversity, its origin, and its subsequent evolution in details unavailable to Vavilov.

Concerning wheat, Vavilov recognized that each of the three ploidy levels of wheat has its own geographical centre of diversity and concluded that each had a separate geographical place of origin. He placed the origin of diploid einkorn wheat (*T. monococcum*, genomes A^mA^m) to Asia Minor, tetraploid durum (*T. turgidum* ssp. *durum*, genomes AABB) and other free-threshing tetraploid wheats to the Eastern Mediterranean and Northeastern Africa, and hexaploid wheat (*T. aestivum*, genomes AABBDD) to a region spanning an area from Afghanistan and Turkmenistan to Transcaucasia (Vavilov 1926, 1992). Vavilov concluded that tetraploid emmer (*T. turgidum* ssp. *dicoccon*, genomes AABB) is an ancient crop that is going extinct and did not identify its geographical place of origin. Finally, he placed the centre of diversity of wild emmer (*T. turgidum* ssp. *dicoccoides*, genomes AABB), which is distributed along the Fertile Crescent from Israel to western Iran, into its south-western tip (Vavilov 1992). In agreement with Vavilov's conclusions, molecular studies placed the origin of diploid einkorn wheat to south-eastern Turkey (Heun *et al.* 1997; Kilian *et al.* 2007), the origin of durum to eastern Mediterranean and north-eastern Africa (Luo *et al.* 2007), the origin of hexaploid wheat to Transcaucasia and north-western Iran (Dvorak *et al.* 1998), and the centre of diversity of wild emmer was indeed found to be in the south-western tip of the Fertile Crescent, in a region including modern Israel, Jordan, Lebanon, and south-western Syria (Luo *et al.* 2007).

When the relationships between the geography of diversity and geography of domestication are examined in detail, relationships between crop diversity and crop origin begin to break down. Consider for example domesticated emmer. Since domesticated emmer is most diverse in the Mediterranean basin (Luo *et al.* 2007) it should be expected that emmer was domesticated in the south-western tip of the Fertile Crescent. However, it appears that emmer was domesticated in the Diyarbakir region in south-eastern Turkey (Ozkan *et al.* 2005; Luo *et al.* 2007) along with einkorn wheat (Heun *et al.* 1997). Detailed examination of the origin of *T. aestivum* reveals a similar dilemma. Vavilov concluded that *T. aestivum* is most diverse in the mountainous areas of Afghanistan, Turkmenistan, and Iran, and to a lesser extent in Transcaucasia. However, the relationships between *Aegilops tauschii* (genomes DD), the diploid ancestor of the hexaploid wheat D genome (Kihara 1944; McFadden & Sears 1946), and the D genome of hexaploid wheat failed to show that *Ae. tauschii* populations in Turkmenistan and Afghanistan played a role in the evolution of hexaploid wheat, making those areas an unlikely geographical centre of *T. aestivum* evolution (Dvorak *et al.* 1998).

To shed light on these conflicts, we estimated here nucleotide diversity in a sample of genes of wild and domesticated emmer and compared it with previously reported estimates of diversity in nucleotide sequences in wild emmer in the Diyarbakir region in south-eastern Turkey and in *T. aestivum* (Akhunov *et al.* 2010). To gain an idea about the distribution of diversity across the geographical distribution of *T. aestivum*, we computed expected heterozygosity from restriction fragment length polymorphism for populations of common wheat and club wheat spanning a region from China to Europe.

MATERIALS AND METHODS

A total of 299 landraces and varieties of *T. aestivum* were used for restriction fragment length polymorphism (RFLP) genotyping at 131 loci (Dvorak & Luo 1999; Dvorak *et al.* 2006; Luo *et al.* 2007). Geographical origins of the lines and RFLP methodology have been described elsewhere (Dvorak & Luo 1999; Dvorak *et al.* 2006; Luo *et al.* 2007) and will not be repeated. Likewise, the analyses of nucleotide sequences of 585 A-genome and 576 B-genome gene fragments in 10 accessions of wild emmer from the Diyarbakir region in Turkey and 590, 584 and 679 A-, B-, and D-genome gene fragments, respectively, from 13 accessions of *T. aestivum* have been reported

and information about the accessions, genes and the sequences can be found therein (Akhunov *et al.* 2010). Wild emmer and *T. aestivum* accessions used by Akhunov *et al.* (2010) were selected to represent genetic diversity of the respective populations (for details see Akhunov *et al.* 2010).

