Mol Gen Genomics
DOI 10.1007/s00438-006-0136-6

ORIGINAL PAPER

Haplotype structure at seven barley genes: relevance to gene pool bottlenecks, phylogeny of ear type and site of barley domestication

Benjamin Kilian · Hakan Özkan · Jochen Kohl · Arndt von Haeseler · Francesca Barale · Oliver Deusch · Andrea Brandolini · Cemal Yucel · William Martin · Francesco Salamini

Received: 18 October 2005 / Accepted: 24 April 2006

Abstract Archaeological remains indicate that the origin of western agriculture occurred in a brief period about 10,500 years ago in a region of the Middle East known as the Fertile Crescent, where the wild progenitors of several key agricultural cereal species are endemic. Domestication entailed the appearance of agronomic traits such as seed size and threshability. For a representative sample of 20 domesticated barley (*Hordeum vulgare*) lines, including 13 two-rowed and 7 six-rowed varieties, we determined the haplotypes at seven loci—*Adh2*, *Adh3*, *Amy1*, *Dhn9*, *GAPDH*, *PEPC* and *WAXY* encompassing 5,616 bases per line—and compared them to the haplotypes at the same loci for 25 wild forms (*Hordeum spontaneum*) collected within and outside the Fertile Crescent. In comparisons of wild versus domesticated barley, the number of haplotypes (70 vs. 17), average nucleotide diversity, π, (0.0077 vs. 0.0028), and Watterson's theta at silent sites (0.0104 vs. 0.0028) was reduced in domesticated lines. Two loci, *Amy1* and *PEPC*, were monomorphic in domesticated lines; *Amy1* and *GAPDH* produced significant values of Tajima's D. At *GAPDH*, π was slightly higher in domesticated than wild forms, due to divergent high-frequency haplotypes; for the remaining six loci, 87% of nucleotide diversity has been lost in the domesticated forms. Bottlenecks acting on neutrally evolving loci either during the domestication process, during subsequent breeding, or both, are sufficient to account for reduced diversity and the results of Tajima's test, without the need to evoke selection at these loci. Phylogenetic networks data uncover distinct wild and domesticated barley genotypes and suggest that barley may have been domesticated in the Jordan valley. Because, based on AFLP data, the domesticated Turkish cultivars had a genetic basis as large as that present in large germplasm collections, all comparisons provided in this paper are of general value more than being restricted to the Turkish barley germplasm.

Electronic Supplementary Material Supplementary material is available to authorised users in the online version of this article at http://dx.doi.org/10.1007/s00438-006-0136-6.

Communicated by R. McCombia

Manuscript information: Sequence data from this article have been deposited in GenBank Data library under accession nos. DQ195928 to DQ196067.

B. Kilian · F. Salamini
Max-Planck-Institute for Plant Breeding Research,
Carl-von-Linné-Weg, 10, 50829 Koeln, Germany

H. Özkan · C. Yucel
Department of Field Crops, Faculty of Agriculture,
University of Cukurova, 01330 Adana, Turkey

J. Kohl · A. von Haeseler
Institute for Bioinformatics, Heinrich-Heine-Universitaet
Duesseldorf, Universitätsstr., 1, 40225 Düsseldorf, Germany

A. Brandolini
Istituto Sperimentale per la Cerealicoltura - CRA,
Via Mulino, 3, 26866 S. Angelo Lodigiano, (LO), Italy

B. Kilian · O. Deusch · W. Martin
Institute for Botany III, Heinrich-Heine-Universitaet
Duesseldorf, Universitätsstr., 1, 40225 Düsseldorf, Germany

F. Barale · F. Salamini (✉)
Fondazione Parco Tecnologico Padano,
Via Einstein-Località Cascina Codazza, 26900 Lodi, Italy
e-mail: francesco.salamini@tecnoparco.org

Keywords Barley domestication · Haplotypes · Nucleotide diversity · Domestication bottlenecks

Introduction

The domestication of grasses began during "the Neolithic revolution", about 12,000 years before present (BP), when humans living as hunter-gatherers became sedentary food producers (Diamond 1997; Salamini et al. 2002). In genetic terms, grass domestication involved human counterselection against unfavorable alleles at loci governing flowering time, rachis brittleness, naked seeds, and seed size, accompanied by the accumulation of small genetic effects at quantitative trait loci (QTLs) that, collectively, confer yield increase, reduction in plant height, changes in tillering, inflorescence, and plant architecture (Salamini et al. 2002). Domestication traits were probably selected as a grass-specific set (Buckler et al. 2001), thus allowing a genetically "convergent" domestication across grasses, as demonstrated for maize, rice and sorghum (Paterson et al. 1995). The conversion from brittle to non-brittle rachis, the transition from distichous to polystichous spike, and the appearance of the naked caryopsis (nude) character (Søgaard and von Wettstein-Knowles 1987; Harlan 1976) were the main genetic and morphological events that accompanied the development of domesticated barley, *Hordeum vulgare*, from its wild progenitor *Hordeum spontaneum*. The domestication of barley is thought to have occurred in the Fertile Crescent (Badr et al. 2000; Salamini et al. 2002; Morrell et al. 2003) although varieties with naked seeds appear quite early in the Himalayan region, probably as the result of human dispersal and introgression (Badr et al. 2000).

Considerable allelic and haplotype diversity is found among wild barley populations (Nevo et al. 1979, 1986a, b; Snow and Brody 1984; Jana et al. 1987; Chalmers et al. 1992; Dawson et al. 1993; Badr et al. 2000; Lin et al. 2001, 2002; Morrell et al. 2003, 2005). Evidence favoring a monophyletic, or possibly diphyletic, domestication of barley in the Jordan valley (Badr et al. 2000; Salamini et al. 2004) would predict a reduction of nucleotide diversity in alleles found among gene pools of domesticated versus wild *Hordeum* accessions in the wake of domestication bottlenecks. Reduced polymorphism following domestication effect can easily be misidentified as a signature of selection (Tenaillon et al. 2004; Wright and Gaut 2005; Wright et al. 2005). Domestication thus involves bottlenecks and the fixation of particular allele combinations during the initial domestication process, but is followed by the further reduction of variability at linked loci that have been selected by breeders since the domestication process. This reduction is a function of the rate of recombination between the selected sites and the linked loci surveyed. This hitchhiking effect can be detected as a skew in frequencies of molecular markers (Vigouroux et al. 2002) when compared to frequencies expected under an equilibrium-neutral model.

Only a few hundred effective meiotic cycles (those resulting from natural hybridization events or intentionally carried out by plant breeders) have occurred during the roughly 10,000 years endured by cereal germplasm in the domesticated condition (Paterson 2002). Accordingly, short chromosomal regions may exist in linkage disequilibrium (Paterson 2002; Rafalski 2002; Morgante and Salamini 2003), indicating that mutant alleles with significant effects on phenotypic traits may have been quickly fixed at several loci both early in the domestication process and during intentional breeding. In such a situation, small DNA regions flanking domestication-related or breeding-related loci are characterized by low levels of diversity (linkage drag). In a gene pool, the existence of domestication and breeding-related effects on the extent of natural variation can be detected by SNP loci, which permit assessment of haplotype diversity at specific loci (Schneider et al. 2001). Here we examine nucleotide diversity and haplotype combinations at seven loci in domesticated barley in comparison to that in *H. spontaneum*, the wild progenitor species.

Materials and methods

Plant material

The correct taxonomical terminology for domesticated and wild barleys is *H. vulgare* subsp. *vulgare* and *H. vulgare* subsp. *spontaneum*, respectively. However, in all recent papers on this subject as well as in the reference book by Zohary and Hopf (2000), the wild is always named as *H. spontaneum*; in this article, we have followed the last taxonomical indication.

The plant material used in this study is listed in Table 1. The 20 domesticated (D) lines represent barley varieties currently cultivated in Turkey, covering a long period of Turkish plant breeding. They were selected among 33 lines available and were chosen on the basis of maximum genetic distance to one another while also considering their morphological variation (see Results and discussion). The 25 *H. spontaneum* lines (W) were those considered by Lin et al. (2001) and reported to span the native range of the wild species, including the

Table 1 Barley lines investigated in the present study

Line no.	Domesticated varieties	Breeding Institute	Year released	Use and rows in ear	Origin
3	Tokak 157/37	FCCRI-A	1937	F 2	Turkish land race
4	Kral 97	BDIWCRI-K	1997	F 6	Land race
5	Avci 2002	FCCRI-A	2002	F 6	Complex cross
6	Yesilköy 387	–	–	F 6	From Zogen 160, land race from Kirklarhi
7	Aydanhanim	FCCRI-A	2002	M 2	Cross of Omega x Tarm92
8	Hamidiye 85	AARI-E	1985	F 2	Tokak mutant
11	Cetin 2000	FCCRI-A	2000	F 6	Line 4875 from Iran
13	Zafer 160	FCCRI-A	–	F 6	Local land race
15	Cumra 2001	AEBMC-K	2001	M 2	Tokak mutant
16	Angora	ABMSIC-K	1999	M 2	Complex cross involving 6 lines
17	Erginel 90	AARI-E	1990	F 6	Cross of Escourgen × Hop21H (France)
20	Karatay 94	BDIWCRI-K	1994	MF 2	Complex cross involving 5 lines
21	Tarm 92	FCCRI-A	1992	MF 2	Land race
23	Yesevi 93	FCCRI-A	1993	F 2	Land race
24	Kalayci 97	AARI-E	1997	F 2	Cross Erginel × Tokak
26	Efes 1	ABMSIC-K	–	ND 2	Unknown pedigree
27	Sladoran	TARI-E	1998	M 2	Introduction from Yugoslavia
30	Sahin 91	SAARI-D	1991	F 2	Unknown pedigree
31	Aday 4	–	–	ND 6	Unknown pedigree
33	Balkan 96	TARI-E	1996	M 2	Unknown pedigree

Line No.[1]	Wild varieties[a]	PI No.[b]	Country of origin	Geographical region
34	2	212305	Afghanistan	E
35	3	212306	Afghanistan	E
36	4	219796	Iraq	Z
37	6	220523	Afghanistan	E
38	9	236388	Syria	W
39	10	253933	Iraq	Z
40	11	254894	Iraq	Z
41	12	268242	Iran	Z
42	13	293402	Turkmenistan	E
43	16	293409	Turkmenistan	E
44	17	293411	Tajikistan	E
45	21	296926	Israel	W
46	22	366446	Afghanistan	E
47	24	401370	Iran	E
48	25	401371	Iran	Z
49	27	406276	Israel	W
50	28	420911	Jordan	W
51	30	420913	Jordan	W
52	32	420916	Jordan	W
53	35	466460	Israel	W
54	36	531851	Israel	W
55	38	531853	Israel	W
56	39	531857	Israel	W
57	43	559556	Turkey	Z
58	44	560559	Turkey	Z

[1] Line number used in this study

More details in supplementary Table 1

Abbreviations are *FCCRI-A* Field Crops Central Research Institute, Ankara; *BDIWCRI-K* Bahri Dagdas International Winter Cereals Research Institute, Konya; *AARI-E* Anatolian Agricultural Research Institute, Eskisehir; *AEBMC-K* Anatolian Efes Beer and Malt Company, Konya; *ABMSIC-K* Anatolian Beer Malt and Southearn Industry Company, Konya; *TARI-E* Thrace Agricultural Research Institute, Edirne; *SAARI-D* Southeastern Anatolian Agricultural Research Institute, Diyarbakir; *F* feed; *M* malting; *ND* not described; *E* Eastern Fertile Crescent; *W* Western Fertile Crescent; *Z* Zagros (Morrell et al. 2003)

[a] Line numbers as listed in the PNAS Supporting Information to Lin et al. (2001)

[b] Plant introduction no.

Jordan valley, the putative site of barley domestication (Badr et al. 2000). Molecular variation found in the D lines, based on AFLP markers, was compared to that present in three groups of domesticated lines: two (20 lines each) were from the 67 lines cited in Castiglioni et al. (1998) and considered by Badr et al. (2000); the third group included the following 21 cultivars, representing a southern Europe gene pool of barley: Alexis, Angora, Apex, Arco, Aura, Betzes, Carina, Cherie, Express, Gitane, Jador, Magda, Mirko, Nudinka, Nure, Onice, Prisma, Proctor, Rebelle, Trebbia and Tremois.

The extent of phenotypic variation across D varieties was evaluated by an experiment carried out in two locations (upland and lowland conditions) in the Adana area (Mediterranean region, 37°21′ N and 35°10′ E), during the 2003–2004 growing season and under rainfed conditions. Each line was grown in 1 m row, in a randomized complete block design with three replications. All traits were recorded on ten individual plants.

Extraction of genomic DNA, generation of PCR primers and PCR amplification

Genomic DNA was isolated from silica-dried single leaves of each line with the Qiagen DNeasy Plant Mini Kit (Qiagen, Hilden, Germany), according to the manufacturers instructions. The Primer3 online software (primer3_www.cgi v 0.2, Whitehead Institute for Biomedical Research, Cambridge, UK) (Rozen and Skaletsky 2000) was used to design primers from published DNA sequences. Oligonucleotides were purchased from OPERON Biotechnologies (Cologne, Germany); their sequences are available in Supplementary Table 2.

DNA amplifications were performed in a 25 µl volume. The reaction mix contained about 100 ng of genomic DNA, 0.4 µM of each primer, 125 µM of each dNTP (AB gene, Surrey, UK), 3 mM $MgCl_2$, 4% DMSO and 1 unit *Taq* DNA polymerase. The reactions were incubated in a PTC-225 Tetrad Thermal Cycler (MJ Research) with the following cycling conditions: 94°C for 3 min, 29–31 cycles of 30 s at 94°C, 40 s at 60.5–64°C, 50–65 s at 72 °C (depending on the gene, see Table 1) followed by a final extension step of 6 min at 72°C. PCR products were separated by agarose gels electrophoresis and recorded as presence (1) or absence (0) of the amplified fragment.