In addition to these data, from 24 to 26 genes were sequenced here in 44 accessions of *Ae. tauschii*, 49 accessions of wild emmer across its entire geographical distribution, and 23 accessions of domesticated emmer. Accessions were selected for sequencing on the basis of their position in a neighbour joining tree, using a strategy described by Akhunov *et al.* (2010). All sequenced genes were also mapped (Luo *et al.* 2009).

Ae. tauschii genes were sequenced using the primer walking strategy along *Ae. tauschii* bacterial artificial chromosome (BAC) clones. BAC clones were selected from *Ae. tauschii* BAC libraries (Xu *et al.* 2002) by hybridization of ^{32}P-labelled cDNA clones with high density screening membranes as described by Akhunov *et al.* (2005). BAC DNAs were isolated with the Qiagen R.E.A.L 96-Prep (Valencia, CA) kit. DNAs were used as templates in the first of several sequencing reactions needed to sequence an entire gene. Sequencing was performed with ABI3730xl. In each sequencing step, a primer located near the end of the sequence was designed for the next sequencing step. Both strands of each gene were sequenced.

For the sequencing of emmer genes, genome-specific PCR primers reported in wheat SNP database were used and gene portions were sequenced as described in Akhunov *et al.* (2010). Briefly, a targeted DNA was PCR amplified from genomic DNA and amplicons were directly sequenced using PCR primers as sequencing primers. Both strands of each gene were sequenced using BigDye v3.1 sequencing chemistry (ABI, Foster City, California) and capillary electrophoresis (ABI3730xl).

The sequences were processed as follows. The phred/phrap (Green 1998) or Staden package (http://staden.sourceforge.net/) programs were used for base calling and assembly of sequencing trace files. The Consed program (Gordon 2004) was used for contig editing. The sequences were trimmed using a phred quality score of 20 or higher and realigned using the Muscle or ClustalX programs. The gaps in the alignments were deleted before the analysis.

Diversity estimates θ_w (Watterson 1975) and θ_π (Tajima 1983) were computed. The former estimate corresponds to the number of segregating sites in a population sample and θ_π corresponds to the average number of nucleotide differences in a pair-wise haplotype comparisons in a population sample. Under a neutral evolutionary history both estimators provide unbiased estimates of $\theta = 4N_e\mu$. The frequency spectrum of single nucleotide polymorphisms (SNPs) in population samples relative to neutral mutation model was assessed using Tajima's *D* statistic (Tajima 1989). Positive Tajima's *D* indicates the presence of too many mutant sites at intermediate frequencies compared to what is expected under the assumption of neutrality while negative Tajima's *D* indicates the presence of too many sites with low frequencies.

RESULTS AND DISCUSSION

Diversity of wild and domesticated emmer and emmer evolution

Estimates of nucleotide polymorphism θ_w, nucleotide diversity θ_π and Tajima's *D* were similar in the A and B genomes of wild emmer (Table 1). Negative estimates of Tajima's *D* indicated a preponderance of rare alleles in both genomes. Earlier reported diversity estimates and Tajima's *D* for the A and B genomes in wild emmer population in the Diyarbakir region and for the A and B genomes in *T. aestivum* were also similar (Akhunov *et al.* 2010). Wild emmer originated about 0.36 million years ago (Huang *et al.* 2002; Dvorak & Akhunov 2005), and the similar amounts of diversity and similar departures from neutrality expectations in both genomes show that there was a sufficient time for a large portion of diversity present in wild emmer to evolve since its origin. Because nucleotide diversity in the A and B genomes is similar, combined A- and B-genome estimates will be used throughout this paper.