Pre-screening for polymorphisms by non-denaturing gel electrophoresis

PCR products were digested, denaturated for 3 min at 94°C, and characterized by SSCP-analysis, as described by Schneider et al. (1999, 2001). Electrophoresis of SSCP-gels (28.5 cm ×25 cm×0.5 mm) was performed at room temperature with a constant power (1.0 W) for 12–16 h, the DNA fragments were visualized by silver staining.

Purification and sequencing of PCR products

Selected lines, representative of each haplotype for each locus, were selected based on SSCP-analysis. The corresponding PCR products were purified by ExoSAP-IT enzyme mixture (USB, Cleveland, USA) according to the protocol provided, and were sequenced directly on both strands on an Applied Biosystems (Weiterstadt, Germany) ABI Prism 3730xL sequencer using BigDye terminators. The amplification conditions for the seven genes in the domesticated barley lines are listed in Table 2. Published sequence data for the seven loci from 25 wild barley lines (*H. spontaneum* C. Koch) were obtained from GenBank. The definition of the molecular state for each haplotype of each gene sequenced was supported by multiple sequencings of the same haplotype from different genotypes.

AFLP genotype fingerprinting

The AFLP procedure of Zabeau and Vos (1993) was adopted. A total of seven primer combinations (Table 3, note 1) were used to amplify *EcoRI*- and *MseI*- digested DNA. Autoradiographs were scored for presence versus absence of polymorphic amplified DNA fragments.

SNP-detection

DNA sequences were processed with AB DNA Sequencing Analysis Software 5.1.1 (Data Collection Software version 2.0) and later manually edited by BioEdit version 7.0.1 (Hall 1999). Sequence alignments were generated in BioEdit, and the allelic haplotypes were defined.

Data analysis and statistics

Nucleotide diversity π (Nei 1987), Tajima's D (Tajima 1989), and Watterson's theta (θ_w) were calculated by DNASP v4.00 (Rozas 2003). Exon and intron sequence positions were derived from published data of *H. spontaneum* (Lin et al. 2001; Morell et al. 2003). DNA sequences of the seven genes for each of the 20 domesticated varieties and of the 25 wild lines were analyzed both individually and as a concatenated data set with a

Table 2 Barley genes and conditions used to amplify them in 20 domesticated barley varieties

Gene	Symbol	Chromosomal location	Accession numbers[a]	bp	Primer combination[b]	Annealing temperature (°C)	Elongation time (s)
Alcohol dehydrogenase 2	Adh2	4H	AY184931-955	837	B104-B106	61	50
Alcohol dehydrogenase 3	Adh3	7H	AF326691-715	825	B201-B202	61	50
Alpha-amylase 1	Amy1	6H	AY349195-219	663	B306-B307	62	50
Dehydrin 9	Dhn9	5H	AY349247-271	753	B503–B504	60.5	50
Glyceraldehyde-3-phosphate DH	GAPDH	6H	AY349298-322	765	B604–B605	61	50
Phosphoenolpyruvate carboxylase	PEPC	Nm	AY349272-297	945	B704–B705	60	65
Granule bound starch synthase	Waxy	7H	AY349323-349	828	B806–B807	64.5	50

The published sequence for each gene and start-end positions for the amplified fragments in brackets are: (*Adh1*) AY184953 (1121-1957); (*Adh3*) AF326715 (109-923); (*Amy1*) AY349219 (86–748); (*Dhn9*) AY349270 (54–800); (*GAPDH*) AY349294 (197–960); (*PEPC*) AY349320 (77–1021); (*Waxy*) AY349344 (327–1151). Sequence data have been deposited in GenBank Data library under accession nos DQ195928 to DQ196067

Nm Not mapped, *Bp* base pairs amplified

[a] As listed in Lin et al (2001, 2002), Morrell et al (2003)

[b] See Supplementary Table 2

Table 3 Interval of variation in two contrasting environments for spike and plant traits and for AFLP molecular markers in 20 domesticated barley varieties considered in this paper, as well as in 21 varieties from a Southern Europe gene pool

Trait	Two-rowed (20 varieties)	Six-rowed (20 varieties)
Spike length	7.4–14.0	5.0–13.2
Spikelets/spike	22–40	39–90
Seeds/spike	19–37	28–83
Grain weight	0.72–2.46	1.32–4.37
Plant heigth	40–134	40–126

AFLP primer combination	% of polymorphic AFLP bands[b]			
	This paper		Badr et al.[a]	
	Turkish gene pool	Southern Europe gene pool	First group	Second group
1	23.7	44.4	44.8	45.8
2	76.0	49.6	50.0	50.5
3	49.1	47.5	57.7	53.1
4	38.7	48.5	46.1	51.3
5	45.2	53.0	51.9	50.0
6	3.1	–	41.6	48.7
7	43.5	–	52.5	53.3
All	44.5	48.6	49.2	50.5

Fingerprints of Turkish domesticated lines were compared to those of the Southern Europe gene pool and to those of two groups of barley varieties studied by Badr et al. (2000)

[a] Two groups of 20 varieties were chosen at random among the 57 considered by Badr et al. (2000)

[b] Primer combinations listed (1–7) were respectively E36M40, E37M34, E37M32, E37M40, E37M45, E40M42 and E40M44 for the lines studied in this paper (Turkish and Southern Europe gene pools), and E36M40, E37M38, E41M40, E41M33, E36M44, E37M33 and E36M36 for the experiment of Badr et al. (2000)

total length of 5,616 bp. Neighbor-Net (NNet) planar graphs (Bryant and Moulton 2004) were constructed from the proportion of nucleotide differences between sequences, which was below 0.04 in all comparisons (Fig. 1a). The distinct advantage of NNet for these data lies in its ability to uncover hybridization-like events, which may occur and which go undetected or are forced to signal-averaging in bifurcating trees (Bryant and Moulton 2004).

Maximum likelihood trees (see Supplementary Fig. 1) were computed with IQPNNI v2.6 (Vinh and von Haeseler 2004) for the concatenated data, stopping at the best tree with a confidence of 95%. For each IQPNNI run the default parameters were chosen, except for the concatenated sequences with a minimum number of 10,000 iterations.

To process the data presented in Fig. 2, a matrix was created of one row for each individual line and one column for each of two possible alleles at each locus into which the major haplotypes (or AFLP data) were written, coded as single ASCII characters each. The Hamming distance (p-distance) between individual lines provides the measure of genetic identity. Individual lines that have the same collection of haplotypes are scored as identical (P=0), those differing at two alleles (of either one or two loci) are more different than those differing at one, and so forth, while those pairs sharing no haplotypes in common assume the maximum distance (P=1). This scores all differences between major haplotypes with equal weight, regardless of whether the haplotypes differ by 2 or by 20 nucleotides. NNet (Bryant and Moulton 2004) as implemented in SplitsTree 4 (Huson 1998) was used to process the p-distance data, to uncover shared similarities.

Also AFLP data from the seven primer combinations were concatenated to build a binary sequence (presence vs. absence of a band) for each of 20 domesticated variety (Fig. 2a). The phylogenetic tree was reconstructed with Tree-puzzle v5.26 (Schmidt et al. 2002) using 10,000 iterations, the two-state substitution

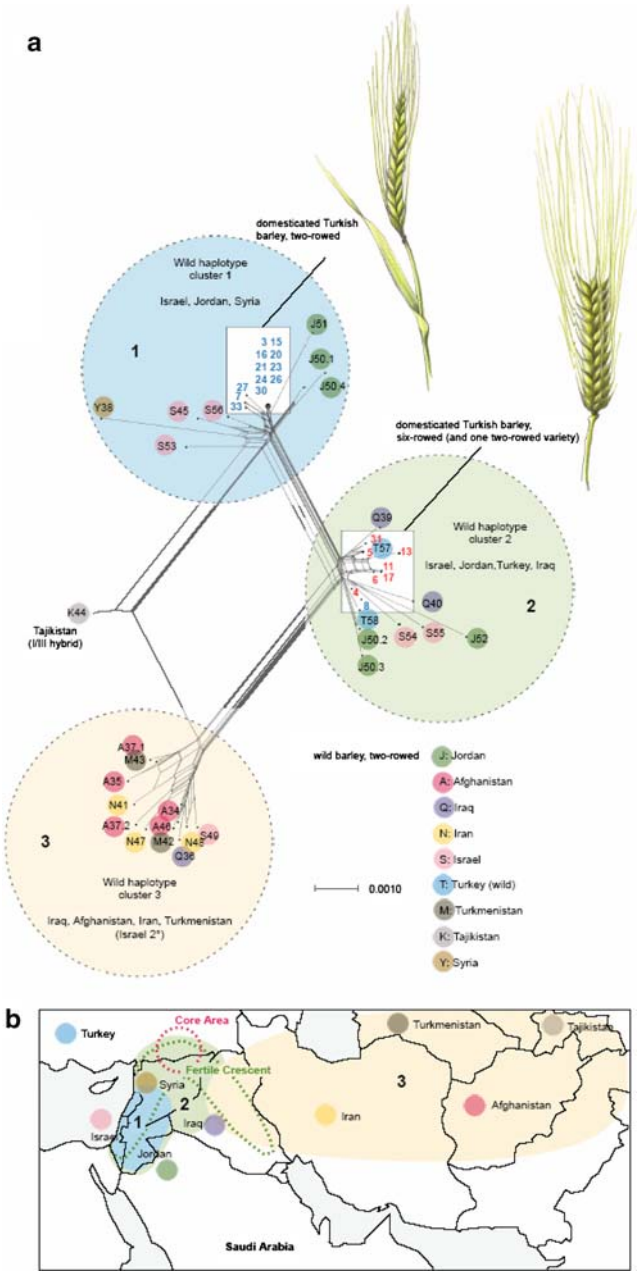

Fig. 1 Haplotype sequence relationships among wild *Hordeum spontaneum* and domesticated *H. vulgare* lines. **a** Neighbor-Net (*NNet*) planar graph of sequence similarity among 20 domesticated and 25 wild barley lines for the concatenated alignment of 5,616 sites. Line numbering corresponds to that in Table 1. Geographical origins are indicated. Domesticated lines are *boxed* and labelled. *Dotted circles* designate WHC1, -2, and -3 (see text). The *scale bar* indicates sequence divergence. **b** Map showing geographical areas relevant to this study. Color coding of regions encompassing WHC1–3 and corresponding to sources of wild lines (Morrell et al. 2003) corresponds to that in (**a**). Barley ear drawings kindly prepared by S. Kilian

model (Felsenstein 1981), assuming a uniform rate heterogeneity, and default settings except the parameter estimation, which was calculated.

Results and discussion

The domesticated barley gene pool

A representative sample of Turkish barley varieties bred during the last 30 years was investigated as the domesticated (D) gene pool. The reason to focus on Turkish varieties stems from the considerations that Turkey includes part of the primary habitats of *H. spontaneum* and shares with the Fertile Crescent the same climatic conditions. Thus, the possibility that allelic frequencies in wild and domesticated gene pools were significantly and differentially modified by environmental factors is minimized.

Out of the 33 varieties listed in the Turkish register of barley varieties, 20 were considered based on their AFLP genetic distances (data not shown). Thus, in the D group only the varieties showing the widest genetic distance were included. Care was also taken to sample representatives of two- and six-rowed varieties, which among European varieties are frequently reported to have different genetic backgrounds.

To assess to which extent the Turkish D gene pool was representative of worldwide existing D pools, two approaches were followed. In the first, both two-rowed and six-rowed varieties were grown in replicated trials, and their morphological traits were recorded. The interval of variation for five such traits within two- and six-rowed varieties is present in Table 3 (top part), pointing to the existence of significant phenotypical differences among genotypes. In the second approach, AFLP fingerprinting data of the 20 selected Turkish varieties were compared to the data recorded by Badr et al. (2000) for a wide spectrum of domesticated barley varieties. Badr et al. (2000) considered 57 out of 67 domesticated accessions described by Castiglioni et al. (1998). Their 67 genotypes were from a collection of 5,842 lines and were characterized by large differences in ear, grain and plant characters. The lines were landraces or old varieties cultivated in the Himalayan region, India, Yemen, Pakistan, Afghanistan, Turkestan, central Asia, Balkans, southern Europe, northern Europe, Ethiopia and Central Africa, America and Australia. The AFLP fingerprinting results (bottom part of Table 3) indicate the percentage of polymorphic bands (295 loci) recorded for each of 7 AFLP primer combinations tested in the 20 Turkish D varieties, compared to similar data from two groups of 20 genotypes, chosen at random among those considered by Badr et al. (2000). The average proportion of polymorphic bands—44.5% for the Turkish lines versus 49.2 and 50.5% for the two groups of 20 lines from Badr et al. (2000)—indicates that in terms of genetic