Nucleotide diversity θ_π was 2.28×10^{-3} in the A and B genomes of wild emmer (Table 1), which was only slightly lower than an estimate of $\theta_\pi = 2.9 \times 10^{-3}$ reported earlier (Haudry *et al.* 2007). In domesticated emmer, θ_π was 1.18×10^{-3}, which is 52% of diversity present in the broad geographical sample of wild emmer. This diversity decline is similar to the previously reported decline to 58%, indicated by RFLP (Luo *et al.* 2007), but is less than a decline to 29.6% previously reported by Haudry *et al.* (2007). The reason for this

discrepancy may be sampling variation or, more likely, non-random sampling of genes studied by Haudry *et al.* (2007), which were mostly located in few chromosome regions, some linked to domestication genes.

Wild emmer in the geographical centre of emmer domestication, the Diyarbakir region in south-eastern Turkey, was less diverse than the geographically broad sample of wild emmer. Nucleotide polymorphism θ_w was 0.73×10^{-3} and nucleotide diversity θ_π was 0.72×10^{-3}; both estimates represent 31% of diversity of wild emmer as a whole (Table 1).

Diversity measurements using RFLP showed that wild emmer diversity was the greatest in north-western Israel, Lebanon, and south-western Syria and declined in the northern and eastern regions of the Fertile Crescent including Turkey, Iraq, and Iran (Luo *et al.* 2007). The wild emmer population in the south-western tip of the Fertile Crescent represents the centre of wild emmer diversity in the sense of Vavilov and the populations in the northern and eastern parts of the Fertile Crescent are peripheral populations. Domesticated emmer was shown to cluster into four populations numbered 4 (India, Oman, and Ethiopia), 5 (Spain, Italy, Palestine, Israel, Syria, and Lebanon), 6 (Balkans, north-western Turkey, and Russia), and 7 (north-eastern Turkey, Transcaucasia, Dagestan, and Iran) (Luo *et al.* 2007). Using RFLP to measure diversity, the most diverse population was the Mediterranean population (No. 5) (Luo *et al.* 2007) and the least diverse were the eastern Turkish and Transcaucasian populations (No. 7).

Domestication results in a diversity bottleneck, and crops are therefore less diverse than their wild ancestors. Crops were reported to retained on average about two-thirds of DNA diversity present in their wild ancestors (Buckler *et al.* 2001; Wright *et al.* 2005; Hyten *et al.* 2006). Diversity of domesticated emmer was lower than diversity of wild emmer; θ_w of domesticated emmer was 40% and θ_π was 52% of that in wild emmer (Table 1). The loss of diversity was slightly greater than the one-third postulated by Buckler *et al.* (2001). However, compared to wild emmer in the Diyarbakir region, domesticated emmer was actually more diverse; θ_w was 129% and θ_π was 164% of the same measures in wild emmer in the Diyarbakir region, respectively (Table 1). This is an obvious paradox. However, the same paradox was observed in domesticated einkorn, which also shows no loss of diversity compared to the ancestral wild population (Kilian *et al.* 2007).

Mutation rates in nucleotide sequences of genes are too low for new mutations to have meaningfully altered nucleotide diversity of emmer since its domestication 10 000 years ago and to account for this paradox. The only other source of diversity that could have altered diversity of domesticated emmer was diversity that existed in wild emmer. If gene flow from wild emmer was an important source of diversity for domesticated emmer, domesticated emmer population 5 should be more diverse than the remaining domesticated emmer populations because it is sympatric with the most diverse wild emmer population. An additional prediction is that genetic distance of domesticated emmer population 5 will be shorter to wild emmer in the south-western tip of Fertile Crescent then to wild emmer in the Diyarbakir region. Both

Table 1. Average nucleotide polymorphism θ_w, nucleotide diversity (θ_π) and Tajima's *D* in wild and domesticated emmer, *Triticum aestivum* and and *Aegilops tauschii*