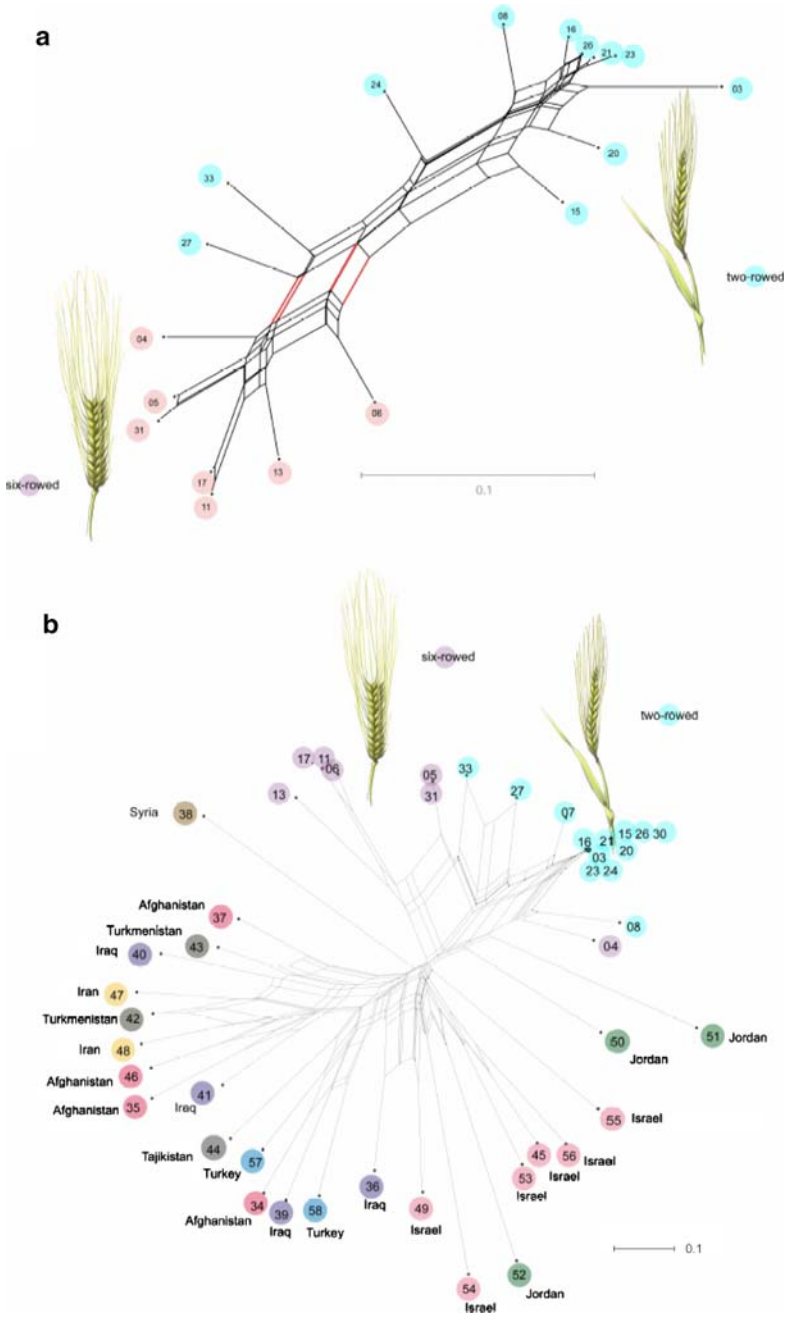

Fig. 2A Neighbor-Net (*NNet*) planar graph of Hamming distances between binaric AFLP data from seven primer combinations, 295 polymorphic loci. The B. NNet planar graph of Hamming distances between haplotypes among wild *Hordeum spontaneum* and domesticated *H. vulgare* lines. Line designations are as in Table 1 and Fig. 1

variability, the 20 domesticated lines studied are a representative sample of the domesticated germplasm available worldwide. To reinforce further this conclusion, we have introduced AFLP data from the AFLP fingerprints (5 primer combinations) of 21 domesticated lines representative of a southern Europe gene pool. The proportion of polymorphic bands was, in this case, 48.6%. The conclusion on the domesticated gene pool sampled in Turkey and studied in this paper is that its interval of molecular variation, as measured by AFLP, is in the range of those typical of large collections of barley, or of a more western varietal gene pool. This provides evidence that the comparison between wild and domesticated lines presented in this paper has a general value, more than being only restricted to the Turkish domesticated germplasm.

Loss of nucleotide diversity in domesticated barley

Gene fragments spanning a length from 663 to 945 nucleotides (Table 2) were amplified in 25 W and 20 D lines for seven loci: *Adh2*, *Adh3*, *Amy1*, *Dhn9*, *GAPDH*, *PEPC* and *WAXY*. Sequence comparisons revealed that for the same gene multiple haplotypes

existed with variable frequencies in W and D lines (Table 4). All PCR amplification products were sequenced twice, no data were included if there were differences between the two. The occurrence of the same haplotype in several lines assigned a given sequence to a specific group, designated with a Roman numeral in Table 4. The fewest number of haplotypes (five) was found at *PEPC*, the largest number (48) was found for *Adh3* (Table 4).

The domesticated lines harbor fewer haplotypes. In total, 70 different haplotypes occur among the 25 wild lines, while only 17 occur among domesticated lines. Wild lines had, on average, ten haplotypes per locus (range 4–17) whereas the domesticated lines had 2.43 (range 1–4). Among the 17 D haplotypes found, six (35%) were not present in the 25 W line sample: *Adh2*-III, *Dhn9*-III, *GAPDH*-II, *GAPDH*-III, *WAXY*-II and *WAXY*-IV. These apparently D-specific haplotypes are very likely to be present in the W gene pool, should more individuals be tested. Their absence at *WAXY*, where 17 wild haplotypes were scored, indicate that the domesticated forms have sequestered a rare wild allele.

All loci sampled revealed a reduction of π and Θ_w in D–W comparisons, except *GAPDH* (Table 5), the exception being due to the divergent haplotypes *GAPDH*-II, *GAPDH*-III, which are not present in the wild haplotype sample. For the remaining six loci, the loss of nucleotide diversity, $1 - \pi_d/\pi_w$ (Tenaillon et al. 2004) ranged from 69.2% at *WAXY* to 100% at *Dhn9* and *PEPC*, with an average of 87%. This is a substantially greater loss of nucleotide diversity than the value of 38% reported for maize domestication (Tenaillon et al. 2004). The loss of haplotype diversity, the corresponding reduction in number of different haplotypes, was 76% in the D–W comparison. Also the value of the d_{DW} statistics indicates a substantial loss of nucleotides diversity passing from wild to domesticated.

The lack of domesticated haplotype variants at *Amy1* and *PEPC* is notable, because both domesticated alleles are common (88 and 52% frequency, respectively) among the phenotypically wild (Salamini et al. 2002) lines sampled (Table 4), hence they cannot be causally associated with the domesticated phenotype. A domestication sweep (human selection) at loci closely linked to *Amy1* and *PEPC* as the possible cause of lacking polymorphism at these loci cannot be strictly excluded but is also unlikely, given the small number of domestication loci known (Salamini et al. 2002) and the circumstance that 2/7 loci sampled had one D haplotype only. Taken together, these findings point to simple bottleneck effects at *Amy1* and *PEPC*, rather than selection. However, whether this bottleneck was incurred during the initial domestication process, or during subsequent barley breeding, cannot currently be determined. Reduction of diversity in the cultivated gene pool of barley has been previously reported by Bundock and Henry (2004), Molina-Cano et al. (2005), Russell et al. (2004), Tanno and Takeda (2004).

Table 4 Haplotypes and their frequencies (%) recorded at seven loci in 25 wild and 20 domesticated lines

Haplotype	*Adh2*		*Adh3*		*Amy1*		*Dhn9*		*GAPDH*		*PEPC*		*WAXY*	
	W	D	W	D	W	D	W	D	W	D	W	D	W	D
I	28	10	4	60	88	100	20	0	15.4	60	52	100	11.1	65
II	4	70	4	15	4	0	8	65	0	10	4	0	0	15
III	0	20	4	25	4	0	0	35	0	30	4	0	3.7	10
IV	28	0	4	0	4	0	4	0	3.8	0	36	0	0	10
V	4	0	4	0	0	0	4	0	38.5	0	4	0	3.7	0
VI	4	0	4	0	0	0	4	0	3.8	0	0	0	3.7	0
VII	4	0	4	0	0	0	4	0	26.9	0	0	0	3.7	0
VIII	4	0	4	0	0	0	4	0	11.5	0	0	0	11.1	0
IX	4	0	28	0	0	0	32	0	0	0	0	0	3.7	0
X	4	0	4	0	0	0	8	0	0	0	0	0	3.7	0
XI	12	0	8	0	0	0	4	0	0	0	0	0	7.4	0
XII	4	0	8	0	0	0	4	0	0	0	0	0	3.7	0
XIII	0	0	8	0	0	0	4	0	0	0	0	0	7.4	0
XIV	0	0	4	0	0	0	0	0	0	0	0	0	3.7	0
XV	0	0	8	0	0	0	0	0	0	0	0	0	3.7	0
XVI	0	0	0	0	0	0	0	0	0	0	0	0	7.4	0
XVII	0	0	0	0	0	0	0	0	0	0	0	0	3.7	0
XVIII	0	0	0	0	0	0	0	0	0	0	0	0	3.7	0
XIX	0	0	0	0	0	0	0	0	0	0	0	0	14.8	0
NoofHaplotypes	11	3	15	3	4	1	12	2	6	3	5	1	17	4

Published sequence data for W lines were from Lin et al (2001); Lin et al (2002); Morrell et al (2003)

W wild lines, *D* domesticated varieties

Table 5 Nucleotide diversity recorded at 7 Barley loci for 25 wild (W) and 20 domesticated (D) lines

Locus	All sites considered							Introns only (silent sites)				
	L	*S*		[a]$\pi\times10^{-3}$		$d^{b}_{DW}\times10^{-3}$	Tajima's D[c]	*L*	*S*		$\theta_w\times10^{-3}$	
		W	D	W	D				W	D	W	D
Adh2	836	17	2	3.98	0.93	3.56 (0.5)	−1.14 n.s.	388	4	1	2.73	0.73
Adh3	809	45	1	20.88	0.62	19.24 (3.9)	+0.88 n.s.	321	21	1	17.33	0.88
Amy1	661	9	0	1.52	0	0.79 (0.5)	−2.07 $P<0.05$	128	3	0	6.21	0
Dhn9	724	14	1	3.59	0.66	2.9 (0.49)	−1.22 NS	473	11	1	6.16	0.60
GAPDH	765	26	22	12.13	14.25	16.18 (1.88)	+2.60 $P<0.05$[d]	548	24	19	11.48	9.77
PEPC	941	3	0	0.68	0	0.47(0.11)	−0.83 NS	384	2	0	1.38	0
Waxy	816	38	10	10.78	3.32	9.17 (1.19)	−0.95 NS	358	38	10	27.54	7.87
Average				7.65	2.79						10.4	2.83

L Number of sites, *S* number of polymorphic (segregating) sites, θ_w Watterson's theta, NS not significant

[a] According to Nei (1987), equation 10.5

[b] Average N° of nucleotide substitutions between D and W according to Nei (1987) using the Jukes and Cantor correction, standard deviation in parentheses

[c] Both W and D sequences considered, *Amy*1 and *GAPDH* have a significant Tajima D test

[d] Within D, the value is 2.84 ($p<0.01$); within W 1.28 (not significant)

Evidence for selection?

Amy1 showed significant negative values of Tajima's D when all wild and domesticated sequences were considered (Table 5), while *GAPDH* gave a significant positive value of Tajima's D in comparisons within domesticated lines (Table 5). In principle, this could potentially indicate a deviation from neutrality, possibly due to positive (*Amy*) or balancing (*GAPDH*) selection. Indeed, there was an a priori expectation that we should be able to detect evidence for selection at *Amy1* because amylase activity is a key component of barley malt production, a trait that was enriched by human breeding. However, as Tajima (1989) has pointed out, the test is contingent upon the assumption that the population(s) in question has been in mutation-drift balance for a long evolutionary time, which is unlikely to apply in the current sample. Indeed, the wild accessions consist of individuals from diverse geographic ranges (not populations in the strict sense). Furthermore, Tajima (1989) has pointed out that if the taxa in question have experienced a bottleneck, Tajima's D can and will produce significantly positive or negative values (Tajima 1989; Wright and Gaut 2005) for genes that are selectively neutral. Hence the results of Tajima's test are also consistent with our hypothesis that bottlenecks due to domestication and breeding are the major determinants of polymorphism loss in the domesticated lines sampled.

If these sequences are in fact evolving neutrally in the wake of a bottleneck, how to account for the lack of nucleotide substitutions among *Amy1*-I and *PEPC*-I haplotypes? If we assume a grass nuclear substitution rate of 6.5×10^{-9} substitutions per site per year (Gaut et al. 1996) and furthermore assume that we have sampled fully 10,000 years per lineage in all *Amy1*-I and *PEPC*-I haplotype comparisons (Salamini et al. 2002), then we would expect to observe about one substitution per 7,500 sites. At *Amy1* and *PEPC* no substitutions were detected in about 15,000 freely mutable sites compared among domesticated lines (5,000 at *Amy1*, 2,560 in introns plus ~2,450 in coding regions; 10,000 at *PEPC*, 7,680 in introns and ~2,450 in coding regions). Thus, even if we had sampled the maximum amount of time possible with *Amy1*-I and *PEPC*-I haplotypes, we would only have anticipated two substitutions where none were observed. It is unlikely that all *Amy1*-I and *PEPC*-I haplotypes diverged 10,000 years ago, hence the lack of segregating sites is still consistent with domestication and breeding bottlenecks.

Revisiting the site of barley domestication

Badr et al. (2000) provided evidence from 400 AFLP loci using 317 wild and 57 domesticated *Hordeum* lines indicating that barley was, most probably, domesticated only once (see also Salamini et al. 2004), and that the Israeli–Jordan area is the region in which barley was brought into culture. This location is well outside the core area in southeastern Turkey (Lev-Yadun et al. 2000; Salamini et al. 2002), which is associated with several other plant domestication events of the Neolithic Near East area. Although the present data only encompass 25 wild (Lin et al. 2001; Morrell et al. 2003) and 20 domesticated lines, albeit at the level of sequences rather than AFLPs, we used it to readdress the site of barley domestication. Individually, the sequences of the seven genes sampled from 45 barley

lines provided only a partial resolution, due to the small number of nucleotide differences both within wild accessions and between wild and domesticated lines (Supplementary Fig. 1).