Species	Population	Genomes	Lines	Loci	$\theta_w \times 10^{-3}$	$\theta_\pi \times 10^{-3}$	Tajima's *D*
Wild emmer	entire taxon	A	49	14	2.15	2.26	–0.32
Wild emmer	entire taxon	B	49	11	2.56	2.32	–0.44
Wild emmer	entire taxon	AB	49	25	2.33	2.28	–0.38
Wild emmer[1]	Diyarbakir	AB	10	1174	0.73	0.72	–0.03
Dom. emmer	entire taxon	AB	23	24	0.94	1.18	0.44
T. aestivum[1]	entire taxon	AB	13	1174	0.58	0.57	–0.05
T. aestivum[1]	entire taxon	D	13	679	0.22	0.18	–0.57
Ae. tauschii	entire taxon	D	44	26	2.44	3.14	0.76

[1]from Akhunov *et al.* (2010)

predictions were experimentally confirmed, and direct evidence was obtained for gene flow between sympatric populations of wild and domesticated emmer (Luo *et al.* 2007). We therefore conclude that the domesticated emmer diversity paradox is caused by gene flow from wild emmer to domesticated emmer that increased domesticated emmer diversity and altered the original diversity pattern. The consequence is geographical superimposition of the most diverse domesticated emmer population on the most diverse wild emmer population and that the centre of emmer diversity does not coincide with the centre of its origin.

Diversity of *T. aestivum* and its evolution

Nucleotide diversity (θ_π) in the *T. aestivum* A and B genomes was 0.57×10^{-3} (Table 1), which is close to $\theta_\pi = 0.8 \times 10^{-3}$ reported previously (Haudry *et al.* 2007). Using our estimates, the *T. aestivum* A and B genomes acquired about a half (48%) of diversity present in domesticated emmer. While this represents a substantial amount of diversity present in the ancestral tetraploid, it is less than that reported by Haudry *et al.* (2007), who found that diversity in the *T. aestivum* A and B genomes equalled that in domesticated emmer. Variation due to sampling and underestimation of diversity present in domesticated emmer pointed out earlier are most likely causes of the finding reported by Haudry *et al.* (2007).

Because *T. aestivum* originated only about 8500 years ago (Nesbitt & Samuel 1996), most of its diversity had to originate during the evolution of its ancestors and be acquired via either (1) multiple origins of hexaploid wheat or (2) gene flow via pentaploid hybrids originating by hybridization of hexaploid wheat with tetraploid wheat. If scenario (1) were true, diversity of the three *T. aestivum* genomes should be proportionate to diversity of the ancestors, domesticated emmer and *Ae. tauschii*. Because nucleotide diversity of *Ae. tauschii* is nearly three times as great as that of domesticated emmer (Table 1), the D genome should be more diverse than the A and B genomes. While the A and B genomes of *T. aestivum* have 48% of nucleotide diversity (measured as θ_π) present in domesticated emmer, the D genome has only 6% of that present in *Ae. tauschii* (Table 1). This great difference between the expected and observed diversity in the *T. aestivum* A, B and D genomes makes the mechanism (1) an unlikely source of *T. aestivum* diversity. In contrast, the mechanism (2) is likely because the magnitude of diversity in the *T. aestivum* genomes parallels the likelihood of hybridization of *T. aestivum* with its ancestors (Akhunov *et al.* 2010).

Tajima's D, which is close to zero in the *T. aestivum* A and B genomes but negative in the D genome (Table 1) suggests that the origin of diversity in the *T. aestivum* A and B genomes on the one hand and in the D genome on the other hand had different dynamics. In the A and B genomes gene flow from tetraploid wheat resulted in allele frequencies consistent with neutral evolutionary history, meaning that gene flow kept pace with the expansion of the *T. aestivum* population. The highly negative Tajima's *D* in the D genome suggests very limited gene flow from *Ae. tauschii*, consistent with rare hybridization events not keeping pace with the expansion of the *T. aestivum* population. Thus,

Table 2. Expected heterozygosity (He) based on RFLP in the A and B genomes of common wheat (*T. aestivum* ssp. *aestivum*) and club wheat (*T. aestivum* ssp. *compactum*)

Species	Germplasm	Geographical region	Lines	He
T. aestivum ssp. *aestivum*	landraces	China	68	0.019
	landraces	Iran, Transcaucasia	77	0.026
	landraces	Turkey	56	0.040
	cultivars	North America	18	0.024
T. aestivum ssp. *compactum*	landraces	Afghanistan, Pakistan, Iran	20	0.021
	landraces	Transcaucasia, Turkey	15	0.030
	landraces	Ethiopia	10	0.023
	landraces and cultivars	Austria, Switzerland	35	0.002

the allelic frequencies in the *T. aestivum* genomes, like the magnitude of diversity, indicate that gene flow from tetraploid wheat to *T. aestivum* was continuous and principal source of *T. aestivum* diversity.