Concatenating the available DNA sequences to an alignment of 5,616 bp per accession has the effect of mixing signals due to recombination or hybridization, which would be highly undesirable in tree-building approaches to sequence relationships. However, the NNet planar graph of sequence differences between individuals reveals three major groups of wild accessions sharing similar haplotype collections, which we designate as wild haplotype clusters (WHC-) 1, 2, and 3 in Fig. 1a. It also uncovers the haplotype-sequence hybrid nature of the wild Tajikistan accession, which has a strong component of shared similarity both with WHC1 and with WHC3 (Fig. 1a). WHC1 comprises wild lines from Israel, Jordan, and Syria. WCH2 contains wild lines reaching further East (Israel, Jordan, Turkey, and Iraq) and into the core area (indicated in Fig. 1a). WHC3 contains lines collected from areas further East still, extending far beyond the primary habitat in the Fertile Crescent (indicated in Fig. 1a), reaching into Turkmenistan (Morrell et al. 2003), but also includes an Israeli line, in agreement with the findings of Badr et al. (2000), who previously reported secondary migrations into Israel in their study of 374 barley lines.

The relationship of the domesticated Turkish lines to the wild lines is twofold: the two-rowed domesticated varieties (except line 8) share the haplotype collection of WHC1, whereas the six-rowed varieties (and the two-rowed line 8) share the haplotypes of WHC2. WHC3 is genetically distinct from the domesticated forms at these loci. Similar relationships are described by the maximum-likelihood tree (Supplementary Fig. 1). However, here the Tajikistan accession n° 44 (PI293411) clusters within sequences belonging to WCH3: the tree shows only one signal, the NNet recovers two.

The WCH1, -2 and -3 clusters represent clusters of shared sequence similarity founded in discrete haplotype distributions and do not correspond to the West, Zagros, and East groups designated by Morrell et al. (2003) on the basis of geographical locations. The countries from which wild accessions were collected (see Morrell et al. (2003) for details) are shown in Fig. 1b together with the ranges observed for members of WHC1–3. WHC1 and 2 both contain lines assigned to the West group, but as Morrell et al. (2003) point out, the main determinant of sequence similarity is not correlated to geographic distance.

With the exception of a single accession from Israel, probably reintroduced (Badr et al. 2000), WHC3 encompasses accessions that were collected outside the primary habitat in the Fertile Crescent and hence, like the wild Himalayan accessions (Badr et al. 2000), likely represent the result of human dispersal. There is no clear correlation between haplotype structure and geographical distance from the primary habitat in WHC3, but it clearly represents a genotype distinct at these loci from those involved in domestication (Fig. 1a).

In concatenated data, the structure of the network (or tree) is determined by distribution of the most divergent haplotypes among individuals, thereby severely skewing the result to reflect the nucleotide divergence signal represented by ancient but randomly assorted alleles. The same problem is encountered when allelic sequence variants at a single locus are analyzed with tree methods (Lin et al. 2001): sequence differences may take millions of years to accumulate but only one generation to reassort into new combinations. Standard measures typically applied to compare populations are inapplicable here, because the plants sampled do not constitute groups of preferentially interbreeding individuals (except perhaps the domesticated forms). In order to examine genotype relationships with deweighted effects from ancient alleles, we calculated the genetic distance between individuals as the proportion of different haplotypes per diploid genotype, thereby scoring haplotypes as either identical or different, regardless of the amount of nucleotide divergence between different haplotypes. This provided a much different picture of the relationships between wild and domesticated barley (Fig. 2b), one in which two-rowed and six-rowed varieties again interleaved. But by scoring haplotypes as either identical or not, the domesticated forms clustered together, yet including to wild accessions from the Jordan valley (lines 50 and 51), in agreement with the independently obtained findings of Badr et al. (2000).

On the basis of haplotype diversity at seven loci in a sample of wild lines, the present data suggest that domesticated barley is genetically more similar to wild lines from the Jordan valley, which lies outside the core area in the Fertile Crescent. The domestication history of two-rowed and six-rowed varieties is unclear, but both types bear haplotypes that predominate in the Western Fertile Crescent. While our findings are in agreement with the previously inferred site of barley domestication in the Jordan valley (Badr et al. 2000), the new data open the possibility that barley domestication might have been diphyletic. Also AFLP data from the 20 D lines processed by the NNet procedure support the clear separation between two and six-rowed genotypes (Fig. 2b). A diphyletic conclusion was previously excluded (Badr et al. 2000), but is favored

by other authors (Molina-Cano et al. 2005 and citation therein) claiming independent barley origins for either two- versus six-rowed ears (Kolodinska Brantestam et al. 2004; Casas et al. 2005; Tanno and Takeda 2004), for brittleness of the rachis (Komatsuda et al. 2004), for hulled-naked caryopsis (Taketa et al. 2004), and for western or eastern cultivated barleys (Komatsuda et al. 2004). This particular matter concerning single versus multiple origins of barley is, however, complicated by the fact that (1) multiple independent introgression of genes from wild relatives to cultivated varieties can mimic multiple domestication events (Abdel-Ghani et al. 2004; Badr et al. 2000, Kanazin et al. 2002); (2) splitting of domesticated genotypes in two alternatives groups may have followed, and not be coeval with, the domestication process.

Acknowledgment This work was supported by the Deutsche Forschungsgemeinschaft.

References

Abdel-Ghani AH, Parzies HK, Omary A, Geiger HH (2004) Estimating the outcrossing rate of barley landraces and wild barley populations collected from ecologically different regions of Jordan. Theor Appl Genet 109:588–595

Badr A, Müller KJ, Schäfer-Pregl R, El Rabey H, Effgen S, Ibrahim HH, Pozzi C, Rohde W, Salamini F (2000) On the origin and domestication history of barley (*Hordeum vulgare*). Mol Biol Evol 17:499–510

Bryant D, Moulton V (2004) Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. Mol Biol Evol 21:255–265

Buckler ES, Thornsberry JM, Kresovich S (2001) Molecular diversity, structure and domestication of grasses. Genet Res Camb 77:213–218

Bundock P, Henry RJ (2004) Single nucleotide polymorphism, haplotype diversity and recombination in the Isa gene of barley. Theor Appl Genet 109:543–551

Casas AM, Yahiaoui S, Ciudad F, Igartua E. (2005) Distribution of MWG699 polymorphism in Spanish European barleys. Genome 48:41–45

Castiglioni P, Pozzi C, Heun M, Terzi V, Muller KJ, Rohde W, Salamini F (1998) An AFLP-based procedure for the efficient mapping of mutations and DNA probes in barley. Genetics 149:2039–2056

Chalmers KJ, Waugh R, Watters J, Forster BP, Nevo E, Abbott RJ, Powell W (1992) Grain isozyme and ribosomal DNA variability in *Hordeum spontaneum* populations from Israel. Theor Appl Genet 84:313–322