Cultivated tetraploid and hexaploid wheat was often grown side by side in landraces and the former is therefore a logical source of enhanced diversity in the *T. aestivum* A and B genomes compared to the D genome. Wild emmer has also been implicated in the origin of genetic diversity of the A and B genomes of *T. aestivum* (Dvorak *et al.* 2006). Sympatry of *T. aestivum* with wild emmer could have taken place principally to the west of the centre of *T. aestivum* origin in Transcaucasia and north-western Iran. The prolonged sympatry with domesticated and wild tetraploid wheat predicts that the greatest diversity in *T. aestivum* will be in Turkey. This was observed both in common wheat (*T. aestivum* ssp. *aestivum*) and club wheat (*T. aestivum* ssp. *compactum*) (Table 2). Gene flow from domesticated and wild tetraploid wheat in Turkey is also a likely candidate for subdivision of *T. aestivum* population into the western and eastern groups indicated by individual loci (Tsunewaki 1966, 1968; Dvorak *et al.* 2006) and clustering with simple sequence repeat (SSR) markers (Balfourier *et al.* 2007). The consequence of gene flow from tetraploid wheat to *T. aestivu*m is that the centre of molecular diversity of *T. aestivum* coincides poorly with the centre of morphological diversity as described by Vavilov.

CONCLUSIONS

Diversity in domesticated emmer and *T. aestivum* was shaped by gene flow from the ancestors, which altered the original diversity pattern present at the time of domestication. New centres of diversity evolved that do not reflect the geography of crop origin but parallel the geography of diversity in the progenitor. The maximum diversity in domesticated emmer is in the Mediterranean basin because domesticated emmer was sympatric with the most diverse population of wild emmer in the eastern Mediterranean. The maximum diversity in *T. aestivum* is in Turkey because there *T. aestivum* was sympatric with domesticated tetraploid wheat and wild emmer. Displacements of maximum diversity in a crop relative to the centre of its origin due to gene flow from the progenitor should be expected every time when a crop originates in a peripheral or specialized population of the progenitor. Such crops will likely be equally or more diverse than the populations from which they descended. That situation exists both in domesticated einkorn and domesticated emmer. Only if domestication takes place in the centre of diversity of the ancestor, geography of crop diversity and geography of its origin will coincide and the diversity pattern may agree with Vavilov's hypothesis.

Acknowledgements. This work has been supported by National Science Foundation Plant Genome Research Program award No. DBI-0321757. The authors express thanks to Dr. A.R. Akhunova for technical assistance with DNA sequencing and Dr. P. Mc Guire for critically reading the manuscript.

References

Akhunov E.D., Akhunova A.R., Dvorak J. (2005): BAC libraries of *Triticum urartu*, *Aegilops speltoides* and *Ae. tauschii*, the diploid ancestors of polyploid wheat. Theoretical and Applied Genetics, **111**: 1617–1622.

Akhunov E.D., Akhunova A.R., Anderson O.D., Anderson J.A., Blake N., Clegg M.T., Coleman-Derr D., Conley E.J., Crossman C.C., Deal K.R., Dubcovsky J., Gill B.S., Gu Y.Q., Hadam J., Heo H., Huo N., Lazo G.R., Luo M.C., Ma Y.Q., Matthews D.E., McGuire P.E., Morrell P.L., Qualset C.O., Renfro J., Tabanao D., Talbert L.E., Tian C., Toleno D.M., Warburton M.L., You F.M., Zhang W., Dvorak J. (2010): Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. BMC Genomics, **11**: 702.

Balfourier F., Roussel V., Strelchenko P., Exbrayat-Vinson F., Sourdille P., Boutet G., Koenig J., Ravel C., Mitrofanova O., Beckert M., Charmet G. (2007): A worldwide bread wheat core collection arrayed in a 384-well plate. Theoretical and Applied Genetics, **114**: 1265–1275.

Buckler E.S., Thornsberry J.M., Kresovich S. (2001): Molecular diversity, structure and domestication of grasses. Genetical Research, **77**: 213–218.

Dvorak J., Luo M.-C. (1999): Evolution of free-threshing and hulled forms of *Triticum aestivum*: old problems and new tools. In: Caligari P.D.S. , Brandham P.E. (eds): Wheat Taxonomy: the Legacy of John Percival. Academic Press, London, 127–136.

Dvorak J., Akhunov E.D. (2005): Tempos of deletions and duplications of gene loci in relation to recombina-

tion rate during diploid and polyploid evolution in the Aegilops-Triticum alliance. Genetics, **171**: 323–332.

Dvorak J., Luo M.C., Yang Z.L., Zhang H.B. (1998): The structure of the *Aegilops tauschii* genepool and the evolution of hexaploid wheat. Theoretical and Applied Genetics, **97**: 657–670.

Dvorak J., Akhunov E.D., Akhunov A.R., Deal K.R., Luo M.C. (2006): Molecular characterization of a diagnostic DNA marker for domesticated tetraploid wheat provides evidence for gene flow from wild tetraploid wheat to hexaploid wheat. Molecular Biology and Evolution, **23**: 1386–1396.

Gokgol M. (1941): Über die Genzentrentheorie und den Ursprung des Weizens. Zeitschrift für Pflanzenzüchtung, **23**: 562–578.

Gordon D. (2004): Viewing and editing assembled sequences using consed. In: Baxevanis A.D. , Davison D.B. (eds): Current Protocols in Bioinformatics. John Wiley & Co., New York, 11.12.11–11.12.43.

Green P. (1998): swat/cross_match/phrap package. Available at http://bozeman.mbt.washington.edu/phrap.docs/phrap.html.

Harlan J.R. (1971): Agricultural origins: centers and noncenters. Science, **174**: 468–474.

Harlan J.R. (1975): Crops and Man. American Society of Agronomy, Madison.

Haudry A., Cenci A., Ravel C., Bataillon T., Brunel D., Poncet C., Hochu I., Poirier S., Santoni S., Glemin S., David J. (2007): Grinding up wheat: a massive loss of nucleotide diversity since domestication. Molecular Biology and Evolution, **24**: 1506–1517.

Heun M., Schafer-Pregel R., Klawan D., Castagna R., Accerbi M., Borghi B., Salamini F. (1997): Site of einkorn wheat domestication identified by DNA fingerprinting. Science, **278**: 1312–1314.

Huang S., Sirikhachornkit A., Su X., Faris J., Gill B.S., Haselkorn R., Gornicki P. (2002): Genes encoding plastid acetyl-CoA carboxylase and 3-phopshoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. Proceedings of the National Academy of Sciences of the United States of America, **99**: 8133–8138.

Hyten D.L., Song Q.J., Zhu Y.L., Choi I.Y., Nelson R.L., Costa J.M., Specht J.E., Shoemaker R.C., Cregan P.B. (2006): Impacts of genetic bottlenecks on soybean genome diversity. Proceedings of the National Academy of Sciences of the United States of America, **103**: 16666–16671.

Kihara H. (1944): Discovery of the DD-analyser, one of the ancestors of *Triticum vulgare* (Japanese). Agriculture and Horticulture (Tokyo), **19**: 13–14.

Kilian B., Ozkan H., Walther A., Kohl J., Dagan T., Salamini F., Martin W. (2007): Molecular diversity at 18 loci in 321 wild and 92 domesticate lines reveal no reduction of nucleotide diversity during *Triticum monococcum* (einkorn) domestication: Implications for the origin of agriculture. Molecular Biology and Evolution, **24**: 2657–2668.

Luo M.C., Yang Z.L., You F.M., Kawahara T., Waines J.G., Dvorak J. (2007): The structure of wild and domesticated emmer wheat populations, gene flow between them, and the site of emmer domestication. Theoretical and Applied Genetics, **114**: 947–959.