Dawson IK, Chalmers KJ, Waugh R, Powell W (1993) Detection and analysis of genetic variation in *Hordeum spontaneum* populations from Israel using RAPD markers. Mol Ecol 2:151–159

Diamond J (1997) Guns, germs and steel. Random House, London

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17:368–376

Gaut BS, Morton BR, McCaig BC, Clegg MT (1996) Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. Proc Natl Acad Sci USA 93:10274–10279

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41:95–98

Harlan JR (1976) Barley. In: Simmonds NW (ed) Evolution of crop plants. Longman, London, pp 93–98

Huson DH (1998) SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics 14:68–73

Jana S, Pietrzak MI, Srivastava JP, Holwerda BC, Thai KM. (1987) Genetic diversity in wild barley (*Hordeum spontaneum*) populations of the Fertile Crescent. Barley Genet 5:63–73

Kanazin V, Talbert H, See D, DeCamp P, Nevo E, Blake T (2002) Discovery and assay of single-nucleotide polymorphisms in barley (*Hordeum vulgare*). Plant Mol Biol 48:529–537

Kolodinska Brantestam A, von Bothmer R, Dayteg C, Rashal I, Tuvesson S, Weibull J (2004) Inter simple sequence repeat analysis of genetic diversity and relationships in cultivated barley of Nordic and Baltic origin. Hereditas 141:186–192

Komatsuda T, Maxim P, Senthil N, Mano Y (2004) High-density AFLP map of *nonbrittle rachis 1* (*btr1*) and 2 (*btr2*) genes in barley (*Hordeum vulgare* L.). Theor Appl Genet 109:986–995

Lev-Yadun S, Gopher A, Abbo S (2000) The cradle of agriculture. Science 288:1602–2603

Lin J-Z, Brown AHD, Clegg MT (2001) Heterogeneous geographic patterns of nucleotide sequence diversity between two alcohol dehydrogenase genes in wild barley (*Hordeum vulgare* subspecies *spontaneum*). Proc Natl Acad Sci USA 98:531–536

Lin J-Z, Morrell PL, Clegg MT (2002) The influence of linkage and inbreeding on patterns of nucleotide sequence diversity at duplicate alcohol dehydrogenase loci in wild barley (*Hordeum vulgare* ssp. *spontaneum*). Genetics 162:2007–2015

Molina-Cano J-L, Russell JR, Moralejo MA, Escacena JL, Arias G, Powell W (2005) Chloroplast DNA microsatellite analysis supports a polyphyletic origin for barley. Theor Appl Genet 110:613–619

Morrell PL, Lundy KE, Clegg MT (2003) Distinct geographic patterns of genetic diversity are maintained in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite migration. Proc Natl Acad Sci USA 100:10812–10817

Morrell PL, Toleno DM, Lundy KE, Clegg MT (2005) Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite high rates of self-fertilization. Proc Natl Acad Sci USA 102:2442–2447

Morgante M, Salamini F (2003) From plant genomics to breeding practice. Curr Opin Biotech 14:214–219

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nevo E, Beiles A, Kaplan D, Starch N, Zohary D (1986a) Genetic diversity and environmental associations of wild barley, *Hordeum spontaneum* in Turkey. Genetica 68:203–213

Nevo E, Zohary D, Beiles A, Kaplan D (1986b) Genetic diversity and environmental associations of wild barley, *Hordeum spontaneum* (Poaceae) in Iran. Plant Syst Evol 153:141–164

Nevo E, Zohary D, Brown AHD, Harber M (1979) Genetic diversity and environmental associations of wild barley, *Hordeum spontaneum*, in Israel. Evolution 33:815–833

Paterson AH (2002) What has QTL mapping taught us about plant domestication? New Phytol 154:591–608

Paterson AH, Lin YR, Li Z, Schertz KF, Doebley JF, Pinson SRM, Liu SC, Stansel JW, Irvine JE (1995) Convergent

domestication of cereal crops by independent mutations at corresponding genetic loci. Science 269:1714–1717
Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94–100
Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497
Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa, pp 365–386
Russell J, Booth A, Fuller F, Harrower B, Hedley P, Machray G, Powell W (2004) A comparison of sequence-based polymorphism and haplotype content in transcribed and anonymous regions of the barley genome. Genome 47:389–398
Salamini F, Özkan H, Brandolini A, Schäfer-Pregl R, Martin W (2002) Genetics and geography of wild cereal domestication in the Near East. Nat Genet Rev 3:429–441
Salamini F, Heun M, Brandolini A, Ozkan H, Wunder J (2004) Comment on AFLP data and the origins of domesticated crops. Genome 47:615–620
Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18:502–504
Schneider K, Borchardt DC, Schäfer-Pregl R, Nagl N, Glass C, Jeppson A, Gebhardt C, Salamini F (1999) PCR-based cloning and segregation analysis of functional gene homologues in *Beta vulgaris*. Mol Gen Genet 262:515–524
Schneider K, Weisshaar B, Borchardt DC, Salamini F (2001) SNP frequency and allelic haplotype structure of *Beta vulgaris* expressed genes. Mol Breed 8:63–74
Snow L, Brody T (1984) Genetic variation of H. *spontaneum* in Israel Eco-Geographical Races, detected by trait measurements. Pl Syst Evol 145:15–28
Søgaard B, von Wettstein-Knowles P (1987) Barley: genes and chromosomes. Carlsberg Res Comm 52:123–196
Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595
Taketa S, Kikuchi S, Awayama T, Yamamoto S, Ichii M, Kawasaki S (2004) Monophyletic origin of naked barley inferred from molecular analyses of a marker closely linked to the *naked caryopsis* gene (*nud*). Theor Appl Genet 108:1236–1242
Tanno K, Takeda K (2004) On the origin of six-rowed barley with brittle rachis, agriocrithon [*Hordeum vulgare* ssp. *vulgare* f. *agriocrithon* (Aberg)]. Theor Appl Genet 110:145–150
Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS (2004) Selection versus demography: a multilocus investigation of the domestication process in maize. Mol Biol Evol 21:1214–1225
Vigouroux Y, McMullen M, Hittinger CT, Houchins K, Schulz L, Kresovich S, Matsuoka Y, Doebley J (2002) Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. Proc Natl Acad Sci USA 99:9650–9655
Vinh LS, von Haeseler A (2004) IQPNN: moving fast through tree space and stopping in time. Mol Biol Evol 21:1565–1571
Wright SI, Gaut BS (2005) Molecular population genetics and the search for adaptive evolution in plants. Mol Biol Evol 22:506–519
Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS (2005) The effects of artifical selection on the maize genome. Science 308:1310–1314
Zabeau M, Vos P (1993) Selective restriction fragment amplification: a general method for DNA fingerprinting. European Patent Application. Publication No:858A1
Zohary D, Hopf M (2000) Domestication of plants in the old world, 3rd edn. Oxford University Press, Oxford