Luo M.C., Deal K.R., Akhunov E.D., Akhunova A.R., Anderson O.D., Anderson J.A., Blake N., Clegg M.T., Coleman-Derr D., Conley E.E., Crossman C.C., Dubcovsky J., Gill B.S., Gu Y.Q., Hadam J., Heo H.Y., Huo N., Lazo G., Ma Y.Q., Matthews D.E., McGuire P.E., Morrell P.L., Qualset C.O., Renfro J., Tabanao D., Talbert L.E., Tian C., Toleno D.M., Warburton M.L., You F.M., Zhang W.J., Dvorak J. (2009): Genome comparisons reveal a dominant mechanism of chromosome number reduction in grasses and accelerated genome evolution in *Triticeae*. Proceedings of the National Academy of Sciences of the United States of America, **106**: 15780–15785.

McFadden E.S., Sears E.R. (1946): The origin of *Triticum spelta* and its free-threshing hexaploid relatives. Journal of Heredity, **37**: 81–89, 107–116.

Nesbitt M., Samuel D. (1996): From staple crop to extinction? The archaeology and history of hulled wheats. In: Padulosi S., Hammer K., Heller J. (eds): Hulled Wheats. Promoting the Conservation and Use of Underutilized and Neglected Crops. 4. Proc. 1st Int. Workshop Hulled Wheats. International Plant Genetic Resources Institute, Rome, Castelvecchio Pacoli, Tuscany, 41–100.

Ozkan H., Brandolini A., Pozzi C., Effgen S., Wunder J., Salamini F. (2005): A reconsideration of the domestication geography of tetraploid wheat. Theoretical and Applied Genetics, **110**: 1052–1060.

Schiemann E. (1939): Gedanken zur Genzentrentheorie Vavilovs. Naturwissenschaften, **27**: 377–383.

Tajima F. (1983): Evolutionary relationship of DNA sequences in finite populations. Genetics, **105**: 437–460.

Tajima F. (1989): Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics, **123**: 585–595.

Tsunewaki K. (1966): Comparative gene analysis of common wheat and its ancestal species. II. Waxiness, growth habit and awnedness. Japanese Journal of Botany, **19**: 175–229.

Tsunewaki K. (1968): Origin and phylogenetic differentiation of common wheat revealed by comparative gene analysis. In: Finley K.W , Shepherd K.W. (eds): 3rd Int. Wheat Genetic Symposium. Australian Academy of Sciences, Canberra, 71–85.

Vavilov N.I. (1926): Studies on the origin of cultivated plants. (Russian) Bulletin of Applied Botany and Plant Breeding, **14**: 1–245.

Vavilov N.I. (1992): Origin and Geography of Cultivated Plants. Cambridge University Press, Cambridge.

Watterson G.A. (1975): On the number of segregating sites in genetical models without recombination. Theoretical Population Biology, **7**: 256–276.

Wheat SNP Database. Available at http://probes.pw.usda.gov:8080/snpworld/Search

Wright S.I., Bi I.V., Schroeder S.G., Yamasaki M., Doebley J.F., McMullen M.D., Gaut B S. (2005): The effects of artificial selection on the maize genome. Science, **308**: 1310–1314.

Xu Z., Deal K.R., Li W., Covaleda L., Chang Y.-L., Dvorak J., Luo M.-C., Gill B.S., Anderson O.D., Zhang H.B. (2002): Construction and characterization of five large-insert BAC and BIBAC libraries of *Aegilops tauschii*, the diploid donor of the wheat D genome. In: Heller S.R. (ed.): 10th Int. Plant and Animal Genome Conference. Scherago International, Inc., San Diego.

Zohary D. (1970): Centers of diversity and centers of origin. In: Frankel O.H., Bennett E. (eds): Genetic Resources in Plants – their Exploration and Conservation. Blackwell Science Publishing, Oxford, 33–42.

Zohary D., Hopf M. (1994): Domestication of Plants in the Old World. Clarendon Press, Oxford.